

SIMS 54th conference

Bergen University College, Norway

October 16-18, 2013

Together, the Bergen University College (HiB) and the Norwegian Society of Automatic Control (NFA) are arranging SIMS 54.

The purpose of the conference is to cover broad aspects of modeling and simulation with a focus on modeling and simulation applied to design of automated solutions for various applications, targeting technical areas in addition to the economy area. It is of interest for model builders, simulator users, scientists, engineers, vendors, etc.

The scientific program includes technical sessions with submitted and invited papers.

Statoil Mongstad



Wednesday 16th October

08.00 Registration and morning refreshments

09.00 Welcome

Keynote

09.10 State of art and challenges in automotive powertrain modeling and simulation

Lars Eriksson

Modeling

Chair: Lars Eriksson

10.00 Modeling for control of run-of-river power plant

Yiyang Ruan, Ingvar Andreassen and Bernt Lie

10.20 PDE Modeling with Modelica via FMI Import of HiFlow3 C++ Components

Kristian Stavåker, Staffan Ronnås, Martin Wlotzka, Vincent Heuveline and Peter Fritzson

10.40 Coffee break

11.00 Modeling of a multi agent system for simulation of battalion operations

Solveig Bruvoll and Rikke Amilde Løvlid

11.20 Intelligent Methods in Modelling and Simulation of Complex Systems

Esko Juuso

11.40 Lunch

Simulation

Chair: Bernt Lie

13.00 Tool-Independent Distributed Simulations Using Transmission Line Elements And The Functional Mock-up Interface

Robert Braun and Petter Krus

13.20 Utilizing the simulator fidelity to leverage fit for purpose in marine and process automation

Ole Andre Gjerpe, Ole Kristian Bjercknes and Øivind Rui

13.40 Coffee break

14.00 Design and simulation of a small scale experimental facility for drilling operations

Emil Cimpan and Gerhard Nygaard

14.20 Simulation and Evaluation of the Drilling Fluid Mixing and Conditioning Process

Gerhard Nygaard and Emil Cimpan

14.40 Coffee break

15.00 Dynamic station-keeping capability analysis using advanced vessel simulator

Dong Trong Nguyen, Luca Pivano, Brede Børhaug and Øyvind Smogeli

General assembly

Chair: Esko Juuso

15.20 SIMS annual General Assembly

16.20 Day 1 Completed

19.30 Conference dinner at Bellevue Restaurant.
Adr. Bellevuebakken 9, 5019 Bergen
Phone: 55 33 69 99

Thursday 17th October

08.30 Registration and morning refreshments

Process data evaluation

Chair: *Gerhard Nygård*

09.00 Methodology for modeling, parameter estimation, and validation of powertrain torsional vibration

Neda Nickmehr, Lars Eriksson and Jan Åslund

09.20 State Estimation of a CO₂ Capture Plant

Sanoja Ariyaratna, Bernt Lie and Morten Melaaen

09.40 To design a novel micromachined electromechanical acoustic sensor to wrap around Roebel bars of large generators to monitor partial discharge

Kourosh Mousavi Takami

10.00 Coffee break

Cases

Chair: *Emil Cimpan*

10.40 Prediction of daily runoff from hydrological catchment area

Bernt Lie, Sobhan Shafiee and Beathe Furenes

11.00 Fuel Efficient Speed Profiles for Finite Time Gear Shift with Multi-Phase Optimization

Xavier Llamas, Lars Eriksson and Christofer Sundström

11.20 Factors Influencing the Estimation of Downhole Pressure far Away From Measurement Points During Drilling Operations

Eric Cayeux and Hans Petter Lande

11.40 Lunch

13.00 Evaluation of various methods for estimating dynamical downhole pressure during drilling operations

Hans Petter Lohne, Gerhard Nygaard, Rune Time and Michael Nikolaou

13.20 Using Probabilistic Analysis with Finite Element Modeling of High Temperature Geothermal Well Casings

Gunnar Skúlason Kaldal, Magnús Þór Jónsson, Halldór Pálsson and Sigrún Nanna Karlsdóttir

13.40 Coffee break

14.00 A Numerical Analysis of a Boiling Front Moving Through Porous Medium

Larus Thorvaldsson and Halldor Palsson

14.20 Electric Transmission and Distribution Loss Reduction Strategies for Iran's Electricity Energy Sector

Kourosh Mousavi Takami, Hassan Gjolinejad, Amin Mousavi Takami, Erik Dahlquist

14.40 Closing remarks

Friday 18th October

Visiting Statoil Mongstad

08.30 Meet in the lobby at the hotel

09.00 Bus leaves Bergen for transport to the Statoil Mongstad refinery, thermal power plant and carbon capture facility

10.00 General presentation of the Mongstad industrial site, by Jon Halstein Tjore, leader communication

10.45 Presentation of a case: Delayed Coker Hydraulic Decoking System: – Automated hydro drilling sequence, by ing. Christian Dreyer Skre

11.20 Lunch

12.00 Guided tour at the industrial area by bus

13.15 Presentation of the CO₂ Technology Centre Mongstad, TCM

15.00 Bus leaves Mongstad

16.00 Bus arrives Bergen

PRACTICAL INFORMATION

Time and place for SIMS 2013:

16th - 18th October

Bergen University College

Nygårdsgaten 112, Bergen, NORWAY

Language:

This conference will be held in English

Registration:

Register for the conference at www.nfaplassen.no, phone +47 90 90 92 94 or email nfa@nfaplassen.no

Registration fee

Members of NFA or SIMS	€ 500 or 4100 NOK
Non members	€ 650 or 5300 NOK
*Students	€ 150 or 1300 NOK
Dinner: Student/extra guest	€ 70 or 570 NOK

(price on registration form is given in NOK)

Full registration fees include admission to all the sessions of the conference, the conference proceedings, refreshments, lunch and the conference dinner as well as excursions on October 18th. (Extra guest for dinner can be booked in registration form)

*The student fees include admission to all the sessions and refreshments. (dinner is not included) If you (student) wish to attend dinner you are welcome. Book dinner in registration form.

All prices are including VAT, and do not include the price of the hotel.

Registration for the conference is binding. There will be no refund of the registration fee.

Organizers in Norway

Emil Cimpan, Bergen University College

Bernt Lie, Telemark University College

Gerhard Nygaard, University of Stavanger

Lars Annfinn Ekornsæter, NFA

Brith Sandell, NFA

International Scientific Committee

Brian Elmegaard - Technical University of Denmark

Lars Eriksson - Linköping University

Esko Juuso - University of Oulu

Tiina Komulainen - Oslo and Akershus University College

Jónas Ketilsson - National Energy Authority of Iceland Magnus

Thor Jonsson - University of Iceland

Tommy Mølbak - DONG Energy A/S

Axel Ohrt Johansson - DONG Energy A/S

Erik Dahlquist - Mälardalen University

Kaj Juslin - VTT

Bernt Lie - Telemark University College

Peter Fritzson - Linköping University

Emil Cimpan - Bergen University College

Gerhard Nygaard - University of Stavanger

Hotel:

We have reserved some rooms at Thon Hotel Bristol. To get the best price you have to mention SIMS 2013.

Address: Thon Hotel Bristol, Torgalmenningen 11, 5014 Bergen

Phone: +47 55 55 10 00

Mail: bristol.bergen@thonhotels.no

Dinner 16th October:

We arrange a very good dinner at Lyststedet Bellevue.

If you are a student the dinner is not included in the fee.

We hope you will attend anyway. You sign up for the dinner using the registration form at our web page www.nfaplassen.no



HØGSKOLEN I BERGEN



Statoil



Modeling for control of run-of-river power plant

Bernt Lie^{†*}, Yiyang Ruan[†], and Ingvar Andreassen[‡]

[†]: Telemark University College, Porsgrunn, Norway

[‡]: Skagerak Kraft AS, Porsgrunn, Norway

September 26, 2013

Abstract

The dynamics of free surface flow of water in shallow channels is of importance in the control and optimal operation of run-of-river hydro power plants. Such dynamics is often described using the Saint Venant model, which is based on mass balance and momentum balance. This combination of free surface and momentum balance makes the problem numerically challenging to solve. Here, the finite volume method with staggered grid is used to illustrate the dynamics of the river upstream from the Grønvollfoss run-of-river power plant in Telemark, Norway, operated by Skagerak Energi. With some tuning and/or further simplifications, the model will be suitable for control design.

Keywords: Run-of-river, Saint Venant model, Finite Volume Method, hydro power, level control.

1 Introduction

Norway has some 1150 hydro power plants with a total capacity of more than 28 GW and an annual production of 135TWh,¹ which constitutes some 98% of the production of electricity. The main turbine types are Pelton, Francis, Kaplan and Bulb. Pelton (heads² above 500 m) and Francis (heads [50, 600] m) turbines are used in high-head plants, while Kaplan and Bulb turbines are common in run-of-river power plants and other plants with head below 50 m.^{3 4} Run-of-river plants normally have little storage capacity/small pondages, and with several such plants along a river, this leads to a strong coupling between the plants. As an example, Skagerak Energi is operating two run-of-river hydro power plants in river Tinnelva flowing out of lake Tinnsjøen (reservoir) in Eastern Telemark. The uppermost, Årlifoss, has

*Corresponding author: Bernt.Lie@hit.no

¹www.regjeringen.no/upload/OED/pdf%20filer/Faktaheftet/EVfakta08/EVfacts08_kap02_eng.pdf;
the actual production varies with the year

²“head” is the pressure across the turbine, measured in water height

³The Kaplan turbine was developed in 1913, and some early run-of-river installations use Francis turbines due to Kaplan’s patent.

⁴www.sintef.no/home/SINTEF-Energy-Research/Project-work/Hydropower

a 16.2 m high water fall, and provides inlet flow to the pondage of Grønvollfoss. Grønvollfoss is situated 5 km downstream from Årlifoss, and has a water fall of 22.6 m. The installed power in Grønvollfoss is 2x16 MW (i.e. two turbines), and the average annual production is 172 GWh. There are three more plants operated by another energy company along the river before it reaches lake Heddalsvannet. In another waterway, Kragerøvassdraget, there are five run-of-river plants in the stretch from lake Toke to the ocean.

For such run-of-river systems, the operation of one plant has an effect on the downstream plants, and it is thus important to consider the optimization of the power production locally (small flow and level variations) and globally along the entire river. An accurate dynamic model plays an important part in controlling and optimizing such a run-of-river system. In [1], the problem of estimating the levels of three cascade river reaches by using the Saint-Venant model has been studied. In [2], the issue of discretization of the Saint-Venant equations using the staggered grid scheme has been. Large rivers with several run-of-river plants constitute large scale control and optimization problems; the application of e.g. Model Predictive Control has recently been studied for such systems, [3].

In this study, a mathematical model of the Grønvollfoss power plant in river Tinnelva is studied. The emphasis is put on the Saint-Venant model of the open surface, shallow channel of water along the river, which is based on the mass and momentum balances with some description of friction along the river bed. Such models are numerically challenging to solve; here the finite volume model with a staggered grid is used. The paper is organized as follows: in Section 2, the principle of the run-of-river power plant is discussed together with the reason why Saint-Venant equations and staggered grid are applied. Section 3 describes the Finite Volume model using staggered grid and the relationship with the Saint-Venant equations. The results of simulations are provided in Section 4. In Section 5, some conclusions are drawn.

2 System description

Figure 1 illustrates the inside and outside of a typical run-of-river power plant. Run-of-river plants typically have much higher flow rates but lower heads than high-head plants. The headwater passes through the turbine and spins the propeller. The turbine is a machine that transfers kinetic hydro power to rotational mechanical power [4]. After giving away its energy, the water will then return to the river via a short tailrace.

The Grønvollfoss power plant is the second plant along river Tinnelva, which means Grønvollfoss does not have a dedicated reservoir. Instead, it uses the outlet flow from the uppermost plant Årlifoss, located some 5 km upstream from Grønvollfoss [6]. Figure 2 illustrates the first two power plants in river Tinnelva. In the figure, \dot{V} indicates volumetric flow along the river, while \dot{V}' indicates volumetric flow per length unit into the river (precipitation, evaporation, side streams). The HRV (highest regulated water level) of Grønvollfoss is 144.50 meters above sea level, while LRV (minimum regulated water level) is not specified. The water level in the dam should be at $\text{HRV} \pm 1$ cm to maximize the head and thus the power production; during irregular conditions, deviations from this water level may be accepted. At a level of 144.52 meters above sea level (2 cm above HRV), the water will flow over the dam, and thus bypass the turbines and constitute a production loss. During the

Hydro power station

A run-of-river power station continuously produces electricity from flowing water in an impounded river. With a flow rate of, for example, 600 cubic metres of water per second it achieves an output of around 26 Megawatt. There are around 600 such power stations in Germany, some are more than 100 years old. Together, they feed more than 2 600 Megawatt electricity into the grid.

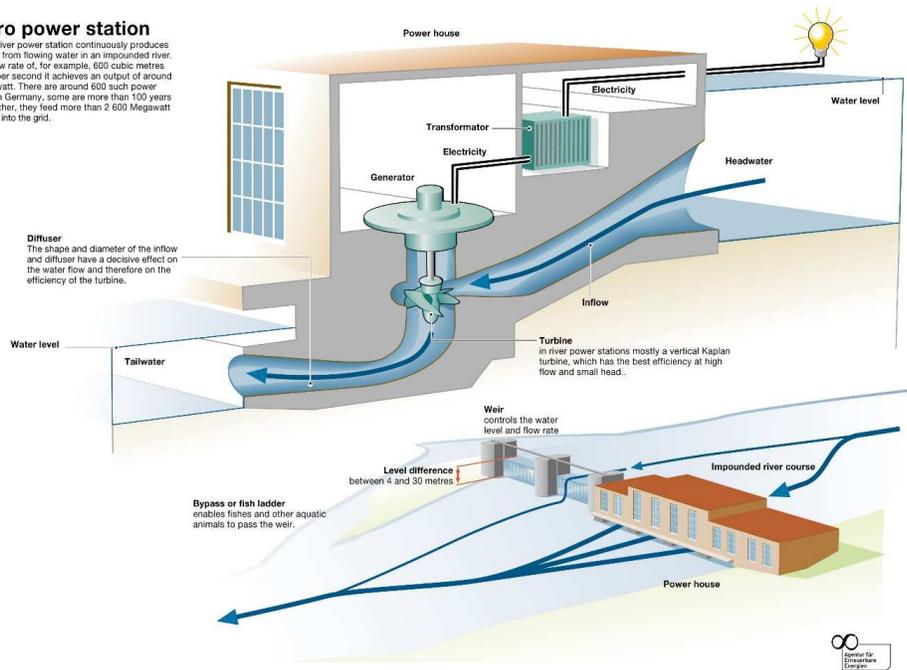


Figure 1: Typical run-of-river power station [5].

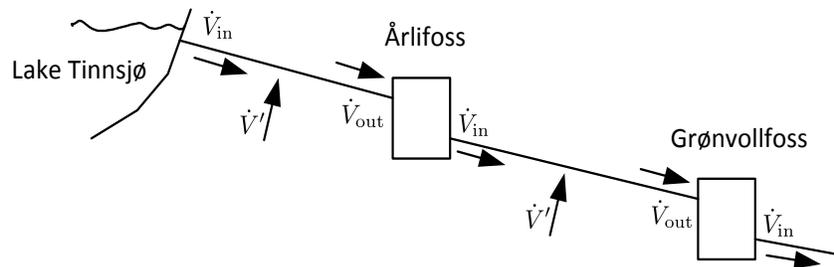


Figure 2: Overview of run-of-river hydro power plants in the upper part of river Tinnelva.

winter, it is important to keep the water level below this level to avoid floating ice falling and damaging equipment and gates.

Thus, in the river reach between Lake Tinnsjø and the Årlifoss power plant, \dot{V}_{in} is given by the gate opening and the water level in Lake Tinnsjø, while \dot{V}_{out} is given by the flow through the turbine of Årlifoss as well as the bypass through the gates at Årlifoss (a function of gate opening and level in the pondage at the inlet of Årlifoss). Similar considerations are valid for the Grønvollfoss power plant.

When controlling open channel flow, a standard goal is to keep a specific water level at a target by smoothly adjusting the controllable structures, for example weirs, gates and pumps. Additionally, the control actions for the structures need to remain within the physically possible capacities, such as maximum gate opening or pump capacity [7]. The controllable structures of Grønvollfoss include segment gate, roller gate, spillway gate, and flood gate. Normally, the segment gate is in use and the other gates remain closed.

From a control perspective, the objective of a controller is to keep the water level in the pondage at the inlet of the power plant at a target value, using the discharge through the segment gate as the manipulatable variable. The input and output variables for a controller are thus related to the boundary conditions of the distributed Saint-Venant equation which describes the behavior of the river.

In this paper, we are concerned with a single river reach, and use the open channel from Årlifoss to Grønvollfoss as a case study. Specifically, we are interested in both the water level h along the open channel, as well as the volumetric flow \dot{V} along the channel. A model description of such a channel involves both the mass balance and the momentum balance. Models involving both mass and momentum balances are tricky numerically, and a common strategy is to use a Finite Volume method with staggered grid; staggered grid means that the grid for the momentum balance is skewed wrt. the grid for the mass balance.

3 Model development

Figure 3 illustrates a Finite Volume discretization of the shallow river system with staggered grid where $i \in \{1, 2, \dots, 11\}$, i.e. the grid for the mass balance (blue control volumes on top of scheme) is skewed compared to the grid for the momentum balance (red control volumes below scheme).

The figure shows the level h and the volumetric flow rate \dot{V} and how these vary with x or i in the interval $x \in [0, L] / i \in \{1, 2, \dots, 11\}$ ⁵. Included is an indication of how the pressure p varies along x, i , and how the river cross sectional area A perpendicular to the flow direction varies with x, i (in the figure, A_{11} is displayed). The width w of the river is assumed to be constant. In a separate box B of the figure is an indication of the slope of the river given by angle θ , and how the river bed (represented by the x -axis) is sloped wrt. gravity g .

The number of control volumes for the momentum balance/flow rate is $N_{\dot{V}}$ — in fig. 3 $N_{\dot{V}} = 5$ and these control volumes all have length Δx . The number of control volumes for the mass balance/level is $N_h = N_{\dot{V}} + 1$; here, the two end volumes have length $\frac{\Delta x}{2}$ while the other $N_h - 2$ volumes all have length Δx .

⁵For simplicity, the figure shows a finite number of control volumes.

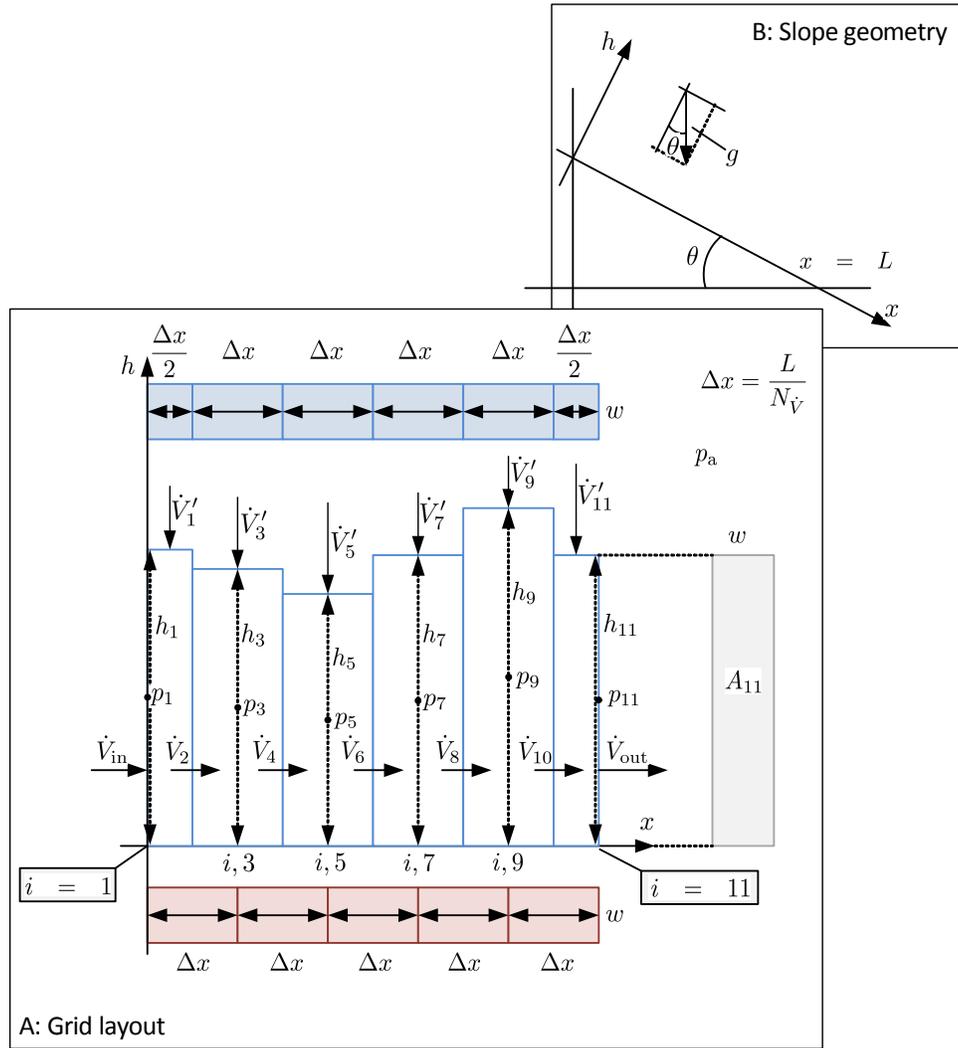


Figure 3: Box A: Geometry of control volumes for mass balance (blue boxes) and momentum balance (red boxes). Box B: Slope of river bed.

The Saint-Venant equations are based on several simplifying assumptions:

1. The water system we are considering is a shallow water river. Thus the depth and width of the river is much smaller than the length. We also assume that the river is not too steep.

This implies that the water level variation of the river is small; thus the vertical variation of the water motion can be replaced by the hydrostatic pressure approximation, and the flow velocity is approximated by the depth-averaged flow velocity [8].

2. The fluid is incompressible and the streamline curvature is small. In addition, formulas related to steady flow (e.g. the Manning Formula and the Chézy Formula) can be used to estimate the effect of friction.
3. Coriolis force is neglected. Only gravity force is taken into consideration.

Stelling and Duinmeijer studied the staggered conservative scheme for the discretization of Saint-Venant equations in [9]. The mass balance leads to the Finite Volume description

$$\frac{dh_i}{dt} = \frac{\dot{V}_{i-1} - \dot{V}_{i+1}}{w \cdot \Delta x} + \frac{\dot{V}'_i}{w}.$$

where $i \in \{3, 5, \dots, 2N_{\dot{V}} - 1\}$. At the ends, we have

$$\begin{aligned} \frac{dh_1}{dt} &= \frac{\dot{V}_{\text{in}} - \dot{V}_2}{w \frac{\Delta x}{2}} + \frac{\dot{V}'_1}{w} \\ \frac{dh_{2N_{\dot{V}}+1}}{dt} &= \frac{\dot{V}_{2N_{\dot{V}}} - \dot{V}_{\text{out}}}{w \frac{\Delta x}{2}} + \frac{\dot{V}'_{2N_{\dot{V}}+1}}{w}. \end{aligned}$$

In the momentum balance, we include pressure forces, the gravity force, and a friction force. The momentum balance leads to the Finite Volume description

$$\frac{d\dot{V}_i}{dt} = \frac{\dot{\mathcal{M}}_{i,i} - \dot{\mathcal{M}}_{i,x}}{\Delta x} + \frac{gw \cos \theta}{2} \frac{(h_{i-1}^2 - h_{i+1}^2)}{\Delta x} + \bar{A}_i g \sin \theta - \frac{g}{\bar{C}_i^2} \frac{\bar{\varphi}_i}{\bar{A}_i^2} |\dot{V}_i| \dot{V}_i$$

for $i \in \{2, 4, \dots, 2N_{\dot{V}}\}$, and where momentum flow per density, $\dot{\mathcal{M}}/\rho$ is given by

$$\begin{aligned} \frac{\dot{\mathcal{M}}_{i,i}}{\rho} &= \begin{cases} \frac{\dot{V}_{\text{in}}^2}{A_1}, & i = 2 \\ \frac{|\dot{V}_{i-2}|}{A_{i-2}} \max(\dot{V}_{i-2}, 0) + \frac{|\dot{V}_{i+2}|}{A_{i+2}} \max(-\dot{V}_{i+2}, 0), & i \in \{4, 6, 8\} \\ \frac{|\dot{V}_8|}{A_8} \max(\dot{V}_8, 0) & i = 10 \end{cases} \\ \frac{\dot{\mathcal{M}}_{i,x}}{\rho} &= \begin{cases} \frac{\dot{V}_i^2}{A_i}, & i \in \{2, 4, \dots, 8\} \\ \frac{\dot{V}_{\text{out}}^2}{A_{11}}, & i = 10. \end{cases} \end{aligned}$$

Here, the barred variables (\bar{A} , \bar{C} and $\bar{\varphi}$) are computed by averaging values in the neighboring nodes. In the model, φ is the wetted perimeter; $\varphi = w + 2h$ when the river width is

Table 1: Parameters for Grønvollfoss run-of-river power plant.

Variable	Value	Unit	Comments
w	100	m	Width of the river
L	5000	m	Length of the river
H	57 m	m	Height drop of river bed
$\sin \theta$	$\frac{H}{L}$	–	Specification of river bed angle
ρ	10^3	kg/m ³	Density of water
g	9.81	m/s ²	The gravitational acceleration
k_S	20	m ^{1/3} /s	Strickler friction coefficient

Table 2: Nominal operating conditions for Grønvollfoss run-of-river power plant.

Variable	Value	Unit	Comments
\dot{V}_{in}	120	m ³ /s	Inlet volumetric flow rate
\dot{V}_{out}	120	m ³ /s	Outlet volumetric flow rate
\dot{V}'	0	m ³ /(s m)	Lateral inflow per length unit
$h(x=L)$	4	m	Steady state level in front of Grønvollfoss dam
\dot{V}	120	m ³ /s	Typical steady volumetric flow rate

constant. C is Chézy's friction coefficient, which is related to Strickler's friction factor k_S by $C = k_S \left(\frac{A}{\varphi}\right)^{1/6}$; Strickler's friction coefficient typically ranges from $20 \frac{\text{m}^{1/3}}{\text{s}}$ (rough stones) to $80 \frac{\text{m}^{1/3}}{\text{s}}$ (smooth concrete or cast steel).

If we let $N_{\dot{V}} \rightarrow \infty$, the Finite Volume model approaches the Saint Venant equations, which here take the form

$$\begin{aligned} \frac{\partial h}{\partial t} &= -\frac{1}{w} \frac{\partial \dot{V}}{\partial x} + \frac{\dot{V}'}{w} \\ \frac{\partial \dot{V}}{\partial t} &= -\frac{\partial}{\partial x} \left(\frac{\dot{V}^2}{A} \right) - gA \cos \theta \frac{\partial h}{\partial x} + Ag \sin \theta - \frac{g}{C^2} \frac{\varphi}{A^2} \left| \dot{V} \right| \dot{V} \end{aligned}$$

with boundary conditions

$$\begin{aligned} \dot{V}_{x=0} &= \dot{V}_{\text{in}} \\ \dot{V}_{x=L} &= \dot{V}_{\text{out}}. \end{aligned}$$

Model parameters are given in Table 1.

Nominal operational parameters are given in Table 2.

4 Simulation

The simulations are started at nominal initial conditions in steady state, which means that \dot{V} will be equal to $\dot{V}_{\text{in}} = \dot{V}_{\text{out}} = 120 \text{ m}^3/\text{s}$ throughout. The steady state levels will be small at the inlets, and will then reach $h(x=L)$ at the outlet. It turns out that it is necessary

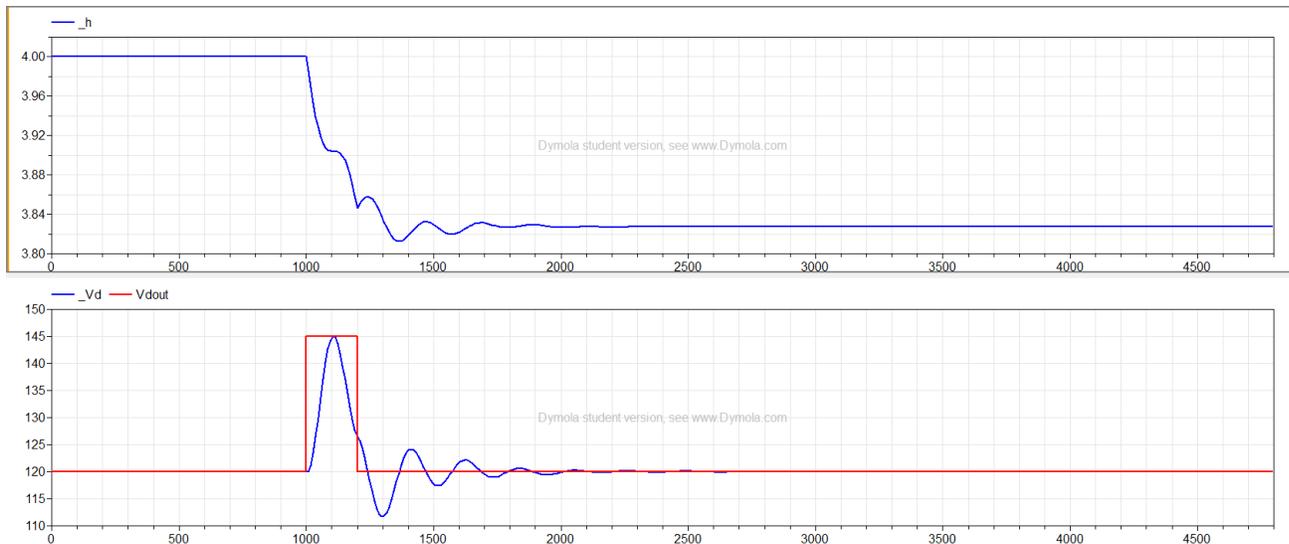


Figure 4: Responses in $h(x=L)$ and $\dot{V}(x=L)$ upon a sudden jump in outlet flow rate \dot{V}_{out} . Strickler friction coefficient $k_S = 20 \text{ m}^{1/3}/\text{s}$.

to set $N_{\dot{V}} \geq 19$ to ensure that all steady state level are positive. This has to do with the “rapid” increase in the steady state level towards the end of the river reach and the use of equal length of the river sections.⁶ The river is simulated for 4800 s, which equals 1 h20 min. Two scenarios are considered.

In the first scenario, the *outlet* flow rate \dot{V}_{out} is suddenly increased from $120 \text{ m}^3/\text{s}$ to $145 \text{ m}^3/\text{s}$ at 1000 s, and is then suddenly reduced back to $120 \text{ m}^3/\text{s}$ at 1200 s. This situation emulates a sudden increase in power production by changing the flow through the turbines. The responses in $h(x=L)$ and $\dot{V}(x=L)$ are shown in fig. 4.

The figure shows that the response is a sudden reduction in the level at the dam, with some oscillation.

In the second scenario, the *inlet* flow rate \dot{V}_{in} is suddenly increased from $120 \text{ m}^3/\text{s}$ to $145 \text{ m}^3/\text{s}$ at 1000 s, and is then suddenly reduced back to $120 \text{ m}^3/\text{s}$ at 1200 s. This situation emulates a preparation for increasing the power production by opening the gates at the inlet to the river reach. The responses in $h(x=L)$ and $\dot{V}(x)$ are shown in fig. 5.

The figure shows that it takes some 1200 s = 20 min from the change in gate opening at the inlet to the reach, before the water level at the dam starts to increase. This time delay is similar to the time delay observed in the actual power plant.

It is of interest to also consider the effect of the friction term. Thus let us consider the same cases as in figs. 4 and 5, but with the Strickler friction coefficient set to $k_S = 40 \text{ m}^{1/3}/\text{s}$. The results are shown in fig. 6 and fig. 7, respectively.

In particular in fig. 6, it looks as if the system is on the border of stability, and probably even slightly unstable. This is clearly not representing physics. It is well known that discretization “adds energy” to a system, and it is a common trick to dampen oscillations after discretization by adding some friction. Thus, the Strickler friction coefficient k_S in the partial differential equation should be increased slightly in the discretized model to give

⁶The model is encoded using the Modelica modeling language, and solved using Dymola.

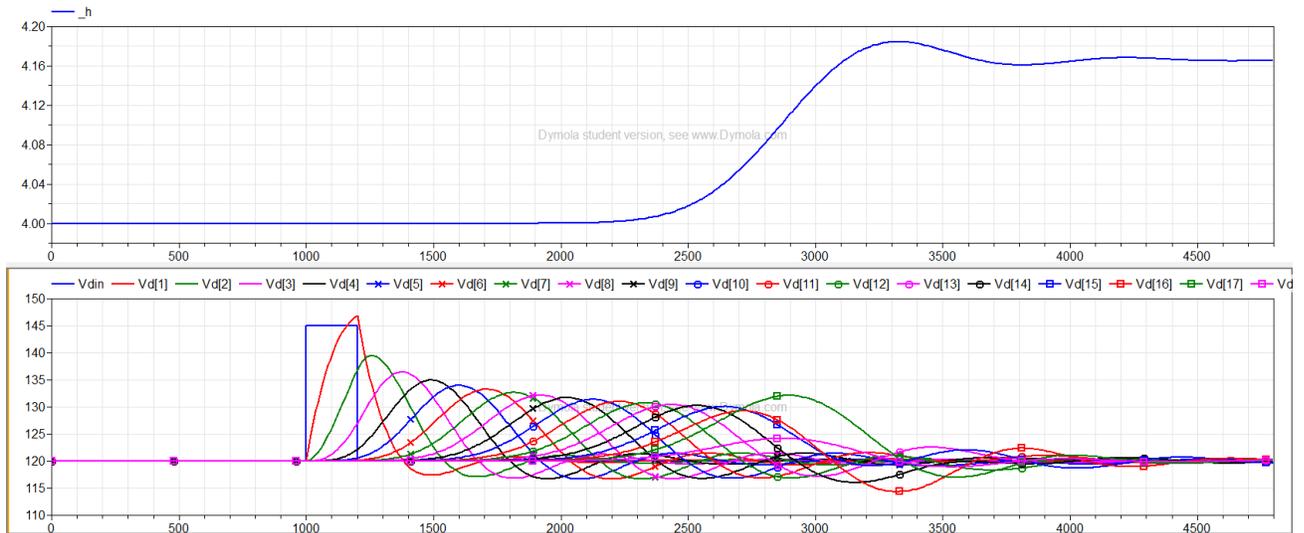


Figure 5: Responses in $h(x = L)$ and $\dot{V}(x)$ upon a sudden jump in inlet flow rate \dot{V}_{in} . Strickler friction coefficient $k_S = 20 \text{ m}^{1/3}/\text{s}$.

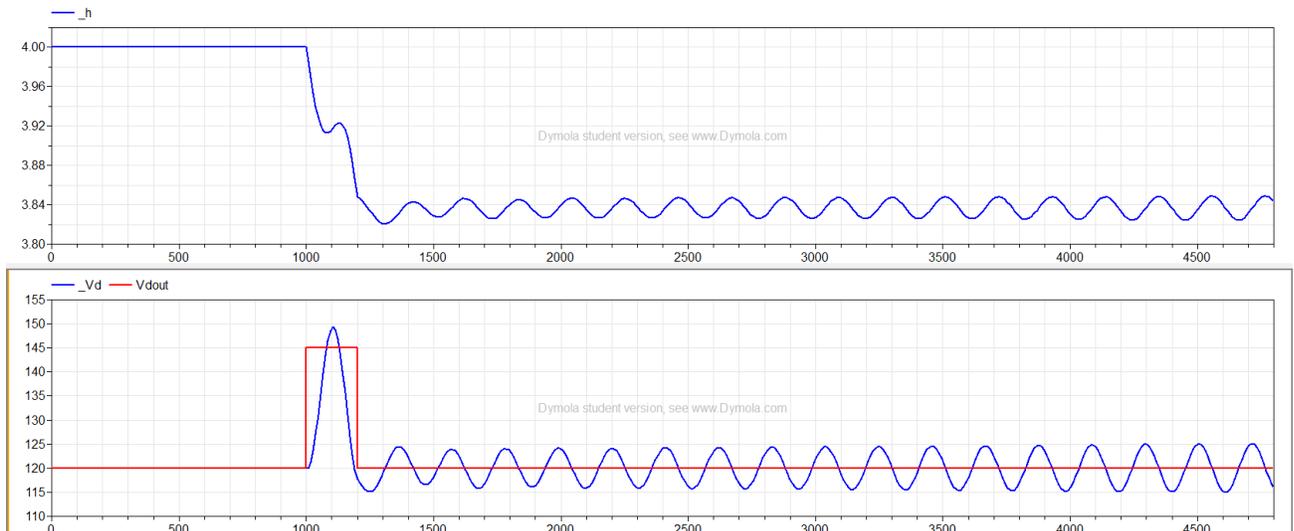


Figure 6: Responses in $h(x = L)$ and $\dot{V}(x = L)$ upon a sudden jump in outlet flow rate \dot{V}_{out} . Strickler friction coefficient $k_S = 40 \text{ m}^{1/3}/\text{s}$.

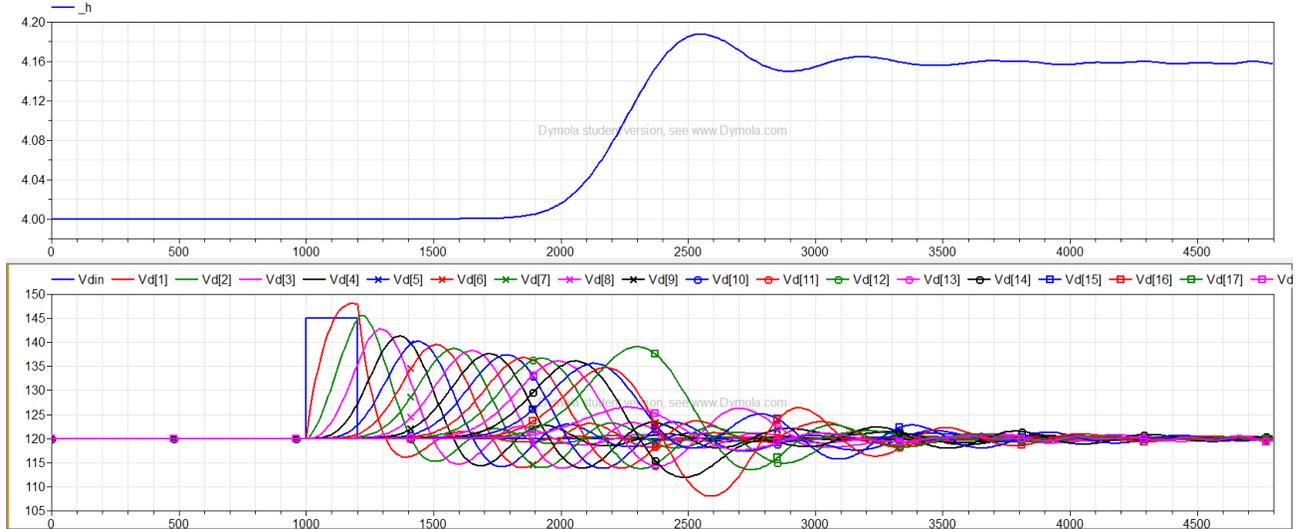


Figure 7: Responses in $h(x = L)$ and $\dot{V}(x)$ upon a sudden jump in inlet flow rate \dot{V}_{in} . Strickler friction coefficient $k_S = 40 \text{ m}^{1/3}/\text{s}$.

correct responses. The same is true for $k_S = 20 \text{ m}^{1/3}/\text{s}$. In all cases, it is difficult to say how much k_S should be increased when the model is discretized. In practice, the friction coefficient should be tuned against experimental data from the river.

By comparing figs. 5 and 7, we also see that the friction coefficient has some effect on the time delay in the river.

5 Conclusions

This paper has explored the application of the Saint-Venant equations in modeling of a run-of-river power plant for control purposes. The Saint-Venant equations were discretized and solved by the finite volume/staggered grid scheme, and simulations have validated that the model gives a reasonable description of reality. It is possible that the model can be further simplified by using a variable sized grid, with higher density towards the end of the river reach where the level rises somewhat.

The model under study is an efficient way of describing the dynamics of water level as well as the discharge. The description simplifies the structure of traditional 2-D models, the finite volume/staggered grid overcomes the numerical difficulty of solving the Saint-Venant equations, and the description is suitable for wide range of shallow water situations, including applications in run-of-river power plants and irrigation systems.

In the simulation process, we have made some unrealistic assumptions about constant width and slope, constant shore profiles along the river, etc. However, even when these assumptions are not perfectly satisfied, it is believed that the simplified model will give results that are qualitatively correct, and it may be possible to fit real systems by tuning some model parameters. Thus, further work should attempt to reconcile the model predictions with real data of shore profile and width, and measurements of rainfall and lateral flow rate, etc.

Acknowledgement Dr. Knut Vågsæther at Telemark University College has given valuable advice regarding the importance of using upstream information in computing the convective terms. Dietmar Winkler at Telemark University College has given valuable help with Modelica.

References

- [1] Farina, M., et al., Moving horizon estimation for distributed nonlinear systems with application to cascade river reaches. *Journal of Process Control*, 2011. 21(5): p. 767-774.
- [2] Xu, M., et al., De Saint-Venant equations-based model assessment in model predictive control of open channel flow. *Advances in Water Resources*, 2012.
- [3] De Schutter, B. and R. Scattolini, Introduction to the special issue on hierarchical and distributed model predictive control. *Journal of Process Control*, 2011. 21(5): p. 683-684.
- [4] Sayers, A.T., *Hydraulic and compressible flow turbomachines*. 1990: McGraw-Hill.
- [5] Energien, A.f.E., *Functioning principles of a hydro power station*, *Hydropower.jpg*, Editor 2011, Agentur für Erneuerbare Energien.
- [6] Ruan, Y., *Model based control of run-of-river power plant Grønvollfoss*, in *Faculty of Technology*, 2013, Telemark University College: Porsgrunn.
- [7] Xu, M., P. van Overloop, and N. van de Giesen, On the study of control effectiveness and computational efficiency of reduced saint-venant model in model predictive control of open channel flow. *Advances in Water Resources*, 2011. 34(2): p. 282-290.
- [8] van't Hof, B. and A.E. Veldman, Mass, momentum and energy conserving (MaMEC) discretizations on general grids for the compressible Euler and shallow water equations. *Journal of Computational Physics*, 2012.
- [9] Stelling, G.S. and S.P.A. Duinmeijer, A staggered conservative scheme for every Froude number in rapidly varied shallow water flows. *International Journal for Numerical Methods in Fluids*, 2003. 43(12): p. 1329-1354.

PDE Modeling with Modelica via FMI Import of HiFlow3 C++ Components

Kristian Stavåker
Programming Environments
Lab
Linköping University, Sweden
kristian.stavaker@liu.se

Staffan Ronnås
Interdisciplinary Center for
Scientific Computing (IWR)
Heidelberg University,
Germany
staffan.ronnas@iwr.uni-
heidelberg.de

Martin Wlotzka
Interdisciplinary Center for
Scientific Computing (IWR)
Heidelberg University,
Germany
martin.wlotzka@iwr.uni-
heidelberg.de

Vincent Heuveline
Interdisciplinary Center for
Scientific Computing (IWR)
Heidelberg University,
Germany
vincent.heuveline@iwr.uni-
heidelberg.de

Peter Fritzson
Programming Environments
Lab
Linköping University, Sweden
peter.fritzson@liu.se

ABSTRACT

Despite an urgent need and desire in academia as well as in industry for modeling Partial Differential Equations (PDEs) using the increasingly popular Modelica modeling and simulation language, there is limited support for this available at the moment. In this work, we propose a solution based on importing PDE models with PDE solvers implemented using the general-purpose parallel finite element library HiFlow3 as models into the Modelica environment using the standard Functional Mock-up Interface. In contrast to methods based on language extensions or automatic semi-discretizations in space, this approach requires no change to the language, and enables the use of specialized PDE solvers. Furthermore, it allows for full flexibility in the choice of geometry, model parameters, and space discretization between simulation runs without recompilation needed. This makes it possible to exploit advanced features of the PDE solver, such as adaptive mesh refinement, and to build complex multi-physics simulations by coupling different models, of both PDE and DAE type, in a straightforward manner using Modelica. We illustrate our method with an example that couples a Modelica Proportional-Integral-Derivative controller to a PDE solver for the unsteady heat equation in a 3D domain.

Categories and Subject Descriptors

I.6.2 [SIMULATION AND MODELING]: Simulation Languages;
I.6.5 [SIMULATION AND MODELING]: Model Development;
G.1.7 [NUMERICAL ANALYSIS]: Ordinary Differential Equations;
G.1.8 [NUMERICAL ANALYSIS]: Partial Differential Equations

General Terms

Languages, Theory

Keywords

Modelica, HiFlow3, OpenModelica, Partial Differential Equations, Finite Element Method, Functional Mock-Up Interface

1. INTRODUCTION

We discuss numerical simulation of models that couple Partial Differential Equations (PDEs) and Differential-Algebraic Equations (DAEs) in the context of the Modelica modeling and simulation language [15, 7]. Modelica originated around the idea of solving complex, coupled dynamical systems, which can be described by systems of Ordinary Differential Equations (ODE) or DAE. Up to now, there is only limited support for working with PDEs, despite the fact that the number of Modelica users in academia and in industry has grown significantly lately. One of the first attempts to incorporate PDE support into Modelica is described in [19], [20], and in Chapter 8 of [7], which investigates two different approaches: (1) expressing the PDEs using a combination of new language constructs and a supporting Modelica PDE library using the method-of-lines; (2) exporting the PDE part to an external PDE FEM C++ tool which solves the PDE part of the total problem. Based on this work, an experimental implementation of PDE support was added to the OpenModelica [4] compiler. However, this implementation has not been maintained, even though there have recently been discussions in the OpenModelica community about re-activating these features. Only one simple PDE operator is currently in the Modelica language specification: spatial distribution for 1D PDEs.

In [3] a Modelica library with basic building blocks for solving one-dimensional PDE with spatial discretizations based on the method of lines or finite volumes is described. Although this approach is attractive due to its simplicity, it is not clear how it could be extended to higher dimensions, without increasing the complexity significantly. Another approach is described in [11], which extends the modeling language with primitives for geometry description and boundary/initial conditions, and uses an external pre-processing tool to convert the PDE model to a DAE based on the method of lines. In both of these two works, the PDE system is expanded early on in the compilation process. In this way important information of the PDE structure is lost, information that could have been used for mesh refinement and adjustment of the run-time solver. Another similar option is to use the commercial MapleSim environment [14]: It means to write the PDEs in a Maple component, to export this component to DAE form using a discretization scheme

and to use the resulting component in MapleSim, which supports the Modelica language. An overview of how to use Maple and MapleSim together for PDE modeling can be found in [10]. Apart from the cost for licenses, this method again has the same drawback, arising from the loss of information regarding the original model.

In this work, we propose a way to allow for PDE modeling with Modelica by importing C++ components, written with the HiFlow3 multi-purpose finite element software [22], into Modelica using the Functional Mock-Up Interface (FMI) [6] import. FMI is a standard for model exchange and co-simulation between different tools. FMI supports only C but with correct linking it is possible to execute with C++ code. We use the OpenModelica [4] development environment but the same approach can be adapted to other Modelica environments. A similar approach was used in [13, 12], which describes a simulation of the energy supply system of a house using Dymola, ANSYS CFD and the TISC co-simulation environment. Some of the products used in that work are proprietary, however, whereas our environment is based on open standards and open source software. Furthermore, we use the “model import” approach for the coupling between components, instead of the “co-simulation” approach applied in those works.

The method described in this paper has several advantages:

1. HiFlow3 is well maintained and has strong support and capabilities for PDE modeling and solving;
2. HiFlow3 and OpenModelica are free to download and use;
3. The PDE structure is not lost but is maintained throughout the actual run-time simulation process. This allows for mesh refinement, solver run-time adjustments, etc.;
4. It is possible to mix PDE and DAE systems in the same system setting. This is also possible in [19].

As a proof-of-concept to demonstrate the Modelica/HiFlow3 integration, we have implemented and tested a coupled model which consists of solving the heat equation in a 3D domain and controlling its temperature via an external heat source. This source consists of a Modelica Proportional-Integral Derivative controller (PID controller), which is taken directly from the Modelica standard library.

The outline of the paper is as follows. In Section 2, we introduce the physical scenario we wish to simulate, and provide a mathematical overview of the two main components in the model. Section 3 describes in detail the realization of the simulation based on coupling of existing software components. In Section 4, we provide numerical simulation results for the example model, followed by a discussion in Section 5, and conclusions in Section 6.

2. SIMULATION SCENARIO

We consider the evolution of the temperature distribution in a rectangular piece of copper. Fig. 1 shows the setup of the system. A heat sensor is attached on the right side of the copper bar, and on the bottom side there is an adjustable heat source. The system is exposed to environmental influences through time-varying boundary conditions at the top and left sides. The goal is to control the temperature in the material by adjusting the heat source, so that a desired temperature is reached at the point of measurement. The

regulation of the heat source is done by a PID controller. It uses the sensor value and a reference temperature to compute the heat source strength. In our simulation, the two entities in this system

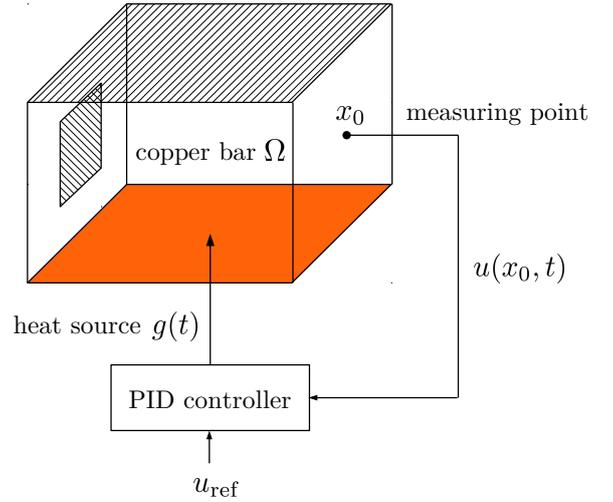


Figure 1: System consisting of a copper bar connected to a temperature regulator based on a PID controller.

are realized by reusing existing software components. The temperature of the copper bar is computed using a HiFlow3 solver, and the “LimPID” controller using a model from the Modelica standard library. The components are coupled by importing the HiFlow3 solver as a Modelica model using FMI, and then creating a third Modelica model, which connects these two components as well as some additional components.

2.1 Computing the Temperature Distribution

The evolution of the temperature distribution is modeled by the unsteady heat equation. In this subsection, we give the mathematical problem formulation, discuss the numerical treatment of the heat equation, and describe the discretization we used in the computations.

2.1.1 Heat Equation

We consider the copper to occupy a domain

$$\Omega := (0, 0.045) \times (0, 0.03) \times (0, 0.03) \subset \mathbb{R}^3,$$

where the boundary of Ω is denoted by $\partial\Omega$. The heat problem formulation for our simulation scenario is the following:

Find a function $u : \Omega \times (0, T) \rightarrow \mathbb{R}$ as the solution of

$$\partial_t u - \alpha \Delta u = 0 \quad \text{in } \Omega \times (0, T), \quad (1a)$$

$$u(0) = u_0 \quad \text{in } \Omega, \quad (1b)$$

$$u = g \quad \text{on } \Gamma_{\text{src}} \times (0, T), \quad (1c)$$

$$u = u_{\text{top}} \quad \text{on } \Gamma_{\text{top}} \times (0, T), \quad (1d)$$

$$u = u_{\text{left}} \quad \text{on } \Gamma_{\text{left}} \times (0, T), \quad (1e)$$

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \Gamma_{\text{iso}} \times (0, T). \quad (1f)$$

The unsteady heat equation (1a) is a parabolic PDE. Its solution, the unknown function u , describes the evolution of the temperature in the copper bar Ω during the time interval $(0, T)$. Here,

$\alpha = 1.11 \times 10^{-4} [\frac{m^2}{s}]$ denotes the thermal diffusivity of the copper. u_0 in equation (1b) is the initial state at time $t = 0$. The sensor is placed at the point $x_0 := (0.045, 0.015, 0.015)$, where the temperature $u(x_0, t)$ is taken as measurement value for time $t \in (0, T)$. The heat source is modeled by the Dirichlet boundary condition (1c). The controlled temperature $g(t)$ is prescribed on the source part of the boundary

$$\Gamma_{\text{src}} := [0, 0.045] \times [0, 0.03] \times \{0\} \subset \partial\Omega.$$

The environmental influence is modeled by the Dirichlet boundary conditions (1d) and (1e). At the top boundary part

$$\Gamma_{\text{top}} := [0, 0.045] \times [0, 0.03] \times \{0.03\} \subset \partial\Omega$$

and the left boundary part

$$\Gamma_{\text{left}} := \{0\} \times [0.01, 0.02] \times [0.01, 0.02] \subset \partial\Omega$$

the temperature is given by the functions $u_{\text{top}}(t)$ and $u_{\text{left}}(t)$ for $t \in (0, T)$, respectively. The rest of the boundary

$$\Gamma_{\text{iso}} := \partial\Omega \setminus (\Gamma_{\text{top}} \cup \Gamma_{\text{left}} \cup \Gamma_{\text{src}})$$

is isolated through the homogeneous Neumann boundary condition (1f).

2.1.2 Variational Formulation

A well-established method for numerically solving PDEs like the heat equation is the finite element method. Here we briefly describe the numerical treatment of problem (1). The methods of this section are taken from [2] and [5].

The finite element discretization is based on a variational formulation, which can be derived as follows. We denote by $C^k(X; Y)$ the set of all k times continuously differentiable functions from X to Y , and by $C_0^\infty(X; Y)$ the set of all smooth functions with compact support. In the common case $X = \Omega$, $Y = \mathbb{R}$, we just write $C^k(\Omega)$. Assuming that there is a classical solution

$$u \in C^1(0, T; C^2(\Omega) \cap C(\bar{\Omega}))$$

of problem (1), equation (1a) is multiplied by a test function $v \in C_0^\infty(\Omega)$ and integrated over Ω :

$$\int_{\Omega} (\partial_t u) v \, dx - \alpha \int_{\Omega} (\Delta u) v \, dx = 0 \quad (2)$$

Green's first identity [9] is applied to the second term of (2), giving

$$\begin{aligned} \int_{\Omega} (\Delta u) v \, dx &= \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, ds - \int_{\Omega} \nabla u \cdot \nabla v \, dx \\ &= - \int_{\Omega} \nabla u \cdot \nabla v \, dx, \end{aligned}$$

where n is the outer unit normal on $\partial\Omega$. The boundary integral vanishes since v is zero on $\partial\Omega$. This leads to

$$\int_{\Omega} (\partial_t u) v \, dx + \alpha \int_{\Omega} \nabla u \cdot \nabla v \, dx = 0. \quad (3)$$

For equation (3) to be well-defined, weaker regularity properties of u and v than in the classical context are sufficient. The problem can be formulated in terms of the Lebesgue space $L^2(\Omega)$ of square-integrable functions and the Sobolev space $H^1(\Omega)$ of functions in $L^2(\Omega)$ with square-integrable weak derivatives. We define the solution space

$$V := \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{top}} \cup \Gamma_{\text{src}}\}$$

and the bilinear form

$$\begin{aligned} a : H^1(\Omega) \times H^1(\Omega) &\rightarrow \mathbb{R}, \\ a(u, v) &:= \alpha \int_{\Omega} \nabla u \cdot \nabla v \, dx. \end{aligned}$$

Note that a is symmetric, continuous and V -elliptic. We denote by $(u, v)_{L^2} = \int_{\Omega} uv \, dx$ the standard inner product in $L^2(\Omega)$. Now we can state a variational formulation of problem (1):

Find $u \in \bar{u} + C^1(0, T; V)$ as the solution of

$$(\partial_t u, v)_{L^2} + a(u, v) = 0 \quad \forall v \in V, \quad (4a)$$

$$u(0) = u_0, \quad (4b)$$

where $\bar{u} \in C^1(0, T; H^1(\Omega))$ is a given function fulfilling the Dirichlet boundary conditions

$$\bar{u} = g \quad \text{on } \Gamma_{\text{left}} \times (0, T),$$

$$\bar{u} = u_{\text{top}} \quad \text{on } \Gamma_{\text{top}} \times (0, T),$$

$$\bar{u} = u_{\text{src}} \quad \text{on } \Gamma_{\text{src}} \times (0, T).$$

This variational formulation admits a unique solution u , which is called a weak solution of the heat equation.

2.1.3 Finite Element Discretization in Space

Let $T_h := \{K_1, \dots, K_N\}$ be a triangulation of Ω with N tetrahedral cells K_i ($i = 1, \dots, N$). We define the finite element space of piecewise linear functions

$$V_h := \{v \in V : v|_K \text{ is linear } (K \in T_h)\}. \quad (5)$$

V_h has the finite dimension $n := \dim(V_h)$. We give the problem formulation for a conforming finite element approximation of (4):

Find $u_h \in \bar{u}_h + C^1(0, T; V_h)$ as the solution of

$$(\partial_t u_h, v_h)_{L^2} + a(u_h, v_h) = 0 \quad \forall v_h \in V_h, \quad (6a)$$

$$(u_h(0), v_h)_{L^2} - (u_0, v_h)_{L^2} = 0 \quad \forall v_h \in V_h. \quad (6b)$$

Let $\{\varphi_1, \dots, \varphi_n\}$ be a basis of V_h . We define the ansatz

$$u_h(x, t) := \sum_{i=1}^n w_i(t) \varphi_i(x)$$

and insert it into (6a), yielding

$$\sum_{i=1}^n \dot{w}_i (\varphi_i, \varphi_j)_{L^2} + \sum_{i=1}^n w_i a(\varphi_i, \varphi_j) = 0 \quad (j = 1, \dots, n).$$

This can be written as

$$M \dot{w} + A w = 0,$$

where

$$M := ((\varphi_j, \varphi_i)_{L^2})_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$$

is the mass matrix and

$$A := (a(\varphi_j, \varphi_i))_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$$

is the stiffness matrix of the problem, and

$$w : (0, T) \rightarrow \mathbb{R}^n$$

is the vector of the time-dependent coefficients. From (6b), an initial condition for w is derived as

$$(Mw_0)_i = (u_0, \varphi_i)_{L^2} \quad (i = 1, \dots, n) \\ \iff w_0 = M^{-1}b,$$

where $b_i = (u_0, \varphi_i)_{L^2}$ ($i = 1, \dots, n$). Thus, the finite element discretization in space leads to the initial value problem

$$M\dot{w} + Aw = 0, \quad (7a)$$

$$w(0) = w_0, \quad (7b)$$

for the coefficient vector w .

2.1.4 Implicit Euler Discretization in Time

As will be discussed in Section 3.6, limitations in the used technology restrict us to use a relatively simple ODE solver for the time discretization. For the heat equation the implicit Euler scheme is suitable, due to its good stability properties [8]. Let $\{0 = t_0 < t_1 < \dots < t_m = T\}$ be a partition of the time interval with step sizes $\delta t_k = t_{k+1} - t_k$ ($k = 0, \dots, m-1$). The implicit Euler time stepping method for problem (7) is defined as

$$[M + \delta t_k A]w(t_{k+1}) = Mw(t_k) \quad (8)$$

for $k = 0, \dots, m-1$. This method is convergent and has first-order accuracy in terms of the step size δt_k .

2.2 PID Controller

A Proportional-Integral-Derivative controller (PID controller) is a form of loop feedback controller, which is widely used to control industrial processes. The controller takes as input a reference value u_{ref} , which represents the desired temperature, and the measured value $u(x_0, t)$. It uses the error $e(t) := u_{\text{ref}} - u(x_0, t)$ to compute the output signal $g(t)$. As the name PID suggests, there is a proportional part that accounts for present errors, an integral part that accounts for the accumulation of past errors, and a derivative part that predicts future errors:

$$g(t) = w_p e(t) + w_I \int_0^t e(\tau) d\tau + w_D \frac{d}{dt} e(t)$$

Here, w_p , w_I and w_D are weight parameters. By tuning these parameters the performance of the controller can be adapted to a specific process. A PID controller is widely regarded as the best controller when information of the underlying process is lacking, but the use of a PID controller does not guarantee optimal control. The tuning of the parameters can be done manually or by using a formal method such as Ziegler-Nichols or Cohen-Coon. There are also software tools available. Sometimes one or several of the parameters might have to be set to zero. For instance, derivative action is sensitive to measurement noise thus this part of the controller might have to be omitted in some situations, resulting in a PI controller. PID controllers are linear and can therefore have problems controlling non-linear systems, such as Heating, Ventilation and Air Conditioning (HVAC) systems. [1]

3. COUPLED IMPLEMENTATION

In this section, we first briefly introduce the technologies used in the present work. We then describe the coupled simulation setup, as well as its two main constituent components in more detail.

3.1 The Modelica Language

Modelica is a language for equation-based object-oriented mathematical modeling which is being developed and standardized through

an international effort in the Modelica Association [15]. The equation parts of Modelica requires a lot of the compiler developer: knowledge in compiler construction, symbolic manipulation of equations and associated mathematics, as well as knowledge of numerical mathematics. The simulation run-time system is also an important part and can be complex; various solver techniques for solving the differential equations can be applied. Modelica allows high-level concepts such as object-oriented modeling and component composition. Multi-domain modeling is also possible in Modelica with the possibility of combining model components from a variety of domains within the same application. There exist several mature and well-maintained Modelica development environments, such as Dymola, OpenModelica, MapleSim, Wolfram SystemModeler, Simulation X, and JModelica.org.

3.2 The OpenModelica Environment

OpenModelica is a modeling and simulation environment, which is developed and supported by an international consortium, the Open Source Modelica Consortium (OSMC) [4]. This effort includes an open-source implementation of a Modelica compiler, a simulator and a development environment for research, education and industrial purposes.

3.3 The HiFlow3 Finite Element Library

HiFlow3 [22] is a multi-purpose finite element software providing powerful tools for efficient and accurate solution of a wide range of problems modeled by partial differential equations (PDEs). Based on object-oriented concepts and the full capabilities of C++ the HiFlow3 project follows a modular and generic approach for building efficient parallel numerical solvers. It provides highly capable modules dealing with mesh setup, finite element spaces, degrees of freedom, linear algebra routines, numerical solvers, and output data for visualization. Parallelism - as the basis for high performance simulations on modern computing systems - is introduced at two levels: coarse-grained parallelism by means of distributed grids and distributed data structures, and fine-grained parallelism by means of platform-optimized linear algebra back-ends.

3.4 The Functional Mock-Up Interface

The Functional Mock-Up Interface (FMI) [6, 16] is a tool-independent standard to support both model exchange and co-simulation of dynamic models which can be developed with any language or tool. Models that implement the FMI can be exported as a Functional Mock-Up Unit (FMU). Such a FMU consists mainly of two parts: (1) XML file(s) describing the interface and (2) the model functionality in compiled binary or C code form. Other tools or models, which also implement the FMI, can import Functional Mock-Up Units. The initial FMI development was done in the European ITEA2 MODELISAR project [17].

3.5 Simulation Overview

Figure 2 gives an overview of the simulation setup. To create the PDE component, we reused an existing HiFlow3 application, which solves the boundary value problem for the heat equation (1). In order to import this code into Modelica, we converted it into a Dynamic Shared Object (DSO), which implements the FMI functions and interface descriptions necessary to build a Functional Mock-Up Unit. We then loaded this FMU via FMI into our Modelica model. The details of the PDE component are described in Section 3.6.

For the PID controller, we used the LimPID component from the

Modelica standard library. This was connected to the PDE component in a new Modelica model, which is described in Section 3.7.

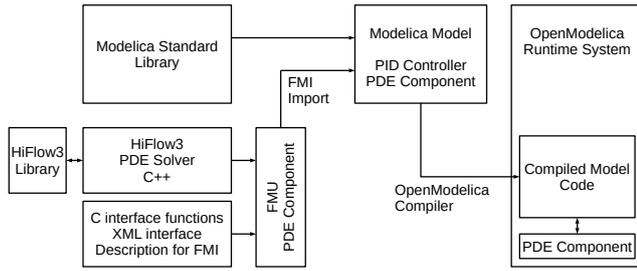


Figure 2: Overview of the creation and coupling of the simulation components.

This model was then compiled with the OpenModelica compiler and executed using the OpenModelica run-time system. By choosing the “euler” solver, the runtime system provides the time-stepping algorithm according to the implicit Euler scheme, and additionally solves the equations for the PID controller component in each time step. Figure 3 illustrates on a time axis the calls made to the compiled model code, which in turn makes calls to the PDE component.

3.6 HiFlow3-based PDE Component

The main sub-part of the PDE component is the `HeatSolver` class, which is a slightly modified version of an existing HiFlow3 application. This class uses data structures and routines from the HiFlow3 library to solve the heat problem (1) numerically. It uses a finite element discretization in space and an implicit Euler scheme in time as described in Sections 2.1.3 and 2.1.4.

Furthermore, this class provides functions for specifying the current time, the controlled temperature of the heat source, the top and bottom temperatures, and for retrieving the temperature at the measuring point. The top level routine of the `HeatSolver` class is its `run()` function, see Listing 1. This function computes the solution of the heat equation.

We prepared the triangulation T_h in a preprocessing step and stored it in a mesh file. When the `run()` function is called for the first time, it reads the mesh file and possibly refines the mesh. It also creates the data structures representing the finite element space V_h from (5), the linear algebra objects representing the system matrix, the right-hand-side vector and the solution vector. Then, the `run()` function assembles the system matrix $M + \delta t_k A$ and the right-hand-side vector $Mw(t_k)$ according to (8). It computes the solution vector $w(t_{k+1})$ for the new time step $t = t_{k+1}$ using the conjugate gradient method [18]. The solution is stored in VTK format [21] for visualization.

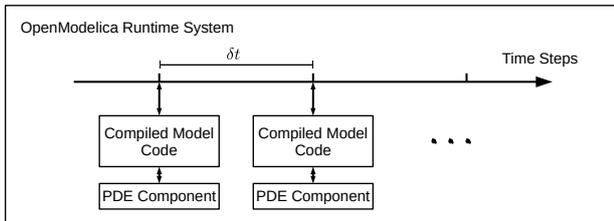


Figure 3: Interaction between the OpenModelica run-time system and the coupled model with the implicit Euler solver.

Listing 1: Run function of the `HeatSolver` class.

```
HeatSolver_run() {
    // if this is the first call
    if (first_call) {
        // read mesh file and eventually refine it
        build_initial_mesh();

        // initialize the finite element space and
        // the linear algebra data structures
        prepare_system();

        first_call = false;
    }

    // compute the system matrix and
    // the right-hand-side vector
    assemble_system();

    // solve the linear system
    solve_system();

    // visualize the solution
    visualize();

    // keep solution and time in memory
    CopyFrom(prev_solution, old_solution);
}
```

It is important to note that the solution vector and the current time are kept in memory inside the PDE component, since this data is required for computing the solution at the next time step. Although it has been planned for a future version, at present the FMI standard does not directly support arrays, which prevents passing the solution vector back and forth between the PDE component and the Modelica environment as a parameter [16]. Although this use of mutable internal state in the PDE component might be preferable from a performance point of view, it has the drawback of making the function calls referentially opaque: two calls with the same parameters can yield different results, depending on the current internal state. This imposes a strong restriction on the solver used, which must assure that the sequence of time values for which the function `run()` is called is non-decreasing. For this reason, we only tested the method with the simple implicit Euler solver, and verified that the calls were indeed performed in this way. For more complicated solvers, such as DASSL, this requirement is no longer satisfied.

The entry point of the PDE component is the `PDE_component()` function, see Listing 2. This function is called within the Modelica simulation loop. When it is called for the first time, it creates a `HeatSolver` object. It sets the input values for the heat source, the temperatures at the top and bottom boundary, and the current time. Then the `run()` function of the `HeatSolver` is called to compute the solution of the heat equation. Finally, the run counter is incremented and the measurement value is returned.

3.7 Modelica Model

Our Modelica model is shown schematically in Figure 4. It contains the PDE component, the PID controller, and four source components. Two of the sources represent the environmental influences, which are given by sinusoidal functions.

Listing 2: Main simulation routine of the PDE component.

```

PDE_component (
double in_Controlled_Temp,
double in_Top_Bdy_Temp,
double in_Bottom_Bdy_Temp,
double in_Time)
{
  // create HeatSolver object if this
  // is the first call
  if (run_counter == 0)
    heat_solver = new HeatSolver();

  // set input values
  heat_solver->set_time(in_Time);
  heat_solver->set_g(in_Controlled_Temp);
  heat_solver->set_top_temp(in_Top_Bdy_Temp);
  heat_solver->set_bottom_temp(
    in_Bottom_Bdy_Temp);

  // run the HeatSolver
  heat_solver->run();

  // increment the run counter
  run_counter++;

  // return the measurement value
  return heat_solver->get_u();
}

```

They are connected to the PDE component to give the top and left boundary temperatures u_{top} and u_{left} in Equations (1d) and (1e), respectively. One source is connected to the PID controller and gives the reference value u_{ref} for the desired temperature. The fourth source is connected to the dummy state variable of the PDE component. The dummy state variable and its derivative are in the model due to the fact that the OpenModelica implementation of FMI 1.0 import does not allow for an empty state variable vector. There is however nothing in the FMI 1.0 model exchange specification that disallows this. Additionally, the measurement value of the PDE component is connected to the input of the PID controller, and the output signal of the PID controller is connected to the heat source input of the PDE component.

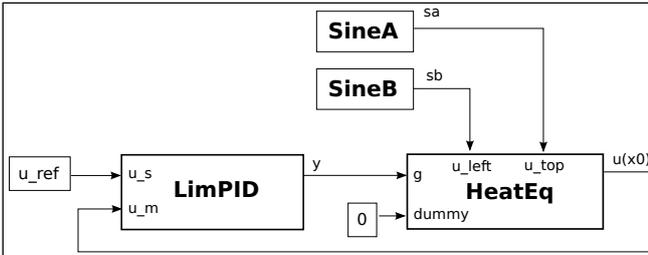


Figure 4: Schematic view of the coupled Modelica model used in the simulation.

The internal constant parameters of the components are summarized in Table 1.

4. RESULTS

We carried out a numerical experiment with the following setting. We took a simulation time of $T = 1500$ seconds and a time step of $\delta t = 1$. We set the initial temperature $u_0 = 0$ everywhere in the computational domain Ω , and we specified the desired temperature as $u_{\text{ref}} = 3$. On the upper part of the boundary Γ_{top} and on the left

Component Parameter	Value
<i>LimPID</i>	
proportional gain w_P	0.05
integral gain w_I	0.2
derivative gain w_D	0.0
<i>HeatEquationFMU</i>	
thermal diffusivity α	$1.11 \cdot 10^{-4} \text{ m}^2 \text{ s}^{-1}$
<i>SineA</i>	
amplitude	$0.5 \text{ }^\circ\text{C}$
vertical offset	$3.5 \text{ }^\circ\text{C}$
start time	150.0 s
frequency	0.001 s^{-1}
<i>SineB</i>	
amplitude	$6.0 \text{ }^\circ\text{C}$
vertical offset	$3.0 \text{ }^\circ\text{C}$
start time	350.0 s
frequency	0.002 s^{-1}

Table 1: Internal parameters of the components in the simulation model.

part of the boundary Γ_{left} we modeled environmental influences by the functions

$$u_{\text{top}}(t) = \begin{cases} 3.5 & \text{if } t < 150, \\ 3.5 + 0.5 \sin\left(\frac{(t-150)\pi}{500}\right) & \text{if } t \geq 150, \end{cases}$$

and

$$u_{\text{left}}(t) = \begin{cases} 3 & \text{if } t < 350, \\ 3 + 6 \sin\left(\frac{(t-350)\pi}{250}\right) & \text{if } t \geq 350, \end{cases}$$

which are shown in Figure 5.

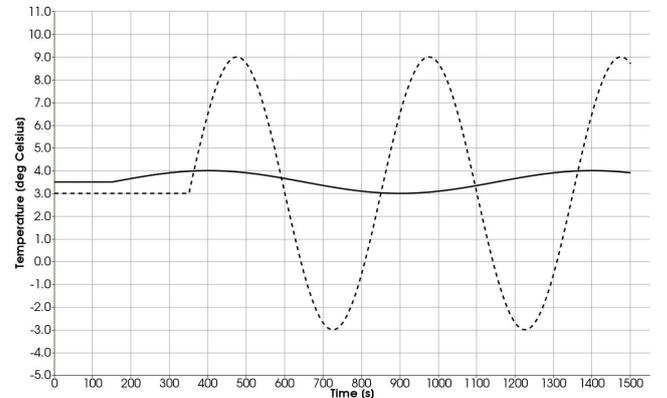


Figure 5: Environmental temperature prescribed on the boundary parts Γ_{left} and Γ_{top} . Dashed: $u_{\text{left}}(t)$, solid: $u_{\text{top}}(t)$.

For comparison, we first performed a simulation run with a constant, uncontrolled heat source $g \equiv 3$ on the lower boundary Γ_{src} . Figure 6 shows that the temperature at the point of measurement deviates from the desired temperature $u_{\text{ref}} = 3$ due to the environmental influences.

The results of our simulation run with a controlled heat source

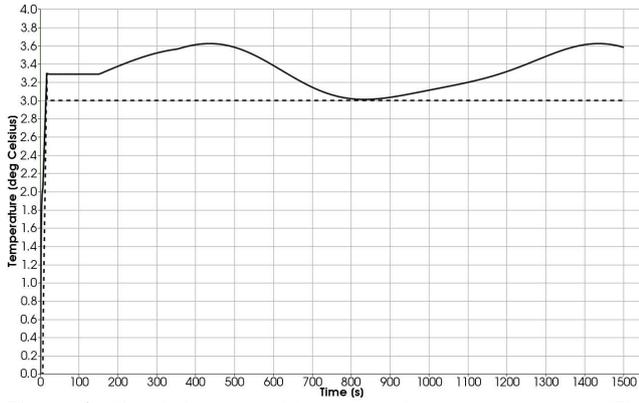


Figure 6: Simulation run with constant heat source $g \equiv 3$. The temperature $u(x_0, t)$ at the point of measurement deviates from the desired value. Dashed: g , solid: $u(x_0, t)$.

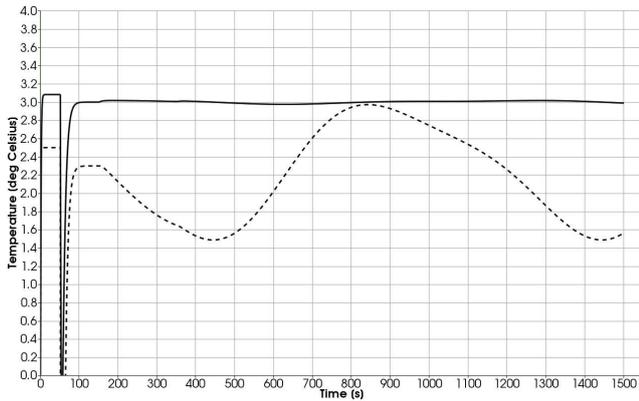


Figure 7: Simulation run with controlled heat source $g = g(t)$. The temperature $u(x_0, t)$ at the point of measurement accurately follows the desired value $u_{ref} = 3$. Dashed: $g(t)$, solid: $u(x_0, t)$.

$g = g(t)$ are shown in Figure 7. At the beginning, we fixed the heat source at $g = 2.5$ to let the temperature distribution in the copper bar evolve from the initial state to an equilibrium, at which the measured temperature is slightly higher than desired. The PID controller was switched on at $t = 50$ to take control over the heat source. The curves show that the controller first cools the bottom to bring the temperature at the point of measurement down to the target value. During the rest of the simulation, the PID controller reacts to the environmental influences and adjusts the heat source dynamically over time, so that the temperature accurately follows the desired state. Figures 8-10 illustrate the temperature distribution in the copper bar.

5. DISCUSSION

The numerical results for the example presented in the previous section show that our method of integrating the PDE solver from HiFlow3 into a Modelica simulation functions correctly. The realization of this particular scenario serves as an illustration of how one can integrate other, more complicated, PDE models into the complex dynamical simulations for which Modelica is especially suited.

The coding and maintenance effort for importing an existing PDE model implemented in HiFlow3 with the method presented here is minimal: in essence only a set of wrapper functions dealing with

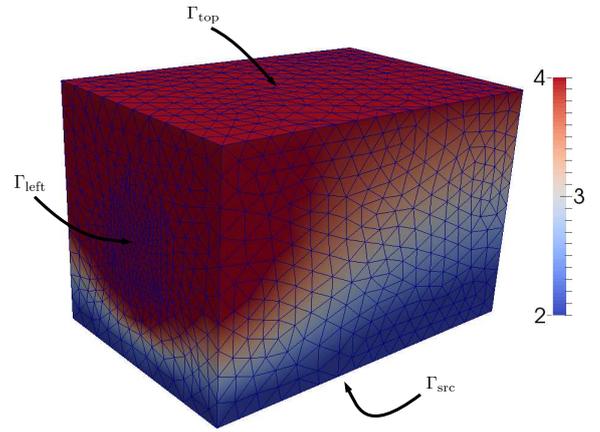


Figure 8: Computational domain of the copper bar with triangulation. The colors indicate the temperature distribution on the surface at time $t = 440$.

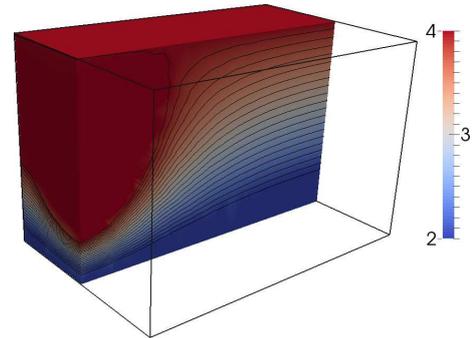


Figure 9: Sectional view with isothermal lines at time $t = 440$.

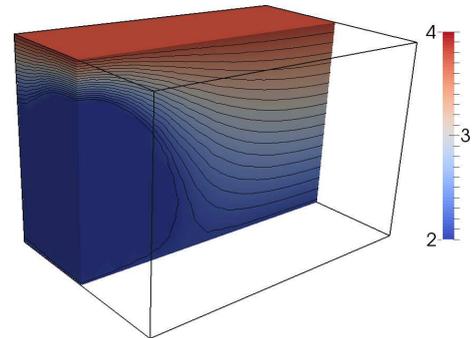


Figure 10: Sectional view with isothermal lines at time $t = 1250$.

input and output of parameters and state variables is all that is required. The fact that HiFlow3 is free and open source software simplifies the process greatly, since it makes it possible to adapt and recompile the code. This is significant, since the FMI model import requires the component to be available either as C source code or as a dynamic shared object, which is loaded at run-time.

Compared to the efforts aiming at extending the Modelica language with support for PDEs, we are working at a different level of abstraction, namely that of software components. The advantage of this is the ability to make use of the large wealth of existing implementations of solvers for various models, in the present case the multi-purpose HiFlow3 library. Extending the Modelica language would also make it considerably more complex, since problems for PDEs generally require descriptions of the geometry and

the conditions applied to the different parts of the boundary. Furthermore, using this information to automatically generate a discretization and a solution algorithm would require a sophisticated classification of the problem, since different types of PDEs often require different numerical methods. A drawback of working at the software level is that the mathematical description of the problem is not directly visible, as it would be if it was part of the language.

In contrast to the use of “co-simulation”, in this work we have chosen to import the PDE component into the OpenModelica environment, and to make use of one of its solvers. The main benefit is again simplicity: very few changes were required to the PDE component itself, and it was possible to maximize the reuse of existing software. On the other hand, co-simulation, where each sub-model has its own independent solver, which is executed independently of the others, also has its advantages. In particular, specialized, highly efficient solution algorithms can be applied to each part of the model, and it is possible to execute the various components in parallel. We are considering to extend the present work to make it usable in a co-simulation setting. Furthermore, we want to investigate the parallelization of the simulation both within and between components.

6. CONCLUSION

In this paper we have investigated a method of incorporating PDEs in the context of a Modelica model, by using FMI to import a PDE solver from the finite element library HiFlow3. Numerical results obtained using a simple coupled model controlling the heat equation using a PID controller demonstrate that this method works in practice. The main advantages of this type of coupling include its simplicity and the possibility to reuse existing efficient and already validated software. This approach allows to make use of more complex PDE models including high-performance, parallel computations. It has the potential of greatly simplifying the development of large-scale coupled simulations. In this case, however, an extension of the method presented here to support co-simulation might be necessary.

7. ACKNOWLEDGMENTS

Funded by the ITEA2 European Union MODRIO Project, by SSF in the EDOP project, and by the Swedish National Graduate School of Computer Science (CUGS).

8. REFERENCES

- [1] Karl Johan Åström and Tore Hägglund. *PID Controllers: Theory, Design, and Tuning. 2nd edition*. Instrument Society of American, 1995.
- [2] Dietrich Braess. *Finite Elemente*. Springer, 2007.
- [3] Farid Dshabarow. *Support for Dymola in the Modeling and Simulation of Physical Systems with Distributed Parameters*. Master thesis, Department of Computer Science, Institute of Computational Science, ETH Zürich, 2007.
- [4] The Open-Source OpenModelica Development Environment. <http://www.openmodelica.org>.
- [5] Alexandre Ern and Jean-Luc Guermond. *Theory and Practice of Finite Elements*. Springer, 2010.
- [6] The Functional Mockup Interface (FMI). <https://www.fmi-standard.org/>.
- [7] Peter Fritzson. *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*. Wiley-IEEE Press, 2004.
- [8] Ernst Hairer, Syvert P. Norsett, and Gerhard Wanner. *Solving Ordinary Differential Equations*. Springer, 2008.
- [9] Harro Heuser. *Lehrbuch der Analysis*. Teubner, 2002.
- [10] Samir Khan. Discretizing pdes for maplesim. adept scientific plc. 2012.
- [11] Zhihua Li, Ling Zheng, and Huili Zhang. Solving pde models in modelica. *2008 International Symposium on Information Science and Engineering. ISISE 08*, 1:53–57, 2008.
- [12] Manuel Ljubijankić and Christoph Nytsch-Geusen. 3d/1d co-simulation von raumlufströmungen und einer luftheizung am beispiel eines thermischen modellhauses. *Fourth German-Austrian IBPSA Conference, BauSIM 2012, Berlin University of the Arts*, 2012.
- [13] Manuel Ljubijankić, Christoph Nytsch-Geusen, Jörg Rädler, and Martin Löffler. Numerical coupling of modelica and cfd for building energy supply systems. *8th International Modelica Conference 2011*, 2011.
- [14] Maple and MapleSim by MapleSoft. <http://www.maplesoft.com/products>.
- [15] Modelica and the Modelica Association. <http://www.modelica.org>.
- [16] Modelica Association. *Functional Mock-up Interface for Model Exchange and Co-Simulation*, v. 2.0 beta 4 edition, 2012. <https://www.fmi-standard.org/downloads> [Accessed 2013-08-06].
- [17] The ITEA2 MODELISAR Project. <http://www.itea2.org/project/index/view/?project=217>.
- [18] Youssef Saad. *Iterative Methods for Sparse Linear Systems*. 2000.
- [19] Levon Saldamli. *PDEModelica - A High-Level Language for Modeling with Partial Differential Equations*. Dissertation 1016, Department of Computer and Information Science, Linköping University, 2006. <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-7281>.
- [20] Levon Saldamli, Bernhard Bachmann, Hans-Jürg Wiesmann, and Peter Fritzson. A framework for describing and solving pde models in modelica. *In Proceedings of the 4th International Modelica Conference, Hamburg, Germany, March 7-8, 2005*.
- [21] Will Schroeder, Ken Martin, and Bill Lorenson. *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*. Kitware, Inc., 2006.
- [22] The HiFlow3 Multi-Purpose Finite Element Software. <http://www.hiflow3.org/>.

A multi agent system for simulation of battalion operations

Solveig Bruvoll and Rikke Amilde Løvliid

FFI - Norwegian Defence Research Establishment
PO Box 25, NO-2027 Kjeller, Norway

Abstract

We have created a system for simulation of battalion operations executed by autonomous battle command agents. Simulations of battalion operations can be useful for example in operational planning and for training of commanders and staff. In this paper we discuss the resolution of our model and its suitability for these applications with respect to three aspects. First, we describe our representation of the military hierarchy, where tactical behavior is modeled only for platoon level and above. Second, we discuss the level of detail of our behavior model, which is based on Context-Based Reasoning (CxBR). Within this paradigm the agents' situational awareness and action rules are divided into contexts to limit the number of possible perceptions and actions the agents need to consider in any situation. We exploit the possibilities in CxBR to gradually extend the detail level of the model with more complex behavior. Finally, we discuss which behavior should be deterministic and which should be stochastic, and whether this choice depends on the application.

1 Introduction

The use of simulators, computer generated forces (CGFs) and serious games for military training and operational planning is becoming increasingly important [1]. Simulation based training can be cost effective and time efficient. An important aspect of simulation based training is the need for realistic computer controlled entities. The simulation systems currently used for training of command and staff consist of humans operating CGFs, which model and simulate the combat entities and systems in the simulation. The operators receive a high level order, for example a battalion order with company tasks, and transform it into lower level tasks for subordinate units, like platoons. Finally, they manually input these tasks into a simulation system.

Our aim is a simulation system that is capable of simulating the execution of a higher level operational order autonomously and to respond dynamically to

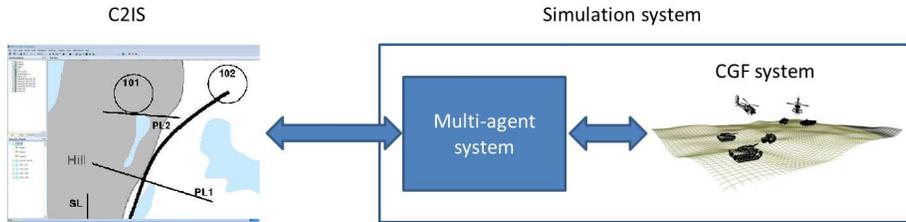


Figure 1: System architecture

occurring events. This would reduce the number of operators needed during training and make it possible to utilize simulation for other task like course of action analysis during operational planning. Current CGFs do not have this capability as they are only able to perform simple task like "move along route" and "move into formation". In order to simulate the execution of higher level orders, the simulation system will have to understand and plan how to execute higher level commands, like "seize area x" or "support unit y by fire", and be able to react to unplanned events according to doctrine.

We have developed a multi agent system (MAS) that models the tactical behavior of a battalion, and have constructed a simulation system consisting of our MAS and the CGF system VR-Forces [2]. This simulation system receives orders from a command and control information system (C2IS), as illustrated in figure 1. C2ISs are used by military officers for issuing commands to subordinate units, and ideally commanding virtual forces should be carried out equivalently. The MAS plans how the received order is to be executed and sends for example rules of engagement, choice of formation, and movement commands to VR-Forces, which simulates the entities' movement through the terrain and determines for example when enemies are observed, when entities are firing, and damage assessment. In this paper we present our battalion tactical behavior model, which plans and executes an offensive military operation and discuss whether the resolution of the model is suitable for our application. A more detailed description of the MAS is found in [3].

We have used the example scenario described in section 2 when developing the MAS. This first version of the behavior model is designed to handle the tasks in the example order and to execute the simulation, handling unplanned events autonomously. We present the behavior model and evaluate it with respect to possibilities for further extension of the model. In particular we focus on three elements of the model: The hierarchy resolution, the behavior model, and the balance between stochastic and deterministic elements. The hierarchy

resolution determines to which level a detailed behavior model is necessary, while the behavior model is artificial intelligence model included in the MAS. The balance between stochastic and deterministic elements influences the degree of predictability in the simulations.

The MAS models the military hierarchy from battalion to platoons, while the execution of platoon tasks and single units is left to the CGF. In section 3 we discuss whether this is suitable for the behavior model in the hierarchy, based on our application, or whether we should have left less or more of the behavior to the CGF system.

The tactical behavior is modeled within the artificial intelligence framework context-based reasoning (CxBR), a modeling paradigm specifically designed for representing human tactical behavior [4, 5, 6]. In section 4 we explain how the tactical behavior model is realized with CxBR, and how CxBR can be used to gradually expand the model to support more ordered tasks and increase the complexity of the included tasks as needed.

Our simulation system consists of stochastic and deterministic elements. Section 5 discusses whether the balance between these elements is suitable for the application and presents possible alternatives.

2 Example scenario

An example scenario was used as a guide to what functionality we needed in order to handle a simple military operation and to limit the behaviours that had to be modelled for a first version of the simulation system.

The example scenario is part of a larger offensive operation, and the goal of the entire operation is to attack an enemy that is situated further to the north-east of the map shown in figure 2. The plan is to approach the enemy along the road that exits the map near the upper right corner. There is an enemy vanguard at area 102, and this area must be seized before the friendly forces can continue towards the main enemy position. Our example scenario considers this first part of the operation, to seize area 102.

The operation is to be carried out by one battalion consisting of the companies RecceCoy1, MechInfCoy3 and MechInfCoy 4. The military hierarchy of the battalion, with its companies and platoons, is shown in figure 3. Each platoon consists of four entities, which are not included in the figure.

The battalion operation order with tasks for each of these companies is shown in table 1. In cooperation with Subject Matter Experts (SMEs) we determined how the tasks in the order should be carried out, and how the agents should handle possible events not described in the order. The following outlines the desired behaviour if there are no unplanned events, e.g. no enemies outside the target areas 101 and 102.

Table 1: The synchronization matrix.

Company	Task	
RecceCoy1	Reconnoiter axis from SL to 102	Keep 102 under surveillance
MechInfCoy3		Seize objective area 101
MechInfCoy4		Support by fire MechInfCoy4 (towards 102)
		Seize objective area 102

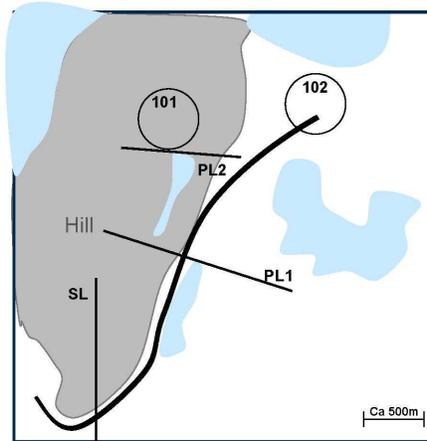


Figure 2: The map with defined areas and phase-lines from the OPORD.

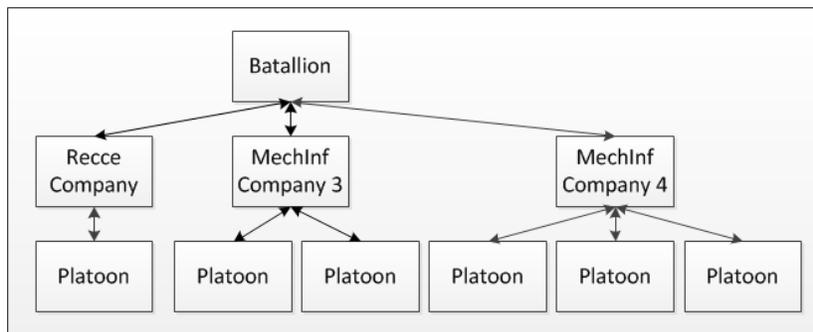


Figure 3: The military hierarchy.

RecceCoy1 RecceCoy1 reconnoiters the axis from the start line (SL) to the objective area 102 while moving along the hillside to area 101. From the hill, the company has a good overview of the road and the surrounding area. The company looks for enemy activities in the area and reports to the other companies. RecceCoy1 also checks the terrain for accessibility, and the findings can be used by other companies for route planning. After passing area 101, it creates an observation post at a suitable place for surveillance of 102.

MechInfCoy3 MechInfCoy3 follows the same route as RecceCoy1, knowing that the route has been checked for enemies and obstacles. When MechInfCoy3 crosses phase line PL2 it prepares to attack potential enemies positioned in the objective area 101. After seizing area 101, it secures the area and moves into position for overlooking the objective area 102, waiting for a signal to support MechInfCoy4 by fire during their attack on area 102.

MechInfCoy4 MechInfCoy4 moves along the road at high speed to PL1. It has moved through the canalizing terrain along the lake and spreads to a wider formation. When area 102 is within weapon range, the attack begins, with support from MechInfCoy3. All platoons move into the objective area 102 when they cannot affect it from outside.

Unexpected enemies If unexpected entities are detected or if an entity is fired upon, the first reaction is to pull back and regroup. Then the behaviour depends on the observation; small enemy unit, large enemy unit or an unidentified unit. If the enemy is sufficiently small compared to the observing platoon, the platoon will attack it. If the enemy is too large for the platoon, but can be taken out by the company, the rest of the company will assist in the attack. If the observed unit is unidentified, the observing entity should try to obtain more information about the unit, but this is out of scope for the current version.

3 A hierarchy of battle command agents

The example order is for a battalion operation, and the structure of the battalion, with its companies and platoons, is shown in figure 3. Our simulation system is developed mainly for operational planning and testing of orders, and the resolution of the hierarchy in the MAS is based on this application. We have chosen to let the agents in the MAS represent commanders and their staff for the battalion, companies and platoons, while the CGF system simulates the combat units of the platoons and their entities. This resolution enables the companies and platoons to be tasked independently in the MAS, while the entities in each platoon are given identical tasks in the CGF.

In the MAS the hierarchy of intelligent agents decomposes the order from company level tasks to low level CGF commands, which are sent to the CGF in a standardized format [7]. The MAS consist of one agent for the battalion, one for each company in the battalion and one for each platoon. The agents represent the commanders and staff of these military units, and model the planning and decision making done by these leaders. The agents in the MAS are ordered in the same hierarchy as the military troops they represent, which makes the design clear and simple. This also makes the agent system and decision making easy to understand for military experts, and simulation of the real chain of command also prepares the system to be used for other task, like studies of communication in the hierarchy.

The battalion agent receives an order with company tasks from the C2IS and makes sure the companies receive their tasks as scheduled in the order. The company agents plan their tasks and command their platoons. The platoon agents break their tasks further down into simple commands for the platoon aggregates in the CGF system. Units below platoons (i.e. squads, vehicles, soldiers) are not represented in the MAS, and the decomposition of platoon tasks to tasks for single entities is handled by the CGF system, which means all units in the platoon will get equivalent tasks.

The CGF platoons report back to the platoon agents about observed enemies, their task status, position, damage, fuel level, etc. This makes it possible for the agents to monitor the simulation and react to occurring events. For example if an unexpected enemy is observed, the agents will change their plans according to the situation, not blindly follow through with the received plan. The agents also send reports to their superiors, and they update the perceived truth shared by all agents. Ground truth and perceived truth of both own and enemy position are reported back to the C2IS ¹.

We have chosen to let the platoons be the lowest represented level. Another alternative was to only include battalion and company agents in the agent hierarchy, and to let the CGF handle companies and platoons. This alternative would require more sophisticated and complex behavior in the CGF, and, depending on the CGF that is used, this may be a good alternative. We have used VR-Forces, which we consider to not have complex enough built-in behavior for company tasks. We want the platoons to be able to perform tasks independently at a higher level than what is possible automatically in VR-Forces, without extending the built-in models. For example when handling unexpected enemies, we want a platoon to destroy it alone if it is sufficiently small, or together with its company if it is larger. For this type of behavior we considered an agent hierarchy which included platoons as the best solution.

Since we have not included behavior in which the entities in a platoon performs different tasks, we decided not to extend the agent hierarchy further to single

¹Ground truth of enemies can be reported because this is a simulation where everything is available. One can then decide in the C2IS what should be visible for different users.

vehicles. These are handled by the CGF, as we concluded that the built-in models in VR-Forces were sufficient for the low level tasks we give to the platoons at this point. If the model is to be used for training of soldiers, the hierarchy would most likely need to be extended to include lower level agents and to allow more complex behavior also at lower levels.

4 Agent behavior modeling with CxBR

The behavior of the agents is modeled with Context-Based Reasoning (CxBR), which is a reasoning paradigm for representation of tactical behavior in agents [4, 5, 6]. The motivation behind CxBR is the realization that people only use a fraction of their knowledge at any given time. The idea is to divide the knowledge into contexts in order to limit the number of possibilities for the action selection process. For example, an agent representing a military platoon will require a different set of capabilities and knowledge when it is performing an attack versus when moving along a road.

The contexts are organized in a context hierarchy consisting of a *mission context*, *major* and *minor contexts*. The mission context is a purely descriptive context, and does not control the agent. It contains information defining the current overall mission of the agent. Such information includes the objective and a plan for reaching it together with parameters like objective area, phase-line, route etc., and a context map, where all possible transitions between major contexts are defined.

Major contexts are the primary element of agent control and represent the major situations an agent can face. There is only one major context in control of the agent at any time, called the active context. A major context basically contains three kinds of knowledge: *action knowledge*, *transition knowledge* and *declarative knowledge*. Action knowledge is knowledge about how the agent should behave in the context. We have implemented this as action rules. As the situation and environment for an agent change, the agent will transition accordingly between contexts. Knowledge of when to switch to another context is collected in the transition knowledge. This knowledge can be contained in transition rules, with criteria for when the agent makes the transitions defined in a *context map*. Declarative knowledge includes context specific parameters and possible minor contexts.

We have two main reasons for choosing CxBR to model the behavior of the battle command agents. First, the modeling paradigm is very intuitive, and there is a natural connection between contexts and military tasks, making it possible to build behavior models that are easy to understand for military experts. Second, the modularity of the CxBR paradigm facilitates a gradual development of the behavior model, starting with simplified behavior and gradually increasing the complexity as it is needed. In this article we will focus on the latter.

To illustrate how to gradually develop more and more complex behavior, we will use the task *Seize*, which is one of the three task (*Reconnoitre*, *Seize* and *Support by fire*) from the order that was expressed as a synchronization matrix in table 1. The modeling of the remaining two task will not be reviewed in this paper, but is documented in a technical report [3].

The military task *Seize* is defined as "a tactical mission task that involves taking possession of a designated area by using overwhelming force" [8]. In our CxBR model all task in the order are mapped to mission contexts, meaning *Seize* is a mission context with an objective area (OA) as argument. Figure 4 illustrates how more complex solutions of this mission context can be realized by expanding the context map step by step.

The simplest way to realize seize would be to use only one context, *Attack*, which has only one action rule "when entering the context, move towards the OA and fire at will" (Figure 4a) . The next step might be to differentiate between moving towards the OA and attacking as this typically will involve at least different preferred formations and possible different rules of engagement. One could realize this extension by adding more action rules to the *Attack* context, but it might be better to add an additional major context *Move* as this behavior probably will be useful in other situations than just when going to perform an attack (Figure 4b).

In the next step (Figure 4c) we want to incorporate proper reactions to enemy encounter during movement towards the OA. Major context *Regroup* is added, and when entering this context the agent will stop and evaluate the strength of the encountered enemy in order to decide whether to attack or ask its superior for help. If the enemy is too large, the agent will move back and wait for assistance. A major context *Wait* is added to function as a more general default context in Figure 4d.

In Figure 4e we illustrate how the major contexts *Move* and *Attack* can be split in two. Major context *Move* represents fast movement as before, where as *Move Cautiously* is used when there is higher probability for encountering enemies and when approaching the OA. The original major context *Attack* does now represent a preplanned attack, where as *Hasty Attack* implements a more spontaneous engagement and is used for unexpected enemy encounters.

Whether to increase complexity in the behavior model by adding action knowledge to an existing context or adding a new context depends on what kind of behavior you want to add. If it fits as a part of one or more existing contexts, you might want to include it in that context or in a sub-context which shared is shared among these contexts. If it on the other hand is a new, distinct military capability, one would rather make a new context. Also it is important to keep in mind that the purpose of splitting the behavior into context is to limit the number of possibilities for the action selection process. Thus, one would not want one context to be too large and complex, if that happens, one might consider splitting it into two or more different context. On the other hand, making

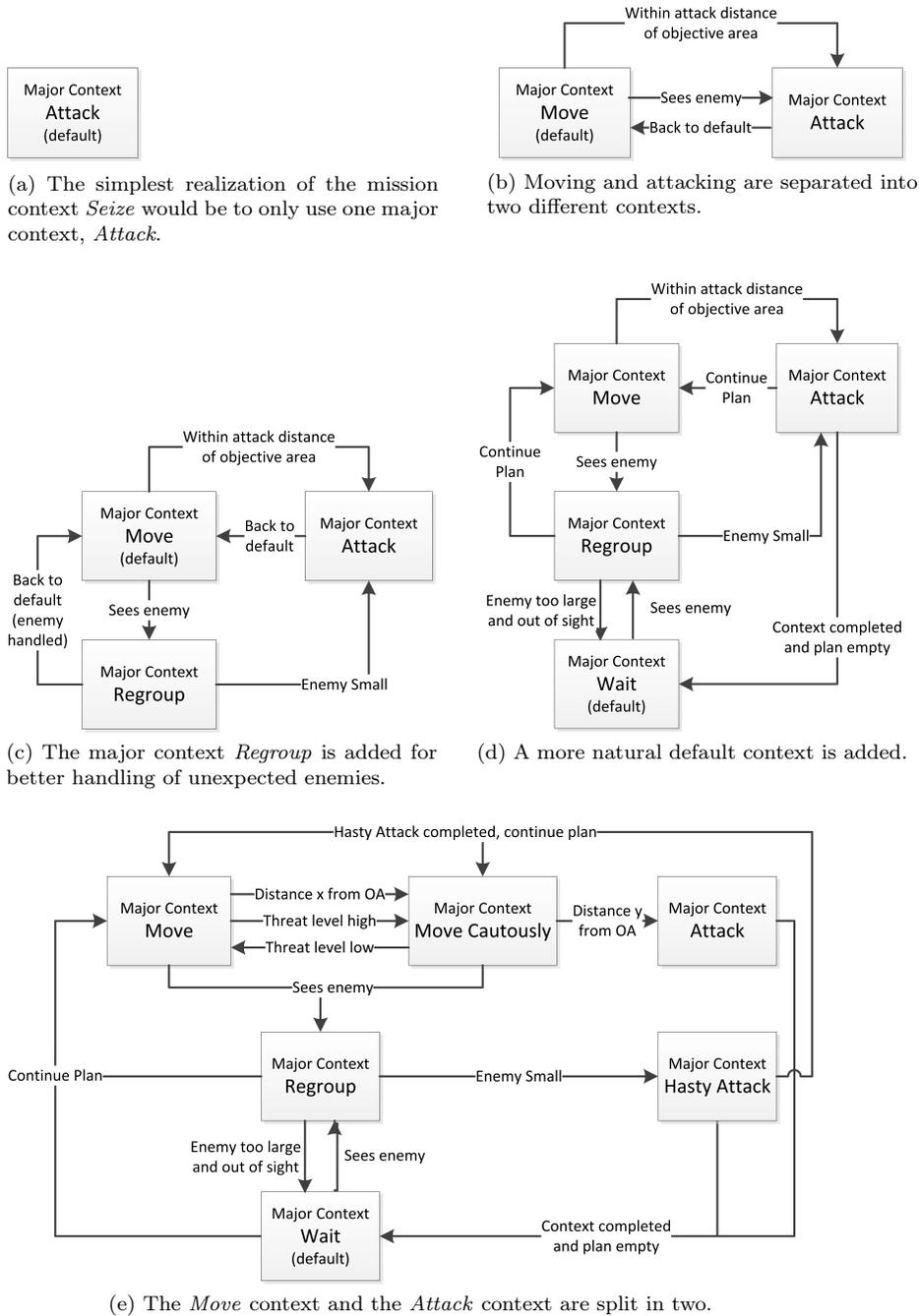


Figure 4: Gradually developing a context map for mission context *Seize*.

more contexts only means that the decision process is moved from the action selection process inside each contexts to choosing which transition rules to fire. The challenge is to balance these two levels of decision making. In both the *Move* and *Attack* examples, one might want to start expanding the contexts with new action rules and then split the contexts when the number of rules becomes too large, meaning when it becomes more efficient splitting the decision making into two parts by first having to choose which *Move/Attack* context and then fire the appropriate action rules within the chosen context.

To summarize the CxBR modeling paradigm makes it possible to gradually add behavior as it is needed. Currently the model is only able to handle the tasks included in the example order, but the modularity should make it easy to both extend the model to cover other military tasks as well as more complex solutions to current tasks. How detailed the behavior needs to be will of course depend on the application. For a coarse visualization of the execution of an order, one might not need to see formation changes or specific movement patterns. However, if the simulation is to be used for more detailed course of action analysis or to play parts of own or enemy forces during training, more detailed behavior will be required. How detailed the behavior needs to be should be considered together with the depth of the agent hierarchy.

5 Stochastic or deterministic simulation

A deterministic simulation produces identical results every time as long as the parameters are unchanged. Stochastic simulations depend on random processes which influence the simulations. Whether a deterministic outcome is desired or not depends on the use of the simulation. In our simulation system we have chosen a deterministic model for the battle command agents, while the CGF system uses stochastic models for example for determination of inflicted damage. Our simulation therefore contains both stochastic and deterministic elements. Since the battle command agents are deterministic, a received order will be planned in the same way in every run. But since the execution of the order in the CGF depends on stochastic processes, the simulation itself may proceed in a different way in each run. In some runs an enemy may be destroyed rather early, while it has time to take cover in other runs. Such differences will influence further events, and may result in different needs for replanning during the execution of the original identical order. Therefore a deterministic battle command model may result in various outcomes when combined with a stochastic CGF agent.

If the simulation is to be used to model the enemies in training of soldiers, a completely deterministic enemy will soon be easy to defeat. In such applications a stochastic element in planning would be advantageous, in order to make the enemy less predictable. For a simulation used to test strengths and weaknesses of a plan, there is not the same need for variation in planning, although variation

in the execution of the simulation may better detect weaknesses in the plan. Another possibility is to remove the stochastic element in the CGF agent. This results in a fully deterministic simulation, which will always produce the same outcome. We would then lose the possibility to rerun the scenario in order to get an impression of the robustness of the plan.

6 Conclusion

We have constructed a simulation system capable of receiving a battalion order and simulating the operation executed by autonomous battle command agents. This simulation system is developed mainly for operational planning and testing of orders. In this paper we have presented the simulation system and evaluated it with respect to hierarchy resolution, behavior model and predictability of the simulation.

The simulation system consists of a Multi-Agent System (MAS) and a Computer Generated Forces (CGF) system. The agents in the MAS are ordered in a hierarchy analogous to the military hierarchy of a battalion. We have chosen to include battle command agents for battalion, companies and platoons in the MAS, while the CGF executes tasks at platoon level and lower. This enables the platoon agents to be tasked individually and the company agents to synchronize cooperation between platoons, which we consider a suitable resolution for our application.

The behavior of the agents is modeled with CxBR, with contexts reflecting the military tasks and orders. In the current model only the tasks relevant to the example order are included. This limits the possible orders the system can execute, but the quality of the behavior for these contexts is sufficient for execution of basic plans. However, the modularity of the CxBR modeling paradigm should make it easy to both extend the model to cover other military task as well as more complex solutions to current tasks.

The processes in the MAS, including the planning of the order, is deterministic, while the events in the CGF are partly stochastic. Since the MAS reacts to the events in the CGF, several runs will not necessarily produce the same final result. This will depend on for example when enemies are observed and destroyed in the CGF.

It is not straightforward to validate the behavior of our model, as it models human tactical behavior in battle. Not only shall the simulation system choose tactics that preferably leads to victory, but the behavior should also appear plausible. We have evaluated the model through demonstrations for military experts, which have offered their opinions of the credibility of the behavior. The feedback was that the behavior of the basic tasks in the current model was sufficiently plausible, but that more variation in behavior is needed in later versions. In the future we will aim to validate the behavior model by comparing

the autonomous behavior with the behavior of CGFs controlled by military experts, or maybe even compare the simulated behavior with data from live exercises or operations.

The simulation system gives promising results in its current form, but is rather basic at the moment. We plan to extend the simulation system further, both to handle a broader set of scenarios, and to improve the behavior model. The hierarchy resolution seems suitable also for an extended version of the model, but we will consider whether we need more advanced behavior at a lower level in the hierarchy when the behavior model is extended. The behavior model needs to include more tasks, which we plan to include gradually, as the need arises. We also intend to increase the complexity of the existing tasks. As the behavior model is extended, we plan to include a stochastic element in the planning process, such that a task is not always planned identically. This would simulate the fact that military officers receiving the same order will not produce identical plans for execution of the tasks.

References

- [1] J. Fletcher, “Education and training technology in the military,” *Science*, vol. 323, pp. 72–75, 2009.
- [2] VT MÅK VR-Forces. [Online]. Available: www.mak.com/products/simulate/computer-generated-forces.html
- [3] R. A. Løvliid *et al.*, “Modelling battle command with context-based reasoning,” Forsvarets forskningsinstitutt, FFI-rapport 2013/00861, 2013.
- [4] A. Gallagher, A. J. Gonzalez, and R. DeMara, “Modeling platform behaviors under degraded states using context-based reasoning,” in *Proceedings of the 2000 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC-2000)*, 2000.
- [5] A. J. Gonzalez and R. Ahlers, “Context-based representation of intelligent behavior in training simulations,” *Transactions of the Society for Computer Simulation International*, vol. 15, no. 4, pp. 153–166, 1998.
- [6] A. J. Gonzalez, B. S. Stensrud, and G. Barret, “Formalizing context-based reasoning: A modeling paradigm for representing tactical human behavior,” *International Journal of Intelligent Systems*, vol. 23, pp. 822–847, 2008.
- [7] A. Alstad *et al.*, “Low-level battle management language,” in *Proceedings of the 2013 Spring Simulation Interoperability Workshop*, no. 13S-SIW-032, 2013.
- [8] Headquarters Department of the Army , “Field manual no. 3-90, tactics,” 2001. [Online]. Available: www.globalsecurity.org/military/library/policy/army/fm/3-90/.

Intelligent Methods in Modelling and Simulation of Complex Systems

Esko K. Juuso*

* *Control Engineering Laboratory
Department of Process and Environmental Engineering
P.O.Box 4300, 90014 University of Oulu, Finland (email:
Esko.Juuso@oulu.fi)*

Abstract: Data mining with a multitude of methodologies is a good basis for the integration of intelligent systems. Small, specialised systems have a large number of feasible solutions, but developing truly adaptive, and still understandable, systems for highly complex systems require domain expertise and more compact approaches at the basic level. This paper focuses on the integration of methodologies in the smart adaptive applications. Statistical methods and artificial neural networks form a good basis for the data-driven analysis of interactions and fuzzy logic introduces solutions for knowledge-based understanding the system behaviour and the meaning of variable levels. Efficient normalisation, scaling and decomposition approaches are the key methodologies in developing large-scale applications. Linguistic equation (LE) approach originating from fuzzy logic is an efficient technique for these problems. The nonlinear scaling methodology based on advanced statistical analysis is the corner stone in representing the variable meanings in a compact way to introduce intelligent indices for control and diagnostics. The new constraint handling together with generalised norms and moments facilitates recursive parameter estimation approaches for the adaptive scaling. Well-known linear methodologies are used for the steady state, dynamic and case-based modelling in connection with the cascade and interactive structures in building complex large scale applications. To achieve insight and robustness the parameters are defined separately for the scaling and the interactions.

Keywords: Linguistic equations, smart adaptive systems, intelligent methods, statistical analysis, modelling and simulation.

1. INTRODUCTION

Models understood as relationships between variables are used for predicting of properties or behaviours of the system. Variable interactions and nonlinearities are important in extending the operation areas of control and fault diagnosis, where the complexity is alleviated by introducing software sensors (Fig. 1).

Adaptive systems can be developed for nonlinear multivariable systems by various statistical and intelligent methodologies, which are in sensor fusion combined with data pre-processing, signal processing and feature extraction (Fortuna et al., 2007). Fault diagnosis is based on symptoms generated by comparing process models and measurements (Frank et al., 2000), signal analysis (Lahdelma and Juuso, 2011), limit checking of measurements (Fortuna et al., 2007) and human observations (Isermann, 1997). All these are used in intelligent control and detection of operating conditions, which introduce reasoning and decision making to the smart adaptive systems, whose hybrid nature is seen in literature where these topics are combined from different perspectives.

The linguistic equation (LE) approach originates from fuzzy set systems (Juuso and Leiviskä, 1992): rule sets are replaced with equations, and meanings of the variables are handled with scaling functions which have close connec-

tions to membership functions (Juuso, 1999). The nonlinear scaling technique is needed in constructing nonlinear models with linear equations (Juuso, 2004). New development methodologies (Juuso, 2009; Juuso and Lahdelma, 2010) improve possibilities to update the scaling functions recursively (Juuso, 2011; Juuso and Lahdelma, 2011).

This paper classifies methodologies and focuses on combining advanced statistical analysis and soft computing in developing LE based applications for complex systems.

2. STEADY-STATE MODELLING

The steady-state simulation models can be relatively detailed nonlinear *multiple input, multiple output (MIMO)* models $\mathbf{y} = F(\mathbf{x})$, where the output vector $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is calculated by a nonlinear function F from the input vector $\mathbf{x} = (x_1, x_2, \dots, x_m)$. More generally, the relationship could also be a table or a graph. Fuzzy set systems, artificial neural networks and neuro-fuzzy methods provide additional methodologies for the function $F(\mathbf{x})$.

Statistical modelling in its basic form uses linear regression for solving coefficients for linear functions. In the *response surface methodology (RSM)*, the relationships are represented with *multiple input, single output (MISO)* models, which contain linear, quadratic and interactive terms (Box

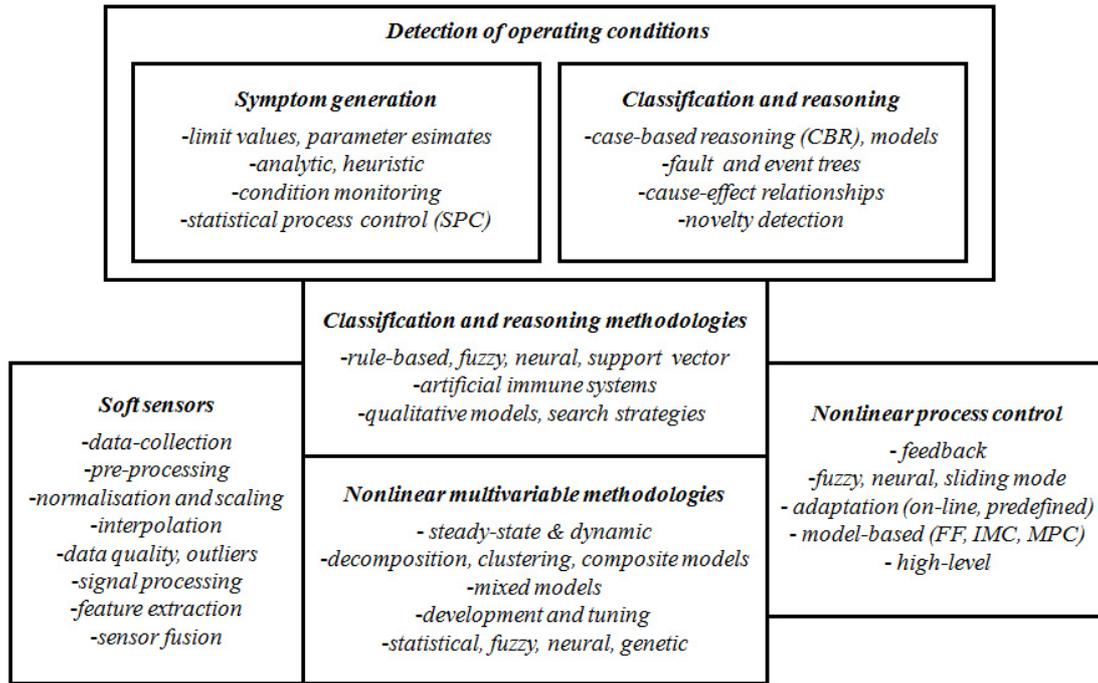


Figure 1. Methodologies for modelling of complex systems.

and Wilson, 1951). Application areas of the linear modelling can also be extended by arbitrary nonlinear models, e.g. semi-physical models, developed by using appropriate calculated variables as inputs, see (Ljung, 1999). *Principal component analysis (PCA)* compresses the data by reducing the number of dimensions: each principal component is a linear combination of the original variables, usually the first few principal components are used. Various extensions of PCA are referred in (Jolliffe, 2002). *Partial least squares regression (PLS)* uses potentially collinear variables (Gerlach et al., 1979).

Fuzzy logic emerged from approximate reasoning, and the connection of fuzzy rule-based systems and expert systems is clear, e.g. the vocabulary of AI is kept in fuzzy logic (Dubois et al., 1999). *Fuzzy set theory* first presented by Zadeh (1965) form a conceptual framework for linguistically represented knowledge. *Extension principle* is the basic generalisation of the arithmetic operations if the *inductive mapping* is a monotonously increasing function of the input. The interval arithmetic presented by Moore (1966) is used together with the extension principle on several membership α -cuts of the fuzzy number x_j for evaluating fuzzy expressions (Buckley and Qu, 1990; Buckley and Hayashi, 1999; Buckley and Feuring, 2000). The fuzzy sets can be modified by intensifying or weakening modifiers (De Cock and Kerre, 2004). *Type-2 fuzzy models* introduced by Zadeh in 1975 take into account uncertainty about the membership function (Mendel, 2007). Most systems based on interval type-2 fuzzy sets are reduced to an interval-valued type-1 fuzzy set.

Linguistic fuzzy models (Driankov et al., 1993), where both the antecedent and consequent are fuzzy propositions, suit very well to qualitative descriptions of the process as they can be interpreted by using natural language, heuristics and common sense knowledge. *Takagi-Sugeno (TS) fuzzy models* (Takagi and Sugeno, 1985), where each

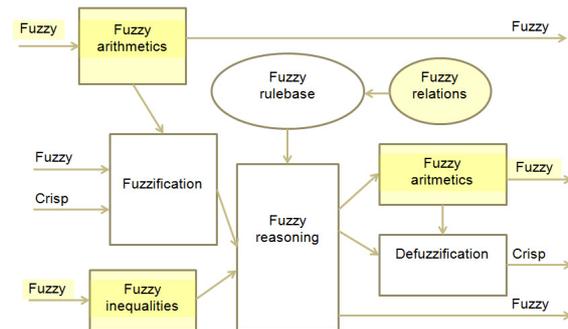


Figure 2. Combined fuzzy modelling.

consequent y_i , $i = 1, \dots, n$, is a crisp function of the antecedent variables x , can be interpreted in terms of local models. For linear functions, the standard weighted mean inference must be extended with a smoothing technique (Babuška, 1998). *Singleton models*, where the consequents are crisp values, can be regarded as special cases of both the linguistic fuzzy models and the TS fuzzy models. *Fuzzy relational models* (Pedrycz, 1984) allow one particular antecedent proposition to be associated with several different consequent propositions,

Several fuzzy modelling approaches are combined in Figure 2): fuzzy arithmetics is suitable both for processing fuzzy inputs for the rule-based fuzzy set system and the fuzzy outputs; fuzzy inequalities produce new facts like $A \leq B$ and $A = B$ for fuzzy inputs A and B ; fuzzy relations can be represented as sets of alternative rules, where each rule has a degree of membership.

Artificial neural networks (ANN) are used as behavioural input-output models consisting of neurons. Network architectures differ from each other in their way of forming the net input, use of activation functions and number of layers. *Linear networks* correspond to the models with

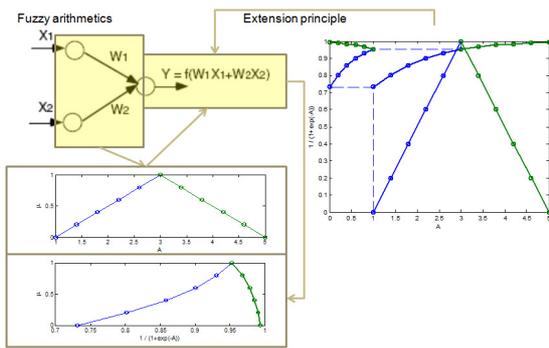


Figure 3. A fuzzy neuron.

linear terms in RSM models. The most popular neural network architecture is the *multilayer perceptron (MLP)* with a very close connection to the *backpropagation learning* (Rummelhart et al., 1986).

Neurofuzzy systems use fuzzy neurons to combine the weight factors and the inputs (Fig. 3). The activation function is handled with the extension principle from the fuzzy input, which is obtained by the fuzzy arithmetics (Fullér, 2000). Also cascade architectures of fuzzy set systems and neural networks are often called neurofuzzy systems. Neural computation is used for tuning fuzzy set systems which can be represented by neural networks, see (Jang, 1993).

A *function expansion* presented in (Ljung, 2008) provides a flexible way to present several types of black box models by using basis functions, which are generated from one and the same function characterised by the scale (dilation) and location (translation) parameters. The expansion can contain, for example, radial basis functions, one-hidden-layer sigmoidal neural networks, neurofuzzy models, wavenets, least square support vector machines, see (Ljung, 1999).

Approximate reasoning based on T-norms and S-norms, also called T-conorms, is an essential part of combining antecedents and rules in fuzzy logic (Driankov et al., 1993). T-norms and S-norms can be used in neurofuzzy systems if the inputs are normalised to the range $[0, 1]$ (Fullér, 2000).

3. DECOMPOSITION METHODOLOGIES

A modelling problem can be divided into smaller parts by developing separate models for independent subprocesses (Fig. 4). Cluster analysis can be used in the data-driven approach. Composite local models can be used, and fuzzy set systems provide feasible techniques for handling the resulting partially overlapping models. The system may also include models based on the first principles. A process plant consists of several processing units interconnected with process streams. *Decomposition* can be continued within process units. In an electric furnace presented in (Juuso, 1980), a cylindrically symmetrical one-electrode model was based on two-dimensional areas defined by overlapping rectangular grids where the amount of detail can be increased in selected parts (Juuso, 1990). In addition to spatial or logical blocks decomposed modelling can be based on different frequency ranges.

Hundreds of *clustering algorithms* have been developed for the data-driven analysis by researchers from a number

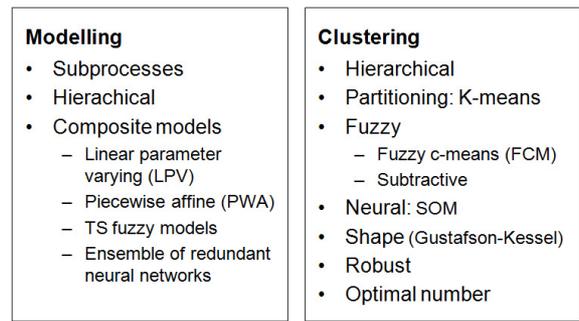


Figure 4. Decomposition for modelling.

of different scientific disciplines. *Hierarchical clustering* groups data by creating a cluster tree, where clusters at one level are joined as clusters at the next higher level. *Partitioning-based clustering* algorithms, e.g. K-means, minimise a given clustering criterion by iteratively relocating data points between clusters until a (locally) optimal partition is attained. (Äyrämö and Kärkkäinen, 2006). Numerous *fuzzy clustering* algorithms have been proposed and applied to a variety of real-world problems (Bezdek, 1981). *Fuzzy c-means (FCM)* clustering is a partitioning-based method: each data point belongs to a cluster to some degree membership. *Subtractive clustering* (Chiu, 1994) is an algorithm for estimating the number of clusters and the cluster centers according to the parameters of the algorithm. *Neural clustering* use competitive networks based on competitive layers, e.g. self-organising maps (SOM) (Kohonen, 1995) have several alternatives for calculating the distance in the competitive layer. The response of a radial basis functions (RBF) neuron is obtained from an exponential function (Chen et al., 1991).

The clustering algorithms have limitations in shape, cluster centres and generalisation of the results. The algorithm with the standard Euclidean norm imposes a spherical shape on the clusters, regardless of the actual data distribution (Babuška, 1998). Gustafson and Kessel (1979) extended the standard by employing an adaptive distance norm to detect clusters of different geometrical shapes. *Robust clustering*, which is based on a spatial median, is aimed for problems where classical clustering methods are too sensitive to erroneous and missing values (Äyrämö and Kärkkäinen, 2006). *Optimal number of clusters* is selected iteratively by using some quality criteria, see (Windham, 1981).

Composite local model approach constructs a global model from local models, which usually are linear approximations of the nonlinear system in different neighbourhoods. If the partitioning is based on a measured regime variable, the partitioning can be used in weighting the local models. *Linear parameter varying (LPV) models*, where the matrices of the state-space model depend on an exogeneous variable measured during the operation, are close related to local linear models (Ljung, 2008). *Piecewise affine (PWA) systems* are based on local linear models, more specifically in a polyhedral partition (Sontag, 1981). The models can be state-space models or parametric models. The model switches between different modes as the state variable varies over the partition (Ljung, 2008).

Fuzzy models can be considered as a class of local modelling approaches, which solve a complex modelling problem by decomposing into number of simpler understandable subproblems (Babuška et al., 1997; Babuška, 1998). The smoothing problem around the submodel borders of TS fuzzy models needs special techniques, e.g. smoothing maximum (Babuška, 1998), or by making the area overlap very strong. *Multiple neural network systems* improve generalisation through task decomposition and an ensemble of redundant networks (Shields and Casey, 2008).

A *mixed approach* using both the rigorous first principles and the black box modelling in an integrated environment is an interesting alternative for complex systems, see (Macías-Hernández, 2004). Ljung (2008) classifies the models as a palette of grey shades from white to black into six categories: first principles, identified parameters, semi-physical models, composite models, block oriented models, and black box models. In semi-physical models, linear modelling used together with nonlinear transformations which are based on process insight.

4. ADAPTIVE NONLINEAR SCALING

Membership definitions provide nonlinear mappings from the operation area of the (sub)system, defined with feasible ranges, to the linguistic values represented inside a real-valued interval $[-2, 2]$. The feasible range is defined by a membership function, and membership functions for finer partitions can be generated from membership definitions (Juuso et al., 1993). The basic scaling approach presented in (Juuso, 2004) has been improved later: a new constraint handling was introduced in (Juuso, 2009), and a new skewness based methodology was presented for signal processing in (Juuso and Lahdelma, 2010).

4.1 Working point and feasible ranges

The concept of feasible range is defined as a trapezoidal membership function. In the fuzzy set theory (Zimmermann, 1992), support and core areas are defined by variable, x_j , specific subsets,

$$\text{supp}(F_j) = \{x_j \in U_j \mid \mu_{F_j}(x_j) > 0\}, \quad (1)$$

$$\text{core}(F_j) = \{x_j \in U_j \mid \mu_{F_j}(x_j) = 1\}, \quad (2)$$

where U_j is an universal set including F_j ; $\mu_{F_j}(x_j)$ is the membership value of x_j in F_j . The main area of operation is the core area, and the whole variable range is the support area. For applications, a trapezoidal function providing linear transitions between 0 and 1 is sufficient (Fig. 5). The corner parameters can be defined on the basis of expert knowledge or extracted from data. The slope can be different on upper and lower part depending on the linearity or nonlinearity of the system. The complement of a fuzzy set is defined as a subset (Zimmermann, 1992)

$$\bar{F}_j = \{x_j \in U_j \mid \mu_{\bar{F}_j}(x_j) = 1 - \mu_{F_j}(x_j)\}, \quad (3)$$

where $\mu_{\bar{F}_j}(x_j)$ is the membership value of x_j in \bar{F}_j . The membership function of the complement corresponds to the highest and lowest membership functions (Fig. 5).

The support area is defined by the minimum and maximum values of the variable, i.e. the support area is $[\min(x_j), \max(x_j)]$ for each variable $j, j = 1, \dots, m$. The central tendency value, c_j , divides the support area into

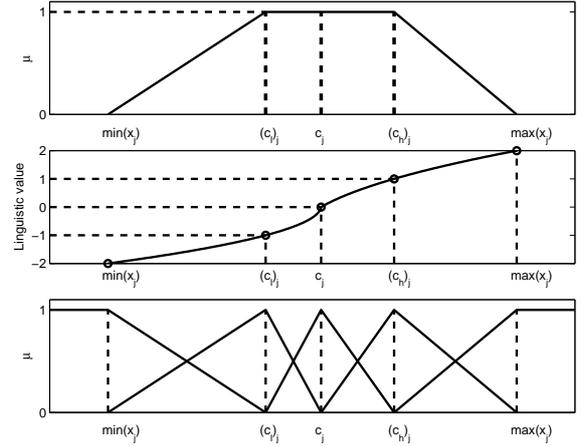


Figure 5. Nonlinear scaling (Juuso, 2004)

two parts, and the core area is defined by the central tendency values of the lower and the upper part, $(c_l)_j$ and $(c_h)_j$, correspondingly. This means that the core area of the variable j defined by $[(c_l)_j, (c_h)_j]$ is within the support area.

The corner points can be extracted from existing rule-based fuzzy systems or defined manually. Feasible ranges should be consistent with membership definitions, and therefore they are defined together in the data-driven approach. Earlier the analysis of the corner points and the centre point has been based on the arithmetic means or medians of the corresponding data sets (Juuso, 2004).

The norm defined by

$$\|\tau M_j^p\|_p = (\tau M_j^p)^{1/p} = \left[\frac{1}{N} \sum_{i=1}^N (x_j)_i^p \right]^{1/p}, \quad (4)$$

where $p \neq 0$, is calculated from N values of a sample, τ is the sample time. With a real-valued order $p \in \mathfrak{R}$ this norm can be used as a central tendency value if $\|\tau M_j^p\|_p \in \mathfrak{R}$, i.e. $x_j > 0$ when $p < 0$, and $x_j \geq 0$ when $p > 0$. The norm (4) is calculated about the origin, and it combines two trends: a strong increase caused by the power p and a decrease with the power $1/p$. All the norms have same dimensions as x_j . The norm (4) is a Hölder mean, also known as the power mean. The generalised norm for absolute values $|x_j|$ was introduced for signal analysis in (Lahdelma and Juuso, 2008).

For variables with only negative values, the norm is the opposite of the norm obtained for the absolute values. If a variable has both positive and negative values, each norm is an average of two norms obtained from data sets made positive and negative. (Juuso, 2011)

The operating area of each variable is defined by a feasible range represented with a trapezoidal membership function whose corner points are $\min(x_j)$, $(c_l)_j$, $(c_h)_j$ and $\max(x_j)$. Warnings and alarms can be generated directly from the degrees of membership of the complement (3).

4.2 Membership definitions

A membership definition is defined as a (nonlinear) mapping of variable values x_j inside its range to $X_j \in [-2, 2]$, denoted as *linguistic range*. It more or less describes the

distribution of variable values over its range. The range $[-2, 2]$ includes the normal operation in the range $[-1, 1]$ and the areas with warnings and alarms. The values X_j are called *linguistic values* since the scaling idea originates from the fuzzy set systems: values -2, -1, 0, 1 and 2 associated to the linguistic labels are defined with membership functions (Fig. 5). The number of membership functions is not limited to five: the values between these integers correspond to finer partitions of the fuzzy set system. The early applications of the linguistic equations used only integer values (Juuso, 1999).

In present systems, membership definitions are used in a continuous form consisting of two second order polynomials: one for negative values, $X_j \in [-2, 0)$, and one for positive values, $X_j \in [0, 2]$. The polynomials,

$$\begin{aligned} f_j^-(X_j) &= a_j^- X_j^2 + b_j^- X_j + c_j, \quad X_j \in [-2, 0), \\ f_j^+(X_j) &= a_j^+ X_j^2 + b_j^+ X_j + c_j, \quad X_j \in [0, 2], \end{aligned} \quad (5)$$

should be monotonous, increasing functions in order to result in realisable systems. The upper and lower parts should overlap at the linguistic value 0. (Juuso, 2004)

The functions are monotonous and increasing if the ratios,

$$\begin{aligned} \alpha_j^- &= \frac{(c_l)_j - \min(x_j)}{c_j - (c_l)_j}, \\ \alpha_j^+ &= \frac{\max(x_j) - (c_h)_j}{(c_h)_j - c_j}, \end{aligned} \quad (6)$$

are both in the range $[\frac{1}{3}, 3]$, see (Juuso, 2009). If needed, the ratios are corrected by modifying the core $[(c_l)_j, (c_h)_j]$ and/or the support $[\min(x_j), \max(x_j)]$. Errors are checked independently for f_j^- and f_j^+ : each error can always be corrected either by moving the corner of the core or the support. In some cases, good results can also be obtained by moving c_j . If these constraints allow a non-empty range, the maximum of the lower limits and the minimum of the upper limit are chosen to define the limits for continuous definitions.

The coefficients of the polynomials can be represented by

$$\begin{aligned} a_j^- &= \frac{1}{2}(1 - \alpha_j^-) \Delta c_j^-, \\ b_j^- &= \frac{1}{2}(3 - \alpha_j^-) \Delta c_j^-, \\ a_j^+ &= \frac{1}{2}(\alpha_j^+ - 1) \Delta c_j^+, \\ b_j^+ &= \frac{1}{2}(3 - \alpha_j^+) \Delta c_j^+, \end{aligned} \quad (7)$$

where $\Delta c_j^- = c_j - (c_l)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if some coefficients α_j^- or α_j^+ equals to one.

For each variable, the membership definitions are configured with five parameters, which can be presented with three consistent sets. The working point (centre point) c_j belongs to all these sets, where the other parameters are:

- the corner points (Fig. 5) are good for visualisation;
- the parameters $\{\alpha_j^-, \Delta c_j^-, \alpha_j^+, \Delta c_j^+\}$ suit for tuning;
- the coefficients $\{a_j^-, b_j^-, a_j^+, b_j^+\}$ are used in the calculations.

4.3 Adaptation of nonlinear scaling

Recursive data analysis facilitates the adaptation of the functions to changing operating conditions, also the orders of the norms are re-analysed if needed. The existing scaling functions provide a basis for assessing the quality of new data: outliers should be excluded, but the suspicious values may mean that the operating conditions are changing. In this research, the scaling functions are extended for analysing outliers and suspicious values to select data for the adaptive scaling. Different operating areas can be analysed with clustering, and statistical process control provide additional tools for detecting changes, anomalies and novelties.

The parameters of the nonlinear scaling functions can be recursively updated by including new equal sized sub-blocks into calculations. The number of samples can be increasing or fixed with some forgetting, and weighting of the individual samples can be used in the analysis. If the definitions should cover all the operating areas, also suspicious values are included as extensions of the support area. In each adaptation step, the acceptable ranges of the shape factors α_j^- and α_j^+ are checked and corrected if needed. The orders $(p_l)_j$, $(p_0)_j$ and $(p_h)_j$ of the corresponding norms are re-analysed if the distribution is changing considerably with new measurements.

5. INTELLIGENT SYSTEMS

Nonlinear models can be constructed by using scaled values in linear modelling based on data and expertise (Juuso, 1999, 2004). Compact model structures are beneficial in building and tuning dynamic and case-based models for complex systems. The recursive analysis provides new tools for both the adaptation of the scaling functions and the model interactions to changing operating conditions. Linear interactions are used in steady-state models and extended to dynamic systems by parametric structures used in identification. Decomposition of the modelling area is used for case-based systems which can include both steady state and dynamic models. The nonlinear scaling is performed twice: first scaling from real values to the interval $[-2, 2]$ before applying linguistic equations, and then scaling from the interval $[-2, 2]$ to real values after applying linguistic equations. Variable selection is needed in large-scale systems.

5.1 LE models

The nonlinear scaling with membership definitions transforms the nonlinear model $\mathbf{y} = F(\mathbf{x})$ to a linear problem. The basic element of a linguistic equation (LE) model is a compact equation

$$\sum_{j=1}^m A_{ij} X_j(t - n_j) + B_i = 0, \quad (8)$$

where X_j is a linguistic value for the variable j , $j = 1 \dots m$. Each variable j has its own time delay n_j compared to the variable with latest time label. Linguistic values in the range $[-2, 2]$ are obtained from the actual data values by membership definitions. The directions of the interaction are represented by interaction coefficients $A_{ij} \in \mathfrak{R}$. In the

original system (Juuso and Leiviskä, 1992), the linguistic labels $\{very\ low, low, normal, high, very\ high\}$ were replaced by numbers $\{-2, -1, 0, 1, 2\}$. The approach was generalized for finer fuzzy partitions in (Juuso et al., 1993). The bias term $B_i \in \mathfrak{R}$ was first introduced as an additional component in fuzzy LE models (Juuso, 1996), and later extended for fault diagnosis systems (Juuso, 2004).

The coefficients A_{ij} and B_i in (8) have a relative meaning, i.e. the equation can be multiplied or divided by any non-zero real number without changing the model. A LE model with several equations can be represented as a matrix equation

$$\mathbf{AX} + \mathbf{B} = 0, \quad (9)$$

where the interaction matrix A contains all coefficients A_{ij} , $i = 1, \dots, n$, $j = 1, \dots, m$, and the bias vector \mathbf{B} all bias terms B_i , $i = 1, \dots, n$. The time delays of individual variables are equation specific. As linear equations, each model can be used in any direction, i.e. the output variable can be chosen freely.

5.2 Hybrid LE systems

Statistical analysis is an essential part of the development and tuning of the LE systems: data-driven development of the scaling functions, which is based on advanced generalised norms and moments, is suitable for different statistical distributions. The linguistic equation approach originates from the fuzzy set systems which keeps the connections of the methodologies strong. Compact LE models provide a good basis for multimodel systems, where local LE models are combined with fuzzy logic, to handle transitions between models, some special situations and uncertainty with fuzzy set systems. Fuzzy reasoning is an important part of the LE based fault diagnosis and the decision making in the recursive adaptation.

The LE based development of fuzzy systems on any partition can be done if a sufficient number of variables are known or varied by selecting membership locations. Fuzzy set systems, which represent gradual changes by interpolating with membership functions, can be handled by membership definitions and linguistic equations, i.e. the system does not necessarily need any uncertainty or fuzziness. Fuzzy set systems have been moved to higher levels in applications, when first modelling and control, and later also the detection of the operating conditions was realised with the LE approach.

Fuzzy numbers can be handled in LE models by the *extension principle* (Juuso, 2007a,b, 2008). LE models are extended to fuzzy inputs with this approach if the membership definitions, i.e. functions f_j^- and f_j^+ defined by (5) and the corresponding inverse functions, are replaced by the corresponding extensions of these functions. The square root functions are used in the *linguistification* part. The extension principle is needed for fuzzy inputs. The result of the fuzzy extension is a nonlinear membership function for the output even if the membership function of the input is linear. The number of membership levels should increase with the growing fuzziness of the input.

Data mining need to be combined with domain expertise to develop practical systems. The LE approach provides a feasible integration framework for practical intelligent applications. The process insight is maintained since all the modules can be assessed by expert knowledge and the membership definitions relate measurements to appropriate operating areas. The nonlinear scaling methodology based on statistical analysis enhanced with domain expertise is the corner stone of the approach, which represents the variable meanings in a compact way to introduce intelligent indices for control and diagnostics. Different statistical and intelligent methods are used together with the LE approach. The weighting of submodels also is based on the scaled values and fuzzy logic. The cascade and interactive model structures are used in building more complex large scale applications.

REFERENCES

- Äyrämö, S. and Kärkkäinen, T. (2006). Introduction to partitioning-based clustering methods with a robust example. Reports of the Department of Mathematical Information Technology Series C. University of Jyväskylä, Software and Computational Engineering No. C. 1/2006. Jyväskylä.
- Babuška, R. (1998). *Fuzzy Modeling and Identification*. Kluwer Academic Publisher, Boston.
- Babuška, R., Setnes, M., Kaymak, U., and Verbruggen, H.B. (1997). Fuzzy Modelling: a Universal and Transparent Tool. In L. Yliniemi and E. Juuso (eds.), *Proceedings of TOOLMET'97 - Tool Environments and Development Methods for Intelligent Systems, Oulu, April 17-18, 1997*, 1–27. Oulun yliopistopaino, Oulu.
- Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function*. Plenum Press, New York.
- Box, G.E.P. and Wilson, K.B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B*, 13(1), 1–45.
- Buckley, J.J. and Feuring, T. (2000). Universal approximators for fuzzy functions. *Fuzzy Sets and Systems*, 113, 411–415.
- Buckley, J.J. and Hayashi, Y. (1999). Can neural nets be universal approximators for fuzzy functions? *Fuzzy Sets and Systems*, 101, 323–330.
- Buckley, J.J. and Qu, Y. (1990). On using α -cuts to evaluate fuzzy equations. *Fuzzy Sets and Systems*, 38(3), 309–312.
- Chen, S., Cowan, C., and Grant, P.M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, 2(2), 302–309.
- Chiu, S. (1994). Fuzzy Model Identification Based on Clustering Estimation. *Journal of Intelligent & Fuzzy Systems*, 2(3), 267–278.
- De Cock, M. and Kerre, E.E. (2004). Fuzzy modifiers based on fuzzy relations. *Information Sciences*, 160(1–4), 173–199.
- Driankov, D., Hellendoorn, H., and Reinfrank, M. (1993). *An Introduction to Fuzzy Control*. Springer, Berlin, Germany.
- Dubois, D., Prade, H., and Ughetto, L. (1999). Fuzzy logic, control engineering and artificial intelligence. In H.B.

- Verbruggen, H.J. Zimmermann, and R. Babuska (eds.), *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, 17–57. Kluwer, Boston.
- Fortuna, L., Graziani, S., Rizzo, A., and Xibilia, M.G. (2007). *Soft Sensors for Monitoring and Control of Industrial Processes*. Advances in Industrial Control. Springer, New York. 270 pp.
- Frank, P.M., Garcia, E.A., and Köppen-Seliger, B. (2000). Modelling for fault detection and isolation versus modelling for control. *Mathematics and Computers in Simulation*, 53, 259–271.
- Fullér, R. (2000). *Introduction to Neuro-Fuzzy Systems*. Advances in Soft Computing. Springer. 289 pp.
- Gerlach, R.W., Kowalski, B.R., and Wold, H.O.A. (1979). Partial least squares modelling with latent variables. *Anal. Chim. Acta*, 112(4), 417–421.
- Gustafson, D.E. and Kessel, W.C. (1979). Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings of IEEE CDC, San Diego, CA, USA*, 761–766.
- Isermann, R. (1997). Supervision, fault-detection and fault-diagnosis methods - an introduction. *Control Engineering Practice*, 5(5), 639–652.
- Jang, J.S.R. (1993). ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665–685.
- Jolliffe, I.T. (2002). *Principal Component Analysis*. Springer, New York, 2 edition. 487 pp.
- Juuso, E. (1980). A computer analysis of temperature distribution and energy consumption in a submerged arc furnace used in the production of high-carbon ferromanganese. *Acta Universitatis Ouluensis, Series C Technica No. 17, Technica Processionum No. 3*. 42 pp.
- Juuso, E. and Lahdelma, S. (2010). Intelligent scaling of features in fault diagnosis. In *Proceedings of the 7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 22-24 June, 2010, Stratford-upon-Avon, UK*, volume 2, 1358–1372. Curran Associates, NY, USA.
- Juuso, E. and Lahdelma, S. (2011). Intelligent trend indices and recursive modelling in prognostics. In *Proceedings The 8th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 20-22 June, 2011, Cardiff, UK*, volume 1, 440–450. BINDT.
- Juuso, E.K. (1990). Multilayer simulation of heat flow in a submerged arc furnace used in the production of ferroalloys. In M. Amouroux and A.E. Jai (eds.), *Control of Distributed Parameter Systems 1989. Selected Papers from the 5th IFAC Symposium, Perpignan, France 26-29 June 1989, IFAC Symposia Series, 1990, Number 3*, 459–464. Pergamon, Oxford, UK.
- Juuso, E.K. (1996). Computational intelligence in distributed interactive synthetic environments. In A.G. Bruzzone and E.J.H. Kerckhoffs (eds.), *Simulation in Industry, Proceedings of the 8th European Simulation Symposium, Simulation in Industry, ESS'96 Genoa, Italy, October 2-5, 1996*, 157–162. SCS International, San Diego, USA.
- Juuso, E.K. (1999). Fuzzy control in process industry: The linguistic equation approach. In H.B. Verbruggen, H.J. Zimmermann, and R. Babuska (eds.), *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, 243–300. Kluwer, Boston. Doi: 10.1007/978-94-011-4405-6_10.
- Juuso, E.K. (2004). Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35, 307–337. Doi: 10.1016/j.ijar.2003.08.008.
- Juuso, E.K. (2007a). Forecasting batch cooking results with intelligent dynamic simulation. In B. Zupančič, R. Karba, and S. Blažič (eds.), *Proceedings of the 6th EUROSIM Congress on Modelling and Simulation, Ljubljana, Slovenia, 9-13 Sept., 2007*, volume 2, 8 pp. University of Ljubljana, Ljubljana, Slovenia.
- Juuso, E.K. (2007b). Intelligent modelling of a fluidised bed granulator used in production of pharmaceuticals. In P. Bunus, D. Fritzon, and C. Führer (eds.), *Conference Proceedings of SIMS 2007 - The 48th Scandinavian Conference on Simulation and Modeling, Göteborg (Särö), 30-31 October, 2007*, 101–108. Linköping University Electronic Press, Linköping, Sweden.
- Juuso, E.K. (2008). Intelligent dynamic simulation of a fed-batch enzyme fermentation process. In *Tenth International Conference on Computer Modelling and Simulation, EUROSIM/UKSim, Cambridge, UK, April 13, 2008*, 301–306. The Institute of Electrical and Electronics Engineers IEEE. Doi: 10.1109/UKSIM.2008.133.
- Juuso, E.K. (2009). Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In M. Kolehmainen (ed.), *Revised selected papers of the International Conference on Adaptive and Natural Computing Algorithms - ICANNGA 2009, Kuopio, Finland, Lecture Notes in Computer Science*, volume LNCS 5495, 161–170. Springer-Verlag, Heidelberg. Doi: 10.1007/978-3-642-04921-7_17.
- Juuso, E.K. (2011). Recursive tuning of intelligent controllers of solar collector fields in changing operating conditions. In S. Bittani, A. Cenedese, and S. Zampieri (eds.), *Proceedings of the 18th World Congress The International Federation of Automatic Control, Milano (Italy) August 28 - September 2, 2011*, 12282–12288. IFAC. [Http://www.ifac-papersonline.net/](http://www.ifac-papersonline.net/).
- Juuso, E.K., Bennavil, J., and Singh, M. (1993). Hybrid knowledge-based system for managerial decision making in uncertainty environment. In N.P. Carreté and M.G. Singh (eds.), *Qualitative Reasoning and Decision Technologies, Proceedings of the IMACS International Workshop on Qualitative Reasoning and Decision Technologies -QUARDET'93, Barcelona, June 16 - 18, 1993*, 234–243. CIMNE, Barcelona.
- Juuso, E.K. and Leiviskä, K. (1992). Adaptive expert systems for metallurgical processes. In S.L. Jämsä-Jounela and A.J. Niemi (eds.), *Expert Systems in Mineral and Metal Processing, Proceedings of the IFAC Workshop, Espoo, Finland, August 26-28, 1991, IFAC Workshop Series, 1992, Number 2*, 119–124. Pergamon, Oxford, UK.
- Kohonen, T. (1995). *Self-Organizing Maps*. Springer, Berlin.
- Lahdelma, S. and Juuso, E. (2008). Signal processing in vibration analysis. In *Proceedings of the Fifth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 15-18 July 2008, Edinburgh, UK*, 867–878. BINDT.
- Lahdelma, S. and Juuso, E. (2011). Signal processing and feature extraction by using real order derivatives

- and generalised norms. Part 1: Methodology. *The International Journal of Condition Monitoring*, 1(2), 46–53. Doi: 10.1784/204764211798303805.
- Ljung, L. (1999). *System Identification - Theory for the User*. Prentice Hall, Upper Saddle River, N.J., 2nd edition.
- Ljung, L. (2008). Perspectives on system identification. In M.J. Chung and P. Misra (eds.), *Plenary papers, milestone reports & selected survey papers, 17th IFAC World Congress, Seoul, Korea, July 6-11, 2008*, 47–59. IFAC. [Http://www.ifac-papersonline.net/](http://www.ifac-papersonline.net/).
- Macías-Hernández, J. (2004). Opportunities of smart adaptive systems in the oil refining industry. In *Proceedings of Eunite 2004 - European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, June 10-12, 2004, Aachen, Germany*, 1–12. Wissenschaftsverlag Mainz, Aachen.
- Mendel, J.M. (2007). Advances in type-2 fuzzy sets and systems. *Information Sciences*, 177(1), 84–110.
- Moore, R.E. (1966). *Interval Analysis*. Prentice Hall, Englewood Cliffs, NJ.
- Pedrycz, W. (1984). An identification algorithm in fuzzy relational systems. *Fuzzy Sets and Systems*, 13(2), 153–167.
- Rummelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rummelhart and J. McClelland (eds.), *Parallel Data Processing*, 318–362. M.I.T. Press, Cambridge, MA.
- Shields, M.W. and Casey, M.C. (2008). A theoretical framework for multiple neural network systems. *Neurocomputing*, 71, 1462–1476.
- Sontag, E. (1981). Nonlinear regulation: The piecewise linear approach. *IEEE Transactions Automatic Control*, 26(2), 346–358.
- Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1), 116–132.
- Windham, M.P. (1981). Cluster validity for fuzzy clustering algorithms. *Fuzzy Sets and Systems*, 5(2), 177–185.
- Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8(June), 338–353.
- Zimmermann, H.J. (1992). *Fuzzy set theory and its applications*. Kluwer Academic Publishers.

Tool-Independent Distributed Simulations Using Transmission Line Elements And The Functional Mock-up Interface

Robert Braun and Petter Krus

Linköping University, Division of Fluid and Mechatronic Systems, SE-58183 Linköping , Sweden

Abstract. This paper describes how models from different simulation tools can be connected and simulated on different processors by using the Functional Mockup Interface (FMI) and the transmission line element method (TLM). Interconnectivity between programs makes it possible to model each part of a complex system with the best suited tool, which will shorten the modelling time and increase the accuracy of the results. Because the system will be naturally partitioned, it is possible to identify weak links and replace them with transmission line elements, thereby introducing a controlled time delay. This makes the different parts of the system naturally independent, making it possible to simulate large aggregated system models with good performance on multi-core processors. The proposed method is demonstrated on an example model. A suggestion of an XML extension to the FMI standard for describing TLM ports is also presented.

Keywords: Functional Mockup Interface (FMI), Functional Mockup Unit (FMU), Transmission Line Element Method (TLM), Parallelism, Co-Simulation

1 Introduction

If different parts of a simulation model can be run on different processor cores, execution time can be considerably shortened. By using the transmission line element method (TLM) with independent distributed solvers, it is possible to achieve natural parallelism. This paper investigates the possibilities to combine this with the Functional Mock-up Interface (FMI), a standardised interface for connecting different simulation environments. Using TLM and FMI together makes it possible to run distributed simulations with sub-models from different tools, which can fully exploit the benefits of multi-core processors.

First, the backgrounds of FMI, TLM, and the Hopsan simulation environment are explained. Then the implementations of import and export routines of FMI are presented. Finally, the validity of the method is confirmed by experiments on an example model.

2 Functional Mockup Interface

The Functional Mock-up Interface (FMI) is an open standardised interface for connecting simulation environments in a variety of ways. It is developed by the MODELISAR consortium, initiated by Dassault Systems [1]. There are four areas where FMI can be used:

- FMI for model exchange
- FMI for co-simulation
- FMI for applications
- FMI for PLM

The basic concept is to create a Functional Mockup Unit (FMU) from a model in one tool and then import it and use it in a target environment. An FMU consists of a compressed ZIP file with the FMU file extension. It contains an XML description of the contents and the simulation code as a set of C functions, either as binary files and/or compilable source code, which can be used by the host program. It can also contain documentation and a graphical icon. The use of a plain C interface makes FMUs compiler independent. They do, however, still depend on platform and architecture. With FMI for co-simulation the solver is included in the FMU, as opposed to FMI for model exchange, where the solver must be provided by host program. This paper focuses only on model exchange, mainly because it is supported by a larger number of vendors. In the upcoming FMI 2.0 standard, the difference between co-simulation and model exchange will be reduced [2]. It is, however, not yet released and is not used in this paper.

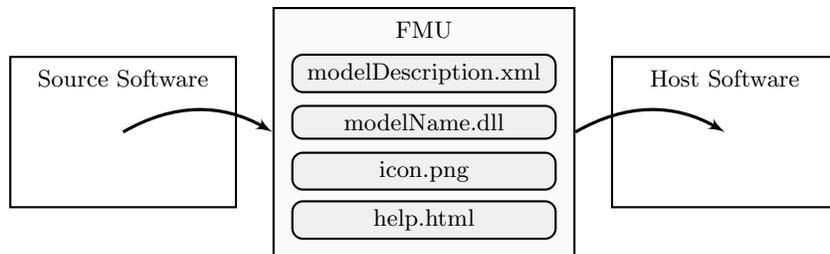


Fig. 1: With the FMI standard, Functional Mockup Units (FMUs) can be used to exchange models among simulation tools.

3 Transmission Line Element Connections

The transmission line element method (TLM) is a method for partitioning models by introducing physically motivated time delays. It is related to the method of characteristics [3] and to transmission line modelling [4]. In physical systems, information propagation is always delayed by capacitances. The concept with transmission line elements is to replace these capacitances in the model with transmission line elements, modelled as characteristic impedances. This method makes it possible to maintain accurate wave propagation, which is not possible by using only pure time delays.

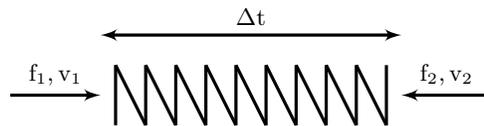


Fig. 2: An example of a transmission line element is the linear mechanical spring.

An example of a TLM element is the linear mechanical spring, as shown in figure 2. As can be seen it is subjected to two forces, f_1 and f_2 , from the left and the right side, respectively. According to equation 1 each force is defined by a function of the velocity at the same side and the delayed force and delayed velocity from the other side. This implies that one end is always independent of the opposite end at the same point of time. This enables the use of distributed solvers, where each sub-component in the model solves its own equations. This approach is very suitable for co-simulation and parallel execution [5].

$$\begin{aligned} f_1(t) &= F(v_1(t), f_2(t - \Delta t), v_2(t - \Delta t)) \\ f_2(t) &= G(v_2(t), f_1(t - \Delta t), v_1(t - \Delta t)) \end{aligned} \quad (1)$$

When communicating between different programs using TLM connections, it is important that the variables in each connection are clearly specified. This can be done manually by the user when importing the model to the host environment, although this can be quite cumbersome. An alternative solution would be to include this information in the XML specification in the FMU. A TLM connection is defined as four variables; intensity, flow, wave variable, and characteristic impedance. Some additional variables may also be necessary depending on the physical domain, such as position and equivalent inertia for mechanical connections. Even though similarities between different domains exist, it is unfortunately not possible to use a general definition; the set of variables will always need to be hard-coded depending on the physical type of the connection. For example, an incompressible fluid needs only one flow variable, while a compressible fluid might need variables for both mass flow and volume flow. It is also necessary to provide information of whether the FMU is a resistive component (Q-type) or a transmission line connection (C-type). In order to transfer this information, the following addition to the XML description is suggested, see listing 3

```
<tlmConnections>
  <tlmConnection type="q" domain="hydraulic">
    <q>q1</q>
    <p>p1</p>
    <c>c1</c>
    <Zc>Zc1</Zc>
  </tlmConnection>
  <tlmConnection type="c" domain="mechanic">
    <F>F2</F>
    <x>x2</x>
    <v>v2</v>
    <me>M2</me>
    <c>c2</c>
    <Zc>Zc2</Zc>
  </tlmConnection>
</tlmConnections>
```

Fig. 3: An addition to the FMU XML description, describing TLM ports, is suggested.

4 Importing Functional Mock-up Units

All experiments in this paper are conducted in Hopsan, a distributed simulation environment developed at Linköping University [6]. It is based on the transmission line element method and uses distributed solvers. Components in Hopsan are quite similar to FMUs; they consist of a pre-compiled shared library file and an XML description file. The library file, however, is linked against Hopsan from where it inherits classes and can thus not be used standalone.

The process of importing an FMU to Hopsan is implemented in the following way, see figure 5. First the files are extracted to a temporary directory. Then the XML is parsed, including TLM specifications if present. Then the source file for the component library (`fmuLib.cc`) and the Hopsan component (`fmuName.hpp`) are generated and compiled to a shared library. A simple solver that uses the forward Euler method is also included, in order to be able to solve the equations. This would not be required with FMI for co-simulation, where the solver is included in the FMU. Finally, the XML description (`fmuName.xml`) is generated. The component can then be loaded from Hopsan and will then be available from the component library.

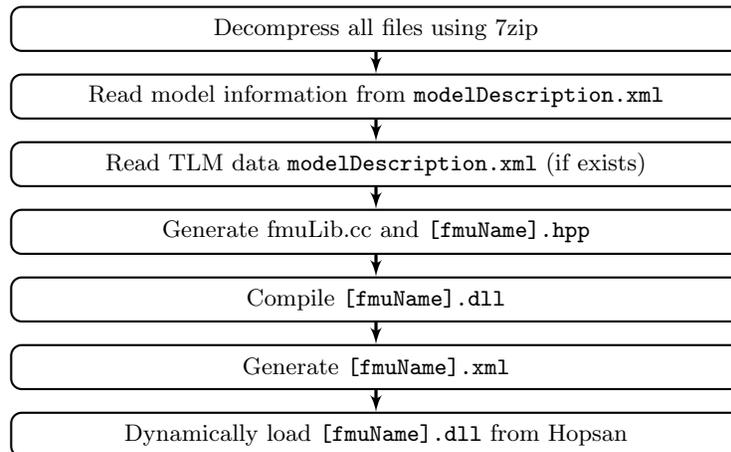


Fig. 4: FMUs are imported to Hopsan by compiling a wrapper library.

Hopsan is an object-oriented simulation environment where everything is pre-compiled. Components are objects that can be connected by node objects [7]. During the simulation, each component solves itself independently. Therefore no compilation prior to simulation is required. The main simulation uses fixed time steps for communication between components and nodes. It is, however, up to each component to decide how to perform its calculations. Thus, it is possible to use variable time steps inside a certain component. Theoretically, it is possible to use variable time steps also for the whole model, but previous experiments show that this is generally not worth the effort [8].

5 Exporting Models to Functional Mock-Up Units

Hopsan does not contain any equations or numerical solvers. The natural method would thus be to export an FMU for co-simulation. For compatibility with other software, however, the model exchange interface is used. The exported FMU does, however, actually solve itself and is basically working as an FMU for co-simulation.

According to the FMI specifications, only one shared library file is allowed. It is thus not possible to link against a pre-compiled Hopsan library file; the simulation core must be compiled into the FMU library. The model file is also included as a string variable in a header file and is loaded during initialisation.

Another factor which must be considered is that the FMI standard requires a plain C interface to ensure cross-compiler compatibility. Because the Hopsan simulation core is written in C++, a wrapper file is used to allow access to all required functions without using C++ features, such as classes and objects.

The export process begins with parsing the model and generating the XML description, together with the TLM extension if required. All source code files are then generated, including the model header file and the wrapper files. These files are then compiled along with the source code from the Hopsan simulation core and the required FMI source code. Finally, the binary and the XML description are compressed using 7zip. It is important to use the “deflate” compression method, to ensure compatibility.

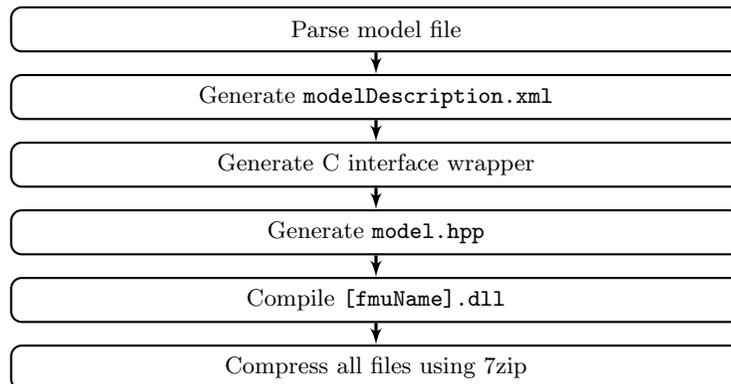


Fig. 5: FMUs are exported from Hopsan by compiling the simulation core with a C interface using a wrapper.

FMUs exported from Hopsan are successfully validated by importing them into OpenModelica, Dymola, FMU SDK, and also back into Hopsan itself. They are verified with FMU Checker version 1.0.2. No errors or warnings are reported.

6 Example Simulation

In order to demonstrate the proposed method, an example model is created, see figure 6. It consists of a four-wheel vehicle with an engine, a mechanical gearbox, and a hydraulic transmission. The hydraulic system is modelled using built-in pre-compiled C++ components in Hopsan. The engine is modelled as a PI-controlled torque source with velocity feedback. As a demonstration, it is exported from Hopsan to an FMU and then imported back into Hopsan. The brake component is created in the same way. Models for the vehicle, the wheels, and the gearbox are all equation-based models created in OpenModelica, an open-source Modelica-based simulation tool [9]. They are then exported to Hopsan as FMUs. The vehicle consists of a linear inertia with a drag coefficient parameter. Wheels are modelled as rotating inertias with ports for drive shaft, brakes, and attachments in the vehicle. The gearbox is modelled as a rotating inertia with changeable gear ratio.

All components are connected through transmission line elements, representing the shafts and mechanical connections. This means that stiffness is replaced by characteristic impedances and a time delay, which results in a decoupled system with good wave propagation accuracy. Pre-defined components in Hopsan are used for this purpose.

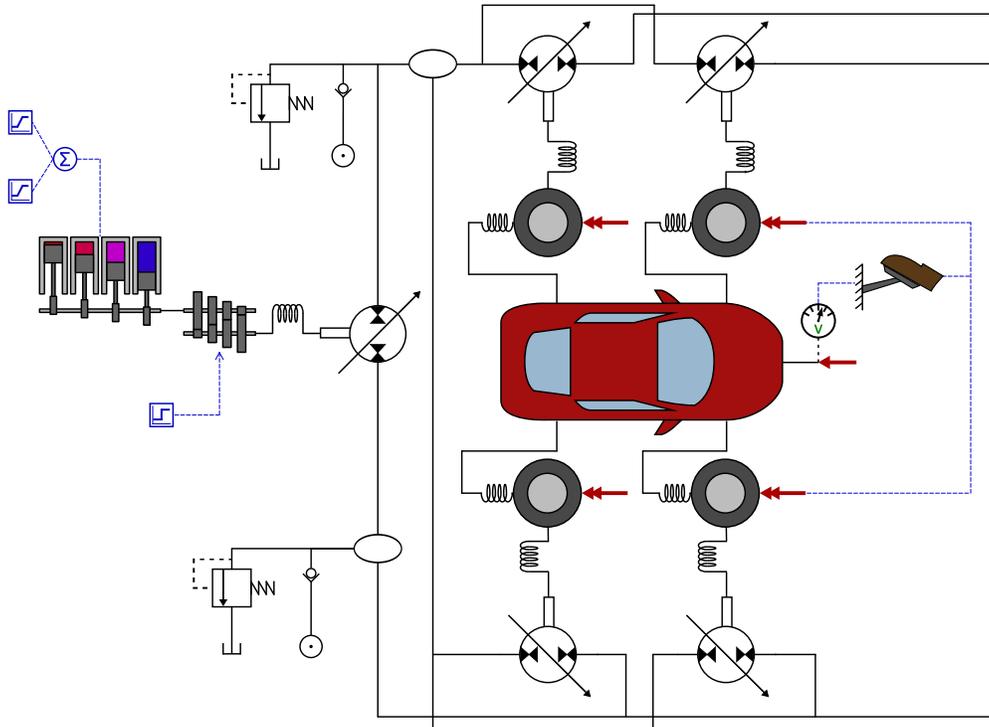


Fig. 6: An example model that describes a four-wheel vehicle with a simple hydraulic transmission is used to verify the proposed method.

In order to verify the functionality of the model, a simple drive cycle is simulated. The vehicle is first accelerated to 50 km/h and then to 70 km/h. It is then slowed down to 30 km/h and finally comes to a stop, see figure 7.

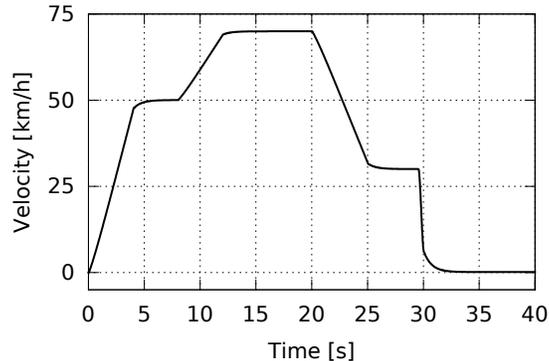


Fig. 7: A simple drive cycle is used to verify the functionality of the example model.

Load balancing is an essential aspect in parallel programming. If the work is not equally distributed over the threads, the speed-up will be limited by the slowest thread. In Hopsan this is solved by an automatic algorithm that measures the simulation time for each component over a few time steps before the actual simulation. This information is in turn used to distribute the components evenly over the simulation threads [5]. The average measured time per iteration for each sub-component type in the example model is shown in table 1. Note that these measurements are made on different sub-models and can thus not be used to compare the performance of different simulation tools.

Sub-model	Time/iteration
Wheel (OpenModelica, FMU)	1.192 μ s
Gearbox (OpenModelica, FMU)	1.083 μ s
Vehicle (OpenModelica, FMU)	0.574 μ s
Brake (Hopsan, FMU)	0.476 μ s
Engine (Hopsan, FMU)	0.189 μ s
Relief Valve (Hopsan, built-in)	0.121 μ s
Pump (Hopsan, built-in)	0.115 μ s
Volume (Hopsan, built-in)	0.024 μ s

Table 1: The simulation time for each sub-component is measured before the simulation, to achieve good load balancing.

The resulting distribution is shown in table 2. Components of C-type are generally much faster than those of Q-type. In this case they only required 5.8% of the total time. For this reason, only the threads for Q-type components are analysed.

	Thread 1	Thread 2	Thread 3	Thread 4
	Wheel 1	Wheel 2	Wheel 3	Wheel 4
	Gearbox	Vehicle	Pump 1	Pump 2
			Relief Valve 1	Relief Valve 2
			Pump 3	Pump 4
			Pump 5	Check Valve 1
				Check Valve 2
Total time:	2.275 μ s	1.765 μ s	1.644 μ s	1.606 μ s

Table 2: Sub-components are automatically distributed over the simulation threads.

As can be seen, a decent although not perfect load balancing is achieved. There are also overhead time costs from time measurements and thread synchronisation. Simulation time is, however, still more than twice as fast with four threads compared to with one thread. See table 3 for simulation times for 10,000 time steps with different numbers of processors. The time reduction from parallel simulation will increase when larger models are used. Theoretical maximum of speed-up is limited by the number of processor cores [5].

Threads:	Simulation time:
1	3307 ms
2	2091 ms
4	1466 ms

Table 3: Parallel execution reduces simulation time.

7 Conclusions

This paper shows that it is possible to combine the FMI standard with the transmission line element method. This makes it possible to simulate large aggregated models, consisting of sub-models from different modelling tools, in parallel on multi-core processors. Simulation time can then be significantly reduced. An interesting continuation could be real-time applications, where simulation performance is a critical aspect. Other possible future work could be to investigate higher level modelling methods for describing aggregated FMI models. Experiments were performed with FMI for model exchange using a simple solver. FMI for co-simulation would be more suitable, but is so far not supported by many simulation tools. Such difficulties will be easier to overcome with the FMI 2.0 standard, where co-simulation and model exchange will be harmonised.

8 Acknowledgements

This work was supported by ProViking research school and the Swedish Foundation for Strategic Research (SSF).

References

- [1] Torsten Blochwitz, M Otter, M Arnold, C Bausch, C Clauß, H Elmqvist, A Junghanns, J Mauss, M Monteiro, T Neidhold, et al. The functional mockup interface for tool independent exchange of simulation models. In *Modelica'2011 Conference, March*, pages 20–22, 2011.
- [2] T Blochwitz, M Otter, J Akesson, M Arnold, C Clauß, H Elmqvist, M Friedrich, A Junghanns, J Mauss, D Neumerkel, et al. Functional mockup interface 2.0: The standard for tool independent exchange of simulation models. In *9th International Modelica Conference, Munich*, 2012.
- [3] D.M. Auslander. Distributed system simulation with bilateral delay-line models. *Journal of Basic Engineering*, pages 195–200, June 1968.
- [4] P.B. Johns and M.A. O'Brian. Use of the transmission line modelling (T.L.M) method to solve nonlinear lumped networks. *The Radio And Electronic Engineer*, 50(1/2):59–70, 1980.
- [5] R. Braun, P. Nordin, B. Eriksson, and P. Krus. High Performance System Simulation Using Multiple Processor Cores. In *The Twelfth Scandinavian International Conference On Fluid Power*, Tampere, Finland, May 2011.
- [6] M. Axin, R. Braun, A. Dell'Amico, B. Eriksson, P. Nordin, K. Pettersson, I. Staack, and P. Krus. Next Generation Simulation Software Using Transmission Line Elements. In *Fluid Power and Motion Control*, Bath, England, October 2010.
- [7] B. Eriksson, P. Nordin, and P. Krus. Hopsan NG, A C++ Implementation Using The TLM Simulation Technique. In *The 51st Conference On Simulation And Modelling*, Oulu, Finland, 2010.
- [8] A. Jansson, P. Krus, and J-O Palmberg. Variable time step size applied to simulation of fluid power systems using transmission line elements. In *Fifth Bath International Fluid Power Workshop*, Bath, England, 1992.
- [9] OpenModelica website. <https://www.openmodelica.org/>, July 2013.

Utilizing the simulator fidelity to leverage fit for purpose in marine and process automation

Ole Andre Gjerpe, Ole Kristian Bjerknæs and Øivind Rui

Kongsberg Maritime, Kongsberg, Norway

Through three decades KONGSBERG has utilized different tools for simulation. This includes maritime operational trainers, process trainers, military trainers, missile and environment simulators, military performance simulators, dynamic process simulators, vessel hydrodynamic simulators and power system simulators, computer instruction set simulators etc. The simulators serve very different purposes ranging from internal development and test, process engineering, military capability evaluation, system acceptance testing and operator training. Investment in a dynamic simulator has become accepted best practice for O&G upstream process facilities with multiple reuse of the model during the asset lifecycle. For maritime automation and control systems simulators has not achieved the same level of acceptance. This might have caused in high level of digital control in these applications. For leased production units and LNG applications dynamic simulators are utilized in establishing process designs.

For dynamic positioning and power management HIL/SIL simulators is increasingly used for testing and verification of control systems. The HIL system's main purpose is the testing of failure tolerance and robustness of the established control system. This need has been identified for the maritime applications in general but the cost of establishing a process simulator model or a HIL test system is substantial and not generally accepted by the yards/ship owners. By combining the basic framework already in place for simulation and the generic production tools for applications one can produce a "simulator model" from the control system as the engineering model. Reuse of the information gathered for building the control system is currently sufficient to build time synchronized field simulators for a selection of control applications. The models for applications vary in accuracy and fidelity, but serve the purpose in the maritime industrial manufacturing chain. End user interest shows that the models also serve the purpose for operational training.

Keywords { Cost effective, Automatic generation of the model, Maritime control systems, Clock synchronized with single step control, Initial conditions for automation control & model, Dynamic models, Field operated software devices }

1 INTRODUCTION

Kongsberg Maritime's automation system AIM (Advanced Integrated Multi function) is a block based system. A customer-supplied control system consists of several software blocks which are connected, parameterized and configured to perform a specific task. This might be opening a valve, start a motor, perform power management, or do safety application tasks. The phases of delivering a control system consist of pre-engineering, engineering, production, installation, commissioning and operation. Much time and effort is placed in the first 3 phases. The first phase is the basis for the vessels operating philosophy. During the next phases, a control application is produced and ends up ready for verification and test.

The verification of the final control applications is performed as a factory acceptance test (FAT) witnessed by the customer. In this early phase the simulator is used as verification tool. The essential in this test is that the verification is performed almost identically to the real life with respect the electrical input/output (I/O) interface of the control system. It is also important that the control applications and simulator applications communicates so that they are synchronized with respect to time. This enables us to execute each computer step (sample time) and study details in the control system logic.

2 DRIVERS

Through three decades KONGSBERG has utilized different tools for simulation. The simulators span maritime operational trainers, process trainers, military trainers, missile and environment simulators, military performance simulators, dynamic process simulators, vessel hydrodynamic simulators and power system simulators, computer instruction set simulators etc.

The simulators serve very different purposes ranging from internal development and test, process engineering, military capability evaluation, system acceptance testing and operator training.

One main reason for utilizing simulators is that real life testing or training is not possible or at least not affordable.

The willingness to invest in a simulation environment depends on several factors that all in the end boil down to one single question – is the simulator investment profitable. In many cases are simulators very low cost compared to real life – for both testing and training. As examples think of maritime operational trainers letting student access a virtual ship and environment instead of having to use both a ship and wait for the correct operational conditions required for training, or a simulator for testing missiles where both the missiles and firing range and other equipment are very expensive in real life testing.

Investment in a dynamic simulator has become accepted best practice for O&G upstream process facilities with multiple reuse of the model during the asset lifecycle. The simulator is initially used for process design and verification. Further the simulator is used for control system check-out in testing phases of a project this might also include tuning activities for regulator control loops normally not possible until connected to the real process. Thereafter an updated process simulator can be a tool for both operator training, modification studies and for online simulation of the process – enabling look ahead, prediction and advanced model based control. This extensive reuse of the simulator model has through projects proved its value and ROI and is established as best practice for upstream Oil & Gas projects.

For maritime automation and control systems, simulators have not achieved the same level of acceptance. For leased production units and LNG applications the building and tuning of control applications for new process concepts has required a dynamic simulator as the best tool available. Examples are boiler control applications, regasification units and reliquefaction plants. The simulator investment has enabled the control system engineers to test control application up front having a higher quality system available for system start-up. Further the simulators have several times been proved useful when also the process design has met last time challenges that are not easily understood.

For dynamic positioning and power management HIL/SIL simulators are increasingly used for testing and verification of control systems. The HIL system's main purpose is the testing of failure tolerance and robustness of the established control system. The value of process simulators is just as obvious in maritime applications but the cost of establishing a high fidelity process model or a HIL test system is substantial and not generally accepted for maritime projects. In cases where new concepts are introduced, that typically are to delivered in series, the model cost can be distributed over several projects and by that capture the early phase benefits related to process verification, concept validation, control solution development.

The later stage benefits as control system check-outs, operator training, test simulation for late control system changes etc. are not as easily captured as the integrating process models for all parts for all ships in a series is not readily available.

By combining the basic framework already in place for simulation and the generic production tools for applications, it is feasible to produce a lower fidelity “simulator model” from the control system as the engineering model. The models for applications vary in accuracy and fidelity, but serve many of the late phase purposes for maritime industrial manufacturing chain. The current simulator production cost is suited also for maritime projects and are believed to fill the gap as visualized in Figure 1 Some selected marine/offshore simulator markets that until now has not been reachable with previous simulation technologies.

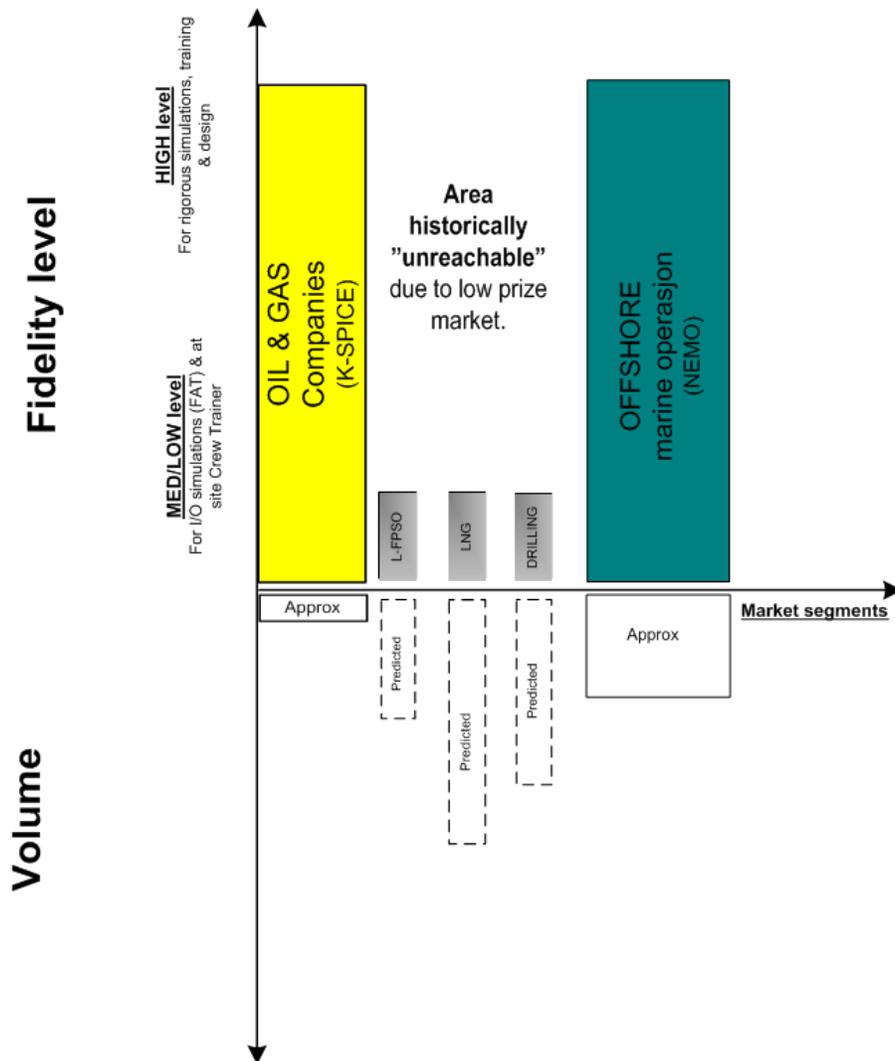


Figure 1 Some selected marine/offshore simulator markets

3 SIMULATION MODEL CONCEPT

The control system applications are developed and manufactured as visualized in yellow block left in the Figure 2 Simulation concept. The control applications communicate with dedicated software blocks reflecting the field equipment to be controlled. These control applications can be executed either on a soft controller or a field installed controller (Remote Controller Unit). In the control system AIM, there is made one corresponding simulator module for each of the real-time controller module. A complete new controller is automatically created and a script is executed in order to "reflect" the control application delivery and the parameter settings based on the data contained in the real time control modules. The result of this process is marked as yellow / green in the same figure.

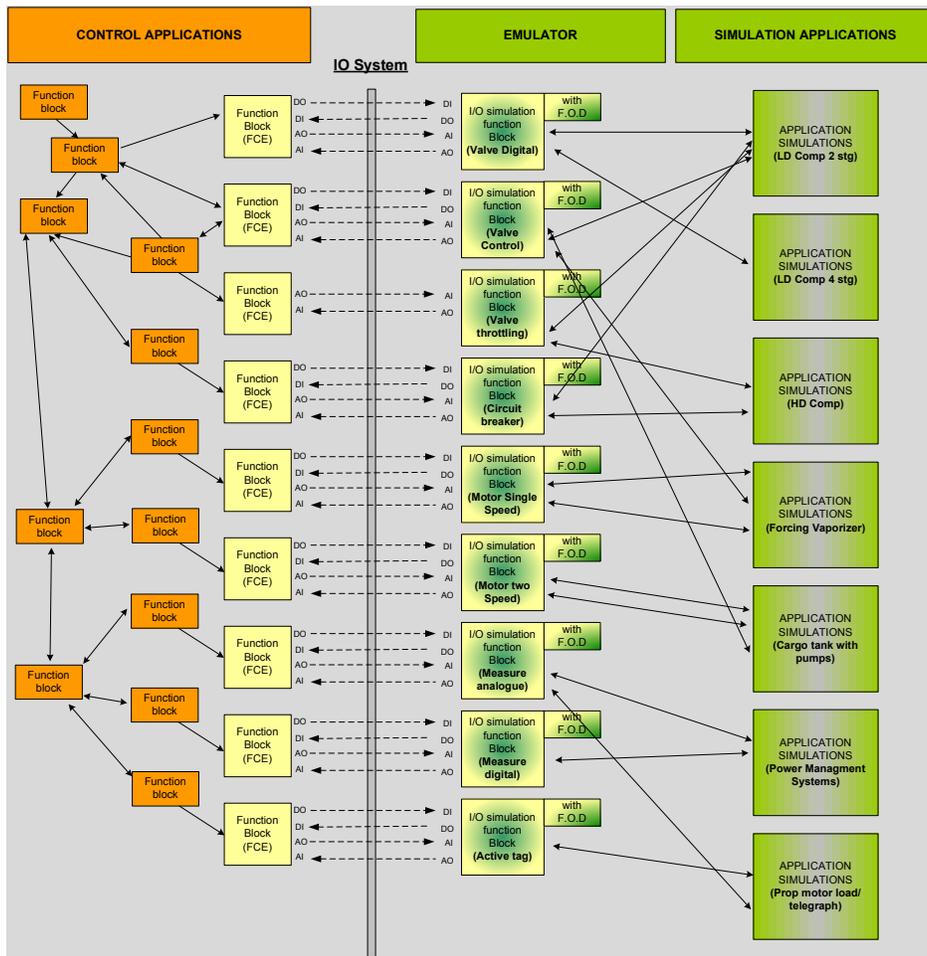


Figure 2 Simulation concept

The real time controller units communicates through the I/O interface. This is time-synchronized within the simulator application (Figure 3 Emulator/Simulator task control). This system feature is the same as the system features used for high fidelity dynamical simulations when integrated to a dynamic process simulator.



Figure 3 Emulator/Simulator task control

In certain simulation cases, it is necessary to create dynamics based on logical blocks (green blocks). While for other applications such as power management or thruster control, most of the dynamics are built into each dedicated simulator module.

When a control application runs for the first time, it is normally controlled into different operating conditions and stored. This enables the simulator system to load up any stored condition (initial condition) from exactly the same state as stored. This can be used for retesting the same scenario after control logic changes or simulator optimisation. On the simulator side, this can typically be updating or adding curves from empirical data or from equipment datasheet.

In time critical control applications such as power management or thruster control, simulation can be performed and executed stepwise (as per real sampletime). This makes a more accurate study of bump less transfer with respect to the dynamic response. In power management this can typically be load sharing, load dependant start / stop, blackout scenarios or loss of diesel generators during critical marine operations (e.g drilling).

A familiar class of control systems for Kongsberg Maritime is dynamic positioning of vessels. The dynamic positioning system for any selected vessel with the vessels automation/safety applications (exact copy) can run together as shown in Figure 4 Safety&Automation with dynamic positioning. This can be linked through network I/O (identical to the actual installation). The DP system has a corresponding simulation model that reflects the I/O and the model is auto parameterized. The data are collected and obtained from the delivery. There are other methods during project which verifies that correct data are collect from the engineering database.

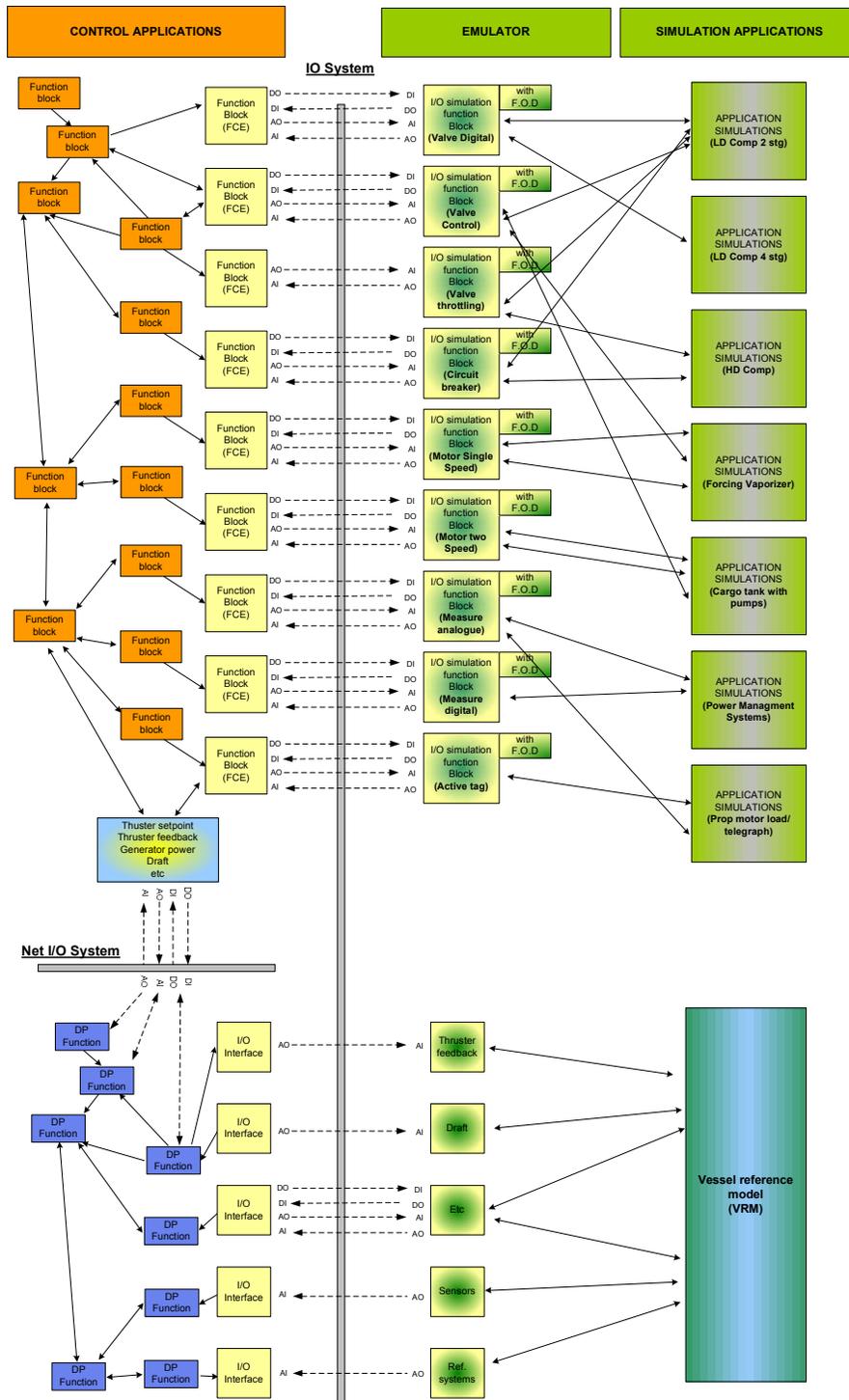


Figure 4 Safety & Automation with dynamic positioning.

The model used for verification in dynamic positioning of vessels, is also a "fit for purpose" model. The model is called Vessel Reference Model (VRM). The intention of the model is to test and verify the most common operating modes on a vessel. Due to standardisation of the control software for automation, safety and dynamic positioning are operating on the same platform, an installation wide verification between the different control applications can be performed in this technical framework.

4 APPLICATION AND DISCUSSIONS

A few selected cases of control applications that are repeatedly delivered and where the emulators have proved useful in the system delivery process. These examples are from advanced LNG applications as SRVs, FSRUs and drilling applications with combinations of power management and Dynamic Positioning.

4.1 Control application: Dual Fuel process

In Figure 5 Verification of a dual fuel ship an automation system for a LNG carrier ship can be seen. Pressurized gas is fed into the engines that drive electric generators. The gas is collected from the boil-off from huge liquefied natural gas (LNG) tanks on the ship. The boil-off gas is increased to 4-5 bar through the compressors. The motors that drive the generators can also be fuelled by marine diesel oil or heavy fuel oil in case of loss of gas pressure or for other operational reasons. This generates power to the main switchboard and is distributed through the power management system to the electrical consumers such as main propulsion of the ship.

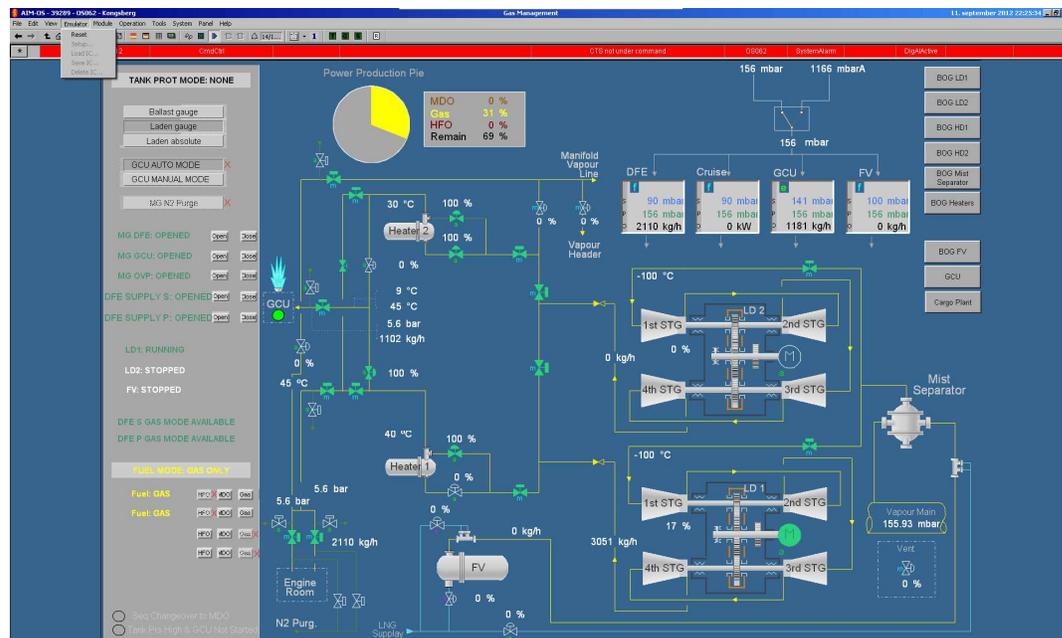


Figure 5 Verification of a dual fuel ship

4.2 Simulator application: Dual Fuel process

When the control application is produced and I/O points are configured, the I/O points are reflected as simulator blocks. This is completely separated from the control logic. The dynamics for the power management are automatically generated. Control applications containing compressors, flow, temperatures, pressures or other thermodynamic properties are applied from a library and adapted through logical blocks such as integrators, Boolean algebra, etc. The adaptations are based on expected behaviour based on field experience from similar applications. In some cases adaptation of dynamical behaviour could be collected from a high fidelity dynamic simulator based on thermodynamics and network solvers.

In Figure 5 Verification of a dual fuel ship, an initial condition is loaded from the simulator system. An initial condition is an operational condition in which the dynamical variables are saved to file. The compressors are running on gas, the gas combustion unit is burning excess gas and the PID controllers have automatically switched to auto and track, etc. The control application (Gas Management System) communicates through the I/O interface with the mirrored modules. Some possible missing I/O from the engineering phase would now be detected as errors/alarms with respect to lack of feedback from the limit switches, position sensors etc.

For the purpose of test and verification, the simulator contains field operated devices (FOD). This is common in industrial simulator technologies. The purpose is to introduce different failures or different scenarios. The variables and parameters are saved together with initial conditions corresponding to the control application under verification. Typical FOD's for field equipment in such simulator applications can be seen in Figure 6 Simulator panel: Analogue measurement, Figure 7 Simulator panel: Speed starter or Figure 8 Simulator panel: Control valve.

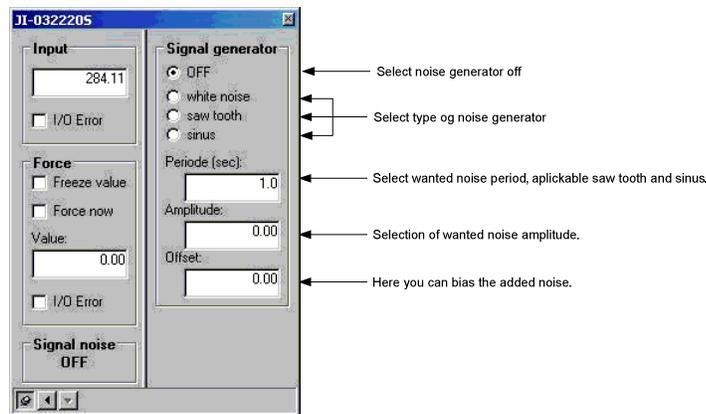


Figure 6 Simulator panel: Analogue measurement

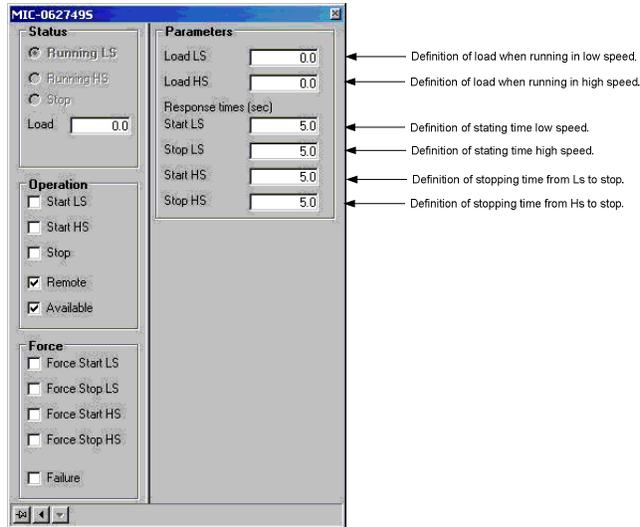


Figure 7 Simulator panel: Speed starter

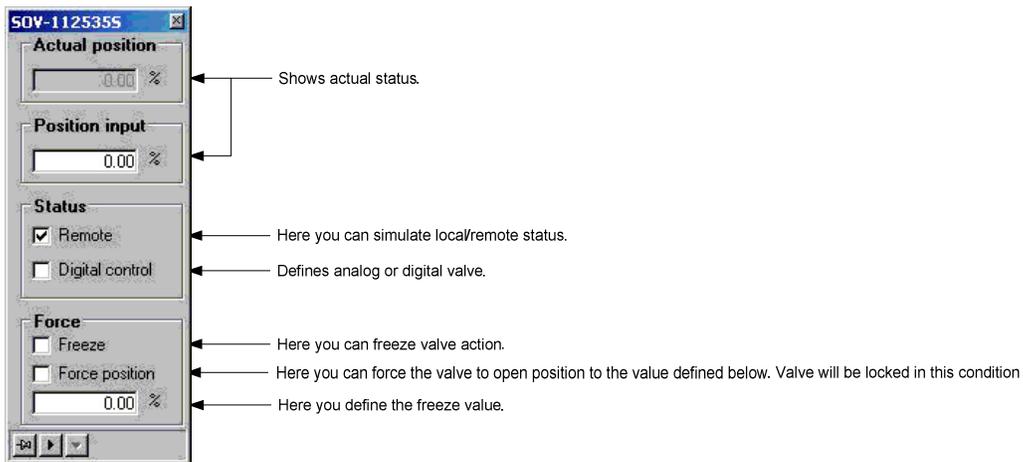


Figure 8 Simulator panel: Control valve

4.3 Control application: Power Management

The task of a power management system on a vessel is to make sure that sufficient electric power is available for the electric consumers on the vessel in different operational scenarios. A power management system under test and verification are shown in Figure 9 Verification of a Power management on a vessel

On a vessel, there exists equipment's such as diesel electric generators, electrical switchboards, electrical breakers, transformers and large electric consumers. The consumers could be thrusters, engines, process equipment or other "hotel load". In critical marine operations, confidence in a power management system is crucial. The system must handle error conditions such as failures in diesel electric generators, error condition in the breakers, unusually heavy loads, short-circuits, etc. A control strategy for failures and alternative electrical routing must be performed and be reliable. Some errors can be detected by the field equipment itself, others must be detected upon real-time model based approaches. A typical action by the controls system could be to get rid of non critical loads or fast re-routing electrical power.



Figure 9 Verification of a Power management on a vessel

4.4 Simulator application: Power Management

The simulator for the power management system is almost fully auto generated. Parameters from the control application is automatically collected from application configuration and passed through a calculator applying rule sets and then written into the simulator modules. These parameters are typically electric power generated by each generator, electric power consumed by consumers, etc.

Failure scenarios and tests related to the power management systems are important on vessels. The reasons are obvious due to the effects of a blackout during critical marine operations. An extensive control application test on an installation can be very demanding for the equipment due to tear & wear on the actual installed equipment. Repeated tests due to control system issues are seldom popular due to the same reason. By introducing field operated devices (FOD) in a simulation environment, the control system application can be tested repeatedly and as many times as required without wearing out the equipment. Cross testing scenarios such as e.g. fall out of generators and at the same time starting heavy consumers onboard can be studied, discussed and tested in-house. Several FOD's are used for this purpose as shown in Figure 10 Simulator panel: Diesel generator, Figure 11 Simulator panel: Thruster panel and Figure 12 Simulator panel: Circuit breaker

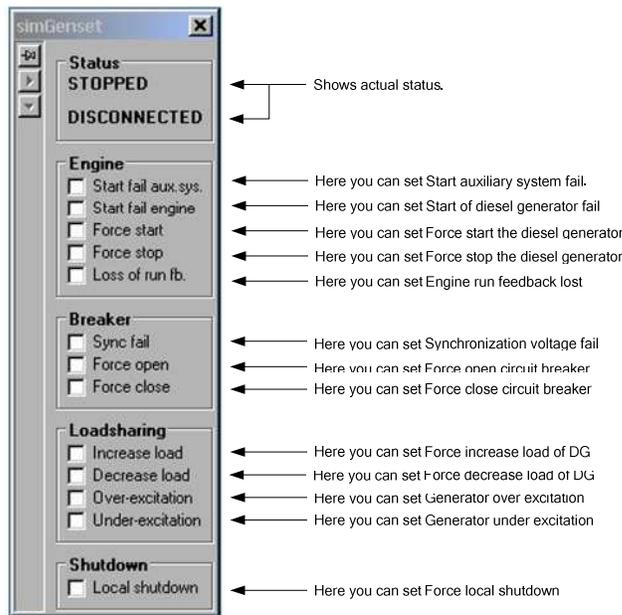


Figure 10 Simulator panel: Diesel generator



Figure 11 Simulator panel: Thruster panel

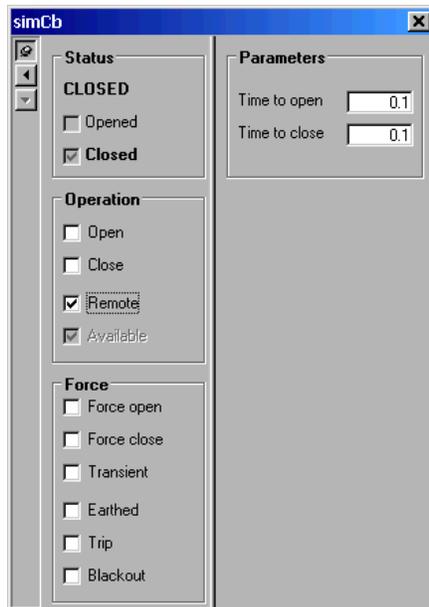


Figure 12 Simulator panel: Circuit breaker

5 CONCLUSIONS

The subject of this paper is model accuracy versus model purpose. KONGSBERG is, in addition to the simulator described here, a supplier of several types of high fidelity simulators for different purposes in the industry. If the purpose of using a simulator is field equipment design, that is at the forefront of which have not previously been physically proven, high fidelity simulators will be the suggested tool. For the purpose of verification and testing of control applications that are not highly dependent on thermodynamic effects or rigorous modelling of equipment a simulator based on logical models are often sufficient hence the term "fit-for-purpose". The choice of fidelity level and what is "fit-for-purpose" is decided together with the customer. The choice is determined by factors such as time, finances, experiences, access to accurate data or access to empirical measurements.

Power management control applications provide a greater degree of digital control. The dynamics in such an application are more related to generators, thrusters and other rotating equipment with a defined inertia. The dynamic behaviour of such equipment can often be described by using curves from vendor or from field collected data. The efficiency of using an "emulator" or a medium fidelity simulator is often larger in such applications compared to a complicated hydrocarbon process application. This is due to the few dependencies between the simulated field components and the applications lack of thermodynamics.

System functionality in the "emulator" such as task control, clock synchronization, single step, loading and storage of initial conditions are important features in order to ensure higher quality in control applications. It allows the user to do control application corrections and repeat the same scenario several times and study detailed control-related transitions. This could be bump less transfer between modes, timing between different control applications, propagating dynamic effect of shutdown scenarios or scenarios that was not thought of in the engineering phase of the vessel. Task control and clock synchronization also allows the user to expand the simulator with a more rigorous simulation model that requires greater computing capacity. This feature is advantageous due to that the control application and the simulator application are always following each other, hence reliable simulation results.

This type of simulator (medium fidelity) has currently been applied in several marine control projects and application areas. In some cases it could be discovered that the engineering substrate is incomplete, incorrect or that control applications can be further optimized. This is normally discovered during or before the internal acceptance test (IAT). A typical action can be re-tuning, optimization by parameterization or revising the control philosophy. The use of this type of simulator has resulted in a noticeable quality improvement and reduced variance in the commissioning phase of field equipment.

6 ACKNOWLEDGEMENTS

We want to thank several of our marine customers and involved yards for valuable input and improvement of the simulator. This input comes due to the tight relationship and mutual understanding of correct quality delivered to the right time. Without this collaboration, the development of the simulator couldn't have been done.

Design and simulation of a small scale experimental facility for drilling operations

Emil Cimpan¹ and Gerhard Nygaard^{2,3}

¹Bergen University College, Nygårdsgaten 112, NO-5020 Bergen, Norway

²University of Stavanger, Department of Petroleum Engineering, NO-4036 Stavanger, Norway

³Cybernetic Drilling Technologies AS, Thormøhlensgate 55, NO-5008, Bergen, Norway

Abstract:

During drilling operations the understanding of the drilling fluid hydraulics dynamics is important for optimizing the pressure gradient in the wellbore. To enable online estimates of the downhole pressure gradient, a small scale experimental facility is designed. The design is based on simulation of the dynamics behavior of the drilling fluid, both for the real full scale drilling facility and the small scale experimental facility.

The results of the simulations enables the possibility of designing a small scale facility that best replicates the full scale facility regarding pressures and fluid flow rates. This enables testing of a novel instrumentation methods referred to as the Instrumented Standpipe System, that is used to estimate the downhole wellbore properties such as pressures and flow rates.

1. INTRODUCTION

The drilling fluid hydraulics dynamics are influencing the downhole pressure conditions during drilling operations. To maintain a correct pressure in the wellbore during drilling, the main parameters such as density and rheology the drilling fluid should be monitored continuously.

Today, the common practice during drilling operations is to identify the various drilling fluid parameters using a sample of the drilling fluid. The density is found by weighing a known volume at atmospheric pressure. A rheology value is typically determined by measuring the time a mud sample takes to run out of the Marsh funnel. The dynamic viscosity of the sample would then be assessed by using a parameter table that correlates the measured flow duration with dynamic viscosity. More accurate rheology measurements would be found by using a standard Fann viscometer.

However, drilling fluids are typically having a non-Newtonian dynamic behavior, and other methods of defining the rheology should be found.

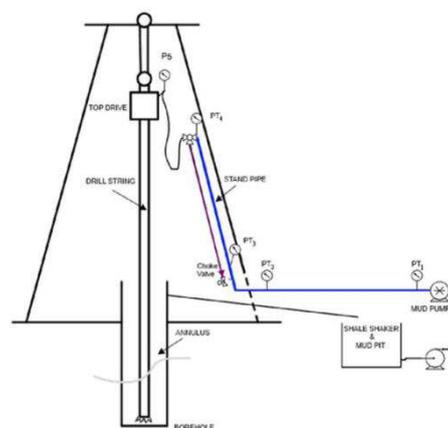


Figure 1. Schematic view of an instrumented standpipe system. The horizontal differential pressure transmitter provides PT1 and PT2 measurements while PT3 and PT4 are given by the vertical differential pressure transmitter. From [Nygaard, 2011].

Direct measurement of rheology of non-Newtonian drilling fluids based in drilling fluid samples is challenging. Another method for providing rheology measurements focusing on differential pressure transducers has been discussed in literature the recent years. In [Nygaard, 2011] the rheology of the drilling fluid is calculated by using differential pressure measurements on both a horizontal section and a vertical section of the standpipe system. A schematic of such a system, often referred to as the “Instrumented Standpipe System” is presented in Figure 1.

The method would benefit from the robustness, the reliability, and the high accuracy of pressure sensors that have been significantly improved over the last decade.

First, by using the horizontal differential pressure readings $P_H = PT_1 - PT_2$ over the separation distance L between the horizontal pressure sensors PT_1 and PT_2 , mounted on a pipe with diameter D , the Darcy friction factor of the drilling mud would be calculated using

$$f_D = \frac{2DP_H}{\rho L v^2} \quad [1]$$

Thus, instead of calculating the friction factor of the mud from measurements of the mud viscosity through manual conventional methods and then calculating the resultant frictional pressure, it is easier and more accurate to measure directly by automation means the mud frictional pressure drop over a certain horizontal section.

Further, by using again the horizontal differential pressure measurements between PT_1 and PT_2 together with the difference in pressure readings $P_V = PT_4 - PT_3$ over the vertical pipe height h between PT_3 and PT_4 (see Figure 2), the density of the drilling fluid mixture could be obtained from equation [1]

$$\rho = \frac{P_V - P_H}{gh} \quad [2]$$

Since the flow velocity v would in principle be determined by measuring the pressure drop over a restriction in the flow path, such as the corner connection between PT_2 and PT_3 , the instrumented standpipe system could be used

for flow measurements in addition to the estimation of the mud characteristics. However, this method of measuring the flow rate was not investigated in this work. The flow velocity of the drilling fluid could be easily measured directly by using the existing stroke counter of the pump.

Since the volume flow rate $Q = Av$, where A is the cross-sectional area of the pipe, the apparent viscosity could also be expressed from [10] as

$$\mu = \frac{\rho Q D}{\text{Re } A} \quad [3]$$

When calculating the apparent viscosity, the flow regime should be identified to obtain the correct values. The friction factor was used further to calculate the Reynolds number. The Reynolds number corresponding to laminar flow is given by

$$\text{Re}_{\text{lam}} = \frac{64}{f_D} \quad [4]$$

The Reynolds number for turbulent flow can be given as

$$\text{Re}_{\text{turb}} = 6.9 \left[10^{(-1/1.8\sqrt{f_D})} - \left(\frac{\varepsilon/D}{3.7} \right)^{1.11} \right]^{-1} \quad [5]$$

However, in the transition region of the flow regime for which $2300 < \text{Re} < 3000$, the flow of the drilling fluid is either laminar or turbulent. Since the friction factor will have approximately the same value over the transition region, information about the flow regime is needed to calculate the Reynolds number when using the instrumented standpipe. Afterwards, the viscosity of the fluid can be calculated using equation [3].

Finally, the characteristics of the drilling fluid, the fluid velocity together with a hydraulic model of the drilling fluid could be employed to assess the BHP of a drilling well.

Some initial simulations and experiments have been performed by Carlsen and fellow authors, [Carlsen et al, 2012a], [Carlsen et al, 2012b] and the PhD dissitation [Carlsen 2012]. Carlens work indicates that this measurement principle should be examined further, both in academia and in industry. In order to train new students in this principle of drilling fluid properties

evaluations, several recent works the last year have also focused on using simulations for designing a small scale experimental facility at both the University in Stavanger and Bergen University College.

This paper presents various efforts in designing and evaluating the drilling fluids measurement concept focusing on small scale experimental facilities at educational institutions. The next chapter presents the work by Hansen [Hansen, 2012], which performed the initial experiments using water as base fluid. Chapter three presents the follow-up work by Krogsæter [Krogsæter, 2013] where a more robust sensor with diaphragm seals were used, evaluating various different pipe dimensions. The next chapter presents the work performed by Kråvik, Sætersdal and Hougaard, [Kråvik et al., 2013] where a calculation and user interface for such a measurement system is simulated and developed. The last chapter concludes the various findings and gives suggestions for further work.

2. EXPERIMENTAL SIMULATIONS USING STANDARD DP SENSORS

At the University of Stavanger, a small-scale drilling rig has been constructed for simulating the hydraulics of a drilling operation. The measurement principle using two differential pressure transmitters have been used, as shown in Figure 2. A standard capillary tubing between the process is used.

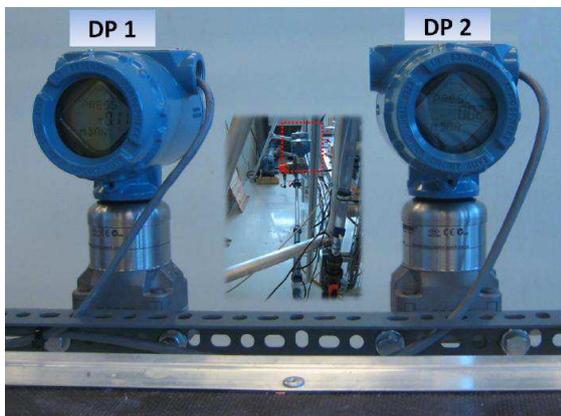


Figure 2 Actual sensor mounting at small scale rig at UiS. From Hansen, 2012.

When evaluating the measurement results and the theoretical findings, it was found that there was only small deviation, as shown in Figure 3. The distance between the process connection point were approx. 1 meter.

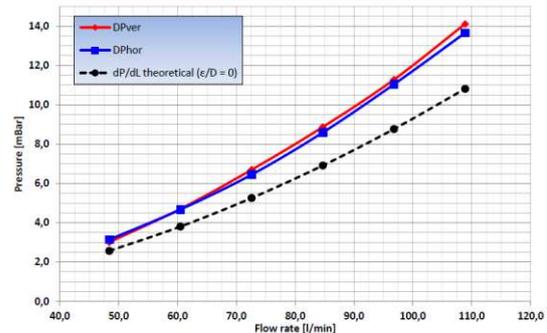


Figure 3 DP measurement results using standard DP sensor. From Hansen, 2012.

To reduce the uncertainty of the interface between the process and the sensor in a real application using drilling fluid, it was decided to upgrade the sensor with factory mounted diaphragm seals, and longer distance between the process connection points.

3. EXPERIMENTAL SIMULATIONS USING DP SENSORS WITH DIAPHRAGM SEALS

A revised and larger setup of the instrumented standpipe setup were designed and implemented by Krogsæter, [Krogsæter, 2013].

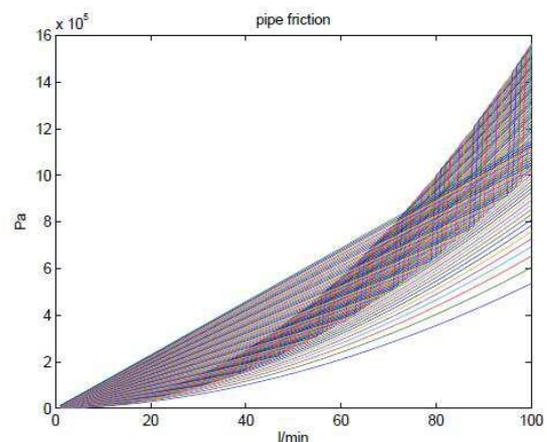


Figure 4 theoretical calculation of differential pressure using various viscosity values.

A theoretical evaluation of the differential pressure loss at various flow rates and various pipe diameters were examined, as presented in Figure 4. The multiple lines represent various viscosity values.

The revised design was constructed and a photo from the laboratory is shown in Figure 5.



Figure 5 Experimental setup using DP sensors with diaphragm seals. From Krogsæter, 2013.

The results show good correlation with the estimated and theoretical values of density, as shown in Figure 6. The various lines indicate different flow rates.

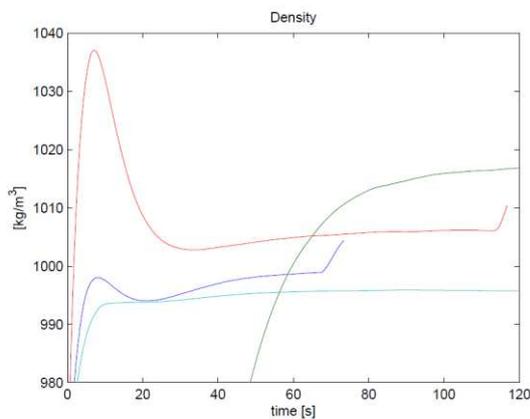


Figure 6 Estimated densities. From Krogsæter, 2013.

The results show also good correlation with the estimated and theoretical values of viscosity, as shown in Figure 7. The various lines indicate different flow rates.

The results s indicated also that the flow sensor of the pump should be improved or calibrated in order to obtain better estimates of both density and viscosity.

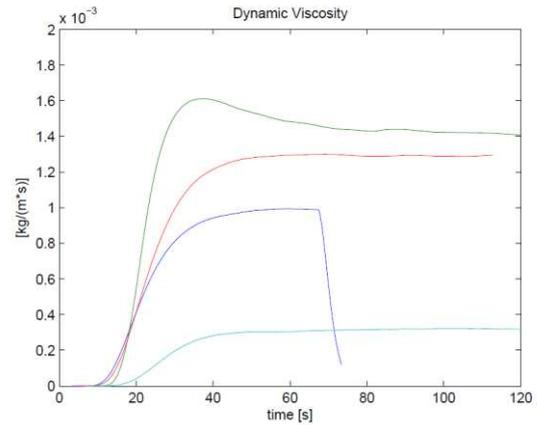


Figure 7 Estimated viscosity, from Krogsæter, 2013.

4. DESIGN AND SIMULATIONS OF REAL-TIME DRILLING FLUID MONITORING SYSTEM

A small-scale drilling rig that is to be used in tests with the instrumented standpipe system is under construction at the Bergen University College. A student group consisting of Kråvik, Sætersdal and Hougaard [Kråvik et al., 2013] designed a drilling fluid monitoring system. Only the physical drilling rig remains to be completed, the whole monitoring system including the data acquisition, transfer, command and control system of the rig implemented in a human machine interface (HMI) have already developed and used in this work.

The physical drilling rig data is simulated using the Kaasa simplified wellbore model, [Kaasa et al., 2012]. The monitoring system is comparing the “physical data” delivered by the differential pressure transmitters placed on the small-scale drilling rig with the simulated data obtained from the drilling process model. Since the drilling rig is under construction, the “physical data” delivered by the differential pressure transmitters PT_1 and PT_2 were not measured but estimated using the formulas developed for the instrumented standpipe system.

4.1. Monitoring system design

The schematic diagram of the control and automation of the small-scale drilling rig

equipped with an instrumented standpipe measurement system is shown in Figure 8 together with some of the physical elements.

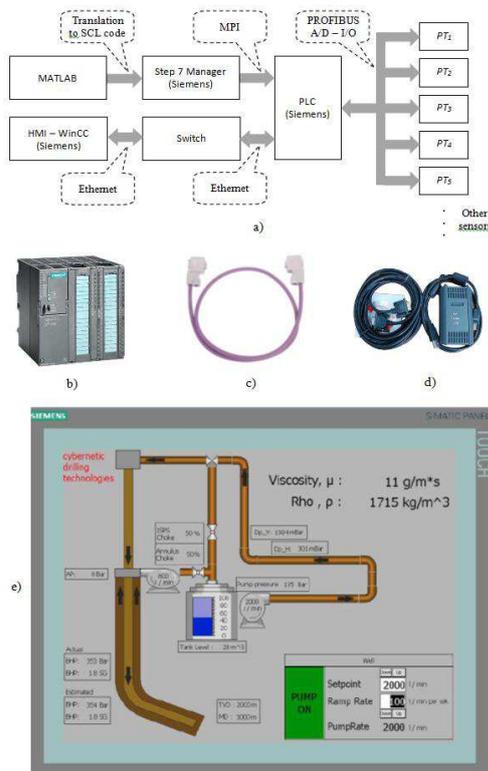


Figure 8 (a) System diagram of the drilling process model implemented in (b) PLC (Programmable Logic Controller) together with the instrumented standpipe system. The formulas were developed in MATLAB and transferred in the PLC through the SCL (Structured Control Language). The communication between the sensors and the PLC is handled by way of either (c) PROFIBUS (Process Field Bus) standard or (d) MPI (Message Passing Interface) Siemens standard. The control interface with the system is implemented in an (e) HMI (Human Machine Interface) panel. Partly from [Kråvik et al, 2013].

PT_1 and PT_2 are the horizontal pressure transmitters, PT_3 and PT_4 are the vertical pressure transmitters and PT_5 measures the swivel pressure (see also Figure 1). Other sensors may be added to the system. The measurement data obtained from all the sensors or the control data to be transmitted to actuators go to and fro between sensors, actuators and the central processing unit represented by a Programmable Logic Controller (PLC). The data transfer is carried out by means of either input/output analog signals lines, the Ethernet protocol, the communication protocol PROFIBUS (Process Field Bus) or the Siemens standard MPI (Message Passing Interface). In this work, a

Siemens CPU314C – 2 PN/DP type PLC was used. The formulas to be used were developed in MATLAB and translated in SCL (Structured Control Language) so that to be transferred in the PLC, which uses the programming language Siemens SIMATIC Manager Step 7. The command and control interface with the system is implemented in the HMI (Human Machine Interface) that was designed using the Siemens program SIMATIC WinCC Flexible 2008 Compact. The HMI interface that in addition to the controlling inputs and the measuring data illustrates the details of the drilling process was implemented by using the 15" TP1500 Basic PN panel.

4.2. Simulation results

A test scenario using the instrumented standpipe system is presented in Figure 9.

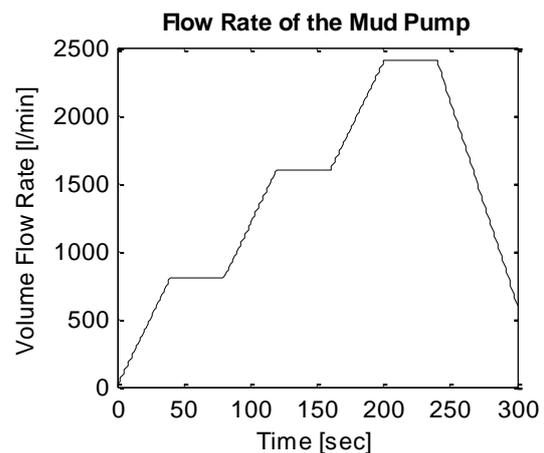


Figure 9 The volume flow rate developed by the mud pump in the ISPS over a time period of 300 time units. From [Kråvik et al, 2013].

The volume flow rate of the drilling fluid pump was changed according to ramp variations in a couple of series over a time period of 300 time units. The corresponding response in pressure in the measuring system while the choke valve is manually manipulated and partly closed in the 50% – 80 % range is shown in Figure 10.

In order to hold constant the pressure in the measuring system while the volume flow rate would be increasing in ramp fashion as shown in Figure 9, a PI (Proportional-Integral) controller was used to control the pressure delivered by the drilling fluid pump. The response of the mud

pump pressure control system over the time period of 300 time units is shown in Figure 11.

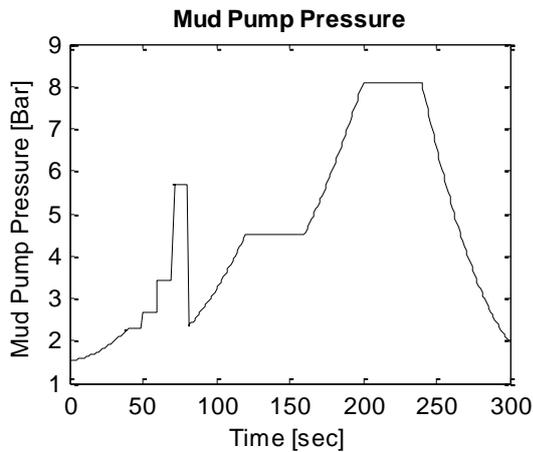


Figure 10 The pressure developed by the mud pump in the ISPS over a time period of 300 time units while the Choke Valve is varied in the (40 – 80) sec range. From [Kråvik et al, 2013].

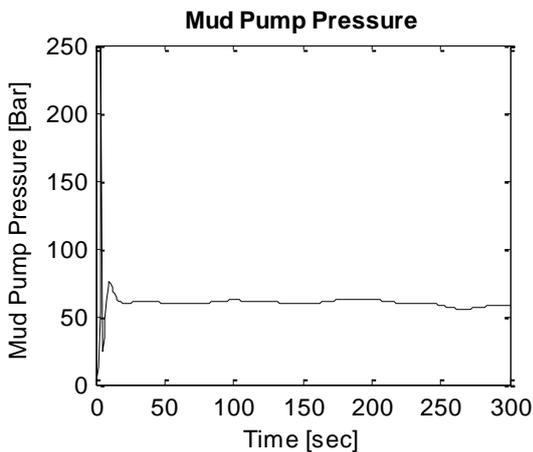


Figure 11 The response of the pressure control system assisted by a PI controller over the time period of 300 time units. From [Kråvik et al, 2013].

The relative higher variation in pressure at the beginning of the testing session is due to low values in both volume flow rate and pressure that challenged the PI controller, and the pressure fluctuations can be reduced by introducing a gain schedule scheme of the PI controller.

Over the same testing period in which the measuring system pressure was hold constant, the density of the drilling mud was estimated as shown in Figure 12. The estimated variation in the viscosity during the testing session over the 300 time units is shown in Figure 13.

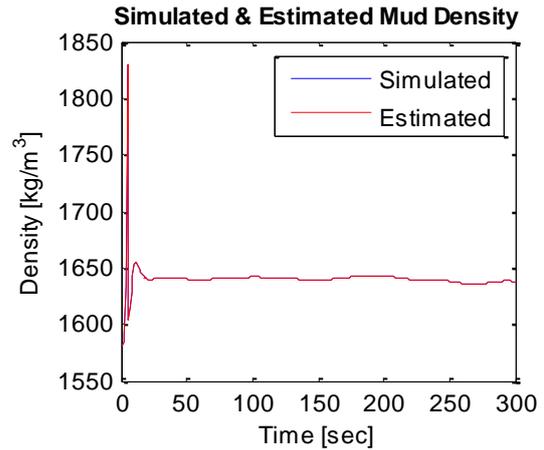


Figure 12 The estimated and the simulated change in the density of the drilling mud within the time period of 300 time units. From [Kråvik et al, 2013].

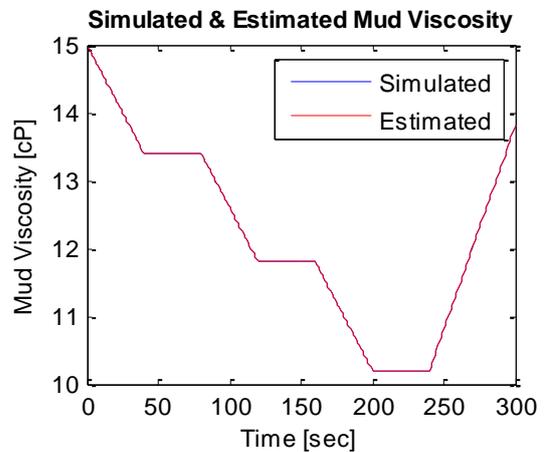


Figure 13 The estimated and the simulated change in the viscosity of the drilling mud within the time period of 300 time units. From [Kråvik et al, 2013].

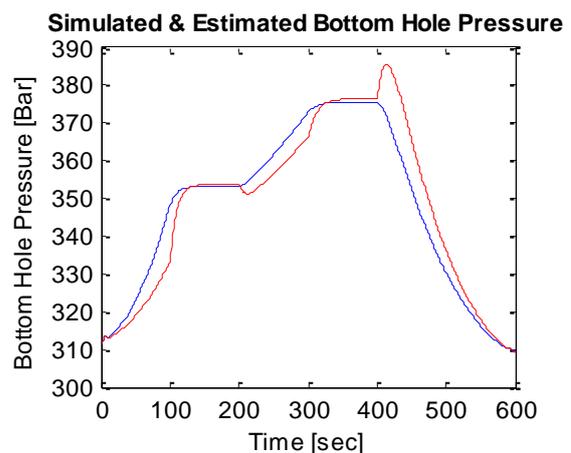


Figure 14 Overall comparison between the simulated values of the BHP by using the Kaasa model and the estimated values of the BHP based on the instrumented standpipe method over a time period of 600 time units. From [Kråvik et al, 2013].

When comparing Figure 13 with Figure 9, it can be seen that the decrease in the viscosity value of the drilling fluid while the volume flow rate is increasing is in accordance with the shear thinning effect characteristic to the pseudoplastic liquid flow, as it is the case for the majority of the mud solutions.

As an overall comparison, the simulated values of the BHP by using the model and the estimated values of the BHP based on the instrumented standpipe method over a time period of 600 time units are shown in Figure 14.

5. CONCLUSIONS

The results of the simulations indicate that the measurement principle of using direct differential pressure transmitters will give the required online measurements required to mix and condition the drilling fluid automatically during the drilling operation.

Further developments of the instrumented standpipe method will be carried out using the small-scale drilling fluid model. The results of the simulations and tests carried out by utilizing the small-scale drilling fluid system will be compared with real field testing performed in offshore drilling systems.

6. REFERENCES

- GO Kaasa, ØN Stamnes, L Imsland and OM Aamo, Simplified Hydraulics Model Used for Intelligent Estimation of Downhole Pressure for a Managed-Pressure-Drilling Control System. *Society of Petroleum Engineers (SPE) Drilling & Completion*, March, 2012;
- G. Nygaard, "Evaluation of automated drilling technologies developed for petroleum drilling and their potential when drilling geothermal wells", presented at GeoEnergy2011, Bergen, 7.-8. September, 2011.
- L.A. Carlsen, G. Nygaard, R. Time, "Utilizing Instrumented Stand Pipe for Monitoring Drilling Fluid Dynamics for Improving Automated Drilling Operations", Proceedings of the 2012 IFAC Workshop on Automatic Control in Offshore Oil and Gas

Production, Norwegian University of Science and Technology, Trondheim, Norway, May 31 - June 1, 2012.

- E. Hansen, "Automatic evaluation of drilling fluid properties", MSc thesis, University of Stavanger, June, 2012.
- L.A. Carlsen, N. L. Rolland, G. Nygaard, R. Time, "Simultaneous Continuous Monitoring of the Drilling-Fluid Friction Factor and Density", *Journal SPE Drilling & Completion*, Vol. 28, No. 1 pp. 34-44, March, 2013.
- L.A. Carlsen, "Automating Well Control Procedures", PhD Thesis, No. 185, University of Stavanger, February, 2013.
- G. Nygaard, J.-M. Godhavn, "Automated Drilling Operations", University lecturing booklet, University of Stavanger and Norwegian University of Science and Technology, April, 2013.
- M.P. Kråvik, H. Sætersdal, L.W. Hougaard, "Drilling fluid monitoring during MPD operations", BSc thesis, Bergen University College, June 2013.
- Kurt Louis Krogsæter, Automatic Evaluation of Drilling Fluid Properties Conventional and MPD Operations, MSc thesis, UiS, July 2013.

Simulation and Evaluation of the Drilling Fluid Mixing and Conditioning Process

Gerhard Nygaard^{1,3} and Emil Cimpan²

¹University of Stavanger, Nygårdsgaten 112, NO-4036 Stavanger, Norway

²Bergen University College, Department of Petroleum Engineering, NO-5020 Bergen, Norway

³Cybernetic Drilling Technologies, Thormøhlensgate 55, NO-5008, Bergen, Norway

Abstract:

During drilling operation the drilling fluid composition and properties are crucial in maintaining a safe and efficient drilling process. Various additives for adjusting the drilling fluid density and rheology are injected into the drilling fluid in order to maintain the required properties of the drilling fluid.

This paper evaluates a novel measurement principle of using direct differential pressure transmitters while mixing and conditioning the drilling fluid. The evaluation is performed by developing a numerical model of the mixing and conditioning process, including non-Newtonian rheology effects and drilling fluid compressibility.

The results of the simulations indicate that the measurement principle of using direct differential pressure transmitters will give the required online measurements required to mix and condition the drilling fluid automatically during the drilling operation.

1. INTRODUCTION

During drilling operations, the drilling fluid evaluation is typically performed manually by measuring various properties of a sample of the drilling fluid at the drilling fluid laboratory on the rig. Based on this manual evaluation, the crew on the rig calculates how the drilling fluid components should be adjusted to maintain certain density and rheology properties of the drilling fluid [Caenn et al, 2011].

The recent years, a novel method of automatically and continuous evaluation of the drilling fluid has been examined in several papers and reports. [Nygaard, 2011], [Carlsen et al, 2012], [Hansen, 2012], [Krogsæter, 2013], [Kråvik et al, 2013]. The method is based on employing dual differential pressure sensors between the drilling fluid pumps and the top drive, as shown in Figure 1. The measurement

setup is often referred to as the “instrumented standpipe” system (ISPS).

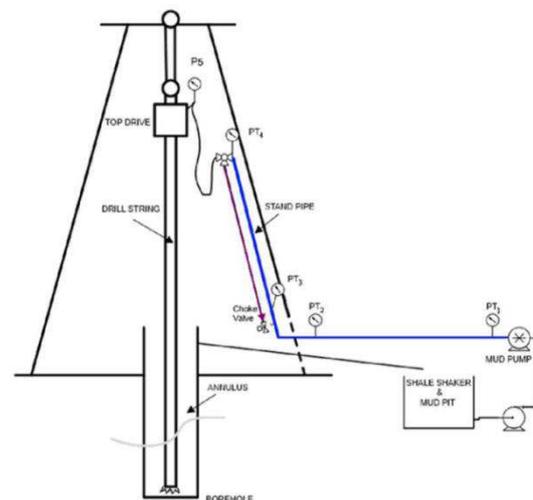


Figure 1 Schematic view of the instrumented standpipe system

Based on the horizontal differential pressure measurement and the vertical differential pressure measurement, both the density and the apparent viscosity may be calculated. Further details are found in [Carlsen et al, 2012].

During drilling operations, there will be a continuous need for adjusting both the density and the viscosity to maintain a certain pressure level in the wellbore. In [Nafikov and Glomstad, 2013] a novel method of controlling the drilling fluid density and rheology is described. However, the method for determining the drilling fluid rheology is not described in detail. To utilize a continuous method for controlling the main properties of the drilling fluid in the drilling fluid tank, continuous measurements of the drilling fluid rheology should be available.

This paper focuses on presenting a method for simulating the drilling fluid mixing and conditioning process during drilling operations including continuous measurements. The following section presents the dynamic model for drilling fluid mixing and conditioning in a drilling fluid tank, including densifying, viscosifying and diluting the drilling fluid. The properties of the tank fluid are measured using the instrumented standpipe setup. The third section presents and discusses the simulations of several operational cases, including both compressible drilling fluids and drilling fluids with non-Newtonian behavior. In section four, the conclusions and suggestions for further work are given.

2. MODELING DRILLING FLUID TANK MIXING AND CONDITIONING

The drilling fluid model is based on the Kaasa wellbore model [Kaasa et al., 2012], with extension of the drilling fluid tank unit and the instrumented standpipe system. The dynamic drilling fluid tank model includes modeling of densifying, viscosifying and diluting fluids.

The density model is similar to [Nafikov and Glomstad, 2013] and represented by

$$V\dot{\rho}_t = -\rho_t\dot{V} + \rho_i q_i + \rho_b q_b + \rho_v q_v + \rho_w q_w - \rho_t q_p$$

where ρ_t is the density of the tank fluid, V is the volume of the tank, ρ_i is the density of the fluid from the wellbore entering the tank, q_i is the volume flow rate of the fluid from the wellbore entering the tank, ρ_b is the density of the

densifying liquid, q_b is the volume flow rate of the densifying liquid, ρ_v is the density of the viscosifying liquid, q_v is the volume flow rate of the viscosifying liquid, ρ_w is the density of the diluting liquid, q_w is the volume flow rate of the diluting liquid, and q_p is the volume flow rate of the liquid out of the tank.

The dynamic viscosity model is based on the Arrhenius method for blending fluids with different viscosity as described in [Sutton and Bergman, 2012]. The Arrhenius method is based on

$$\ln(\mu_m) = \sum_{i=1}^n X_i \ln(\mu_i)$$

where μ_m is the mixture viscosity measured in centiPoise (cP), X_i is the volume fraction of the fluid with viscosity μ_i .

The term $\ln(\mu_i)$ can be expressed using the term "Viscosity Blend Index" (VBI), denoted I_i as in [Sutton and Bergman, 2012]. The dynamics of the viscosity in the tank, can then be given by

$$V\dot{I}_t = -I_t\dot{V} + I_i q_i + I_b q_b + I_v q_v + I_w q_w - I_t q_p$$

where I_t is the VBI of the tank fluid, I_i is the VBI of the fluid from the wellbore entering the tank, I_b is the VBI of the densifying liquid, I_v is the VBI of the viscosifying liquid, I_w is the VBI of the diluting liquid. The resulting viscosity could then be found using

$$\mu_t = e^{I_t}$$

where μ_t is the viscosity of the tank. Additional effects of the drilling fluid caused by temperature are not included in the present setup, but could be added in a future implementation of the model.

In addition to the mixing of density and viscosity, the compressibility effects of the drilling fluid in the wellbore are included using the simple relation

$$\rho = \rho_0 + \frac{\Delta\rho}{\Delta p} p$$

where ρ_0 is the density at atmospheric pressure and p is the pressure downstream the pump.

Similarly, the non-Newtonian effect of the drilling fluid in the wellbore is included in the model using the simple relation

$$\mu = \mu_0 + \frac{\Delta\mu}{\Delta q}q$$

where μ_0 is the viscosity at the nominal flow rate and q is the volume flow rate of the drilling fluid in the drillstring.

3. SIMULATING DRILLING FLUID DYNAMICS

To evaluate the mixing dynamics of the drilling fluid in the tank, the viscosity models and the density models were implemented in Matlab. The case used was a standard wellbore equipment setup in addition to managed pressure drilling equipment.

The various parameters of the drilling fluid model are presented in Table 1.

Table 1 Drilling fluid mixing model parameters

Wellbore vertical depth	2000 m
Wellbore length	3000 m
Tank volume	10 m ³
Drilling fluid initial density	1580 kg/m ³
Drilling fluid initial viscosity	0.023 Pa s
Densifying liquid density	4300 kg/m ³
Densifying liquid viscosity	0.023 Pa s
Viscosifying liquid density	2000 kg/m ³
Viscosifying liquid viscosity	0.100 Pa s
Dilluting liquid density	1000 kg/m ³
Dilluting liquid viscosity	0.001 Pa s

This section presents three different scenarios. The first scenario shows how the drilling fluid properties were measured at constant flow rate while adding densifier, viscosifier and dilluting the liquid. The second scenario shows how variations in the drilling fluid volume rate

change the density and the apparent viscosity due to compressibility and non-Newtonian behavior. The third scenario shows how the drilling fluid measurements vary, when the drilling fluid flow rate is changed, showing drilling fluid compressibility and non-Newtonian effects, in addition to adding densifier, viscosifier and dilluting liquid.

To obtain a realistic measurements scenario in the calculations, measurement noise of +/-0.5 mBar was added to the differential pressure simulations.

3.1. Conditioning drilling fluid during constant pump rate.

In this scenario the drilling fluid pump flow rate is kept constant at 2000 liters/minute. At time 10 minutes after the start, the densifier liquid is added to the tank at a flow rate of 10 liters/minute. At about 26 minutes, the viscosifier liquid is added at a flow rate of 100 liters/minute. At about 46 minutes, the diluting liquid is added at a rate of 10 liters minute.

The resulting simulations are shown in the following figures as follows: Figure 2 presents the pump pressure at the constant flow rate, Figure 3 presents the differential pressure readings, Figure 4 and Figure 5 are a zoomed view of these readings. Figure 6 and Figure 7 present the calculated and simulated densities and apparent viscosities.

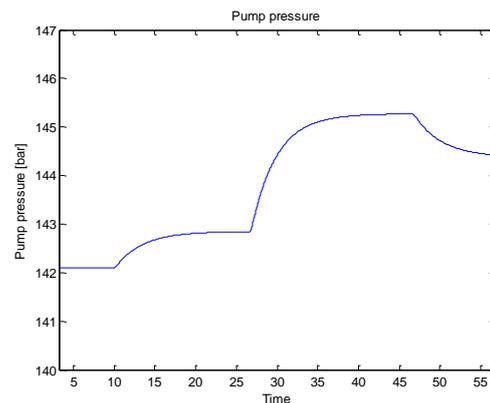


Figure 2 Pump pressure during constant flow rate

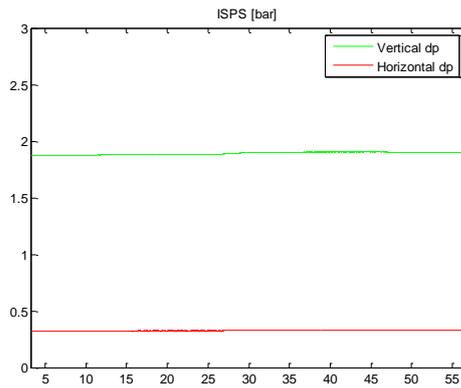


Figure 3 Differential pressure measurements.

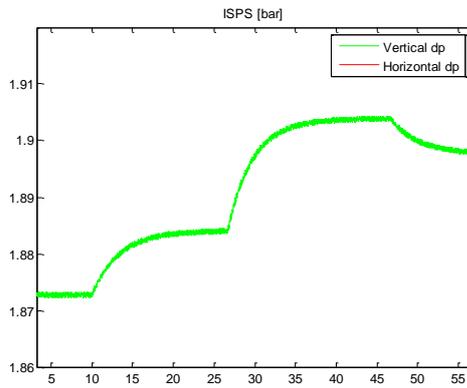


Figure 4 Zoomed view of the vertical differential pressure sensor

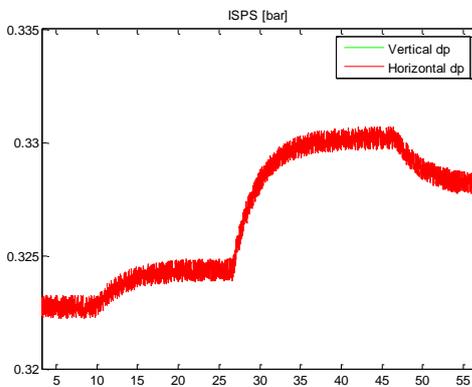


Figure 5 Zoomed view of the horizontal differential pressure sensor

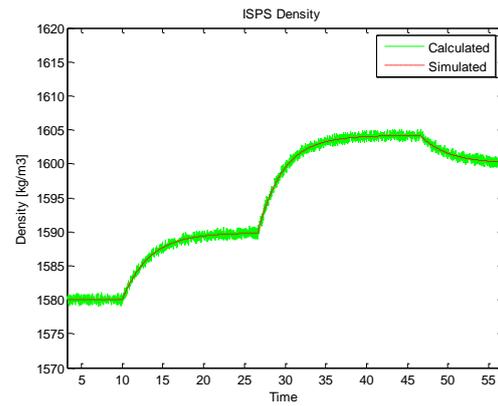


Figure 6 Calculated and simulated density

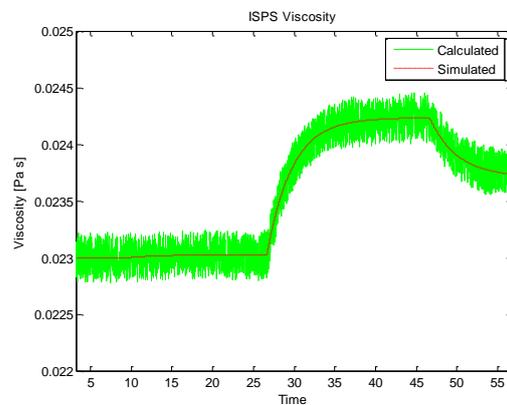


Figure 7 Calculated and simulated apparent viscosity

The results show that the differential pressure readings can be used for calculating the resulting densities and apparent viscosities during drilling fluid mixing and conditioning.

3.2. Various drilling fluid pump flow rates

The scenario in this section presents how variations in the drilling fluid volume flow rate change the density and apparent viscosity due to compressibility and non-Newtonian behavior. No mixing or conditioning is performed. Initially the drilling fluid pump flow rate is 2000 liters/minute. After 20 minutes, the drilling fluid pump flow rate is changed to 2500 liters/minute, and at 40 minutes, the drilling fluid pump flow rate is changed to 3000 liters/minute.

The results of the simulations are shown in the following figures. In Figure 8 the pump pressure is given, and in Figure 9 the pump flow rates is presented. The differential pressure readings are shown in Figure 10 and a zoomed view of the vertical differential pressure measurement is

shown in Figure 11. The zoomed view of the horizontal differential pressure is shown in Figure 12. Calculated and simulated density is presented in Figure 13 and in Figure 14 the apparent viscosity calculations are shown.

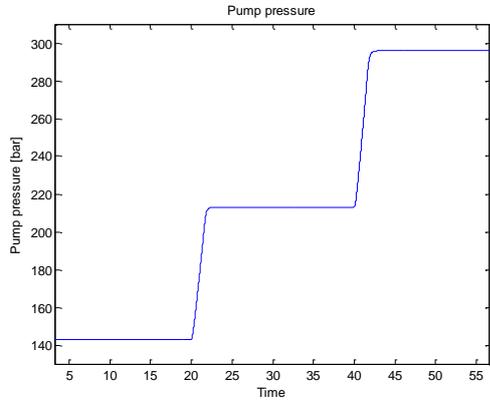


Figure 8 Pump pressure at varying drilling fluid flow rates

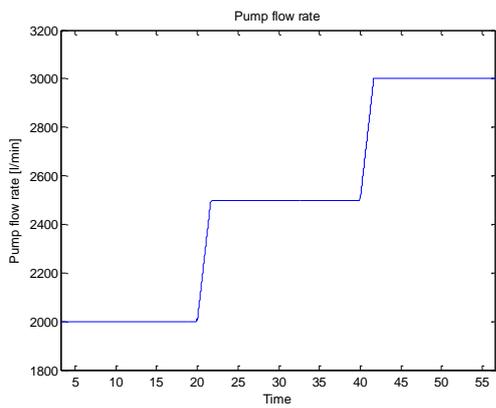


Figure 9 Pump flow rates

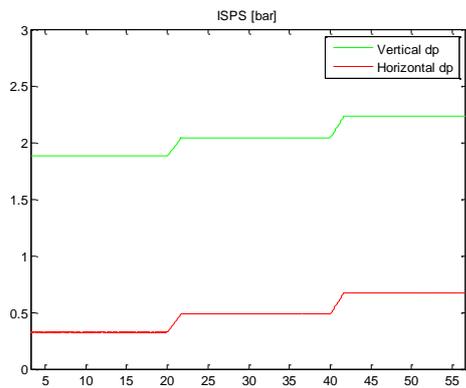


Figure 10 Differential pressure measurements

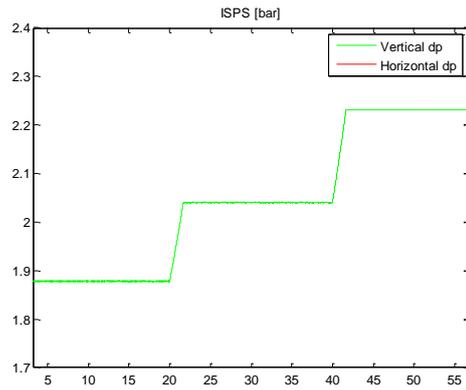


Figure 11 Zoomed view of vertical differential pressure sensor

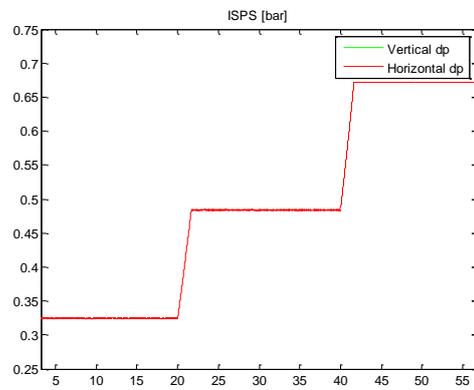


Figure 12 Zoomed view of horizontal differential pressure sensor

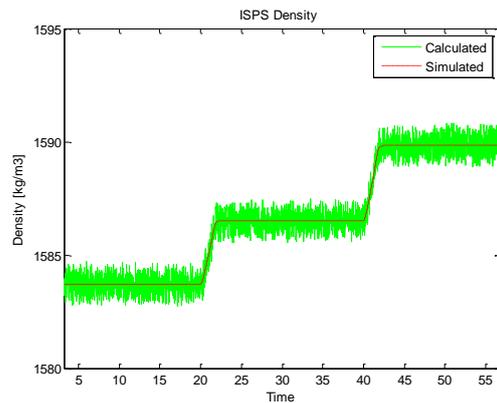


Figure 13 Calculated and simulated density at various pressures

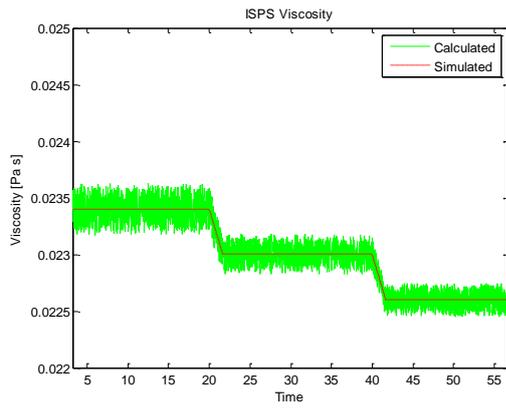


Figure 14 Calculated and simulated apparent viscosity at various flow rates

The results show that the differential pressure readings are able to be used for calculating the resulting densities and apparent viscosities during flow rate changes, including compressibility and non-Newtonian effects.

3.3. Various drilling fluid pump flow rates during conditioning of the drilling fluid

The third scenario is obtained by combining the two previous scenarios, showing how the drilling fluid property measurements varies when the drilling fluid flow rate is changed due to the effects of the drilling fluid compressibility and non-Newtonian effects. In the same scenario, the drilling fluid is adjusted by adding densifier, viscosifier and diluting liquid.

Initially the drilling fluid pump flow rate is 2000 liters/minute. At time 10 minutes, the densifier liquid is added to the tank at a flow rate of 10 liters/minute. At 20 minutes, the drilling fluid pump flow rate is changed to 2500 liters/minute. At about 26 minutes, the viscosifier liquid is added at a flow rate of 100 liters/minute. At 40 minutes, the drilling fluid pump flow rate is changed to 3000 liters/minute. At about 46 minutes, the diluting liquid is added at a rate of 10 liters minute.

The results of the simulations are shown in the following figures. In Figure 15 the pump pressure is given, and in Figure 16 the pump flow rates is presented. The differential pressure readings are shown in Figure 17 and a zoomed view of the vertical differential pressure measurement is shown in Figure 18. The zoomed view of the horizontal differential pressure is shown in Figure 19. Calculated and simulated density is presented in Figure 20 and

in Figure 21 the apparent viscosity calculations are shown.

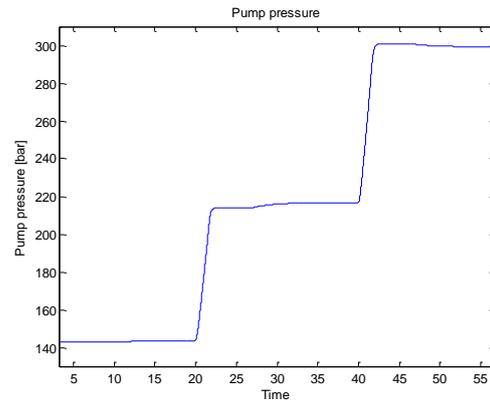


Figure 15 Pump pressure at various flow rates and drilling fluid properties

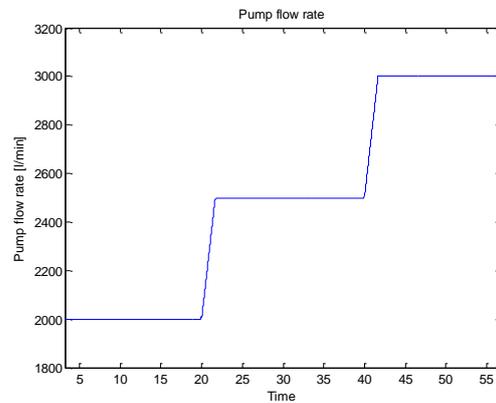


Figure 16 Pump flow rates

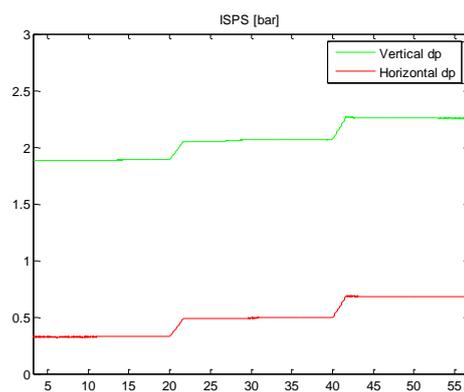


Figure 17 Differential pressure sensor measurements

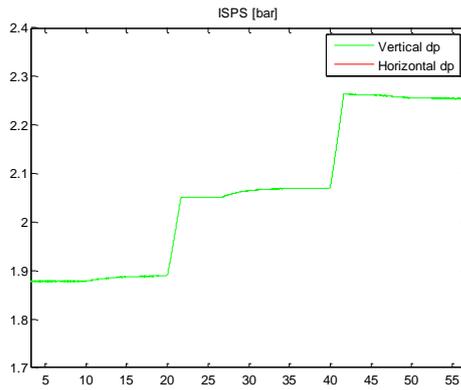


Figure 18 Zoomed view of vertical differential pressure measurements

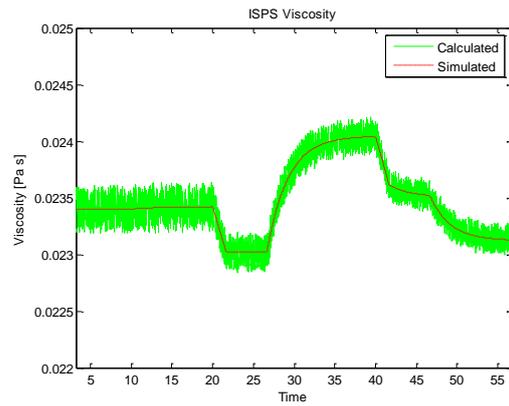


Figure 21 Calculated and apparent viscosity at various flow rates and pressures

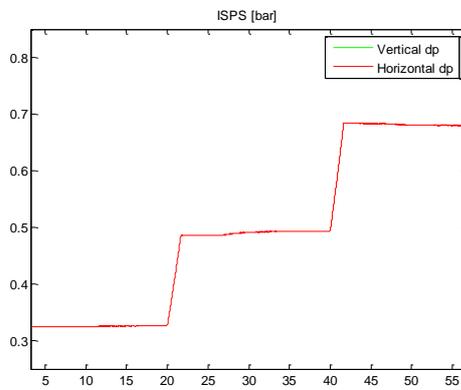


Figure 19 Zoomed view of horizontal differential pressure sensor values

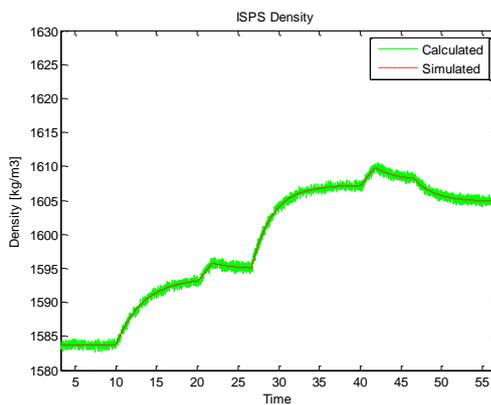


Figure 20 Calculated and simulated density at various pressures and flow rates

The results show that the differential pressure readings can be used for calculating the resulting densities and apparent viscosities during flow rate changes while performing mixing and drilling fluid conditioning.

4. CONCLUSIONS

Models for mixing and conditioning the drilling fluid density and viscosity is presented and simulated. The simulations show that the models are capable to both describe drilling fluid compressibility and drilling fluid non-Newtonian rheology. In addition, the mixing and conditioning of the drilling fluid can be modeled using the described dynamics.

The described model might be expanded in future works by including a temperature model to account for temperature dependent rheology effects.

5. REFERENCES

R Caenn, HCH Darley , and GR Gray, *Composition and properties of drilling and completion fluids*, 6th edition, Gulf Publishing Company, 2011.

RP Sutton and DF Bergman, *Application of the Bergman-Sutton Method for Determining Blend Viscosity*, SPE Production & Operations, SPE-117711-PA, Vol. 27, No 1, pp. 106-124, February 2012.

GO Kaasa, ØN Stamnes, L Imslund and OM Aamo, *Simplified Hydraulics Model Used for Intelligent Estimation of Downhole Pressure for a Managed-Pressure-Drilling Control*

System. *Society of Petroleum Engineers (SPE) Drilling & Completion*, March, 2012;

- G. Nygaard, "Evaluation of automated drilling technologies developed for petroleum drilling and their potential when drilling geothermal wells", presented at GeoEnergy2011, Bergen, 7.-8. September, 2011.
- E. Hansen, "Automatic evaluation of drilling fluid properties", MSc thesis, University of Stavanger, June, 2012.
- L.A. Carlsen, N. L. Rolland, G. Nygaard, R. Time, "Simultaneous Continuous Monitoring of the Drilling-Fluid Friction Factor and Density", *Journal SPE Drilling & Completion*, Vol. 28, No. 1 pp. 34-44, March, 2013.
- Kurt Louis Krogsæter, Automatic Evaluation of Drilling Fluid Properties Conventional and MPD Operations, MSc thesis, UiS, July 2013.
- M.P. Kråvik, H. Sætersdal, L.W. Hougaard, "Drilling fluid monitoring during MPD operations", BSc thesis, Bergen University College, June 2013.
- RR Nafikov and MS Glomstad, Automatic Mud Mixing, SPE163473-MS, 2013 SPE / IADC Drilling Conference and Exhibition, Mar 05 - 07, Amsterdam, The Netherlands, 2013.

Dynamic station-keeping capability analysis using advanced vessel simulator

Nguyen Dong Trong, Pivano Luca, Børhaug Brede, and Smogeli Øyvind
Marine Cybernetics, Trondheim, Norway
Corresponding Author: ntd@marinecyb.com

Abstract

Estimating the vessel station-keeping performance has always been a challenge for vessel operation. The traditional dynamic positioning capability analysis (in short DPCap) as described in the International Marine Contractors Association (IMCA M140, [6]) specification, which is the current industrial standard for analyzing a vessel's station-keeping capability, might not be as accurate as the reality. Dynamic station-keeping capability (DynCap) analysis has been developed as a new method to estimate more accurate the station-keeping capability. The objective of this paper is to present the use of systematic time-domain simulations in the DynCap analysis. The simulator used in DynCap analysis consists of with a sophisticated 6-degree-of-freedom vessel model, including dynamic wind and current loads, 1st and 2nd order wave loads with slowly-varying wave drift, a complete propulsion system including thrust losses, power system, sensors, and a dynamic positioning (DP) control system model. Most of the limiting assumptions needed for the traditional DPCap analysis are removed, yielding results much closer to reality, hence increasing the confidence in planning the operation. It is also possible to tailor the acceptance criteria in the analysis to the requirements for each vessel and operation, such as station-keeping footprint, sea-keeping criteria, dynamic power load, and transient motion after failure. Case study with a supply vessel, a shuttle tanker, and a drilling semisub rig is carried out to demonstrate the simulation aspects in the DynCap analysis. The DynCap analysis will increase the confidence in operating the vessel in harsh environments or in degraded conditions; hence increasing safety and reducing non-productive time.

Keyword: Dynamic, station keeping, capability, simulation, dynamic positioning system.

1 Introduction

Station keeping is one of the critical aspects for marine vessels especially in offshore operation such as sub-sea installation and intervention, offloading, diving, drilling, and laying of pipes. For these operations safety and cost effectiveness are considered as primary. In particular, when conducting marine operations in deep water and in harsh environments, the operations are more time consuming, and hence more sensitive to changes in sea states. Therefore for planning of marine and offshore operations, it is essential to determine the weather operational window where the vessel can maintain its position in intact and degraded conditions, typically after a single failure.

Station keeping can be performed by two ways (Sørensen, 2005): position mooring without or with thruster assistance systems, or exclusively by only thrusters known as dynamic positioning (DP). The traditional DP Capability Analysis (DPCap) as described in the International Marine Contractors Association (IMCA M140) specification is the current industrial standard for analyzing a vessel's station-keeping capability. These analyses are used for vessel design, charter agreements and operational planning. A DPCap analysis is inherently quasi-static, meaning that all dynamic effects must either be neglected or handled by safety factors. Hence, the DPCap analysis can only balance the mean environmental forces with the mean thruster forces, and cannot account for e.g. dynamic thrust losses, dynamic environmental loads, the transient conditions during a failure and recovery after a failure. For early design verification and concept development such an analysis may be adequate, since it is fast, relatively simple, and requires limited model knowledge. For a detailed study of a station-keeping capability in realistic dynamic conditions, however, the traditional DPCap analysis comes short.

Estimating the more accurate station-keeping capability has always been a challenge for vessel operation ([8]). Dynamic station-keeping capability (DynCap) analysis has been proposed by Pivano and Smogeli [9] as a next level of DP capability. Later, Børhaug [2] has validated DynCap results for a supply vessel towards model-scaled experiments.

The objective of this paper is to present the DynCap analysis by means of systematic time-domain simulations. The simulator used in DynCap analysis consists of with a sophisticated 6-degree-of-freedom (6DOF) vessel model, including dynamic wind and current loads, 1st and 2nd order wave loads with slowly-varying wave drift, a complete propulsion system including thrust losses, power system, sensors, and a DP control system model. Most of the limiting assumptions needed for the traditional DPCap analysis are removed, yielding results much closer to reality, hence increasing the confidence in planning the operation. It is also possible to tailor the acceptance criteria in the analysis to the requirements for each vessel and operation, such as station-keeping footprint, sea-keeping criteria, dynamic power load, and transient motion after failure. The DynCap analysis will increase the confidence in operating the vessel especially in harsh environment and in degraded conditions; hence increasing safety and reducing non-productive time.

The DynCap results can be provided in various formats such as wind envelope, thrust envelope, yearly operability, yearly power and fuel consumptions based on different power setups, e.g. closed bus and split bus, and different thruster configurations (from intact to degraded conditions). Case study with a supply vessel, a shuttle tanker and a drilling semisub is carried out to demonstrate the proposed DynCap analysis including a comparison with a traditional DPCap analysis.

2 DynCap theoretical background

This section presents the theoretical background to calculate the dynamic station-keeping capacity. The main purpose of the DynCap analysis is to calculate the station-keeping capability of a vessel based on systematic time-domain simulations. The station-keeping capacity is calculated by searching for an environment limit at which the vessel motion is still able to satisfy a set of acceptance criteria. In the IMCA specification for DP capability ([6]), the acceptance criterion is being able to keep the position. Both wind and current include fluctuations in direction and speed for increased realism, and the wave forces include slowly-varying wave drift effects. The analysis can be performed for collinear

environmental loads (wind, current and waves attacking from the same direction) or non-collinear loads.

The computation is typically performed for intact condition where all the equipment is available and for the worst case single failure (WCSF) condition. In addition the analysis can be run with any thruster and power setup to evaluate the station-keeping capability when not all the equipment is available (for example due to maintenance).

2.1 Analysis of simulation results

One of the advantages of the DynCap analysis, compared to a traditional DPCap, is that the limiting environment can be computed by applying a set of acceptance criteria. The position and heading excursion limits can be set to allow a wide or narrow footprint, or the acceptance criteria can be based on other vessel performance characteristics such as sea keeping, motion of a crane tip or other critical point, dynamic power load, or tension and/or angle of a hawser or riser. In this way the acceptance criteria can be tailored to the requirements for each vessel and operation. An example of position and heading acceptance criteria is shown in Figure 1. In this case, the station-keeping capacity is found by searching for the maximum wind speed in which the vessel footprint stays within the predefined position and heading limits.

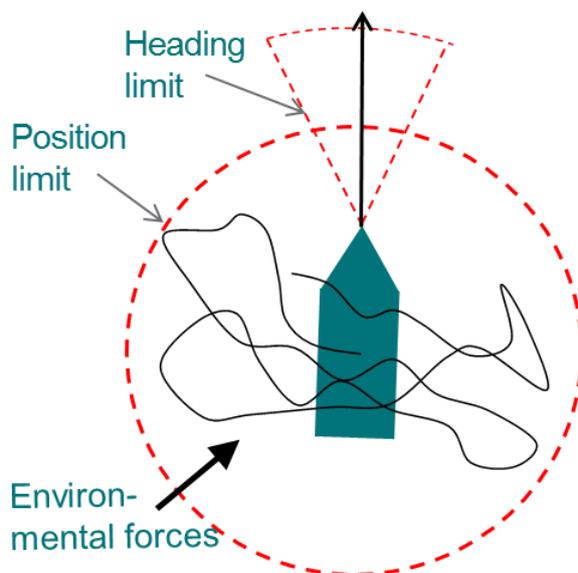


Figure 1: Example of heading and position acceptance limits

By considering the complete vessel dynamics it is also possible to identify temporary position and/or heading excursions due to dynamic and transient effects. As an example, a vessel may stay in position without one thruster according to the traditional DPCap, but the loss of that thruster during station-keeping may cause a temporary excursion outside the positioning acceptance limits.

During DP operation, the vessel position and heading motion is characterized by two components:

- The motion displayed on a DP screen is checked towards positioning limits in the DP (watch circles). This is a filtered, low-frequency motion, which is due to the mean wave drift, thruster, wind and current forces. In literature, this is also referred to as the low-frequency (LF) motion.
- The harmonic motion due to first-order wave loads, which is oscillating about the LF motion. In literature, this is also referred to as the wave frequency (WF) motion.

The actual motion of the vessel is the sum of these two components (Figure 2).

Depending on the requirements to the operation, either the LF motion or the total vessel motion (LF+WF motion) can be used to check if the position acceptance criteria are satisfied in the DynCap analysis.

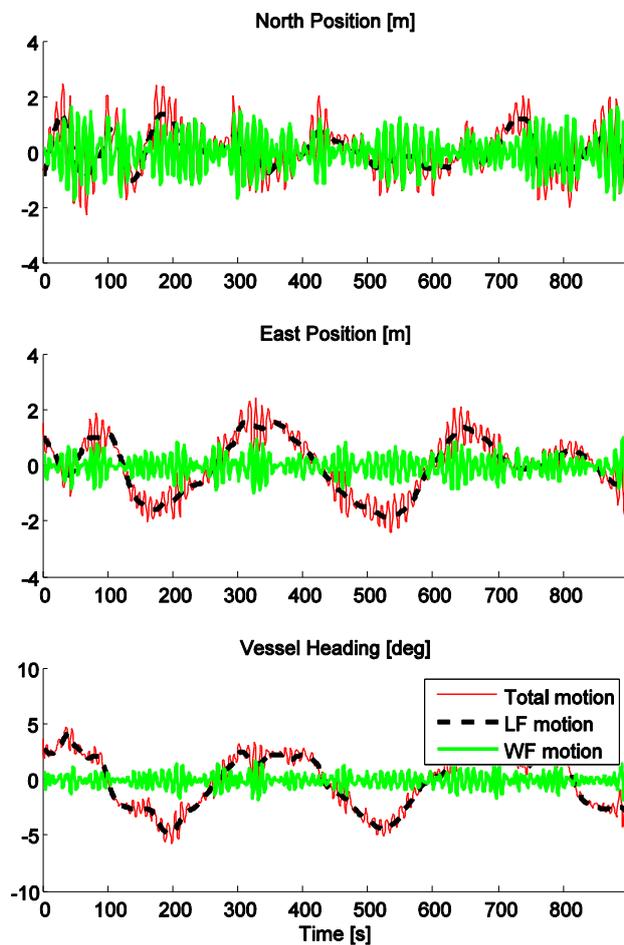


Figure 2: Dynamic motion of the supply vessel

2.2 Presentation of DynCap results

Wind Envelope

The wind envelope is presented as a polar plot where the maximum wind speed at which the vessel can maintain position and heading is plotted for each angle of attack, typically given with 10 deg intervals. This plot is usually utilized to determine the limiting weather window at which the vessel can maintain the position and heading. Figure 3 and Figure 4 show the flowchart for calculate maximum wind speed by means of numerous simulations.

The wind envelope can also be plotted for different acceptance criteria as described in the previous section. For example the plot can show the maximum wind speed where the vessel heave motion is less than for example 3 m, the roll RMS motion is less than 4 deg, or the angle of a riser joint is less than 4 degrees. Combination of station-keeping and motion criteria can be used as well.

Thrust Envelope

For a given design environmental condition the thruster utilization (in %) for maintaining station-keeping is found for each angle of attack. This polar plot is usually employed to determine the robustness to changes in heading.

Both the wind and thrust envelope plots are directly comparable to the similar plots from a conventional DPCap analysis.

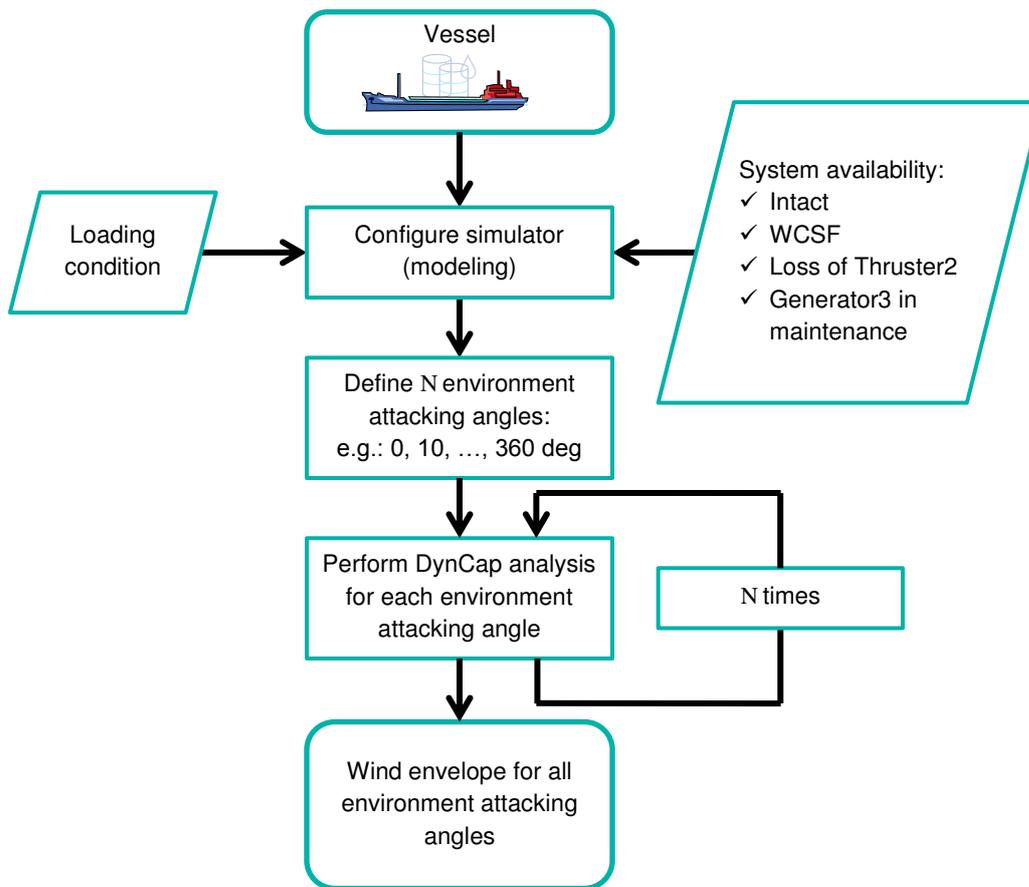


Figure 3: Flowchart for DynCap wind envelope analysis

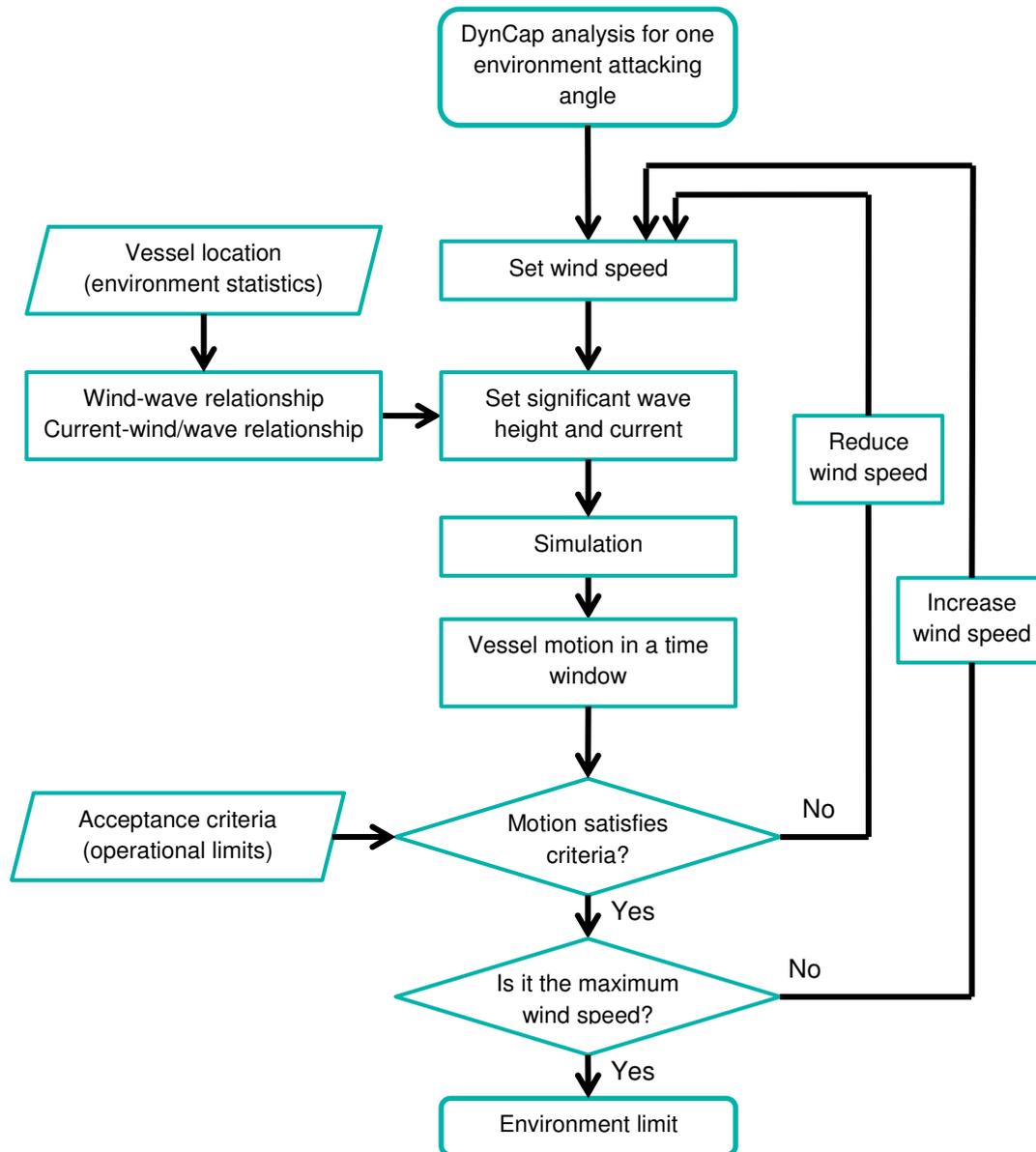


Figure 4: DynCap flowchart for each environment attacking angle

3 Modeling of Marine Vessels

The simulations for DynCap are based on the CyberSea Vessel Simulator. The modeling used in simulations is able to simulate as close as possible the real physics of vessel's dynamics including process disturbance, sensor outputs and control inputs to evaluate the performance of the closed-loop system. By considering the vessel, environmental loads and DP system dynamics, it is not necessary

to reserve a certain amount of thrust for dynamic loads as for the traditional DPCap analysis. DynCap utilizes all the available thrust capacity like the vessel would do in real life.

3.1 Vessel hull kinematics and dynamics

3.1.1 Kinematics

In station keeping, the vessel motions are defined and measured with respect to some reference frames or coordinate systems as shown in Figure 5.

- *The Earth-fixed reference frame* is denoted as the $X_E Y_E Z_E$ -frame, in which the vessel's position and orientation coordinates are measured relative to a defined origin (center of the Earth). Each position reference system (e.g. GPS, hydro acoustics, etc.) has its own local coordinate system, which has to be transformed into the common Earth-fixed reference frame.
- The body frame is fixed to the vessel and thus moving along with it. For convenience, the body frame is often chosen at the mid-ship and waterline.
- The hydrodynamic frame $X_h Y_h Z_h$ -frame is generally moving along the path of the vessel with the x-axis being positive in the forward direction, y-axis positive to the starboard, and z-axis positive downwards. The $X_h Y_h$ -plane is assumed fixed at and parallel to the mean water surface. The vessel is assumed to oscillate with small amplitudes about this frame such that linear theory may apply when modeling the perturbations. In station keeping, operations about the coordinates x_d , y_d , and ψ_d , the hydrodynamic frame is Earth-fixed and denoted as the reference-parallel frame $X_R Y_R Z_R$. It is rotated to the desired heading angle ψ_d , and the origin is translated to the desired x_d and y_d position coordinates for the particular station keeping operation studied. As such, a frame for this reference position and orientation is defined as reference-parallel frame.
- Figure 5 shows also definition of environment (wind, wave and current) direction relative to the bow of the vessel.

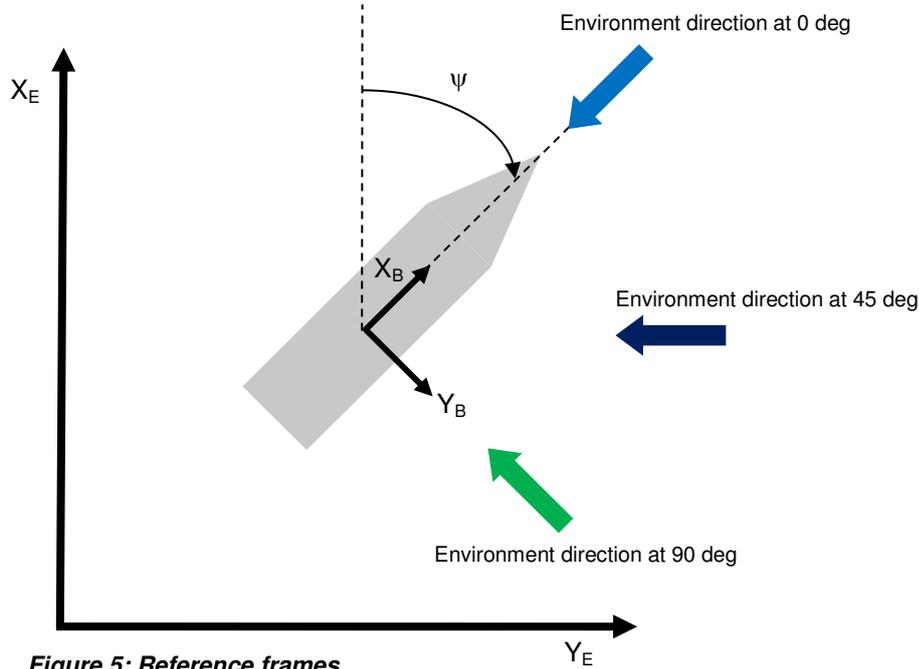


Figure 5: Reference frames

The relationship between the Earth-fixed position and orientation of a floating structure and its body-fixed velocities is given through a transformation matrix ([4]), $J(\eta) \in R^{6 \times 6}$, according to

$$\dot{\eta} = J(\eta)v \quad (1)$$

3.1.2 Dynamics

In modeling the dynamics of floating structure, a simplification can be made by considering two possible models [1], [4], and [10], namely, low frequency (LF) model and wave frequency (WF) model. WF model is primarily for motions due to first-order wave loads whereas LF model accounts predominantly for motions due to second-order mean and slowly varying wave loads, currents and wind loads.

a. Linear WF model

The coupled equations of the WF motions in surge, sway, heave, roll, pitch, and yaw are assumed to be linear, and can be formulated as [10]

$$\mathbf{M}(\omega)\ddot{\eta}_{RW} + \mathbf{D}(\omega)\dot{\eta}_{RW} + \mathbf{G}\eta_{RW} = \boldsymbol{\tau}_{wave1} \quad (2)$$

$$\dot{\eta}_w = J(\psi_d)\dot{\eta}_{RW} \quad (3)$$

where $\eta_{RW} \in R^6$ is the WF motion vector in the hydrodynamic frame; $\eta_w \in R^6$ is the WF motion vector in the Earth-fixed frame; $\psi_d \in R$ is the desired heading angle; $\boldsymbol{\tau}_{wave1} \in R^6$ is the first-order wave excitation vector, which will be modified for varying vessel headings relative to the incident wave

direction; $M(\omega) \in R^{6 \times 6}$ is the system inertia matrix containing frequency-dependent added masses in addition to the rigid mass of the vessel; and $D(\omega) \in R^{6 \times 6}$ is the damping matrix including wave radiation (potential) damping, linear viscous damping, and nonlinear viscous damping (including the effect of antiroll tanks, bilge keel, lift surfaces [5], [7]). The linearized restoring coefficient matrix $G \in R^{6 \times 6}$ is due to gravity and buoyancy affecting heave, roll, and pitch only. Frequency-dependent hydrodynamic added mass, potential damping, and restoring forces, motion RAOs, and wave drift coefficients are computed based on the provided hull 3D model with hydrodynamics software, e.g. WAMIT, ShipX, Octopus, etc. For anchored vessels, it is assumed that the mooring system will not influence the WF-induced motions [11].

b. Nonlinear LF model

The equations of motion for nonlinear LF model of a floating vessel is given by

$$M\dot{\mathbf{v}} + C_{RB}(\mathbf{v})\mathbf{v} + C_A(\mathbf{v}_r)\mathbf{v}_r + D(\mathbf{v}_r) + G(\boldsymbol{\eta}) = \boldsymbol{\tau}_{env2} + \boldsymbol{\tau}_{ext} + \boldsymbol{\tau}_{thr} \quad (4)$$

where $C_{RB}(\mathbf{v}) \in R^{6 \times 6}$ and $C_A(\mathbf{v}_r) \in R^{6 \times 6}$ are the skewsymmetric Coriolis and centripetal matrices of the rigid body and the added mass; $G(\boldsymbol{\eta})$ is the generalized restoring vector caused by the buoyancy and gravitation; $M \in R^{6 \times 6}$ is the body mass matrix including added mass; $\boldsymbol{\tau}_{env2}$ is the dynamic wind, second-order slowly-varying wave and ice (if any) load vector; $\boldsymbol{\tau}_{thr}$ is the force and moment vector produced by the thruster system; and $\boldsymbol{\tau}_{ext}$ is the force and moment vector by external factors (if any), e.g. mooring, offloading, etc. $D(\mathbf{v}_r)$ is the damping vector which is a function of the relative velocity vector $\mathbf{v}_r \in R^6$ between the vessel and current. The current load is included in this damping vector. Details of these terms can be found in [10].

The nonlinear viscous damping and current loads are computed from experimental coefficients or based on:

- ITTC quadratic drag framework/added resistance in surge
- Nonlinear cross-flow drag in sway and yaw using 2D (Hoerner's curve) or 3D current coefficients
- Munk moment from potential coefficients (WAMIT)
- Higher order nonlinear damping terms in heave, roll, and pitch

3.2 Propulsion system model

The propulsion system model is used to represent the complete thrusters/propellers onboard the vessel with motor, frequency converter, pitch mechanism, azimuth mechanism, and local thruster controller. The thruster model is able to simulate the following main features:

- Calculation of the velocity and submergence of the thruster relative to the water, accounting for vessel motion, waves, current, and induced velocities from neighboring thrusters
- Calculation of various thrust loss effects: loss due to changes in advance velocity, ventilation and in-and-out-of-water effects, speed/suction losses for tunnel thrusters due to forward speed, cross-coupling loss for open and ducted propellers due to transverse flow, and Coanda losses for azimuth thrusters
- Motor/shaft dynamics

- Pitch dynamics
- General propeller hydrodynamics model valid for: open and ducted propellers, fixed pitch propeller (FPP) and controllable pitch propeller (CPP), tunnel thrusters, pods and azimuth thrusters
- Azimuth dynamics
- Calculation of thruster noise based on shaft speed and pitch

3.3 Power system model

The power system consists of power generation and distribution models. For the sake of station-keeping simulation, the power system simulates high- and medium-voltage switchboards, buses, generator sets and consumers. The simulator calculates the power balance of each high- and medium-voltage power bus, including the main generator sets and main consumers. The power loads on the consumers correspond to the actual time-varying loads on the thrusters and other large consumers. The simulator produces and transmits the relevant power system signals to the DP computer system, such as thruster power consumption and generator power consumer feedback signals.

It is possible to simulate all available operational configurations of the power system by changing the status of the switches, circuit breakers, and bus-ties during simulation. Power management functionality like pitch reduction and power limitation are simulated if implemented in the real system.

3.4 DP system

The DP system model is able to simulate an industrial standard system. It includes a vessel observer, a DP controller, a joystick module, a thrust allocation routine, and a sensor handling block.

The vessel observer takes position and heading measurements from the sensors as inputs, and generates filtered estimates of vessel position, heading, and velocities. If the measurements contain wave-frequency motion, the observer will filter this to reproduce the low-frequency motion of the vessel. The position and heading measurements from the sensor handling block should be lever-arm compensated.

The DP controller takes filtered position and velocity estimates from the observer as input, compares these to the current setpoints, and outputs the desired forces and yaw moment needed to stay on or converge to the setpoints. The DP controller contains also feedforward compensation from wind and external force measurements.

The thrust allocation takes the desired forces and yaw moment from the DP controller as input, and calculates the necessary commands to each individual thruster (in terms of RPM, pitch, and azimuth angle) to obtain these forces. The thrust allocation is able to handle rudders and other relevant propulsion devices.

The DP system has blackout prevention functionality. The limit for enabling blackout prevention is usually set higher than the load dependent start limit configured in the power management system (PMS).

3.5 Position reference and sensor systems

The simulator generates the sensor and position reference signals sent to the DP system. These signals are contaminated with the nominal noise level that is characteristic for each particular sensor and position reference system type installed on the vessel under the simulated environmental conditions.

Some position reference systems have inputs from specific sensors, such as a gyro compass or a VRU. In such cases, the correction functions within the position reference system are simulated with input from the simulation of those specific sensors.

Redundant sensor and position reference systems (e.g. three different gyro-compasses on the same vessel) are simulated independently with uncorrelated noise and independent failure modes, with possible exception of redundant position reference systems that may share common components.

All sensor and position reference system outputs are referenced to their antenna/transducer positions on the simulated vessel.

3.5.1 Position reference system

Global navigation satellite systems (GNSS) systems consist of several components, the space segment, transmission medium, antenna, receiver components, and usually differential correction. All components can experience failures that degrade or completely destroy positioning performance.

Hydro-acoustic position reference systems are divided into several components: transponders, transmission medium, and transducers.

Range/bearing position reference systems (such as CyScan, Fanbeam, Artemis, RADius, and RadaScan) are based on range and bearing/azimuth measurements using some kind of reflector/transponder at a fixed location.

Other position reference systems include taut-wire based on range (wire length) and two angular measurements, DARPS (Differential Absolute and Relative Positioning Sensor) based on one GPS antenna on the target and one on the vessel, and riser angle reference systems.

3.5.2 Sensors

The outputs of the following sensors are simulated: Gyro-compasses, VRUs, wind sensors, thruster feedback (pitch, RPM, azimuth, thrust), power feedback from thrusters, switchboards and generator sets. Other signals used by the DP system are also simulated, such as diesel fuel rack position, draught sensor, and tension of cable/pipe-lay/crane/hose/hawser.

4 Case study

4.1 Overview

In this section, a comparison between DPCap and DynCap wind envelopes is presented as a case study. Three types of vessel considered for this study are a supply vessel, a shuttle tanker and a drilling semisub rig. The three vessels have different masses and propulsion capacities. The main particulars of the analyzed vessels are described in Table 1.

Table 1: Vessel main particulars

	Supply vessel	Shuttle tanker	Drilling semisub rig	
Length between perpendiculars, L _{pp} [m]	82.0	267.0	114.4	
Breadth, B [m]	19.0	49.0	76.7	
Draught, T [m]	6.6	16.15	23.15	
Displacement, m [tons]	7556	169468	55800	
Propulsion system				
Thruster 1:	type bollard pull	Bow tunnel 1: 172 kN	Bow tunnel 1: 458 kN	Fore stbd azimuth: 1016 kN
Thruster 2:	type bollard pull	Bow tunnel 2: 172 kN	Bow tunnel 2: 458 kN	Mid stbd azimuth: 1016 kN
Thruster 3:	type bollard pull	Bow azimuth: 143 kN	Bow azimuth: 539 kN	Aft stbd azimuth: 1016 kN
Thruster 4:	type bollard pull	Main azimuth port: 280 kN	Stern azimuth: 539 kN	Aft port azimuth: 1016 kN
Thruster 5:	type bollard pull	Main azimuth stbd: 280 kN	Stern tunnel: 279 kN	Mid port azimuth: 1016 kN
Thruster 6:	type bollard pull	-	Main propeller: 1823 kN	Fore port azimuth: 1016 kN

The hydrodynamic coefficients such as added mass, potential damping, hydrostatic coefficients, 1st and 2nd-order wave load coefficients are computed using WAMIT [12]. WAMIT is a 3D potential theory computer program capable of analyzing wave interactions with offshore platforms and other structures or vessels. The input to the program is a 3D geometry file represented by panels given in Figure 6.

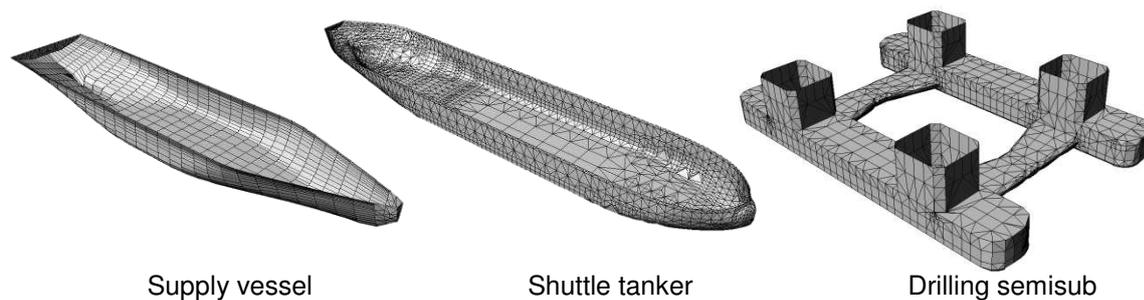


Figure 6: 3D hull geometry

The waves in this analysis are simulated using JONSWAP spectrum with the peak parameter $\gamma = 3.3$ and spreading factor for spreading function $s = 1$ as recommended by DNV [3].

The wind-wave relationship is adopted from the North Sea data in IMCA M 140 ([6]). It is assumed that the current speed is 0.0 m/s. The environmental loads will be set as collinear (wind, current and waves have the same direction).

The wind and current magnitudes are modeled by considering an average speed and a random effect (wind gusts and current fluctuations). Wind and current coefficients for each type of vessel are based on data from a similar vessel of that type.

The power operational philosophy is split switchboard and all generators are online for the healthy switchboards. This means that all generators are online in intact condition and that the generators connected to a disabled switchboard in WCSF are offline. The limit for DP blackout prevention is chosen to be 85% of the generator load.

In this case study, the low-frequency LF motion (section 2.1) is used to check whether the vessel is able to stay within position and heading limits. The limits are given by

- DynCap option 1: position and heading limits: 5 meters and 3 degrees, respectively
- DynCap option 2: position and heading limits: 10 meters and 5 degrees, respectively.

The simulations are carried out with several random seed numbers and the final result is the average of all the DynCap runs. This is to eliminate the random effect of the dynamic process. In this paper, five random seed numbers are used to obtain each DynCap wind envelope.

4.2 Results

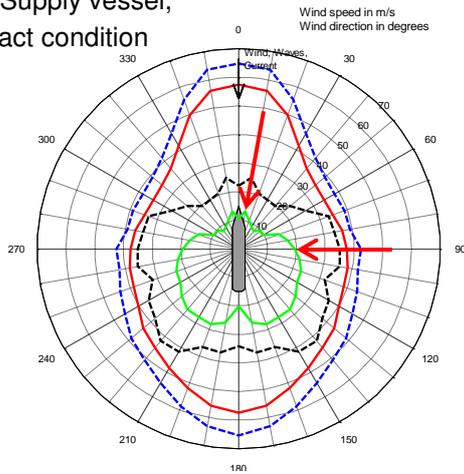
This section shows the DPCap and DynCap wind envelopes and explains the meaning behind the DynCap results. As mentioned above, the DPCap analysis usually reserves a certain amount of thrust for vessel and thrust dynamics. Therefore, the DPCap analysis in this section consists of 2 cases: with 0% and 20% thrust reserved for dynamics. The DynCap analysis consists of 2 acceptance criteria: position and heading limits of 5m/3deg and 10m/5deg. Figure 7 shows the results for three types of

vessel, i.e. the supply vessel, the shuttle tanker and the drilling semisub. It is noted that the environment attacking angle is defined in Figure 5.

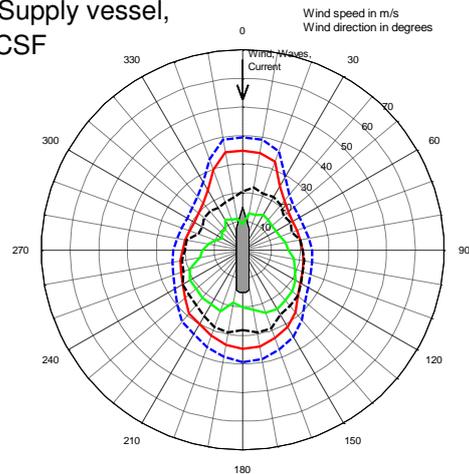
Each DynCap wind envelope is found by running through the flowchart shown in Figure 3. That means each point of the DynCap wind envelope is calculated according to the flowchart shown in Figure 4. The vessel performance is demonstrated in the environment condition above the wind envelope by showing the LF position trace together with footprint and the time series of LF distance from setpoint, LF position and LF heading. The LF position trace plots include also the position acceptance limits. In the DynCap wind envelope with acceptance limit of 5m/3deg for the supply vessel, demonstration are carried out with two points (indicated by arrows in Figure 7) corresponding to the environment attacking angles of 10 deg and 90 deg. Figure 8 and Figure 9 show the vessel performance in such conditions. In addition Figure 2 shows the vessel LF, WF and total motion for the case of attacking angles of 10 deg

The time series for the environment coming from 10 deg case (Figure 8) indicates that the supply vessel 'cannot keep the position' because the LF heading (maximum approximately 5 deg) is out of the acceptance criteria (3 deg). For the environment coming from 90 deg case (Figure 9), the vessel 'cannot keep the position' because both position and heading are outside the acceptance criteria.

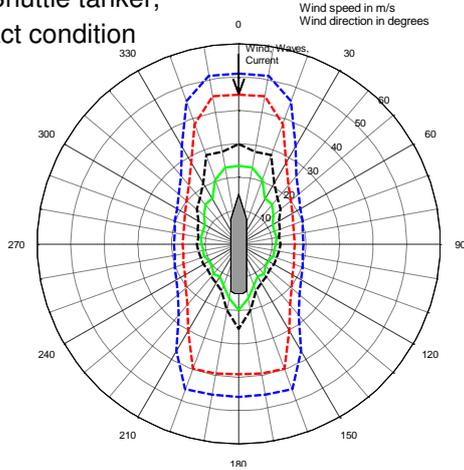
a. Supply vessel,
intact condition



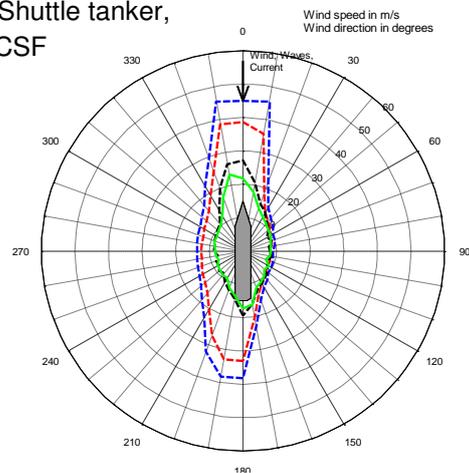
d. Supply vessel,
WCSF



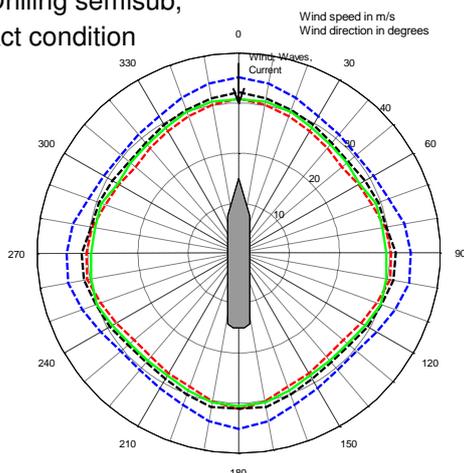
b. Shuttle tanker,
intact condition



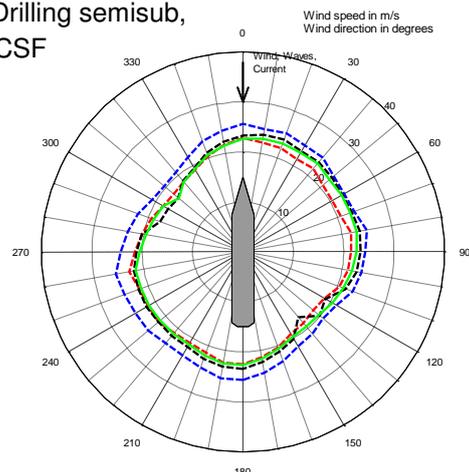
e. Shuttle tanker,
WCSF



c. Drilling semisub,
intact condition



f. Drilling semisub,
WCSF



Current speed: 0.00 (m/s)

- Case 1: Intact DPCap 0% thrust reserved for dynamics
- Case 2: Intact DPCap 20% thrust reserved for dynamics
- Case 3: Intact DynCap, acceptance limit: 10m/5deg
- Case 4: Intact DynCap, acceptance limit: 5m/3deg

Current speed: 0.00 (m/s)

- Case 1: WCSF DPCap 0% thrust reserved for dynamics
- Case 2: WCSF DPCap 20% thrust reserved for dynamics
- Case 3: WCSF DynCap, acceptance limit: 10m/5deg
- Case 4: WCSF DynCap, acceptance limit: 5m/3deg

Figure 7: DPCap and DynCap of different types of vessel in intact and WCSF

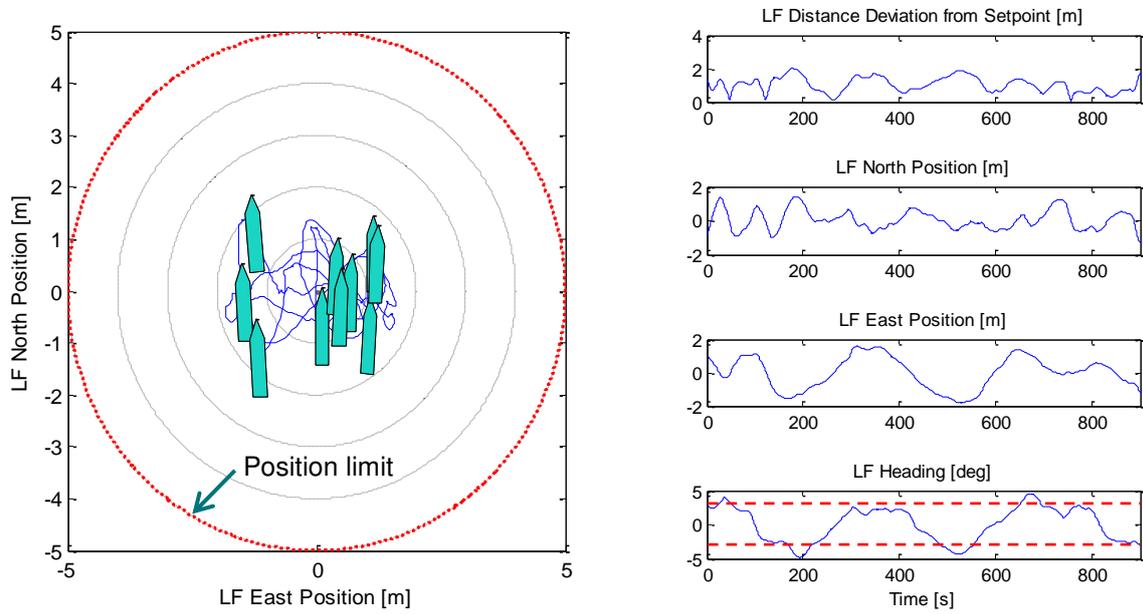


Figure 8: Vessel performance in environment condition above limit, attacking angle = 10 deg

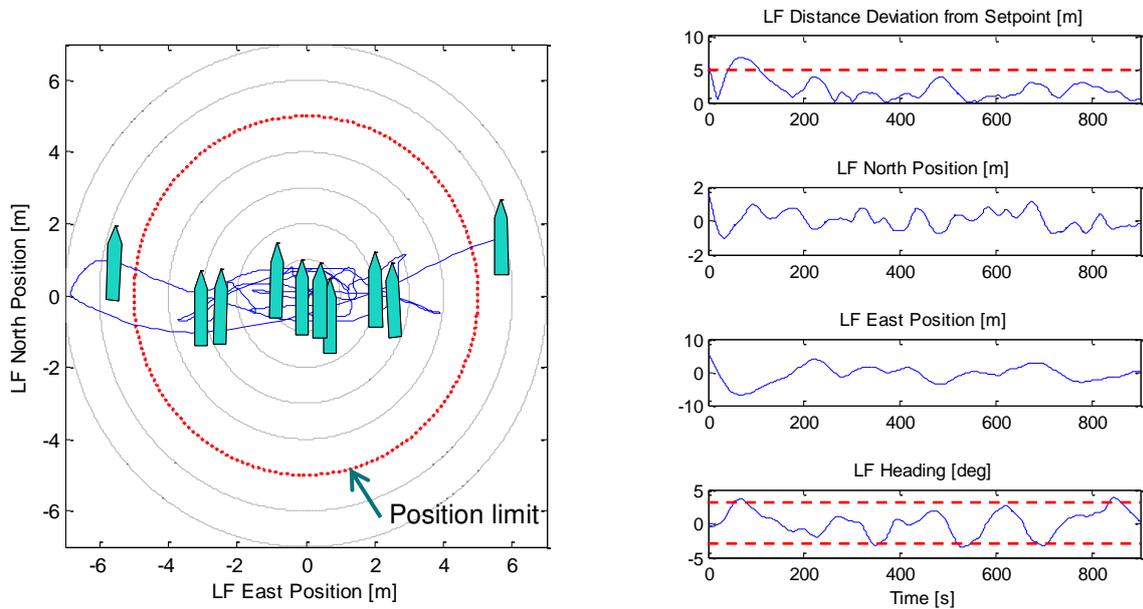


Figure 9: Vessel performance in environment condition above limit, attacking angle = 90 deg

4.3 Discussions

This section discusses the difference between DPCap and DynCap results.

The DPCap with 0% thrust reservation yields the maximum environment limits; and the one with 20% thrust reservation as expected results in smaller environment limits.

For the supply vessel and the shuttle tanker, the environment limits from the DPCap analysis seems to be unrealistic, e.g. 45 to 65 m/s wind speed for head sea in intact condition. The DynCap results seem to be more realistic, e.g. 20 to 30 m/s wind speed for the shuttle tanker. There is a remarkable difference between the DPCap and DynCap results. The difference in the supply vessel case can be due to its small mass; hence the motion is highly affected by the dynamic environmental loads. The difference in the shuttle tanker case may be due to high thrust loss from the main propeller. This means that operation planning for these vessels should not depend on the DPCap analysis.

For the drilling semisub, the difference between DPCap and DynCap is not pronounced. This is due to the dynamic characteristics of this semisub indicating that it might not come back to the setpoint after a small excursion.

In general the DynCap result is closer to DPCap one if the acceptance limit is relaxed.

5 Concluding remarks

This paper presented the use of the advanced simulator for the newly-developed DynCap concept. The simulator consisted of a sophisticated 6DOF vessel model, including dynamic wind and current loads, 1st and 2nd order wave loads with slowly-varying wave drift, the complete propulsion system including thrust losses, power system, sensors, and the DP control system model. Case study with the supply vessel, shuttle tanker and drilling semisub showed that the DynCap analysis yielded more realistic results than the DPCap one. This confirmed that by using the DynCap analysis, the vessel operation can be planned more efficient and safer.

6 References

- [1] J. G. Balchen, N. A. Jenssen, E. Mathisen, and S. Sælid, "A dynamic positioning system based on Kalman filtering and optimal control," *Model. Identif. Control*, vol. 1, no. 3, pp. 135–163, 1980.
- [2] Børhaug, B (2012). *Experimental Validation of Dynamic Stationkeeping*. Master thesis, Norwegian University of Science and Technology, Department of Engineering Cybernetics.
- [3] DNV. Det Norske Veritas. *Recommended practice DNV-RP-C205: Environmental Conditions and Environmental Loads*. October 2010.
- [4] T. I. Fossen, *Marine Control Systems: Guidance Navigation and Control of Ships Rigs and Underwater Vehicles*. Trondheim, Norway: Marine Cybernetics, 2002, ch. 2 and 11.

- [5] Ikeda, Y. et.al. On roll damping force of ship - effect of hull surface pressure created by bilge keels. Technical Report 00402, Dep. of Naval Arch., University of Osaka Prefecture, 1978.
- [6] International Marine Contractors Association (IMCA). Specification for DP Capability Plots. IMCA M 140 Rev. 1 June 2000.
- [7] Kato, H. On the frictional resistance to the rolling of ships. Journal of Zosen Kiokai, 102:115, 1958.
- [8] Pivano, L and Børhaug, B and Smogeli, Ø. Challenges in estimating the vessel station-keeping performance. European Dynamic Positioning Conference, 20-21 June 2013, London.
- [9] Pivano, L. and Smogeli, Ø. (2012). DynCap – The Next Level Dynamic DP Capability Analysis, in Proceedings from Marine Operations Specialty Symposium (MOSS 2012), Singapore, August 6-8.
- [10] A. J. Sørensen, Marine Cybernetics: Modelling and Control, 5th ed. Trondheim, Norway: Dept. Marine Technology, NTNU, 2005.
- [11] M. S. Triantafyllou, "Cable mechanics with marine applications," Lecture Notes, Dept. Ocean Eng., Massachusetts Inst. Technol., Cambridge, MA, May 1990.
- [12] WAMIT. WAMIT User Manual. Versions 6.4, 6.4PC, 6.3S, 6.3S-PC. WAMIT Inc.

Methodology for modeling, parameter estimation, and validation of powertrain torsional vibration

Neda Nickmehr, Lars Eriksson, and Jan Åslund
Dep. of Electrical Engineering, Linköping University, Sweden
neda.nickmehr, larer, jaasl @isy.liu.se

Abstract

A vehicular powertrain is a lightly damped dynamic system that transfers the engine torque to the driving wheels through a number of inertias and elastic elements. Therefore, it is prone to vibrate and emit noise when disturbances are applied. Providing a methodology, for powertrain vibration modeling and simulation, is one of the key steps in various research topics in the field of automobile engineering. Verification of the engine crankshaft torsion and vibration model, as a subsystem of the powertrain, is proposed in this paper. This is achieved by constructing a rotational multi-body system in MATLAB and utilizing nonlinear least squares method for estimation of the model parameters. The simulated engine angular velocity is compared to the measured data, from a car, which shows a good agreement.

1 Introduction

There are different applications in automotive industry where powertrain vibration modeling is needed. Two of the more important cases are as follows:

- Passenger comfort is important for the customers, and consequently the car manufacturers. Powertrain dynamics is one of the main sources of noise, vibration, and harshness (NVH) inside the cars, and reducing its torsional vibration, to an acceptable level, is desirable. A typical powertrain NVH spectrum for a passenger car contains a considerable range of frequencies, from 2 Hz to 5000 Hz [1]. To achieve suitable ride quality, there is a need for better understanding of the dynamics, and a good model is a valuable instrument.
- To distinguish the effects of different excitation sources on the angular velocities of the powertrain parts, is not a trivial task. This causes difficulties for any type of miss-behavior detection. Examples of possible input disturbances are, combustion variation such as, misfire, cold start and cylinder variations; crankshaft torsional vibrations; road roughness; underinflated tires, etc. By using an appropriate powertrain model, and studying the distinct disturbances influences on the simulated outputs, such as angular velocities, it is possible to gain understanding and detection of undesirable modes. Examples of such applications of a model are misfire and underinflated tires, which are investigated by considering the simulated flywheel and driving wheel velocity signals, respectively [2-3].

Powertrain vibration can be modeled by torsional elastic elements. Basic models are linear lumped spring-mass-damper systems which can be extended by adding details for the dynamics of the various parts [4]. Rabeih developed a 14-degrees-of-freedom lumped parameter model, from the engine to the driving wheels, for torsional vibration analysis of the powertrain with a four-cylinder engine and manual transmission [5]. The proposed model was good enough for simulating free, steady, and transient situations. However, no experimental evaluations were done. Crowther used a 6-degrees-of-freedom model to perform numerical simulations for transient vibrations [6]. A powertrain test rig, that was used for measuring torsional vibration, was also presented. Furthermore, powertrain nonlinear torsional phenomena such as, gear backlash and a multi-stage nonlinear clutch, have been studied by Couderc *et al.*, where an experimental test rig was also developed to verify the simulation results [7]. In all the mentioned works, the system parameters are approximate to the passenger cars and no details are provided. In contrast to the previous mentioned studies, the contribution of this paper contains the following three elements. Firstly, building a slightly different model (using dampings between all the inertias). Secondly, focusing on the parameter estimation procedure for the nonlinear engine block model by considering idling condition and using the measurements from a car. Thirdly, performing sensitivity analysis to study the behaviour of the system output with respect to changing of different parameters. This will provide the answer for the question, *how important are different parameters in a model.*

There is an enormous literature about parameter estimation methods and system identification for linear and nonlinear systems, e.g., [8-9]. The procedure for estimating unknown states and parameters of dynamical systems, from noisy measurements, consists of three main tasks [8]:

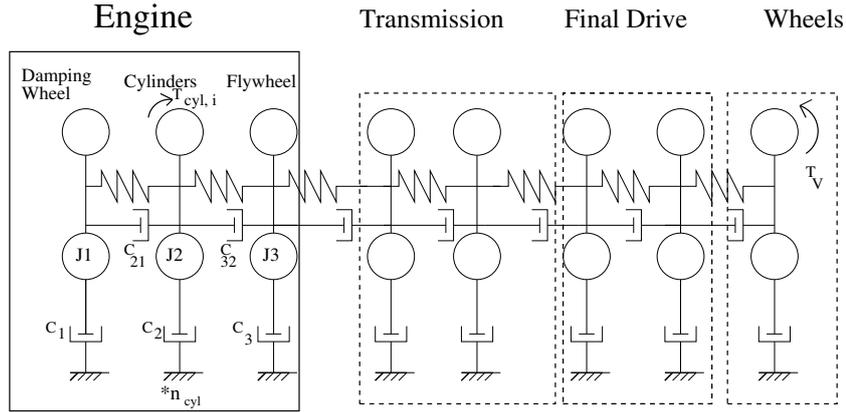


Fig. 1: Powertrain model including engine block.

1. Three basic entities:
 - (a) Measurements of the inputs and the outputs.
 - (b) A model structure.
 - (c) A rule by which candidate models can be assessed using the measurements.
2. Model validation which is done using separate data from the estimation data.
3. The system identification loop.

In this work, in-cylinder pressures, and the engine angular velocity, are measured for different working cycles at the idling condition, step 1.(a). Then, the engine-block torsional vibration model is constructed with unknown parameters which have physical interpretations, namely, spring, mass, and damper components, step 1.(b) in the above description. This is called gray-box model. Further, the nonlinear least squares is used as the rule to estimate the model unknown parameters, step 1.(c), by utilizing the experimental data for one working cycle, from step 1.(a). This is named estimation cycle. Finally, the estimated model is validated by simulating the system, at another cycle, and then comparing the engine angular velocity with the validation data, which is step 2. It is possible that the obtained model is not able to pass validation test, then it is necessary to go back and revise different steps of the procedure, which is step 3.

2 Engine-block model

The proposed powertrain model is a nonlinear lumped parameter system consisting of rotating masses, friction, damping, and stiffness elements, which is shown in Fig. 1. There exists different subsystems, i.e., engine block, transmission, final drive, and the wheels. The total friction in each subsystem is modeled by a number of dampings, which are connected to the ground for all the rotating masses. The modular structure of the model makes it flexible enough to add (remove) components with respect to different vehicle configurations such as front-wheel, rear-wheel, or four-wheel drive cars. In order to achieve better identification performance, the suggestion of this work is to disengage the clutch, i.e. idling condition, so that there is no connection between the flywheel and the transmission. Thus the model is simplified to the engine-block and consequently the delivered torque at the flywheel is zero, since, the engine friction will cancel out the produced torque by the cylinders.

2.1 Mathematical modeling

According to Newton second law, the equations of motion for six different rotating masses, J_i , in the engine block which is a multiple-degrees-of-freedom system, are given by:

$$J_i \ddot{\theta}_i = C_{i+1,i}(\dot{\theta}_{i+1} - \dot{\theta}_i) - C_i \dot{\theta}_i + K_{i+1,i}(\theta_{i+1} - \theta_i), \quad \text{for } i = 1, \text{ Damping wheel} \quad (1)$$

$$J_i \ddot{\theta}_i = C_{i+1,i}(\dot{\theta}_{i+1} - \dot{\theta}_i) - C_i \dot{\theta}_i + K_{i+1,i}(\theta_{i+1} - \theta_i) - C_{i,i-1}(\dot{\theta}_i - \dot{\theta}_{i-1}) - K_{i,i-1}(\theta_i - \theta_{i-1}) + T_{cyl,i}, \quad \text{for } i = 2, 3, 4, 5, \text{ Four cylinders} \quad (2)$$

$$J_i \ddot{\theta}_i = -C_i \dot{\theta}_i - C_{i,i-1}(\dot{\theta}_i - \dot{\theta}_{i-1}) - K_{i,i-1}(\theta_i - \theta_{i-1}), \quad \text{for } i = 6, \text{ Flywheel} \quad (3)$$

where θ_i , $\dot{\theta}_i$, and, $\ddot{\theta}_i$ are angular position, angular velocity, and angular acceleration of the inertia at position i , respectively. Furthermore, C_i is the friction coefficient for element i , and $C_{i,i-1}$ and $K_{i,i-1}$ are the damping and stiffness coefficients, respectively, between two following inertias. The resulting torque, $T_{cyl,i}$, from cylinder i , which imposes the nonlinearity on the system, is calculated as a function of compression pressure force, $F_{c,i}$, and the piston mass force, $F_{p,i}$:

$$T_{cyl,i}(\tilde{\theta}_i) = \left(r \sin(\tilde{\theta}_i) + \frac{r^2 \sin(2\tilde{\theta}_i)}{2\sqrt{l^2 - r^2 \sin^2(2\tilde{\theta}_i)}} \right) \times (F_{c,i}(\tilde{\theta}_i) + F_{p,i}(\tilde{\theta}_i)) \quad (4)$$

where $\tilde{\theta}_i = \theta_i + \delta\theta_i$, and $\delta\theta_i$ is the offset for each cylinder according to ignition order, here, 1342. Also, r and l are crankshaft radius and connecting rod length, which are the geometrical characteristics of the cylinder model. More details on the equations and the model for the cylinder pressures and reciprocating torques can be found in Eriksson *et al.* [2].

2.2 State-space equations

The equations (1)-(3) can be transferred to the state-space form by performing two steps. First, utilizing the experimental data of the in-cylinder pressures, at idling, to obtain the delivered torque, $T_{cyl,i}(\tilde{\theta}_i)$, of cylinder i at different angular positions [2]. Second, defining new variables for angular velocities, $\dot{\theta}_i$, of each rotating mass, J_i . Therefore, there will be totally 12 states, including 6 angular positions, θ_i , and 6 angular velocities, $\dot{\theta}_i$. Then, the engine-block mathematical model in vector form, as a nonlinear ordinary differential equation, (ODE), can be written as follows:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}), \quad y = \mathbf{x}(2) \quad (5)$$

where \mathbf{x} is the state vector, y is the model output, which is damping wheel angular velocity $\dot{\theta}_2$, and \mathbf{p} is the vector of unknown parameters. The complete powertrain model contains 42 unknown parameters. With the aid of symmetry and the proposed method of using idling conditions for parameters identification, the number of parameters, to be estimated in the engine-block, is reduced to 10. In Fig. 1, 8 parameters are noted. The 2 remaining are damping between two cylinders, C_{xx} , and the piston reciprocating mass, m . The vector of parameters, \mathbf{p} , can be categorized in two groups, i.e., 1) inertias and piston reciprocating mass, 2) friction coefficients and dampings. The crankshaft system is very stiff, therefore the stiffness coefficients are limited to high values and are hence not estimated.

3 Engine-block parameter estimation

One of the most well-known parameter estimation procedures, in nonlinear systems, is to write the estimation problem as a nonlinear optimization problem, where the goal is to minimize the difference between the predicted output, from the model, and the observations. The nonlinear least squares method, which is applied in this paper, is a commonly used optimization formulation. It searches the parameters values which minimize the squared errors.

3.1 Nonlinear least squares (NLS) formulation of the problem

A nonlinear regression model is described by an equation of the form:

$$z_i = g(\mathbf{u}_i, \mathbf{p}) + \varepsilon_i \quad (6)$$

where z_i denotes the observations, \mathbf{p} is the set of unknown parameters that are to be estimated, \mathbf{u}_i is a $1 \times k$ vector of known values, and g is a nonlinear function [10]. The least squares estimate is computed by minimizing the following objective function:

$$V_N(\mathbf{p}) = \sum_{i=1}^{i=N} (z_i - g(\mathbf{u}_i, \mathbf{p}))^2 = \sum_{i=1}^{i=N} \varepsilon_i(\mathbf{p})^2 \quad (7)$$

where, for the engine block estimation problem, $\varepsilon_i(\mathbf{p})$ is the discrepancy between the simulated and the measured damping wheel angular velocity. In general, the solution of the above mentioned nonlinear minimization problem, is not available analytically, thus numerical nonlinear optimization algorithms are used [10].

3.2 Forward sensitivity analysis (FSA)

Sensitivity analysis is a procedure for studying the influence of a parameter value perturbation on the model behaviour. Forward sensitivity analysis is a local method [11] which can be performed by introducing the following initial value problem with n states and m parameters \mathbf{p} :

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \mathbf{p}) \quad (8)$$

and the augmented system with sensitivity equations is given as:

$$\dot{\mathbf{w}} = \begin{pmatrix} \mathbf{f}(t, \mathbf{x}, \mathbf{p}) \\ \mathbf{P}'_i \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{P}_i \end{pmatrix}, \quad \mathbf{P}'_i = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{P}_i + \frac{\partial \mathbf{f}}{\partial \mathbf{p}}, \quad \mathbf{P}(0) = 0 \quad (9)$$

where $\mathbf{P} = \frac{\partial \mathbf{x}}{\partial \mathbf{p}}$ is an $n \times m$ matrix function and \mathbf{P}_i points to column i in matrix \mathbf{P} which corresponds to one parameter in \mathbf{p} .

In this paper, FSA is to investigate the engine-block behaviour for small changes of the parameters near estimated values found by NLS.

4 Results and Discussion

4.1 Estimation results by using numerical solver

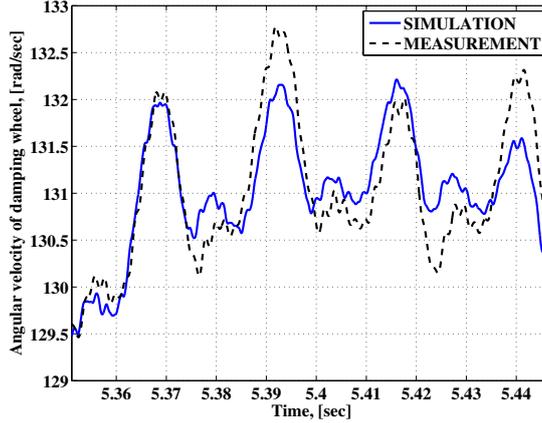
The `lsoptim` is a package, which is written in MATLAB by Eriksson [12], for solving non-linear unconstrained least squares optimization problems. The solver uses the Levenberg Marquardt method, as the optimization numerical algorithm. It is second order and thus has good local convergence properties. The iterative optimization, which is applied in this solver, is based on three stages:

- Start: providing an initial guess for parameters which are to be estimated.
- Iteration: simulation of the ODE in (5) to find the damping wheel angular velocity, which is one of the states in the state vector, \mathbf{x} . Later, the residual, $\varepsilon_i(\mathbf{p})$, can be determined by subtracting the simulated value of this state from the measured data. The residual is used in each iteration to define the direction for updating the set of parameters.
- Stop: The search algorithm will be stopped by a criterion based on the difference between the values of the objective function in two following iterations. In other words, the final iteration will not improve the objective function more than a certain degree. It is worth to mention that `lsoptim` only converges to a local minimum (which might be the global minimum as well).

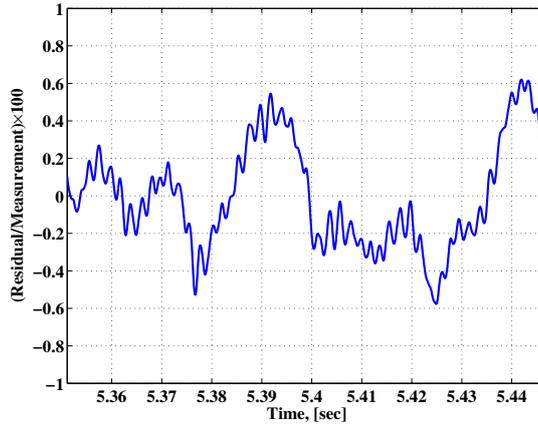
In Fig. 2(a), the simulated and the measured damping wheel angular velocities, are compared by applying the estimated parameters provided by the numerical solver. High resolution measurements, for a 4-cylinder car, with angular resolution of 0.5 degree, are used. It is seen the model is not only able to capture the main frequency from the engine and follow the trend, but also, it resembles the measurement data at higher frequencies. The ratio of the residual value, $\varepsilon_i(\mathbf{p})$, over the real data, in each time sample i , is shown in Fig. 2(b). It provides how far the model result is from the reality, which shows the maximum value of $\sim 0.6\%$, when the engine mean angular velocity is approximately 130 rad/sec.

4.2 Model validation

According to Section 1, the second step in the procedure of the estimation, is model validation, in which the model output is compared with validation data. In other words, for the engine model, the damping wheel angular velocity observations, at validation cycle, is plotted against the simulated results. Therefore, it is determined whether the estimated model accurately captures the system dynamics or not. Fig. 3(a) shows the measured versus simulated data plot for validation data, and Fig. 3(b) shows the ratio of the residual value over the real data, in each time sample. It is seen that there exists a good agreement between the validation data and the output from the model, which is a good evaluation for the estimated model.



(a): Damping wheel angular velocity, measurement vs. simulation.



(b): The ratio of the residual value $\varepsilon_i(\mathbf{p})$ over the real data, in each time sample i .

Fig. 2: Comparison between simulation and empirical data for *estimation cycle*.

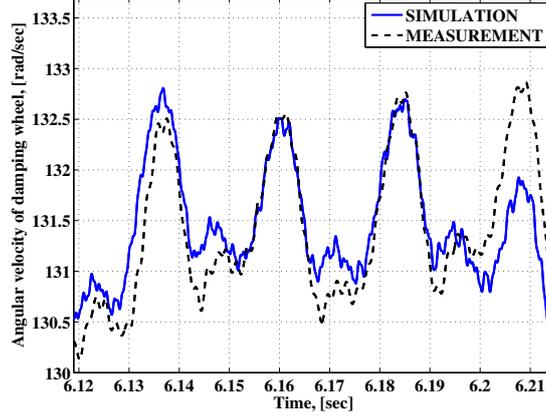
4.3 Results of sensitivity analysis

Here, the effects of the 10 parameters perturbations, mentioned in Section 2.2, are considered. The primary system equations, given in (8), were written in Mathematica and then the sensitivity equations were computed. The final ODE system (9) has been solved in MATLAB.

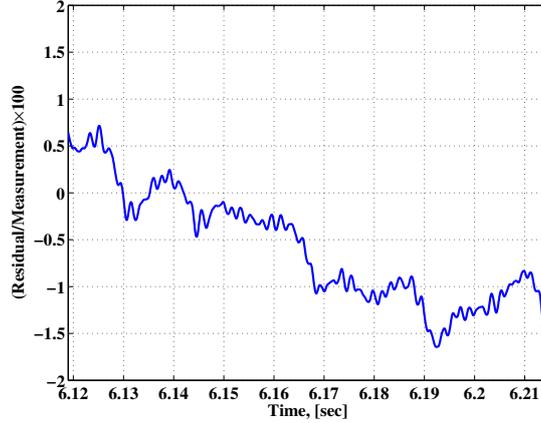
Fig. 4 shows the sensitivities of the damping wheel angular velocity, (rad/sec), to the three friction coefficients, $C_1 - C_3$, and to the three damping coefficients, C_{21} , C_{32} , and C_{xx} . Looking at the values of all the plots on the vertical axis (rad/sec), it is understood that the sensitivities to friction coefficients are in the same level and significantly higher than the sensitivities to the damping coefficients. Possible interpretation is that, friction coefficients are important to follow the trend and changing them has a high influence on increasing or decreasing the engine velocity (rad/sec). However, from the basic vibration knowledge, it is known that the damping coefficients only control the amount of oscillations and have nothing to do with the mean value of the angular velocity. There also exist interpretations for two different shapes of the sensitivities, which are described as follows:

- The reducing shape for the sensitivities of the engine angular velocity, (rad/sec), to the three friction coefficients, $C_1 - C_3$, is the result of friction growth. In other words, by increasing friction coefficient, the system is slowed down, and the slope, $\dot{\omega}$, can be obtained by the relation $\Delta C\omega \approx -(\sum J)\dot{\omega}$, in which ω is the current angular velocity of the system and $\sum J$ is the sum of the inertias in the model. The reduction rate, which is found from this relation, is the same as the one seen in the sensitivity plot. The slope for C_2 is four times of the other coefficients, namely, C_1 and C_3 . The reason is that C_2 is repeated for the four cylinders. Therefore, by perturbation of C_2 , all the four cylinders frictions will be perturbed.
- The damping coefficients will not alter the mean value of the output, and furthermore, it is seen from the plots that after a while, the amplitude of the oscillations becomes lower since the damping coefficients are perturbed.

Fig. 5 contains the sensitivities of the damping wheel angular velocity, (rad/sec), to the three inertias, i.e., damping wheel J_1 , each cylinder J_2 , flywheel J_3 , and to the piston reciprocating mass, m . The inertias are important for catching



(a): Damping wheel angular velocity, measurement vs. simulation.



(b): The ratio of the residual value $\varepsilon_i(\mathbf{p})$ over the real data, in each time sample i .

Fig. 3: Comparison between simulation and empirical data for *validation cycle*.

the amplitudes of the main frequencies in the system. This is exactly what is seen from the shape of the sensitivities to $J_1 - J_3$. The oscillatory motion of the reciprocating mass of the piston adds an oscillatory torque to the pressure torque. This causes oscillations in the angular velocity of the engine as well, which is also presented in the sensitivity plot.

5 Conclusion

A powertrain model, suitable for studying torsional vibrations, was presented. The model is in modular form, and thus flexible enough, to be adapted for various powertrains structures, i.e., front-wheel, rear-wheel, and four-wheel drive systems with different number of cylinders. In order to get better parameter estimation performance, for the engine-block, the idling situation was considered and the clutch was disengaged. Therefore, the number of parameters to be estimated decreased to 10. The results from the model for damping wheel angular velocity, using the estimated parameters, validated versus the validation measured data, which showed significant similarity. This proves the ability of the model to resemble the main behaviours of the real system. Furthermore, the performed sensitivity analysis was helpful to understand the importance of different parameters in the model, besides how the model responds to perturbation of these parameters.

Acknowledgements

Financial support for this research was provided jointly by CADICS, a Linnaeus research environment for Control, Autonomy, and Decision-making in Complex Systems, and Linköping University, Sweden.

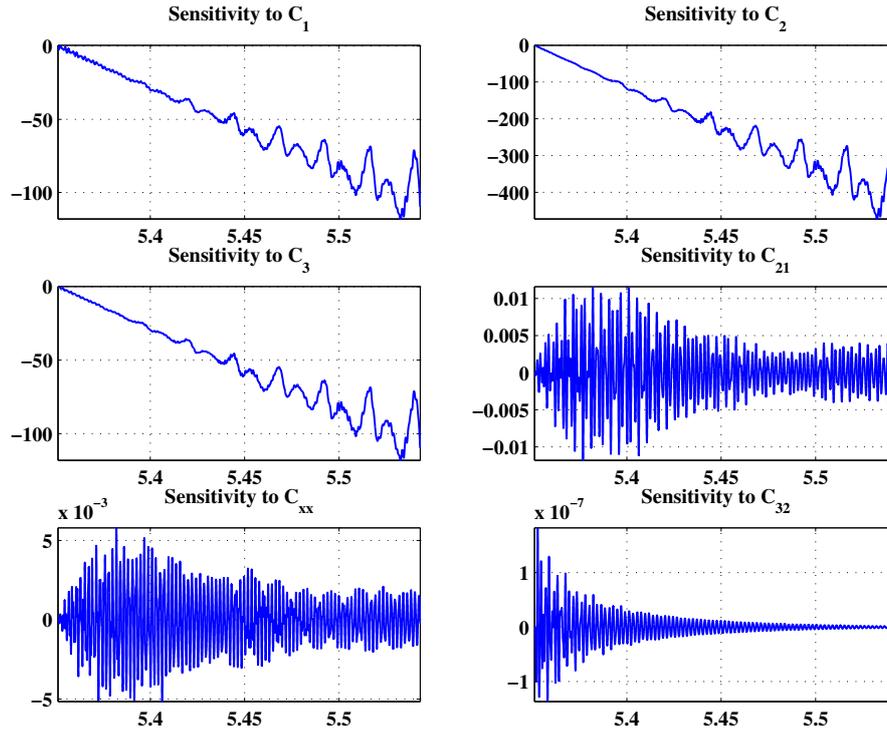


Fig. 4: Sensitivities of the damping wheel angular velocity to 3 friction coefficients, $C_1 - C_3$, and 3 damping coefficients, C_{21} , C_{32} , and C_{xx} .

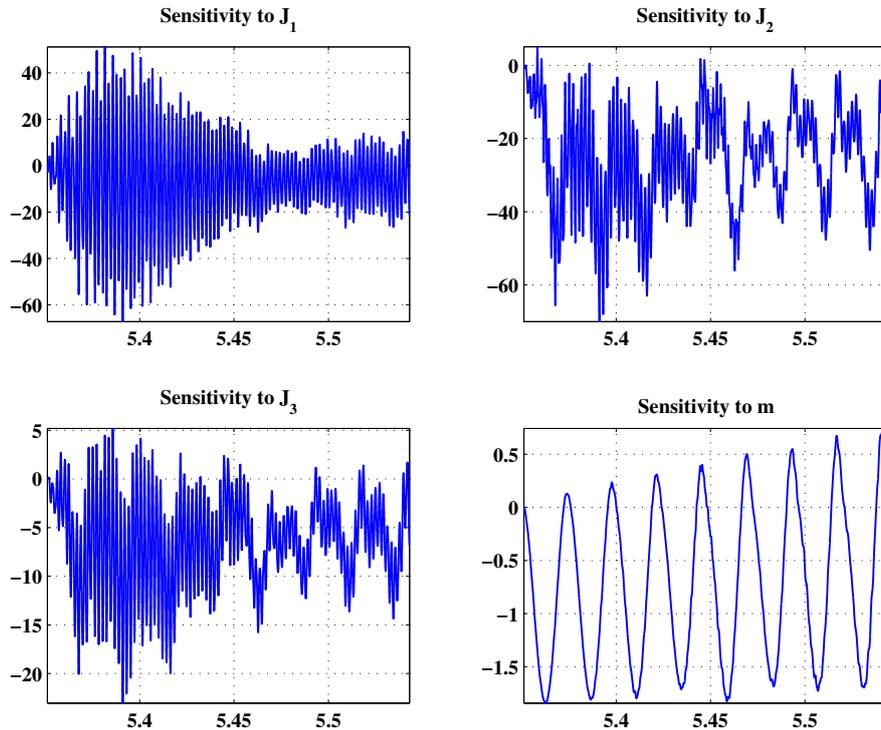


Fig. 5: Sensitivities of the engine angular velocity to 3 inertias, i.e., damping wheel J_1 , each cylinder J_2 , flywheel J_3 , and to the piston reciprocating mass m .

References

- [1] Rahnejat, H. (1998). *Multi-body dynamics vehicle, machines, and mechanisms*. Published by Society of Automotive Engineering, Inc.

- [2] Eriksson, D., Eriksson L., Frisk, E., and Krysander, M. (2013). *Flywheel angular velocity model for misfire simulation*. 7th IFAC international symposium on advances in automotive control (AAC). Tokyo, Japan.
- [3] Johansson, R. (2012). *Modeling of the engine and driveline related disturbances on the wheel speed in passenger cars*. Master Thesis, Linköping University, Sweden.
- [4] Wu, J.-S., and Chen, C.-H. (2001). *Torsional vibration analysis of gear-branched systems by finite element method*. Journal of sound and vibration 240 (1) 159-182.
- [5] Rabeih, E. (1997). *Torsional vibration analysis of automotive driveline*. Ph.D. Thesis, The university of Leeds. U.K.
- [6] Crowther, A. (2004). *Powertrain vibration: Modeling, simulation and testing*. FISITA world automotive congress. Barcelona, Spain.
- [7] Couderc, P., Callenaere, J., Der Hagopian, J., and Ferraris, G. (1998). *Vehicle driveline dynamic behaviour: experiment and simulation*. Journal of sound and vibration 218 (1) 133-157.
- [8] Ljung, L. (1999). *System identification: Theory for the user, Second edition*. Prentice Hall Information and System Sciences Series.
- [9] Schön, T. (2006). *Estimation of nonlinear dynamic systems- theory and applications*. Ph.D. Thesis, Linköping University, Sweden.
- [10] Heij, C., Boer, P., Franses, P., Kloek, T., and Dijk, H. (2004). *Econometric methods with applications in business and economics, Chapter 4*. Oxford University Press.
- [11] Cacuci, D. (2003) *Sensitivity and uncertainty analysis*. Chapman & Hall/CRC.
- [12] Eriksson, L. (2004). *Minimal manual to lsoptim*. Linköping University, Sweden.

State Estimation in a CO₂ Capture Plant

Sanoja A. Jayarathna^a, Bernt Lie^a, Morten C.
Melaen^{a,b}

a. Telemark University College, Kjølnes ring 56, P.O. Box 203, N-3901,
Porsgrunn, Norway.

b. Tel-Tek, Kjølnes ring 30, N-3918, Porsgrunn, Norway

sanoja.ariyaratna@hit.no; bernt.lie@hit.no; morten.c.melaen@hit.no.

Corresponding Author: Morten. C. Melaen.

Phone: +47 3557 5286. Fax: +47 3557 5001.

Abstract

Post combustion CO₂ capturing holds an important position in the area of carbon capture and sequestration (CCS). Research in this area range from experimental work to modeling work. Dynamic models are interesting since these describe the plant operation during variations, up-stream or down-stream, and due to their usefulness in control design. To take full advantage of state space models in control design, it is necessary to have on-line knowledge of all states, also states that are not measured directly. Techniques for state estimation, such as Kalman filter based methods, thus form key technology for advancing control solutions. But state estimation is also of interest in its own right for making available on-line knowledge of states. In this study, a dynamic model of an amine based CO₂ capture plant is used as a basis for a state estimator. A high order version of the model is used to represent the “real” plant. A reduced order model of the plant is then used for state estimation, and the Ensemble Kalman filter is used.

Key words: CO₂ capture; State Estimation; Ensemble Kalman Filter, MATLAB.

1 Introduction

Power generation via fossil fuel-fired power plants is known to be the largest single source of CO₂ emission in the world [1]. The development of capture technologies targeting such sources therefore is important for achieving the goals in CO₂ emission reduction. Post-combustion capture, pre-combustion capture and oxy-fuel combustion are the three main technologies available at present [2], and much research is done with the prospect of developing those techniques further. Post combustion capture is still the best known technique, possibly due to the large number

of existing power plants, and the promising developments that are available. CO₂ capture by amine absorption and stripping is currently considered to be the most feasible option for the removal of carbon dioxide from the power plants' exhaust gases [3].

Modelling work related to CO₂ capture technologies plays an important role with respect to the design, control and optimization of the capture process. Steady state models are important for design and optimization purposes, and dynamic simulation models are important for control applications. Several dynamic models for simulating the amine based CO₂ absorption plants are presented in literature [4]-[6].

A model consisting of a set of first order differential equations to represent the system is referred to as a state space model. To take full advantage of state space models in control design, it is necessary to have on-line knowledge of all states, also states that are not measured directly. Techniques for state estimation, such as Kalman filter based methods, thus form a key technology for advancing control solutions. But state estimation is also of interest in its own right for making available on-line knowledge of states.

The Kalman filter based methods vary from basic Kalman filter (KF) to its extensions and generalizations such as the Extended Kalman filter and Unscented Kalman filter. The basic Kalman filter is applicable for linear dynamic systems, while the extensions and generalizations of the method are there to be used with the nonlinear dynamic systems [7]. The Ensemble Kalman filter (EnKF) is another alternative to the traditional Kalman filter for better handling of nonlinear models with large number of states [8].

Use of traditional KF methods for models with high-dimensional state vectors is computationally difficult as an error covariance matrix for the model states needs to be stored and propagated in time. The Extended Kalman filter uses a linearized equation for the error covariance evolution when the model dynamics are nonlinear. This linearization can result in unbounded linear instabilities for the error evolution [8]. These two problems can be solved to a great extent by using the Ensemble Kalman filter.

In this study, a dynamic model of an amine based CO₂ capture plant is used as a basis for a state estimator. The model as developed and published by Jayarathna et al. ([6], [9] - [12]) is used in this study as the plant model. A high order version of the model is used to represent the "real system" due to the absence of real plant data. A reduced order model of the plant is then used for state estimation, and the Ensemble Kalman filter is used.

2 Theory

The EnKF algorithm is presented in details with the derivation by Evensen [13]. Initialization of the estimator is done by providing values for the initial ensemble. When the number of simulations in an ensemble is N, the values of the initial ensemble are given according to the eq. 1.

$$x_{0|0} \sim \mathcal{N}(\bar{x}_0, P_0) \quad (1)$$

Having the initial ensembles available, the state estimator runs through a propagation step and a measurement update step at each time step. The propagation step consists of three consecutive

steps, the ensemble propagation, the estimated state-output propagation and the covariance calculation. For each simulation i ($i = 1, 2, \dots, N$), the ensemble propagation step is given by eqs. 2 and 3.

$$x_{k|k-1}^i = f_{k-1} \left(x_{k-1|k-1}^i, u_{k-1}, w_{k-1}^i \right) \quad (2)$$

$$y_{k|k-1}^i = h_{k-1} \left(x_{k|k-1}^i, v_{k-1}^i \right) \quad (3)$$

Here w_{k-1}^i and v_{k-1}^i are the model disturbances ($w \sim \mathcal{N}(w_0, W)$) and the measurement noise ($v \sim \mathcal{N}(0, V)$).

The propagation of the estimated state and output is given by eqs. 4 and 5.

$$\hat{x}_{k|k-1} = \frac{\sum_{i=1}^N x_{k|k-1}^i}{N} \quad (4)$$

$$\hat{y}_{k|k-1} = \frac{\sum_{i=1}^N y_{k|k-1}^i}{N} \quad (5)$$

The covariance calculation is done according to eqs. 6 and 7.

$$e_{x,k|k-1}^i = \left(x_{k|k-1}^i - \hat{x}_{k|k-1} \right) \quad (6)$$

$$P_{k|k-1} = \frac{\sum_{i=1}^N \left(e_{x,k|k-1}^i \right) \left(e_{x,k|k-1}^i \right)^T}{N - 1} \quad (7)$$

After completing the propagation step, the state estimator updates the predictions using the available measurements. The measurement update step consists of two consecutive stages in the used algorithm, Kalman gain calculation and state-out-covariance update. The Kalman gain calculations is performed according to the eqs. 8-11.

$$e_{y,k|k-1}^i = \left(y_{k|k-1}^i - \hat{y}_{k|k-1} \right) \quad (8)$$

$$P_y = \frac{\sum_{i=1}^N \left(e_{y,k|k-1}^i \right) \left(e_{y,k|k-1}^i \right)^T}{N - 1} \quad (9)$$

$$P_{xy} = \frac{\sum_{i=1}^N \left(e_{x,k|k-1}^i \right) \left(e_{y,k|k-1}^i \right)^T}{N - 1} \quad (10)$$

$$K_k = P_{xy} P_y^{-1} \quad (11)$$

The state-out-covariance update step is given by eqs. 12-14.

$$x_{k|k}^i = x_{k|k-1}^i + K_k \left(\left(y_k + v_k^i \right) - y_{k|k-1}^i \right) \quad (12)$$

$$\hat{x}_{k|k} = \frac{\sum_{i=1}^N x_{k|k}^i}{N} \quad (13)$$

$$P_{k|k} = P_{k|k-1} - K_k P_y K_k^T \quad (14)$$

Here $y_k = g(x_k)$.

Updated values of the states ($x_{k|k}^i$) are then taken as the initial values to run the estimator for the next time step. The averaged values ($\hat{x}_{k|k}$) are the estimates for the time step. The covariance matrix ($P_{k|k}$) can be used to get an idea about the uncertainty of the predictions.

3 Implementation

In the implementation of the state estimator, the number of ensembles is taken to be 60 ($N = 60$). The Parallel Computing Toolbox in MATLAB is used to work with 12 threads at the same time. This way, 12 simulations are performed simultaneously to complete each ensemble, and the time required for completing the simulations for an ensemble is reduced by five times. Initial state values (x_0) are taken from a simulation with sufficiently long simulation time, thus steady state is assumed. The covariance matrix of the initial states (P_0) is given according to the magnitude of the x_0 values (as a rule of thumb, the values of the diagonal matrix (P_0) are taken as a fraction of the initial state values).

A higher order model, i.e. with a higher number of control volumes than in the state estimator, is used to represent the “real” system due to the absence of real plant data. According to the assumptions made during the model development an infinite number of CVs should be used to represent the columns in the real system. But a finite number of CVs has to be used in practice. Use of a finite number of control volumes introduces diffusion into the column models. Diffusion is a phenomenon that occur in the absorption and stripping columns. Therefore, it is acceptable to have a finite number of control volumes in the model that represents the real plant.

Several compositions and temperatures are taken as the measurements; we assume a total of 26 measurements. Measured compositions are the composition of the cleaned gas and the gas leaving the stripping column (can be measured by gas chromatography). The amount of dissolved CO₂ (total CO₂) and MEA (total MEA) in the solvent streams leaving the absorption tower, stripping tower and the buffer tank are also taken as measurements. Mass flow rate of the amine solutions leaving the absorber, stripper and the buffer tank without the mass flow rate of the dissolved CO₂ are also included in the measurements. Temperatures of the liquid and vapor leaving the absorption tower, temperatures of the liquid and vapor streams leaving the stripping column, temperatures of the liquid streams leaving the heat exchangers and the temperature of the buffer tank are taken as the temperature measurements. Measurement noise is assumed to be white noise.

4 KF Predictions

An analysis related to the number of control volumes used in the tower models, given in figure 1, showed that the execution time increases quadratically with the increasing number of control volumes. Therefore the number of control volumes to be used in the state estimator is chosen to be around 50 control volumes.

When the sensitivity of the model predictions to the number of control volumes used in the tower discretization is considered, from figure 1 it is noticeable that the model predictions

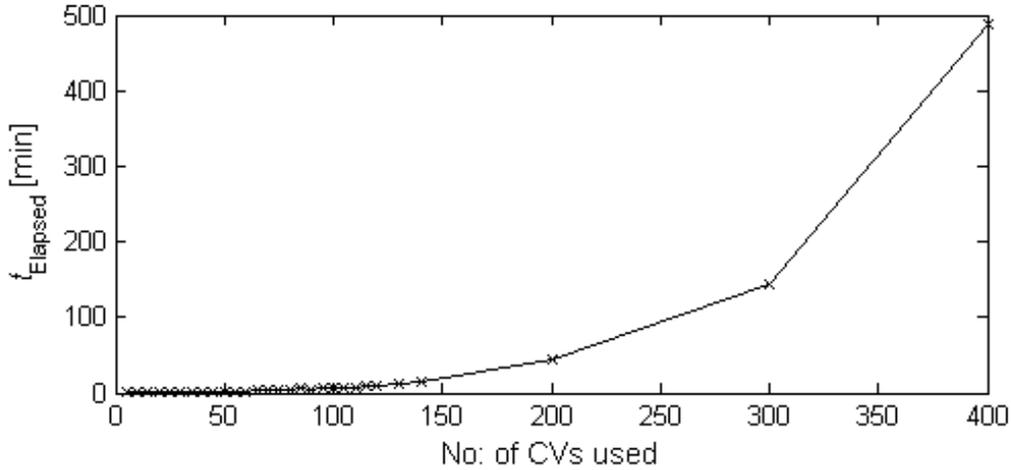


Fig. 1. Execution time of the simulations with the number of control volumes used in the tower discretizations.

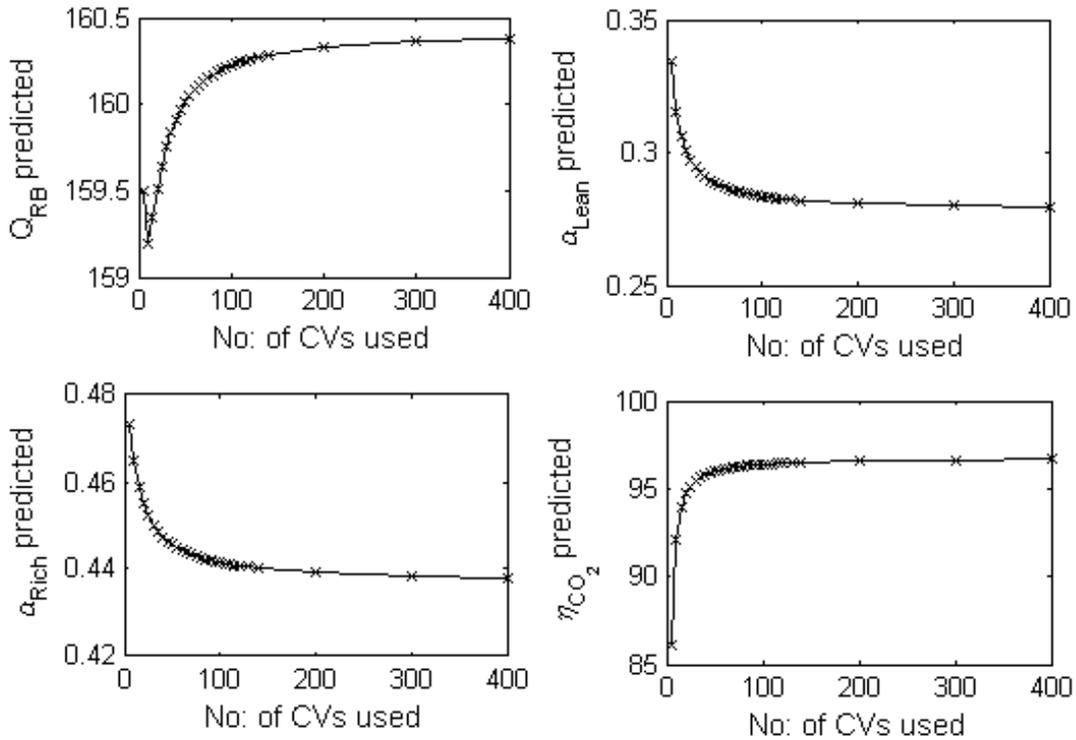


Fig. 2. Dependency of the model predictions on the number of control volumes in the tower discretizations. Q_{RB} predicted is the predicted re-boiler duty, α_{Lean} predicted is the predicted lean loading value, α_{Rich} predicted is the predicted rich loading value and η_{CO_2} predicted is the predicted CO_2 removal efficiency.

improves with the increasing number of control volumes up to about 100 CVs in towers. The predicted values remains with very little variations for higher number of control volumes than 100, but as can be seen from figure 1 the execution time for the simulations increases very much for higher number of CVs than 100. Therefore, 100 CVs is chosen as a rough limit for the maximum number of CVs that are used in the towers for the model that is used for representing the real system. The dependency of the model predictions with the number of control volumes used is shown in figure 2.

Several different cases are studied to analyze the estimates of the EnKF to different scale of

Table 1

Details of the cases used for analyzing the sensitivity to the model error.

Case no:	CVs in Model	CVs in Real System	Ratio	Computation time
1	15	15	1	2 h
2	50	50	1	20 h
3	40	50	1.25	6 h
4	40	100	2.5	6 h

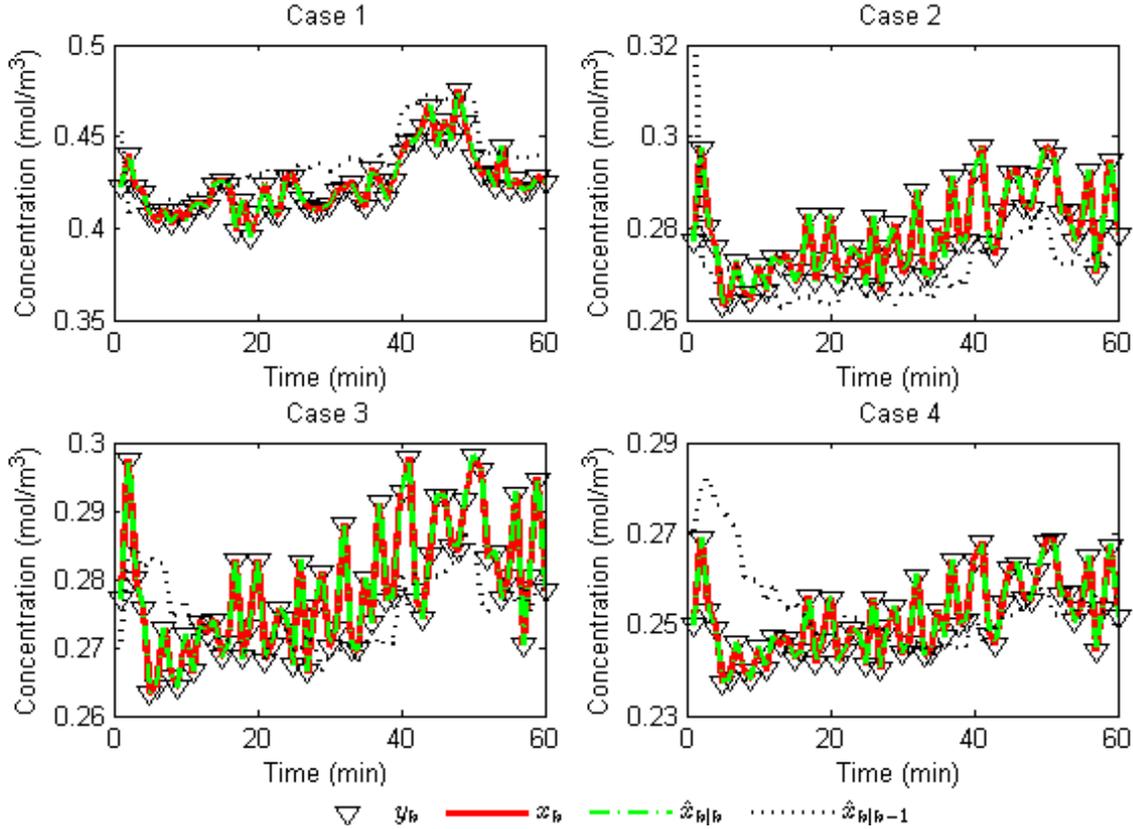


Fig. 3. CO₂ concentration in the cleaned gas leaving the absorption tower. y_k : measurement, x_k : real state, $\hat{x}_{k|k}$: a posteriori state, $\hat{x}_{k|k-1}$: a priori state.

model errors. The model error was increased by increasing the difference between the number of control volumes used in the model that represents the real plant and the model used in the state estimator. One hour of the plant operation is simulated. Information about the cases performed for analyzing the sensitivity of the state estimator for the model error are given in the Table 1 . Selected states are presented in figures 3 - 8 for the four cases of Table 1 to analyze the quality of the estimates. Figures 3 - 5 show three measured states and figures 6 - 8 show three unmeasured states.

Prediction of CO₂ concentration in the cleaned gas leaving the absorption tower (figure 3), which is a measured state, appear to be a good match with the real state in all four cases. Further, it can be seen that the measurement noise of this state is very small, and that should be the reason for this high quality prediction in all the cases. The other two measured states presented, the temperature of the rich amine steam leaving the absorber (figure 4) and the temperature of the rich amine leaving the cross heat exchanger (figure 5), are predicted with larger errors. The pattern of the predictions appears to follow the pattern of the real states, and the error of the prediction appears to be increasing with the model error.

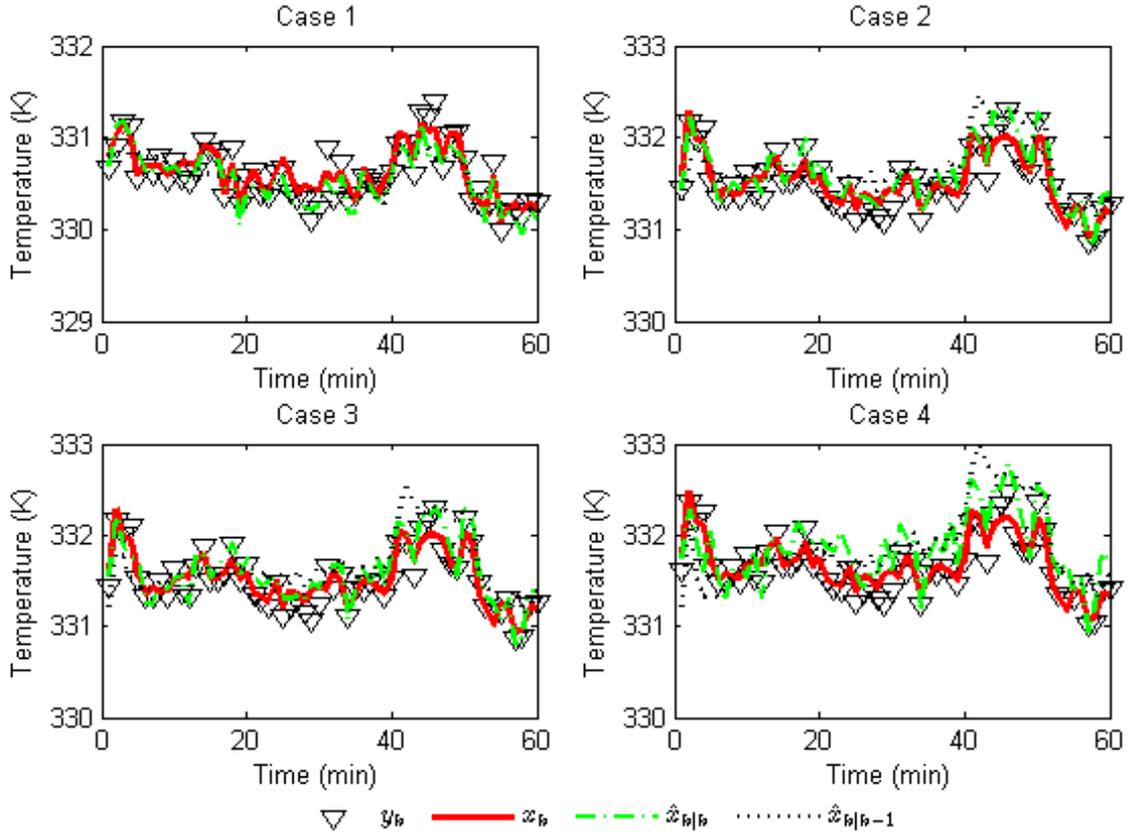


Fig. 4. Temperature of the amine streams leaving the absorption column. y_k : measurement, x_k : real state, $\hat{x}_{k|k}$: a posteriori state, $\hat{x}_{k|k-1}$: a priori state.

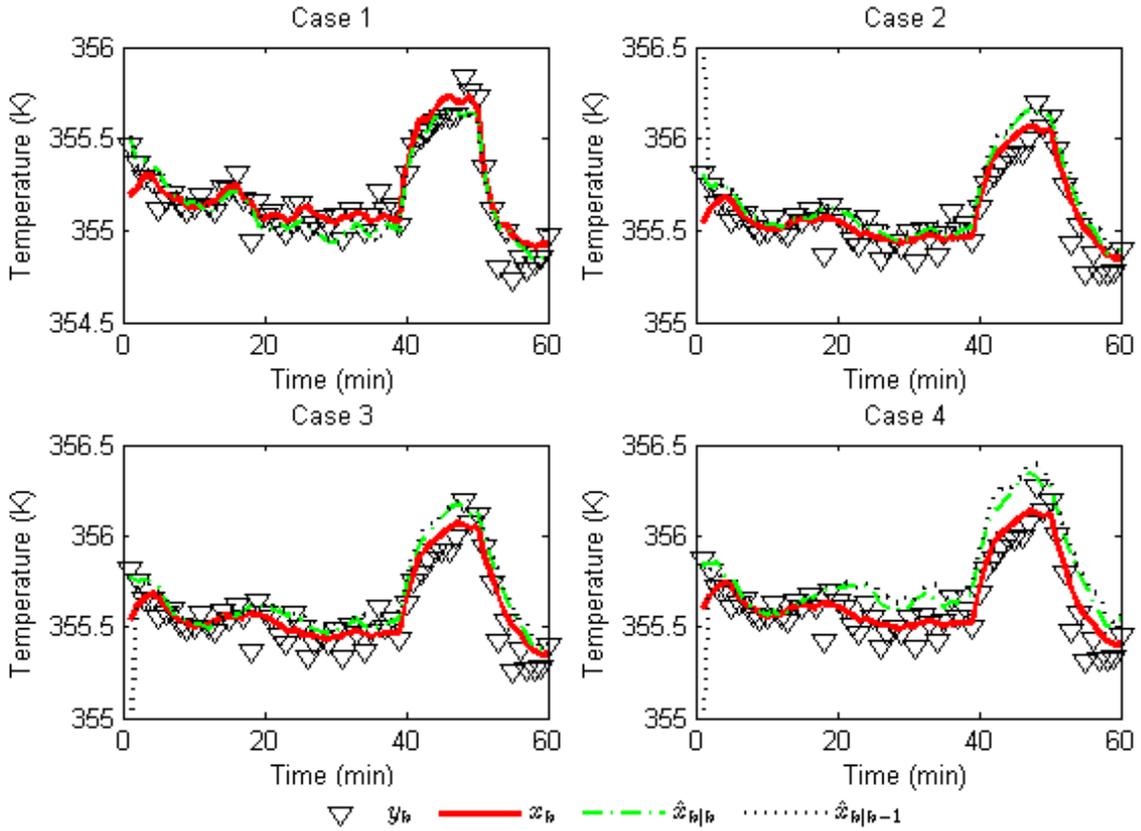


Fig. 5. Temperature of the rich amine stream leaving the cross heat exchanger. y_k : measurement, x_k : real state, $\hat{x}_{k|k}$: a posteriori state, $\hat{x}_{k|k-1}$: a priori state.

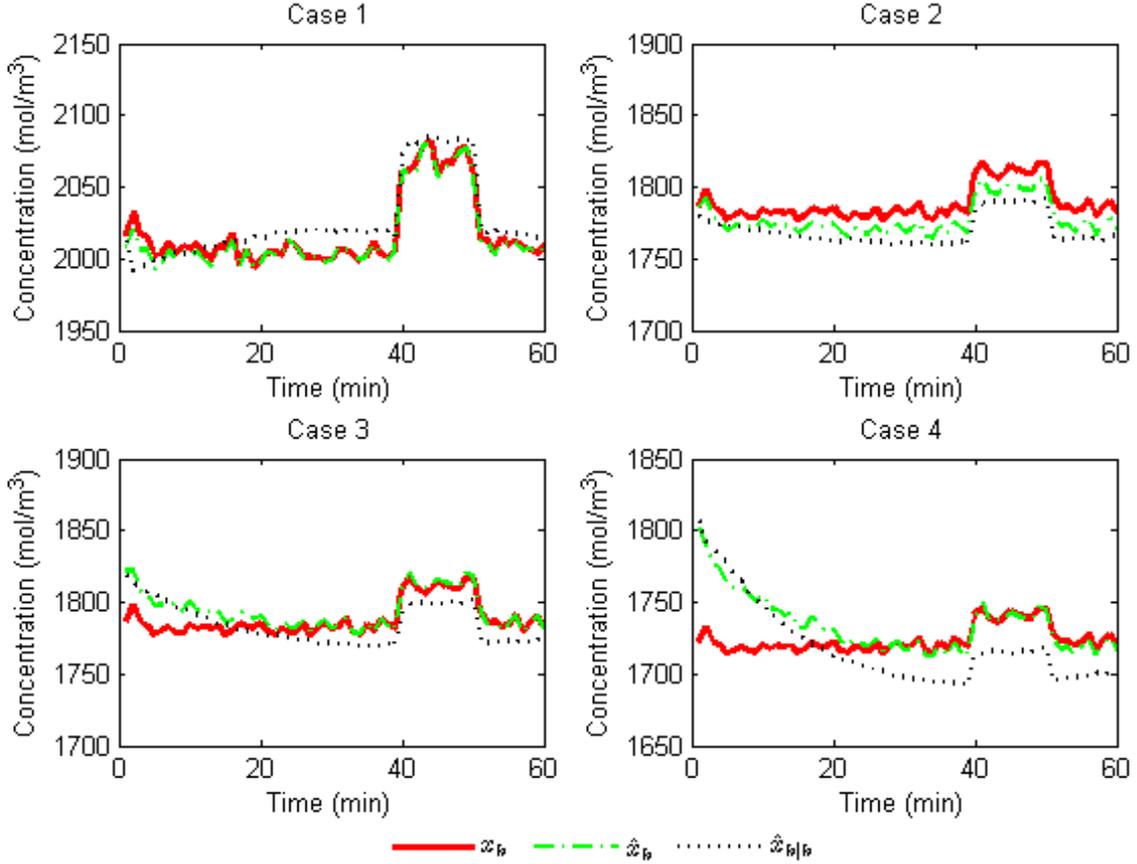


Fig. 6. Total CO₂ concentration in the amine stream at 1/3 of the packing height of the absorption tower. x_k : real state, $\hat{x}_{k|k}$: a posteriori state, $\hat{x}_{k|k-1}$: a priori state.

Estimates of the states that are not measured are also in good accordance with the pattern of the real states (figures 6 - 8). Similar to the observation with the measured states, it can be seen that the error in the estimated values are increasing with the increase in the model error.

When the computation time needed for each of the cases is considered, it can be seen that there will always be a time delay before the estimates are available.

5 Conclusions and recommendations

A state estimator for making on-line predictions of the states of a CO₂ capture plant is developed. The Ensemble Kalman Filter algorithm is used in the state estimator, due to the suitability of the algorithm to handle nonlinear models with a large number of states. A simple model (with small number of control volumes) is used in the estimator while a high order model (with large number of control volumes) is used to represent the real system. Absence of real plant data is the reason for using a high order model as the real system. According to the assumptions of the model development, an infinite number of control volumes should be used in the tower discretization to represent the real system. In practice a finite number of control volumes is used, though. Tower discretization into finite number of control volumes introduces diffusion into the system, which is also there in reality. Therefore, use of a finite number of control volumes in tower discretization is justified.

Sensitivity of the estimates to the model error is analyzed by performing simulations with

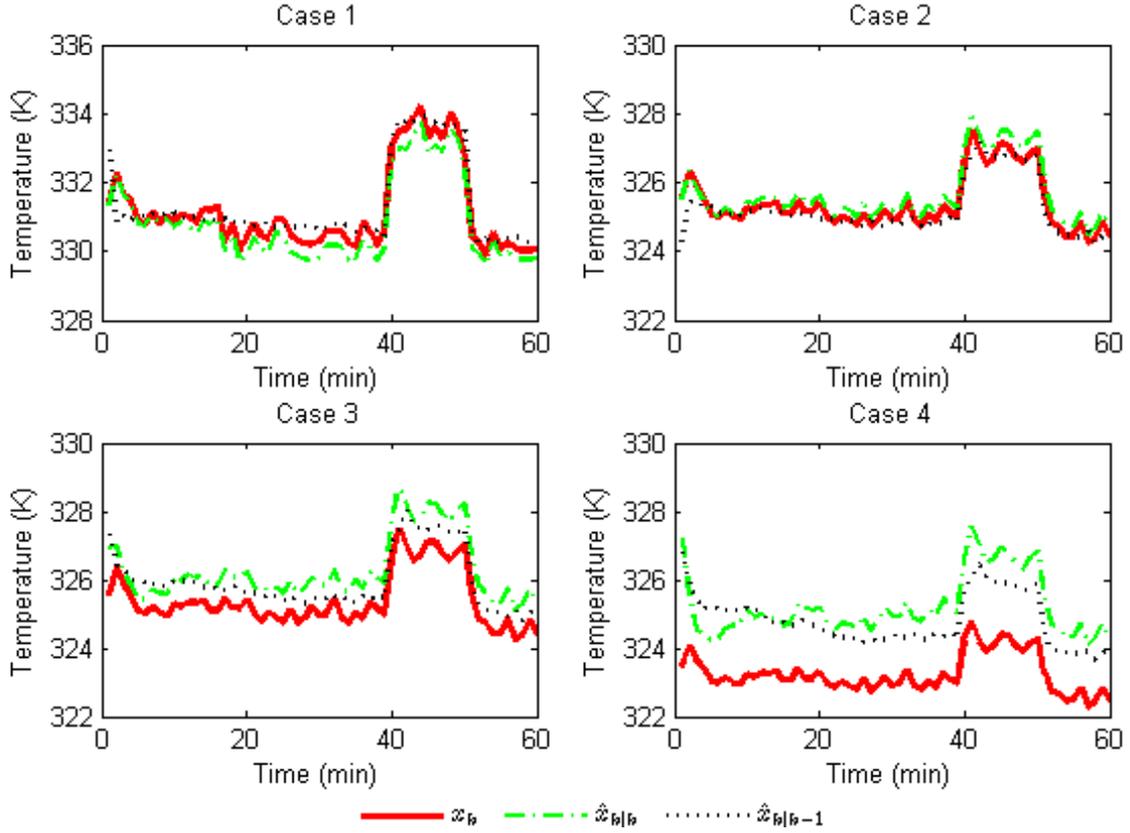


Fig. 7. Liquid phase temperature at 1/3rd of the absorber packing height. x_k : real state, $\hat{x}_{k|k}$: a posteriori state, $\hat{x}_{k|k-1}$: a priori state.

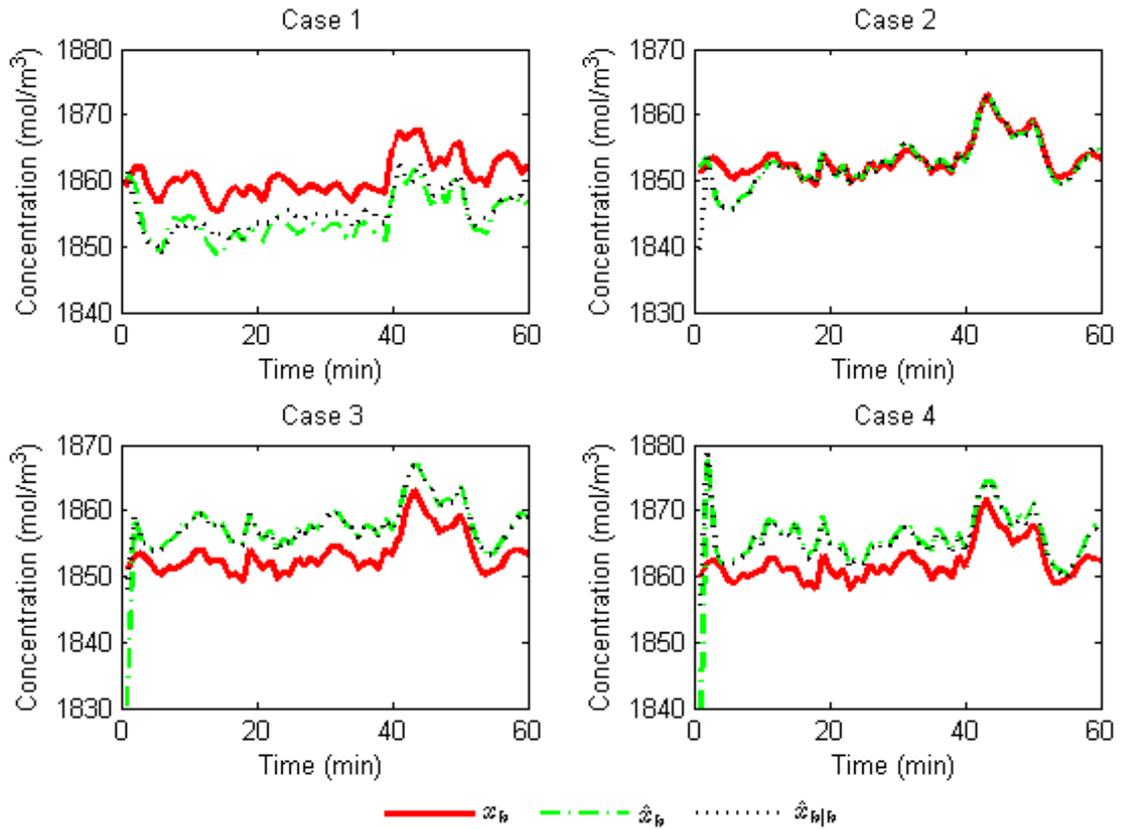


Fig. 8. Total CO₂ concentration at 1/3rd of packing height of the stripping tower. x_k : real state, $\hat{x}_{k|k}$: a posteriori state, $\hat{x}_{k|k-1}$: a priori state.

increasing model error (by making the difference between the model and the system to increase). Some predictions seem very good and some predictions seems to be poorer. The observation from the results of the sensitivity analysis, in general, is that the deviation of the estimates from the real states increases with the increasing model error.

According to the simulation time needed for each of the cases considered in the sensitivity analysis, it can be seen that the estimates are available always with a time delay. This high computational time of the estimator is a problem for making available the timely estimates of the unmeasured states. The use of iterations with the flash calculations in the model can be a reason for the computation time of the estimator. Making table look-up values available to be used instead of performing flash calculations, will be an option to speed up the estimator.

When the tendency to increase the error in the estimates with the increase of model error and the time delay of the estimates are considered, it is concluded that some restructuring of the model is required before the estimator is ready for on-line use.

References

- [1] Freund P. Making deep reductions in CO₂ emissions from coal-fired power plant using capture and storage of CO₂. *Proc IMechE Part A, J Power Energy*, 2003, 217: 1-7.
- [2] Cormos A.M., Gáspár J., Padurean A., Cormos C.C. Techno-economical analysis of carbon dioxide absorption in mono-ethanolamine by mathematical modelling and simulation, in Proc. ESCAPE20, Ischia, Italy. 2010, 81-1 - 81-6.
- [3] Plaza J.M., Wagener D.V., Rochelle G.T. Modelling CO₂ capture with aqueous monoethanolamine. *Energy Procedia*, 2009, 1: 1171-1178.
- [4] Biliyok C., Lawal A., Wang M., Seibert F. Dynamic modelling, validation and analysis of post-combustion chemical absorption CO₂ capture plant. *International Journal of Greenhouse Gas Control*, 2012, 9: 428-445.
- [5] Lawal A., Wang M., Stephenson P., Koumpouras G., Yeung H. Dynamic modelling and analysis of post-combustion CO₂ chemical absorption process for coal-fired power plants. *Fuel*, 2010, 89: 2791-2801.
- [6] Jayarathna S.A., Lie B., Melaaen M.C. Amine based CO₂ capture plant: Dynamic modeling and simulations. *International Journal of Greenhouse Gas Control*, 2013, 14: 282 - 290.
- [7] Simon D. *Optimal state estimation*. John Wiley & Sons, USA, 2006.
- [8] Evensen G. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 2009, 29: 83 - 104.
- [9] Jayarathna S.A., Lie B., Melaaen M.C. NEQ rate based modeling of an absorption column for post combustion CO₂ capturing. *Energy Procedia*, 2011, 4: 1797 - 1804.
- [10] Jayarathna S.A., Lie B., Melaaen M.C. Dynamic modelling of the absorber of a post-combustion CO₂ capture plant: Modelling and simulations. *Computers and Chemical Engineering*, 2013, 53: 178 - 189.
- [11] Jayarathna S.A., Weerasooriya A., Lie B., Melaaen M.C. Dynamic Operations of the Stripping Column of a CO₂ Capture Plant. "Proceedings of SIMS53", Reykjavik, Iceland, October 4-6, 2012.
- [12] Jayarathna S.A., Lie B., Melaaen M.C. Development of a dynamic model of a post combustion CO₂ capture process. "Proceedings of GHGT11", Kyoto, Japan, November 18 - 22, 2012.

[13] Evensen G. *Data Assimilation*, second ed. Springer, USA, 2009.

To design a novel micromachined electromechanical acoustic sensor to wrap around Roebel bars of large generators to monitor partial discharge

Kourosh Mousavi Takami
k.takami@aol.com

Abstract— One of the most effective methods to detect insulation faults in large generators is the method of partial discharge. This predictive maintenance tool is useful and very effective, if it interpreted by an experienced engineer.

Industry statistics by IEEE and EPRI show that more than 40% of all rotating machine faults arise due to damage of the stator winding and core insulation.

Today, electric and magnetic methods are used to detect partial discharges which in spite of all advantages are unreliable. Improper interpretation by engineers, error of measuring devices, short time measuring, etc are such problems.

The author has presented a micromachined piezoelectric multi sensor. Hall Effect magnetoresistive sensor can be used with piezoelectric sensor to verify results. It is wrapped around the final insulation or conductive tape of Roebel bar or semi conductive around overhangs. These sensors together with a pro system enable continuously to monitor PD.

Test performed on a pre damaged Roebel bar and PD results in AE and apparent charge methods were registered. Results show that AE method is more effective even PD occur in a point close to neutral.

A FEM simulation carried out and PD frequencies in each layer of insulation were measured. Simulation showed that deeper damages have lower frequencies.

Keywords—MEMS, sensor, PD, monitoring

I. INTRODUCTION

Rudolf (1) investigations on the categorizing of generator faults illustrated that 56% of the failed machines showed insulation damage, other major types being mechanical, thermal and bearing damages. As shown in figure 1, PD contribution is about 22% from all (2).

Several electrical methods are used to monitor of PD activities. All of those have significant inaccuracy. Acoustic PD detection is a noninvasive method to detect of partial discharge activities in Roebel (3) bars insulation and core insulation as well. It is as complimentary to the electrical or electromagnetically techniques of primary PD detection in insulation.

Complex insulating system (4), (5) in bars and strands are used. Those are such as votafix1 (resined tape to make homogeneity on the bar surface), Calmicaglass (main insulating tape including of mica, glass and resin for straight bars which are located in slots), Calmicafab that is like Calmicaglass but more improved, Calmicaflex (is very flexible resin rich and used somewhere that cannot go under press and have limitation to do with heater and press machine).

A conductive tape is wrapped around final insulating Calmicaglass with trade name of Contafel. It protects the bar against corona by increasing of surface resistance 150 to 650. A semi conductive tape, EGSB- semi conductive varnish, is wrapped around the overhang bars (some part of bars that is located in the out of slots and machine).

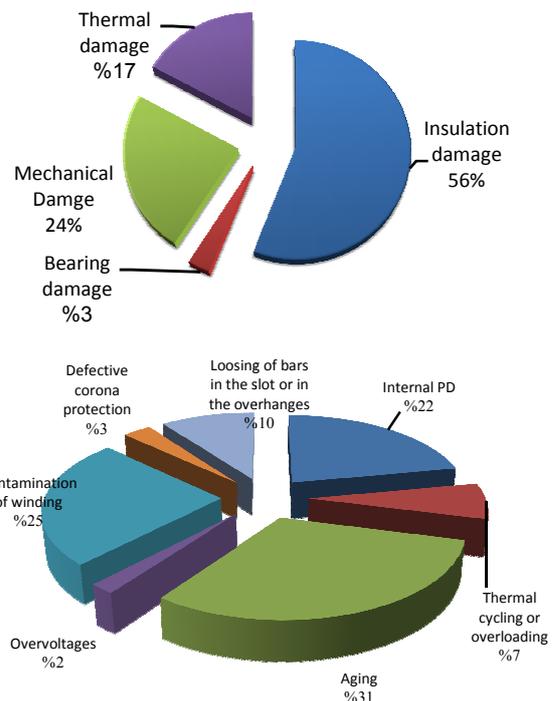


FIGURE 1A AND 1B. DAMAGES OF HYDROGENERATORS (LEFT) AND ROOT CAUSES OF INSULATION DAMAGES (RIGHT).

This is very important that conductor tape around the bar where is located in the slot and so semi conductor around the overhang have overlap. This over lap and length of conductive and semi conductive that put in other side can decrease electrical fields in those area.

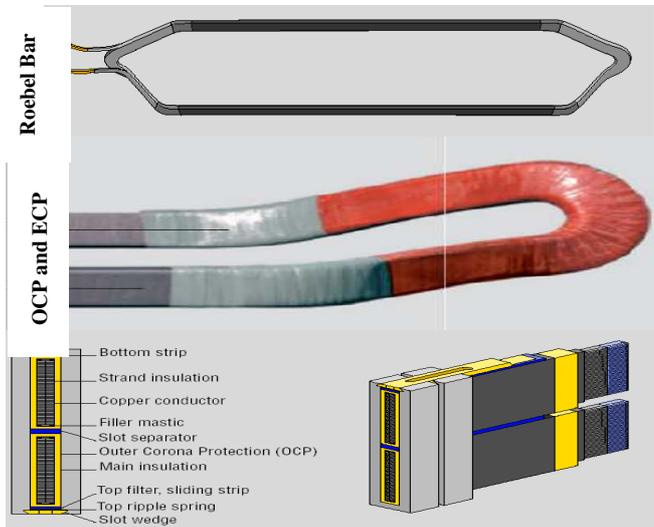


FIGURE 1. OUTER (OCP) AND END (ECP) CORONA PROTECTION OF BAR AND CORE IN A SLOT (6).

To protect of overhang side against dust and humidity, isoseal is used. Isoseal P713 is one of it, which wrapped around overhangs. To protect of overhang against outer corona, Epoflex is used for outer corona protection (OCP).

Total insulation thickness is 2.5kV/mm. It means for a 15 kV bar, thickness is about 6mm.

Nevertheless, PD is one of the most sources of insulation breakdown. It emits sound, mechanical pressure, light and chemical odors. Electrical methods are used to detect the phenomena. Because of several problems electric method is not responsible. Author suggested acoustic emission method that is more efficient if find high technology to produce sensors with low thickness and well efficiency.

Electromagnetic wave velocity is 3×10^8 m/s, but the velocity of sound wave propagation in air is as slow as about 344m/s (at 20°C). At these slower velocities, wavelengths are short, meaning that higher resolution of distance and direction can be obtained.

Because of the higher resolution, it is possible to get larger accuracy. Velocity of sound wave propagation “c” is expressed by the following formula.

$$c = 331.5 + 0.607t \text{ (m/s) where } t = \text{temperature } (^\circ\text{C}).$$

It shows in 80 °C (average temperature of bar), sound velocity is 380 m/s.

Now it needs to introduce Silicon based micro machined MEMS microphone to detect PD sound.

Since recently decades piezoelectric and capacitance type microphones more developed. In piezoelectric, piezoresistive transducer, optical and capacitive (based on the capacitive detection principle) microphones, pressure or displacement translated into a voltage by signal conditioning electronics. Polyvinylidene fluoride, PVDF, $(\text{CH}_2\text{-CF}_2)_n$ transducers have higher sensitivity as well. Acquired data is in form of a voltage which has: amplitude, frequency, and phase of pressure wave.

The first piezoelectric microphone with MEMS method was made by ZnO introduced by Royer et al. in 1983. (7) its microphone dimensions were: 3 mm diameter, 30 μm thick diaphragms, and 3 μm thick ZnO layer, with a sensitivity of 25 μV/μbar and 66 dB (A) for noise.

Other piezoelectric microphone was made by Horowitz et al. for aeroacoustic measurements. The microphone utilized a PZT ring around the circumference of the diaphragm, where stresses are largest. Microphone dimension was: 1.8 mm diameter and 3 μm thick PZT layer with 1.66 μV/Pa sensitivity, 35.7–169 dB re 20 μPa dynamic range and bandwidth up to 59 kHz. (8) This is currently the highest performing piezoelectric MEMS (Micro-Electro-Mechanical-Systems) aeroacoustic sensor. The generalized piezoelectric electric equations are illustrated in the following:

$$D_i = d_{ik} \cdot \sigma_k + \epsilon(d)_{ik} \cdot \sigma_k E_i$$

$$\epsilon_k = d_{ik} \cdot E_i + S_{ik}^E \cdot \sigma_k$$

where, D_i is electric displacement, C/m^2 , d_{ik} is the piezoelectric constant (charge density/applied stress) expressed in C/N , σ_k is the mechanical stress applied to the piezoelectric material [Pa], $\epsilon(d)$ is the electric permittivity at constant stress, E_i is the electric field [V/m], and S_{ik}^E is the mechanical compliance for a constant electric field. In the presented equations i is the indices of the components of polarization and k is the indices of the component of mechanical stress and strain (9).

E.C. Wentz designed the first capacitive, electrostatic, or condenser microphones in 1916 (10), see Fig. 2. High sensitivities and large dynamic ranges are the advantages of capacitive sensors against piezoelectric ones. Piezoelectric sensors fabrication is difficult and is temperature dependant in their operation capacitance sensors. High impedance which reduces the electrical signal, if a buffer is not close to the sensing element, external power is required, and the capacitance is the big disadvantage of capacitance unit. The capacity of capacitance nonlinearly changes with displacement of diaphragm which makes another disadvantage.

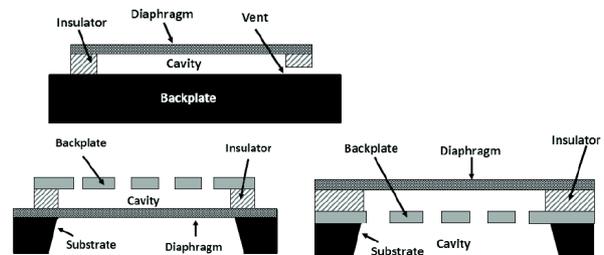


FIGURE 2. CROSS-SECTIONAL VIEW OF GENERIC MEMS CAPACITIVE SINGLE BACKPLATE MICROPHONES. LEFT TO RIGHT: L(A) SOLID BACKPLATE WITH SUSPENDED DIAPHRAGM. (B)PERFORATED BACKPLATE SUSPENDED ABOVE THE DIAPHRAGM. (C)DIAPHRAGM SUSPENDED ABOVE THE PERFORATED BACKPLATE. $C = E A/D$. $E = (6.5-4.5) \times 8.85 \times 10^{-18}$ & $C = 5-50$ pF.

Sean T. Hansen et al. (11) studied on a wideband micromachined acoustic sensors with radio frequency detection. A 1 mm^2 area microphone (capacitive micromachined ultrasonic transducer (CMUT) and membranes and radio frequency (RF) detection) consisting of 45 rectangular $70\text{-}\mu\text{m}$ by $190\text{-}\mu\text{m}$ membranes is connected as part of the phase detection circuit shown in Fig. 3. The remainder of the circuit consists of a 2.8 GHz dielectric-resonator oscillator signal source with output power of 20 dBm, a Minicircuits ZAPD-4 power splitter, a Marki Microwave M1-0204NA mixer, and a dc-coupled amplifier with 60 dB of gain constructed using the AD797 operational amplifier. The signal and noise levels are measured at the amplifier output using an SRS760 FFT spectrum analyzer.

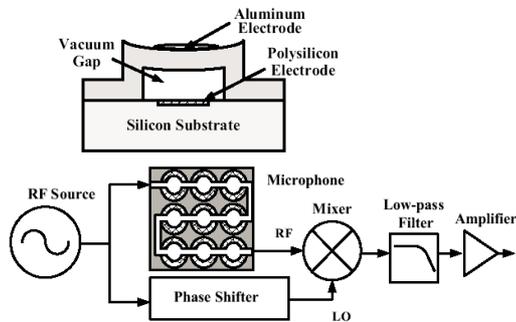


FIGURE 3. RIGHT: BLOCK DIAGRAM SHOWING THE MICROPHONE AS PART OF A PHASE DETECTION CIRCUIT TO RECOVER THE INCIDENT ACOUSTIC SIGNAL. LEFT: STRUCTURE OF A SINGLE CMUT MEMBRANE (11).

The measured sensitivity was 65 dB/Pa at 1 kHz, increasing to over 73 dB/Pa for frequencies above 10 kHz in a 1-Hz band.

A. Problems within PD measuring by electrical methods

Partial discharge has to release by an interference free object. Disturbance is rejected by using of electromagnetic compatibility devices and by measurement over 5MHz. To tune of system for removing of the all noises, so signal processing for time domain measurements and tunable narrow band width filter in frequency domain when signal to noise ratio is high are used. Capacitive and inductive coupling between stator phase windings make cross talk phenomenon, both methods have to be enable to make discrimination between PD and noises.

Several sensors are installed on the generator terminals, earth side, between slot and winding etc. Those can be installed permanently or when is necessity.

Coupling capacitors, capacitor that located in the insulators of isolated phase bus (IPB) or bus support capacitors,

capacitors that protects Gen C.B., single turn C.T. that get around the HV side of generator's terminal, pickup loop etc.

Here is a question, which one is more suitable. After precise investigation by individual researchers they found that bus support capacitors or bus coupling capacitors have a few better results by higher sensitivity.

There is some problem to trap and distinguish of PD signals by the mentioned electrical-capacitance methods. Each kind of sparking in setup transformer, PT and CT those are connected to generator side of transformer make noise. Capacitors of IPB with small degradation can make false noise that is captured by PD system.

Source of partial discharge in generators depend to generator life, insulation quality, load shock, ambient conditions etc. But in several cases false PD signal is received by electrical PD sensors that this signal is not related to generator PD. Such as: any sparking in step up transformer, fault in bleeding resistor of generator C.B., every sparking closed to current sensors, small shortage in bus support capacitor and something like those. Although, transient time between two different sensors in each phase and reflection of PD signals are used to make discrimination between PD and false signals. Also pickup loop is used as additional sensor to locate abnormal PD.

Stator winding act as a transmission line and PD marked itself after a short transient time at the generator terminal. The capacitive and inductive coupling in the end regions of generator caused a second propagation mode with higher frequencies and so cross talk phenomenon between phases is appeared. It needs to a good spectrum to distinguish between these abnormal conditions and PD signal. It means that each PD signals that trapped by that phase PD sensor can be originated by other phases. Also, amplitudes of signal are very damped if discharges occur faraway of the measuring terminal. Then, discharges at near to neutral points cannot be accurately detected because of that lower voltage have inherently poor PD signals and their location is far from PD sensors. PD location has significant dependence to total measured charges at the terminals. PD distribution depends to generator construction and external connections as well.

By the mentioned characteristics of PD signals a narrow bandwidth detector with time domain measurement need to detect electrically PD signals. It means that a special width of signals can be detected.

Because of fault occurrence in capacitors and making of short circuit in generator terminals, continuous condition monitoring is not available. Tuning of bandwidth in narrow bandwidth detector need to complicated methods and in most cases is not accurate. Sometimes need to record signals in more than 30 different tuning frequencies to find and verify of the real PD signals. Experienced personnel have to study on the information.

B. PD's Acoustic Emission

With PD in insulation or core of generator, sound is emitted with frequency from 20 to 1000 KHz. Normally; this frequency is about 150-300 KHz.

When current streamer is formed, insulation is carbonized or vaporized and a sound is emitted, it can be audible or not. This vaporization can lead to a mechanical energy in form of pressure.

By installation of external or internal acoustic sensors, AE is detected. In power transformers, AE sensors mounted to the outside to detect the acoustic signal. In generator it has to be installed in to the slot and over the bars. It is complicated but has several advantages. Multiple location installation of AE sensors get position information and it help to find the severity and type of PD. Acoustic signals are immune for electromagnetic interference (EMI) and it leads to higher signal to noise ratio. Higher SNR means fewer false alarms.

Mechanical vibrations and environmental acoustic noises in power plant are the primary source of acoustic noises but those frequencies are sufficiently lower than PD signal frequency.

Acoustic emission signal normally have very low intensity and often attenuated when release and receive to sensors. Then sensors must be very sensitive to sound pressure amplitude. Resistive sensors are chipper and smaller but are more sensitive than capacitance sensors to temperature.

The measured rms sound pressure in transformer is about 0.2 Pa at 100 millimeter from a 1-pC discharge and a meter away it will be 0.02 Pa in oil. But in generator its value is unknown and has to test with real time measurement. If PD released pressure in the bar is about 1 Pa, transferred voltage will be few mV. Because of that oscilloscopes read in volt ranges, preamplifier amplification has to have a gain from 10^3 to 10^4 times.

Another problem is related to sensor complicated technology to build it. Installation of or based sensors around the final bar insulation is very important.

Acoustic emission sensors can be built in two ways: piezoelectric sensors or microphone based sensors. Microphone based sensors can transfer signals to a pair of headphones, transfer to electrical signal etc. piezoelectric sensors are thicker and cheaper. To better understanding for electric and acoustic signals, figures 4 are presented.

There are several kinds of acoustic sensors in markets to use in outer side of transformer tank; all of those can be used to PD test for single bar of generator in offline mode. Those are such as: R15 alpha (50-200 kHz, -63 dB in volt/ μbar = $10^{63/20}$ = 7.08×10^{-4} v/ μbar =7.08 mV/Pa, is made by Physical acoustics), WS alpha (100-1000 kHz, -62 dB in volt/ μbar =7.94 mV/Pa, is made by Physical acoustics), SPM0204UD5 (10-65 kHz, -47 db=4.47 mV/Pa, (15) is made by Knowles Acoustics:PM0204UD5, Surface Mount Wide-Band Ultrasonic Acoustic Sensor, Operating Temperature Range: -40°C to +100°C, 105°C for 1000 hours. Output

impedance: 300 Ω , current consumption 0.1-0.25 mA, SNR@1KHz:55-59, supply voltage:1.5-3.6 V, dimensions: 1.4x4.72x3.76mm, humidity:85%@85°C for 1000 hours.) and MA40MF14-0B (40 kHz, -87 dB sensitivity=0.44 mV/Pa, is made by Murata). Those are not made by micro electromechanical System and occupy a larger place.

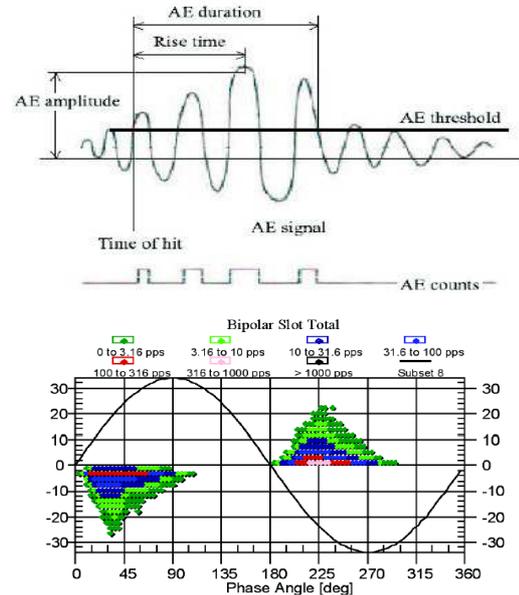


FIGURE 4. A: ACOUSTIC EMISSION WAVE FORM (12), B: SLOT DISCHARGE IN ELECTRICALLY CAPTURE OF PD (13)

Many companies, such as Knowles, Kulite, Akustica, Freescale Semiconductor, Murata (14) and Honeywell all offer MEMS microphones on the micro-scale.

All of mentioned sensors work in temperature less than 105°C but temperature in bar or slot of generator reach to 140°C in some times.

Author presented two papers in SIMS 2011 (16) and 2012 (17) around PD detection by FEM simulation. At now author presented an acoustic Stator Bar-Slot Couplers (ASBSCs) that uses the suggested MEMS sensors in large generators (larger than 100 MW) by taping around final insulation of each bar. It monitors partial discharge activity occurring within the stator bar-slot and at the end winding region of the generator windings while eliminating all types of acoustic or other noise. The proposed standard package includes 1 ASBSC's per each bar of generator, a termination box, coaxial cables, and a hydrogen penetration (if required).

II. A PROPOSED SETUP FOR ACOUSTIC EMISSION

Acoustic emission of partial discharge detector machine has to monitor PD. To achieve to automatically data acquisition need to an Acoustic Stator Bar-Slot Couplers (ASBSC). See Fig.5 and Diagram I. This device can be added to bars with wrapping around final insulation. A hundred or more semi conductive AE dies or sensors per each meter of bar is one choice.

The unit has to be equipped with: a programmable PD data collection based on a combination of user programmable triggers, such as load, temperature, operating voltage, NQN, etc.

This system need to a feature for saving and download the collected data, manage and analysis of trending and a possibility to make comparisons of trends. Test results have to be accessed remotely by a modem or a network connection. The comparison possibility solves the problem of performing PD analysis under changing load and temperature conditions.

The system consists of Data Acquisition Units (one DAU for each monitored machine) and a common system controller. The AETG (acoustic emission turbo guard) unit is equipped with semi conductor based acoustic sensors, amplifier system, noise canceller and its auxiliary software is programmed to collect PD data. Noise canceller with adaptive controller is used to tune of noise canceller for each kind of noise by adaption of canceller parameters.

An applied interface allows users to integrate partial discharge data from continuous on-line PD monitoring instruments, through a LAN (equipped with user selectable RS232, USB or LAN communication ports), into external applications for alarming, displaying and/or trending. The applied interface eliminates the need of a custom software application to display PD data.

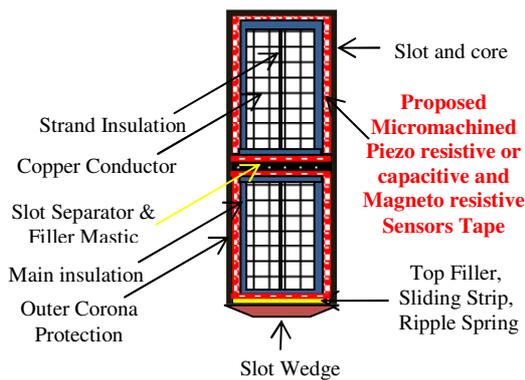


FIG. 5. AUTHOR'S PRESENTED NEW MODEL FOR ROEBEL BAR COVERS

III. DESIGN AN ACOUSTIC SOUND PRESSURE MICROMACHINED SENSOR

Semiconductor materials more developed in the recent century. Micromechanical devices use semiconductors to measure pressure, acceleration, and vibration. Silicon substrate is a good choice to design semiconductor based material. It have a good sensitivity and operating in range of -50° to $+200^{\circ}\text{C}$, have a low leakage, and can be mass-produced with tight tolerances.

piezoresistive effect can be used in silicon for making strain gauges. Also silicon based device have no fatigue, and has high strength and low density, making and it is suitable for micromechanical devices.

Micromechanical sensing devices are used for several sensors such as: Pressure; Force and strain; Acceleration;

Vibration; Flow; Angular rate sensing; Frequency filters. Those are very small, low mass and generally can be subjected to high overloads without damage. It is suitable to use around the bars those sense different tensions.

A. A brief introduction on the piezoelectric Devices

Interaction between electrical and mechanical properties makes piezoelectric effect. Voltage is produced when mechanical force applied on the piezoelectric material, and vice versa. It concerned to non regular charge distribution within the crystal structure of this micromechanical substrate. Electric field can change charges to align them with the electric field. This alignment change shape of the crystal. In response of mechanical force and according to the mentioned process in semiconductor material, voltage is produced across the crystal. Piezoresistive and piezocapacitance are two different common MEMS pressure sensors. Quartz, Rochelle salt, lithium sulphate and tourmaline exhibits the piezoelectric effect.

Another significant set of piezoelectric materials are the piezoelectric ceramics, such as lead-zirconate-titanate (PZT), lead-titanate, lead-zirconate, and barium-titanate.

Crystalline quartz and PZT have different characteristics. Quartz has good stability, minimal temperature effects, and high Q, is appreciate for timing device. Low Q, much higher dielectric constant, higher coupling factor, and piezoelectric charge constant, gives a much higher performance as a transducer to PZT. These features made PZT useful for micromotion actuators, or micropositioning devices.

PZT has a unique relation between pressure and voltage, a pressure is easily converted into a voltage. PZT devices are more sensitive than quartz, it is about 100 times. PZT is used for dynamic sensing from about 1 Hz to well over 10 kHz, but it is not good for static measurements. This is small and cost-effective, and can be used for vibration, shock, and acceleration measurement.

B. Bulk micromachining

Crystalline textile is etched in direction of the crystal planes. A wet anisotropic etch, such as potassium hydroxide is used in the etching of bulk silicon to build pressure sensors, accelerometers, micropumps, and something like this devices.

Pressure sensor is used to detect of partial discharge activities, in form of sound pressure sensing. Backside of the wafer is etched and over 100 dies or sensors are made. Sensor patterns are defined by photolithographic process. The backside of the wafer is covered with an oxide and a layer of light-sensitive resist. Predefined parts of the resist are masked and light is exposed to parts that not masked. With using of buffered hydrofluoric acid the oxide is wet etched. The resist will define the pattern in the oxide, which in turn is used as the masking layers for the silicon etch. To find the process see Fig. 6, 7 and 8. Now, the silicon is etched within an angle of 54.7° in direction of the crystal plane, and a thin layer of silicon is left during etching process, look at to the cross section to find its new configuration. After this process, top

side of the silicon wafer is masked and etched using a same mentioned process for the backside of the wafer.

A micro machined magnetic field sensor such as magnetoresistive element can be used to verify PD measured by piezoelectric. A magnetic field propagated by PD is sensed by Hall Effect magnetoresistive element.

Thickness for each side of sensors without integrated circuits to amplify of pressure is about $2 \mu\text{m}$. The length is chosen by designer and depends to generator slots length.

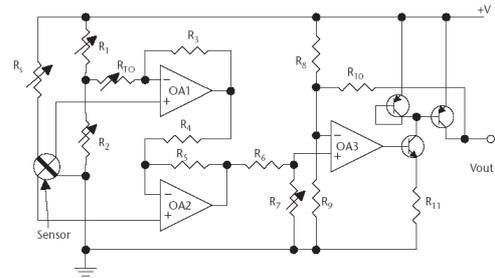


FIG. 9. INTEGRATED CIRCUIT

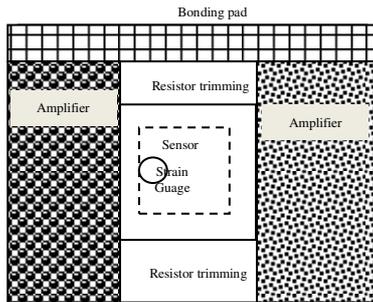


FIG. 6. POSITION OF CONDITIONING CIRCUITS ON THE TOP SIDE OF THE DIE

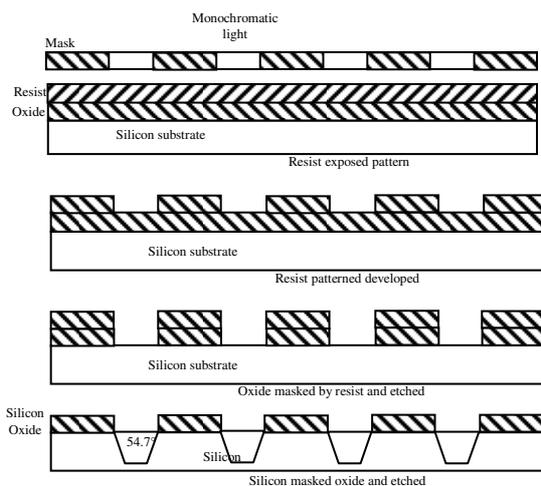


FIG. 7. WAFER ETCH PROCESS.

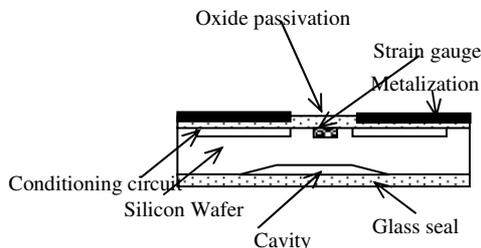


FIG. 8. CROSS SECTION OF MICROMACHINED

Gathering higher sound pressure levels requires a larger gap that produces higher voltage. The thickness and quality of the insulating oxide determines the maximum voltages the model (piezoelectric or capacitive) can withstand before breakdown. The theoretical breakdown voltage of SiO_2 is about $1000 \text{ V}/\mu\text{m}$; Wygant (18) observed breakdown voltages closer to $400 \text{ V}/\mu\text{m}$ for devices fabricated in their facilities. For the presented model it was suggested an oxide thickness of $40 \mu\text{m}$. Diagram 1 shows equipments sequences to arrange.

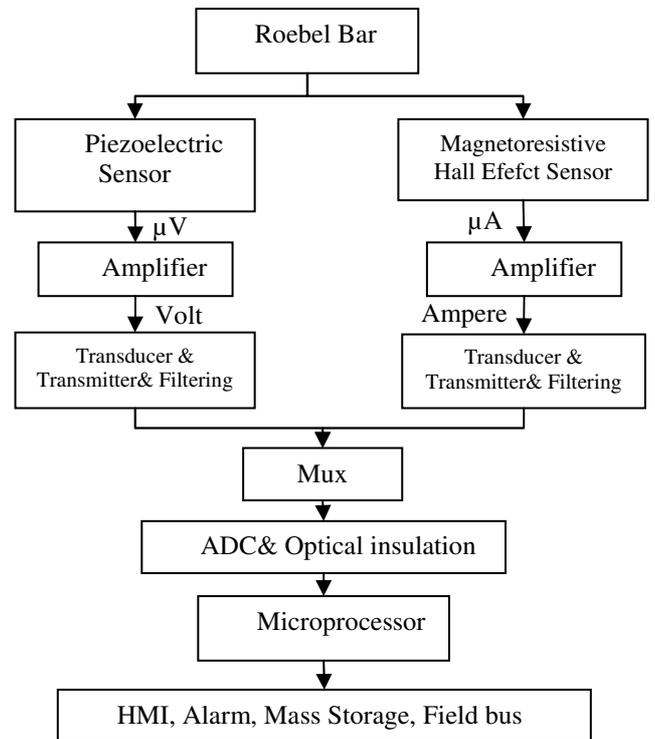


DIAGRAM1. DEVICES POSITION

C. Sampling

Sample rate have to be in degree that the PD signal is well characterized and the amount of information lost is minimized. Sampling rate must be two times greater than sampled signal frequency, to avoid of aliasing effect. The frequency components originated from noise is filtered. Nyquist theorem is used to prevent of aliasing. By applying the Nyquist model,

each analog signal is exclusively reconstructed, without error, from samples taken at equal time intervals. In corrupted signal by noise, aliasing can still occur. However, to get an accurate picture of the PD signal, need a sampling rate of roughly 10 to 20 times the highest frequency.

D. Noise

Noise (19) can be generated within the electrical components of the input amplifier (internal noise), or it can be added to the signal as it travels down the input wires to the amplifier (external noise). Techniques that you can use to reduce the effects of noise are described below. Internal noise arises from thermal effects in the amplifier. Amplifiers typically generate a few microvolts of internal noise, which limits the resolution of the signal to this level. The amount of noise added to the signal depends on the bandwidth of the input amplifier. To reduce internal noise, should select an amplifier with a bandwidth that closely matches the bandwidth of the input signal. External noise arises from many sources. Noise is added to the acquisition circuit from these external sources.

E. Filtering

Filtering also reduces signal noise. For many data acquisition applications, a low-pass filter is beneficial. The cut-off frequency of the filter must be compatible with the frequencies present in the signal of interest and the sampling rate used for the A/D conversion.

A low-pass filter that's used to prevent higher frequencies from introducing distortion into the digitized signal is known as an anti aliasing filter if the cut-off occurs at the Nyquist frequency. That is, the filter removes frequencies greater than one-half the sampling frequency. These filters generally have a sharper cut-off than the normal low-pass filter used to condition a signal.

IV. THEORETICAL SOLUTION FOR PD SOUND PROPAGATION

Wave's different amplitude due to different pressure has different frequency. Higher pressure in source has higher frequency than lower pressure PD waves. To model of PD wave in different media three different equations presented: The Fubini's solution, the Burgers equation and the KZK equation.

Fubini's solution expressed by an infinite sum of weighted Bessel functions: $P(\sigma, t') = P_0 \sum_{n=1}^{\infty} \frac{2}{n\sigma} J_n(n\sigma) \sin(n\omega(t - \frac{x}{c_0}))$, $t' = t - \frac{x}{c_0}$. Where t' is retarded time and P_0 is ambient pressure of media.

The Burgers equation describes the combined effects of nonlinearity and losses for plane progressive waves: $\frac{\partial p}{\partial x} - \frac{\delta}{2c_0^3} \frac{\partial^2 p}{\partial t'^2} = \frac{\beta}{2\rho_0 c_0^3} \frac{\partial^2 p^2}{\partial t'^2}$.

KZK (Khokhlov-Zabolotskaya-Kuznetsov) nonlinear parabolic wave equation account all spectrum of different frequencies obtained with basically the same type of

$$\text{experimental set-up: } \frac{\partial^2 p}{\partial z \partial t'} = \frac{c_0}{2} \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \left(\frac{\partial}{\partial r} \right) \right) p + \frac{\delta}{2c_0^3} \frac{\partial^3 p}{\partial t'^3} + \frac{\beta}{2\rho_0 c_0^3} \frac{\partial^2 p^2}{\partial t'^2}.$$

Where: p denotes the sound pressure, z is the coordinate along the axis of the beam, r is the transverse radial coordinate (the sound beam is assumed to be axisymmetric), c_0 is the sound speed, δ the diffusivity of sound, β is the coefficient of nonlinearity of the bar insulation and medium, J_n is the Bessel function of order n and $\sigma = x/\square$ where \square is the chock formation distance, meaning that the equation is valid as long as $\sigma < 1$.

V. AE DESCRIPTOR

A test has been done on a single bar. The method is same an online monitoring by ASBSC sensors.

To cognate of PD within the bars of generators need to AE sensors, voltage supply from the range of 0-2.5 U_N , a registration system with the range of 2.5 MHz, AE descriptor, Earth electrodes, differential preamplifier, amplifier with amplification of 20-42 db, (sampling frequency 1 MHz in a fragments of 2 s) and band filter.

The preamplifier is used to amplify a very small signal, within a few microvolts, with amplification of hundred times without adding interference. To prevent of risk of diminish the desired signal (is very weak) beyond measurability no filtering is done in this stage. Also, input impedance to amplifier is about 300 ohms and has to be 100 times or more.

To detect PD by microphone, a multiplier is used to make two frequencies with using a reference wave by known frequency and amplifier's signal, the result are frequency $f_2 + f_1$, and the second with the frequency $f_2 - f_1$. Later frequency is chosen a signal in the audible range². The amplifier can be modelled using the Orcad Pspice simulator. Gain and phase can be obtained by this simulator.

Advantech card was installed on PC to make interface for test set up and computer. A simple computer aided code provided to processing, handling and analyzing of the test results.

A coupling capacitor, narrow band filter and data acquisition system has been used to register of electrically PD activities. Apparent charges are derived by the results.

For each selected voltage that varies from 0 to 2.5 U_N , AE samples are registered for 2 seconds and apparent charges are measured.

It is predicted that duration and magnitude of AE signals decrease when source voltage is reduced. Also to ignite of PD

² For use as a multiplier in this application the AD633JNZ produced by Analog devices can be one option. The audio amplifier can be LM386.

need to a threshold voltage. In other hand it is obvious that local properties of registered AE depend to PD location and achieved impulse shall be conditioned. AE impulses vary in periodical clusters.

Some points within the insulation of bar subjected by PD activities generate an AE elastic waves which have a good correlation with released electrical apparent charges. With find a relation for AE propagation in insulation, a good relation can be found between Q and AE.

By registering of AE from PD signals in different voltages, for example 0-2.5 U_n , several issues of acoustic amplitude can be calculated, such as rate of acoustic emission counts by dN/dt , $d(dN/dt)/dU_t$ and derivative of amplitude distributions of AE power in function of discrimination threshold dP/dU_t .

Everywhere in bars when number of charges exceed from the predefined breakdown values, PD occur. Breakdown in void is occurred when the void voltage exceeds over the breakdown voltage, 2kV/mm. PD source generate acoustic signal and the AE elastic signals change during propagation and acquisition.

It sense that must find a good relation between charges and AE. Test sensors are located in a predefined location of bars and PD is measured in sensor location. There are 2 stages bars in most generators. If sensors locate between beneath bars and core it can't measure all sound. Author's suggestion is to wrap sensor strip around each bar.

Amplitude distribution of acoustic emission counts and power of AE signals are deeply investigated by test results. An AE count characterized by AEC and AE power is showed by AEP abbreviation. The correlation factor for AEP and AEC is calculated for several measured analyzed situations. To find the AEC and AEP, slope factor of approximated line use to define the AE descriptors. Lines slope are found by test result and are shown in next figures.

A damaged bar with PD activity has been studied. Seven sensor installed on the bar to detect PD and different voltages applied to the bar.

Figure 10 show a set of acoustic emission amplitude for 1.8-2.5 times of nominal voltage level.

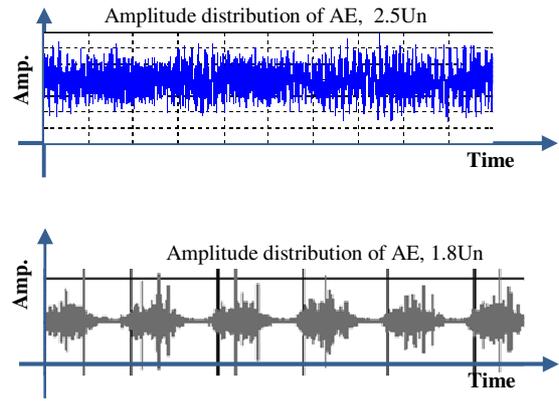


FIG. 10. PD ACTIVITIES TRACED BY AE GRAPHS.

Figure 10 as well shows that in 1.8 U_n , with increasing of voltage to PD initiation level, discharges are started and sounds are silenced when it reaches to extinguishing level. Figure 11 shows amplitude distribution for rate of counts. It shows higher voltages have higher rate of counts. Curves are coming down when discrimination threshold rises and are more flat for greater voltages; it means higher deformation is created.

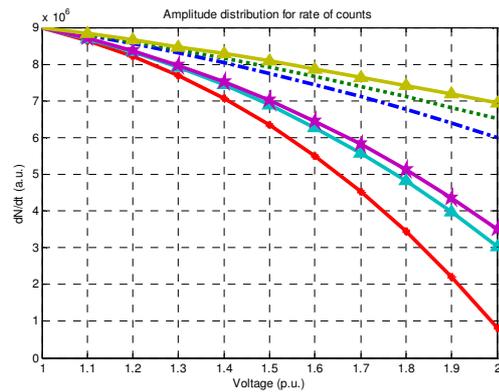


FIG. 11. AMPLITUDE DISTRIBUTION FOR RATE OF COUNTS.

Basic amplitude distribution for rate of counts and power and so derivative of basic amplitude of AE signals and counts in function of discrimination threshold can be obtained.

By calculating of apparent charges can make a comparison between charges and AE signals. Really there is correlation between those. Properties of sensors and insulations are known and results can be verified by theoretical calculations as well. This is better to evaluate shape of AE signal; because of that values have little deviation in each period. Neighboring AE sensors have similar shapes if AE amplitude distributions trace in a logarithmic scale. It helps to better understanding of curves.

Curves slope factor can describe AEC and AEP. Calculated AE descriptors, AEP and AEC, together with respective correlation factors and apparent charges for each measuring situation can be derived. Discrimination threshold is obtained by taking a differential from AEP or AEC against to voltage. It shows in which voltage PD is initiated. Deformation process is shown by AEP and AEC. AE method is able to find the location of PD activity more or less.

By a neural network process or other robust trainable methods can get a good response to find the accurate point for PD activities. Weights in NN are proportional to distance between source and AE sensor.

Lower slope for AEP and AEC show the PD activity. According to test registered results in table 1, so is interpreted that based on AE method locations 1 and 6 have PD activities but just location 1 is proved by apparent charges method. It is true and verify the AE method, because of that any PD in end part of bars where are closed to neutral point cannot sense by electrical methods. Its cause is obvious; voltage is less in neutral point and is close to zero.

VI. CONCLUSION

Acoustic emission method applied to a pre damaged 20 kV Roebel bar. In this method voltage changed in range of 0-2.5 Un. AEC, AEP and apparent charge are registered. Test results show that AE method could find the PD in both end part that closed to neutral point and terminal part of bar.

A block for micromachined piezoelectric sensor together with magnetoresistor which are wrapped around the final insulation and semiconductor of Roebel bar was presented.

Simulation was done by a robust FEM simulator and PD activity monitored by a simulated method. Simulation results can verify the measured data by micromachined piezoelectric sensors.

BIBLIOGRAPHY

- [1] Insulation Failure Mechanisms of Power Generators. Bruetsch, Rudolf. IEEE Electrical Insulation Magazine, pp. 1-7.
- [2] CIGRE. Hydrogenerator Failures – Results of the Survey. s.l.: CIGRE Study Committee SC11, EG11.02, 2003.
- [3] RoshanfekarFard, Poopak. ROEBEL WINDINGS FOR HYDRO GENERATORS. Göteborg, Sweden : Chalmers University, 2007. Thesis for the Master of Science (MSc) degree .
- [4] CONDUCTOR INSULATION FOR ROTATING MACHINES. Dominique Rey, Rudolf Brütisch. Birmingham, UK : s.n., May 24-26, 2006. 10th INSUCON International Electrical Insulation Conference. pp. 1-5.
- [5] J., Gibney Jams. GE generators- an overview. s.l.: GE industrial&power systmes, 1994. GER3688B.
- [6] VoithHydro. Mica elastic insulation for high voltage hydro generator. www.voithhydro.com. [Online] Voith. [Cited: July 15, 2013.]
- [7] ZnO on Si integrated acoustic sensor. Sensors and Actuators. M. Royer, J. O. Holmen, M. A. Wurm, O. S. Aadland, and M. Glenn. 1983. 4:357–362.
- [8] Development of a micromachined piezoelectric microphone for aeroacoustics applications. Stephen Horowitz, Toshikazu Nishida, Louis Cattafesta, and Mark Sheplak. 2007, Journal of the Acoustical Society of America. 122(6):3428–3436.
- [9] Krause, Joshua Steven. MICROMACHINED MICROPHONE ARRAY ON A CHIP FOR TURBULENT BOUNDARY LAYER MEASUREMENTS. DOCTOR OF PHILOSOPHY IN MECHANICAL ENGINEERING. MEDFORD: TUFTS UNIVERSITY, AUGUST 2011.
- [10] A History of Engineering and Science in the Bell System. Fagen., M.D. s.l.: Bell Telephone Laboratories,, 1975.
- [11] Wideband micromachined acoustic sensors with radio frequency detection. Sean T. Hansen, A. Sanlı Ergun, and Butrus T. Khuri-Yakub. [ed.] Edward M. Carapezza V. Stanford CA, USA : Unattended Ground Sensor Technologies and Applications ,Proceedings of SPIE , 2003. Vol. 5090, pp. 42-50. SPIE : 0277-786X/03.
- [12] JANUS, PATRICK. Acoustic Emission Properties of Partial Discharges in the time-domain and their applications. Stockholm, Sweden : KTH University, 2012. Masters' Degree Project. XR-EE-ETK 2012:004.
- [13] TGA-STM, Iris Power. Periodic On-Line Partial Discharge Monitoring Using a Portable Instrument for Large High Speed Turbine Generators. Canada : s.n., 2010.
- [14] Ultrasonic Sensor. www.murata.com. [Online] 2008. Cat.No.S15E-5.
- [15] Acoustics, Knowles. ULTRASONIC ACOUSTIC SENSOR.

TABLE I. AEP, AEC AND Q MEASURED VALUES.

Sensor location No.	U (kV)	AEP	AEC	Q	σ_{AEP}
1	20	-2.8	-2.7	9.7	0.99
2	15	-1.65	-3.9	7	0.99
3	10	-7.4	-3.5	0.5	0.98
4	5	-5.8	-1.9	4.7	0.99
5	3	-1.3	-4.9	2	0.99
6	1	-2.7	-2.6	0.2	0.997

In other hand frequency of waves are different when PD occurs in insulation tapes. Its frequency in the closed part to surface of insulation is higher than when PD created in insulation closed to bar. Because of that waves travelling in a nonlinear medium, it will shift its frequency spectrum to lower frequencies and its magnitude is reduced as it propagates away from the source. This phenomenon is observed by simulation, when is run in AVI mode. A snap of simulation movie has been shown in Fig. 12. Author's results confirmed the achievements of Patrick (20).

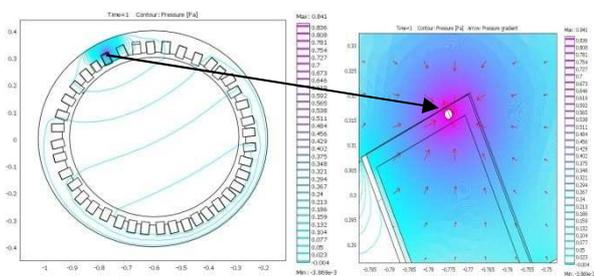


FIG. 12. CONTOUR AND GRADIENT OF PRESSURE BY FEM.

- [16] Etiology of Rey Generator stator core failure and study of its rehabilitation integrity. Takami, Kourosch. Vasteras : Malardalen University, SIMS 2011, 30th September 2011.
- [17] Partial discharge simulation and modeling in large scale generators using FEM method. Takami, Kourosch. Reykjavik, Iceland : SIMS 2012, 3-5 October 2012.
- [18] 50 kHz Capacitive Micromachined Ultrasonic Transducers for Generation of Highly Directional Sound with Parametric Arrays. Wygant, Ira O. 1, s.l.: IEEE, JANUARY 2009, IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL, Vol. 56, pp. 193-204. 10.1109/TUFFC.2009.1019.
- [19] Mechanical thermal noise in micromachined acoustic and vibration sensors. Gabrielson, Thomas. 5, s.l. : IEEE, May 1993, IEEE transaction on electron devices, Vol. 40, pp. 903-909. 9208061.
- [20] JANUS, PATRICK. Acoustic Emission Properties of Partial Discharges in the time-domain and their applications. Stockholm : KTH University, 2012. Masters' Degree Project. XR-EE-ETK 2012:004.

PREDICTION OF DAILY RUNOFF FROM HYDROLOGICAL CATCHMENT AREA

Sobhan Shafiee
Faculty of Technology,
Telemark University
College, Porsgrunn, Norway
sobhan.shafiee@gmail.com

Beathe Furenes
Skagerak Kraft, AS
Porsgrunn, Norway
Beathe.Furenes@skagerakenergi.no

Bernt Lie
Faculty of Technology,
Telemark University College,
Porsgrunn, Norway,
Bernt.Lie@hit.no

Abstract

Prediction of runoff from a catchment to a water reservoir is an important factor when maximizing the use of water in hydro power production. In this paper, the HBV hydrology model formulation which was introduced by Fjeld and Aam [1] has been slightly re-formulated and detailed. Although the vector field has elements which are piecewise-defined functions and an ODE solver with event detection should be considered to find an accurate solution, a standard ODE solver in MATLAB seems to work well. The model was validated using information from the Eggedal catchment in Norway, provided by Skagerak Kraft and the Norwegian Water Resources and Energy Directorate (NVE). A basic parameter fitting had been performed, and sensitivity analysis has been conducted to find which model parameters should be tuned to further improve the model fit.

Key words: hydrology, hydropower, prediction model, HBV model.

1. Introduction

In order to improve the production planning in a hydropower system, forecast/prediction of runoff/inflow of water to a reservoir in a catchment is important. This is in particular important during transitions to and from flooding, where good planning may minimize the amount of water bypassing the turbines.

The HBV model (HBV = Hydrologiska Byråns Vattenbalansavdelning, developed in Sweden in the 1970s), is a lumped precipitation-runoff model, and is extensively used for making runoff/inflow forecasts to hydropower systems where precipitation is stored in the form of snow. In Fjeld and Aam (1980), a mechanistic description was given of the HBV model based on mass balances.

The HBV model has many tuning parameters, and the catchment system is characterized by few measurements. To get realistic forecasts, it is important to update the model on-line. Some model implementations require a manual updating of the model, while others support more automatic model updating. In Fjeld and Aam, it was proposed to use the Kalman filter for state (and possibly parameter) estimation; they also provided an analysis of observability and pointed at difficulties due to few measurements. Modern implementations of the HBV model often involve input-output formulations to get around the observability problem.

It is of interest to reconsider a mechanistic understanding of the HBV model with a new look at the observability problem, and how more modern measurement technology can make the system observable. It is also of interest to use some new ideas wtr. parameter identifiability and how to automatically update model parameters and states.

The paper is organized as follows. In Section 2, a simple process description is given and the dynamic mechanistic model is given. In Section 3, a validation of the model is given. In Section 4, identifiability of model parameters is carried out. In Section 5, some conclusions are drawn.

2. Process description

2.1. Functional description

The system is divided into 4 zones: snow zone, ground zone, soil zone, and basement zone. For the snow zone, a further detailing is given wrt. altitude, see below. The catchment system topology with inputs and outputs is as indicated in Figure 1.

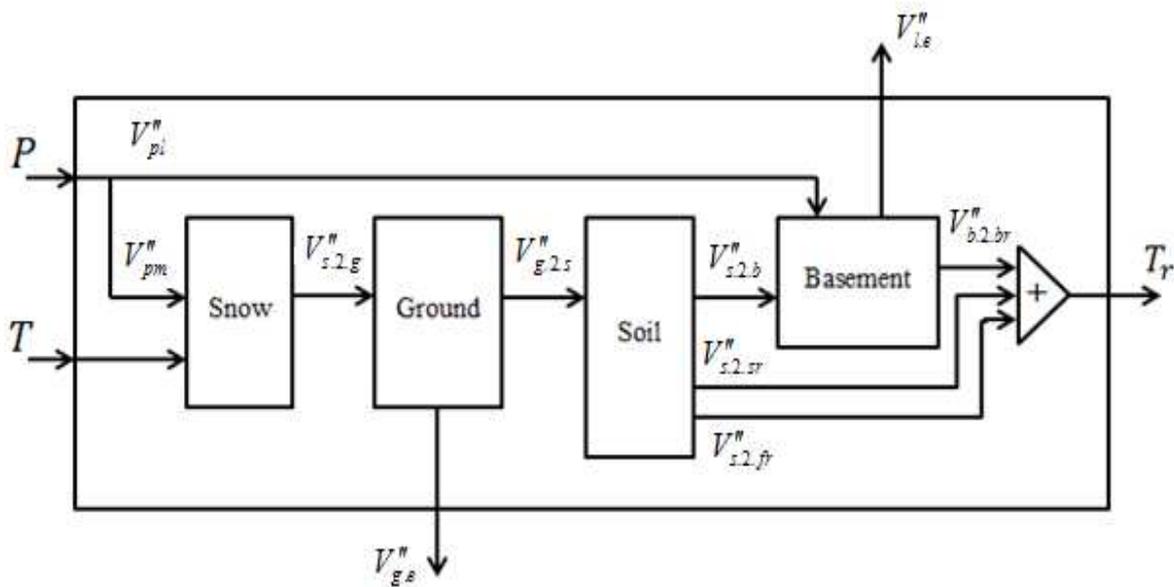


Figure 1: Functional description with topology of catchment.

Condensed water vapor precipitates in the form of snow or rain (P) depending on atmospheric temperature (T) and could be stored as snow/ice or water in the snow zone, the accumulated water in the snow zone can increase to the saturation level by precipitation as rain or by melting snow, then water can runoff to the ground zone ($\dot{V}''_{s,2,g}$). When the ground moisture level reaches the ground saturation level, it is removed by evapotranspiration ($\dot{V}''_{g,e}$) or runoff into the soil layer ($\dot{V}''_{g,2,s}$). If the soil moisture level reaches the soil saturation level, the infiltrated water could percolate to the basement zone ($\dot{V}''_{s,2,b}$), flow out as surface flow ($\dot{V}''_{s,2,fr}$), or leave as Fast runoff ($\dot{V}''_{s,2,fr}$). Percolated water flows into the basement zone, which constitutes the last compartment ($\dot{V}''_{b,2,br}$); water is either stored in the basement zone or joins into the total runoff (T_r). Due to water transformation between lakes and the basement zone, maximum evapotranspiration ($\dot{V}''_{l,e}$) should be considered for the lake surface if there is any lake in catchment area.

2.2. Model zones

The whole model consist of snow zone, ground zone, soil zone and basement zone that are connected to each other in a cascade form. Each zone has its own model which connects inputs to outputs by a specific sub-model which is based on mass balance.

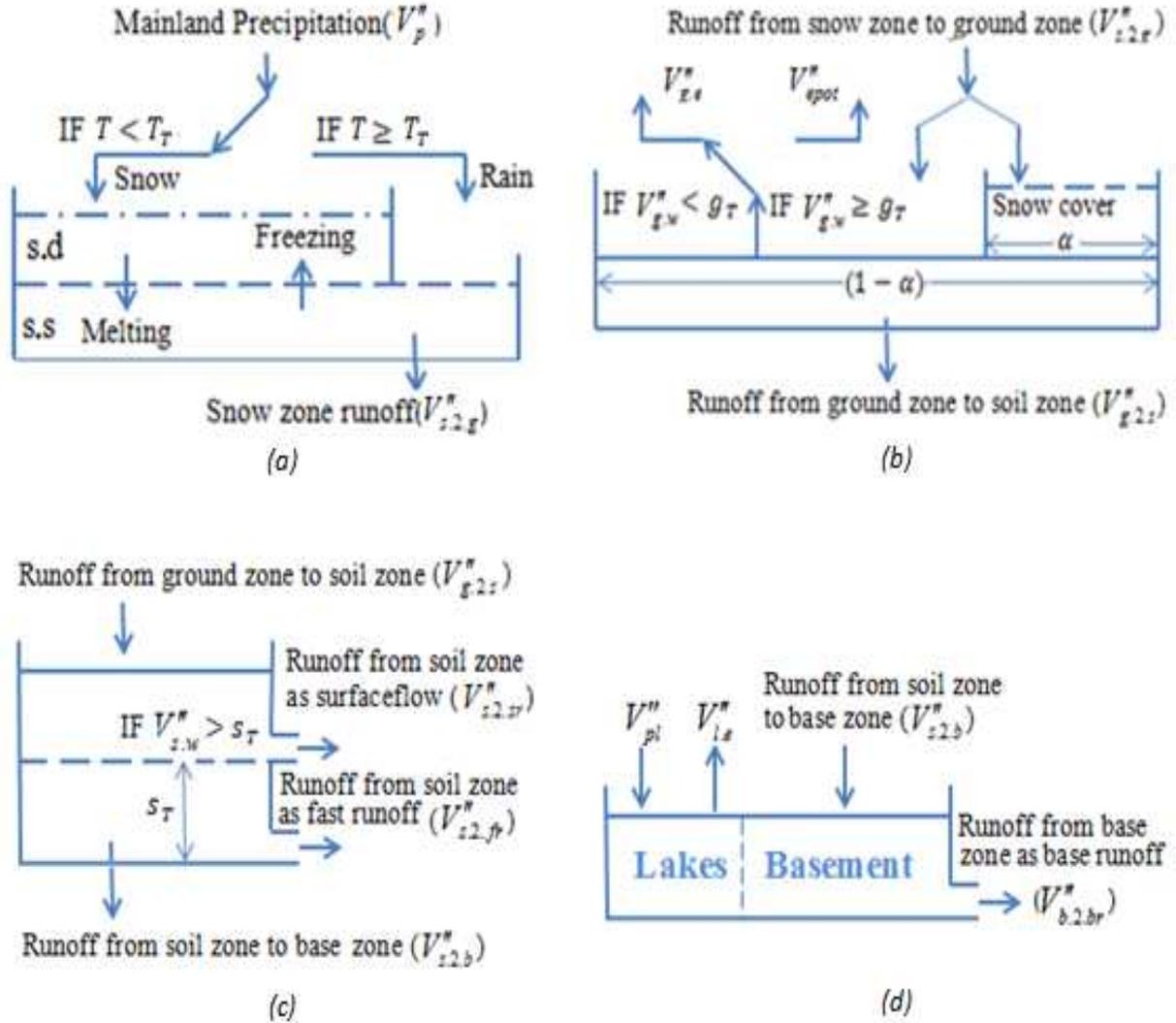


Figure 2: Snow zone model (a), Ground zone model (b), Soil zone model (c), Basement zone model (d).

The result is a state space model with a moderate number of ODEs as follows:

$$\frac{dV_{s,d}''}{dt} = \dot{V}_{p,s}'' + \dot{V}_{w2,d}'' - \dot{V}_{d2,w}'' \quad (1)$$

$$\frac{dV_{s,s}''}{dt} = \dot{V}_{p,r}'' - \dot{V}_{w2,d}'' + \dot{V}_{d2,w}'' - \dot{V}_{s2,g}'' \quad (2)$$

$$\frac{dV_{g,w}''}{dt} = \dot{V}_{s2,g}'' - \dot{V}_{g2,s}'' - (1-\alpha)\dot{V}_{g,e}'' \quad (3)$$

$$\frac{dV_{s,w}''}{dt} = \dot{V}_{g2,s}'' - \dot{V}_{s2,b}'' - \dot{V}_{s2,sr}'' - \dot{V}_{s2,fr}'' \quad (4)$$

$$\frac{dV_{b,w}''}{dt} = \dot{V}_{s2,b}'' + \dot{V}_{pl}'' - \dot{V}_{b2,br}'' - \dot{V}'' \quad (5)$$

Here, the transportation between zones is given by piecewise-defined expressions:

$$\dot{V}_{p,s}'' = \begin{cases} (1-\alpha_L)\dot{V}_p'' & \text{if } T \leq T_T \\ 0 & \text{if } T > T_T \end{cases} \quad (6)$$

$$\dot{V}_{d.2.w}'' = \begin{cases} k_m(T - T_T) & \text{if } T > T_T \text{ and } V_{s.d}'' \geq 0 \\ 0 & \text{other wise} \end{cases} \quad (7)$$

$$\dot{V}_{p,r}'' = (1 - \alpha_L)\dot{V}_p'' - \dot{V}_{p,s}'' \quad (8)$$

$$\dot{V}_{w.2.d}'' = \begin{cases} k_m(T_T - T) & \text{if } T \leq T_T \text{ and } V_{s.s}'' \geq 0 \\ 0 & \text{other wise} \end{cases} \quad (9)$$

$$\dot{V}_{s.2.g}'' = \begin{cases} \dot{V}_{p,r}'' & \text{if } T > T_T \text{ and } V_{s.s}'' = V_{s.d}'' = 0 \\ (1 + \alpha_w)\dot{V}_{d.2.w}'' + \dot{V}_{p,r}'' & \text{if } V_{s.s}'' > 0 \text{ and } V_{s.s}'' = (\alpha_w \cdot V_{s.d}'') > 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$\dot{V}_{g.2.s}'' = \begin{cases} \left(\frac{V_{g.w}''}{g_T}\right)^\beta \times \dot{V}_{s.2.g}'' & \text{if } 0 < V_{g.w}'' < g_T \\ \dot{V}_{s.2.g}'' & \text{if } V_{g.w}'' \geq g_T \end{cases} \quad (11)$$

$$\dot{V}_{g.e}'' = \begin{cases} \left(\frac{\dot{V}_{g.w}''}{g_T}\right) \times \dot{V}_{epot}'' & \text{if } V_{g.w}'' < g_T \\ \dot{V}_{gpot}'' & \text{if } V_{g.w}'' \geq g_T \end{cases} \quad (12)$$

$$\dot{V}_{s.2.b}'' = (1 - a_L) \cdot \text{PERC} \quad (13)$$

$$\dot{V}_{s.2.sr}'' = \begin{cases} a_1(V_{s.w}'' - s_T) & \text{if } V_{s.w}'' > s_T \\ 0 & \text{if } V_{s.w}'' \leq s_T \end{cases} \quad (14)$$

$$\dot{V}_{s.2.fr}'' = a_2(V_{s.w}'') \quad (15)$$

$$\dot{V}_{pl}'' = a_L \cdot \dot{V}_{pm}'' \quad (16)$$

$$\dot{V}_{b.2.br}'' = a_3(V_{b.w}'') \quad (17)$$

$$\dot{V}_{l.e}'' = a_L \cdot \dot{V}_{epot}'' \quad (18)$$

The total catchment runoff rate is obtained by summing $\dot{V}_{b.2.br}''$, $\dot{V}_{s.2.fr}''$ and $\dot{V}_{s.2.sr}''$ as follows:

$$\frac{d\dot{V}_{tot}}{dt} = \begin{cases} [a_1(V_{s.w}'' - s_T) + a_2(V_{s.w}'')] \times (1 - \alpha_L)A + Aa_3(V_{b.w}'') & \text{if } V_{s.w}'' > s_T \\ a_2V_{s.w}''(1 - \alpha_L)A + Aa_3(V_{b.w}'') & \text{if } V_{s.w}'' \leq s_T \end{cases} \quad (19)$$

Table 1: Nominal model parameters and states.

State/ Parameter	Value	Unit	Description
$V''_{s,d}$	zero (on 1st of August)	mm	Dry snow
$V''_{s,s}$	zero (on 1st of August)	mm	Soggy snow
$V''_{g,w}$	zero (on 1st of August)	mm	Ground zone water
$V''_{s,w}$	zero (on 1st of August)	mm	Soil zone water
$V''_{b,w}$	zero (on 1st of August)	mm	Base zone water
a_1	0.547	day ⁻¹	discharge frequency for surface runoff
a_2	0.489	day ⁻¹	discharge frequency for fast runoff
a_3	0.0462	day ⁻¹	discharge frequency for base runoff
A	309.42	km ²	catchment area
PERC	0.25	$\frac{\text{mm}}{\text{day}}$	percolation from soil zone to base zone
k_m	5.2	$\frac{\text{m}^3}{\text{°C}\cdot\text{day}}$	melting factor
T_r	0.8	°C	threshold temperature
α_w	0.08	-	saturation coefficient
α		-	snow surface fraction
g_T	50	mm	ground saturation threshold
β	2	-	ground zone shape coefficient
s_T	20	mm	soil saturation threshold
α_L	0.032	-	lake surface fraction

2.3. Model development

According to hypsographic data and a topographic map, the surface of the catchment is divided into several elevations sub models; the reason for this is that temperature will change with the elevation, leading to precipitation in the form of snow or rain depending on the elevation. Thus, each altitude range has its own dry snow and wet snow values which are known as snow zone states ($V''_{s,d}$, $V''_{s,s}$). The other states ($V''_{g,w}$, $V''_{s,w}$, $V''_{b,w}$) are similar for all altitudes. Consequently, total runoff from the snow zone to the ground zone is obtained by summation of the runoff for all altitudes.

3. Model validation

The accuracy of the model as a representation of the real system was checked to find out how much the simulated model is imitating the real system. These parameters and initial states were chosen by adjusting the values in Fjeld and Aam's paper to the case of the Eggedal catchment, with information provided by Skagerak Kraft and NVE. The adjusted values are as listed in Table 1.

Figure 3 illustrates the simulated runoff and the real runoff from the Eggedal catchment. As can be seen, the main trends in the real data are predicted by the simulation.

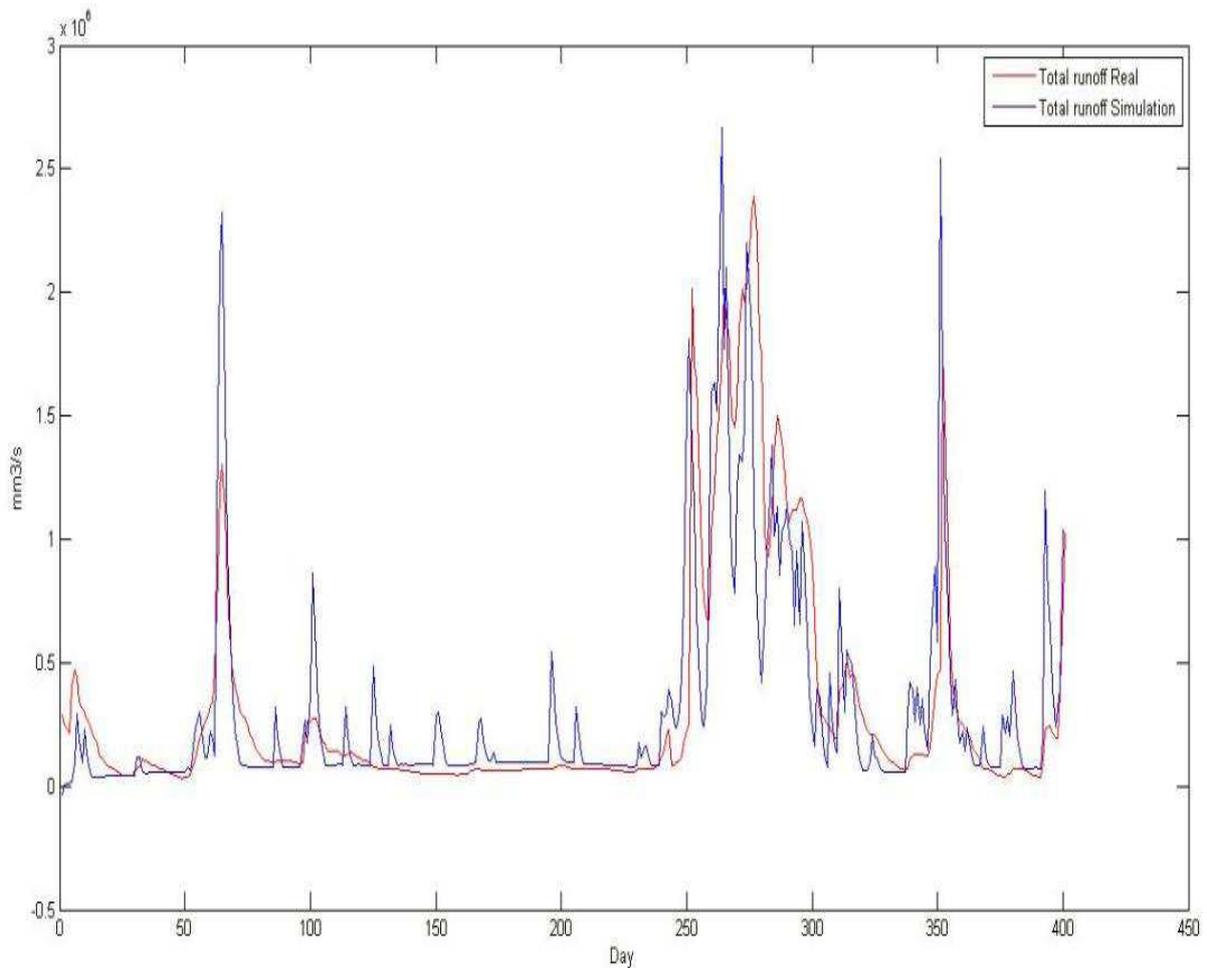


Figure 3: Comparing the real runoff with simulated runoff in Eggedal catchment.

In part of the runoff simulation (day 100 to day 240), some undesirable fluctuations are observed. Further analysis indicates that a coarse recording of temperature and precipitation inputs to the system may explain this: changes in the sign of the temperature seems to cause the observed fluctuations in the model. A possible solution to this problem is to measure temperature and precipitation more often, leading to smoother simulations.

Another difference between the simulated runoff and real runoff is due to poor values of some parameters which have been used in the model. In order to achieve a more accurate result, the model parameters need to be calibrated: it is necessary to adjust parameters specifically for each catchment, and parameter tuning is necessary for the Eggedal catchment model. For nonlinear systems, parameters could be estimated using different calibration methods such as the least squares method, the maximum likelihood method, and the prediction error method (PEM, which includes a state estimator).

The accuracy of the runoff prediction also depends on the accuracy of data obtained by measurement devices. Currently Skagerak gathers measurements of precipitation (based on accumulated rain or snow over 24 hours), temperature (given as average daily value), and runoff density (field variable measured in $\frac{\text{m}^3}{\text{sec}}$). [2]

4. Identifiability analysis

In hydrological systems, setting up the experiments to measure all state variables might not be possible. It is therefore possible that certain parameters do not affect the available measurements, or only affect the measurements weakly. Such parameters that have no/little influence on the available measurements are denoted as unidentifiable.

To be identifiable, the parameter subset has to meet two conditions. First, the model output has to be sufficiently sensitive to the parameter change (sensitivity). Second, changes in the model output that are caused by a single parameter change can be distinguished from changes by other parameter changes (collinearity). [3]

The sensitivity function for each parameter could be obtained by calculating the model output twice: with the nominal parameter set, and with a parameter set perturbed from the nominal set.

4.1. Sensitivity analysis by Principle Component Analysis

This method has been developed to discard insignificant parameters by firstly ranking all of them, and secondly by determining which one is identifiable or unidentifiable according to their ranks. The method was carried out in the following manner:

1. Construct the normalized sensitivity matrix:

$$S_{k,j}^i = \frac{p_j}{Tr_i} \cdot \frac{\partial Tr_i(t_k, p)}{\partial p_j} \quad (20)$$

where

i - denotes the index of system outputs

j - denotes the index of parameters

k - denotes the index of measurement time points

2. Run the PCA for each month to find out insignificant parameters by ranking all parameters according to their component coefficients.

As listed in Table 1, there are 13 parameters in our model. For conducting the sensitivity analysis, only 9 of them were considered – in particular those parameters which values are poorly known, but where estimated ranges exist from studies of other catchments.

As illustrated for specific time ranges: the model output is sensitive to certain parameters, which means that those parameters play a crucial role in comparison to the other parameters. The reason for this is obvious: during a snow-less summer, it will be impossible to estimate parameters related to snow melting, etc. Consequently, it is reasonable to assume that both identifiability and sensitivity are time-varying characteristics. In general, a dynamic nature of sensitivity should be accounted for to find out the specific periods where parameters have higher identifiability and crucial role in representing model runoff. [4]

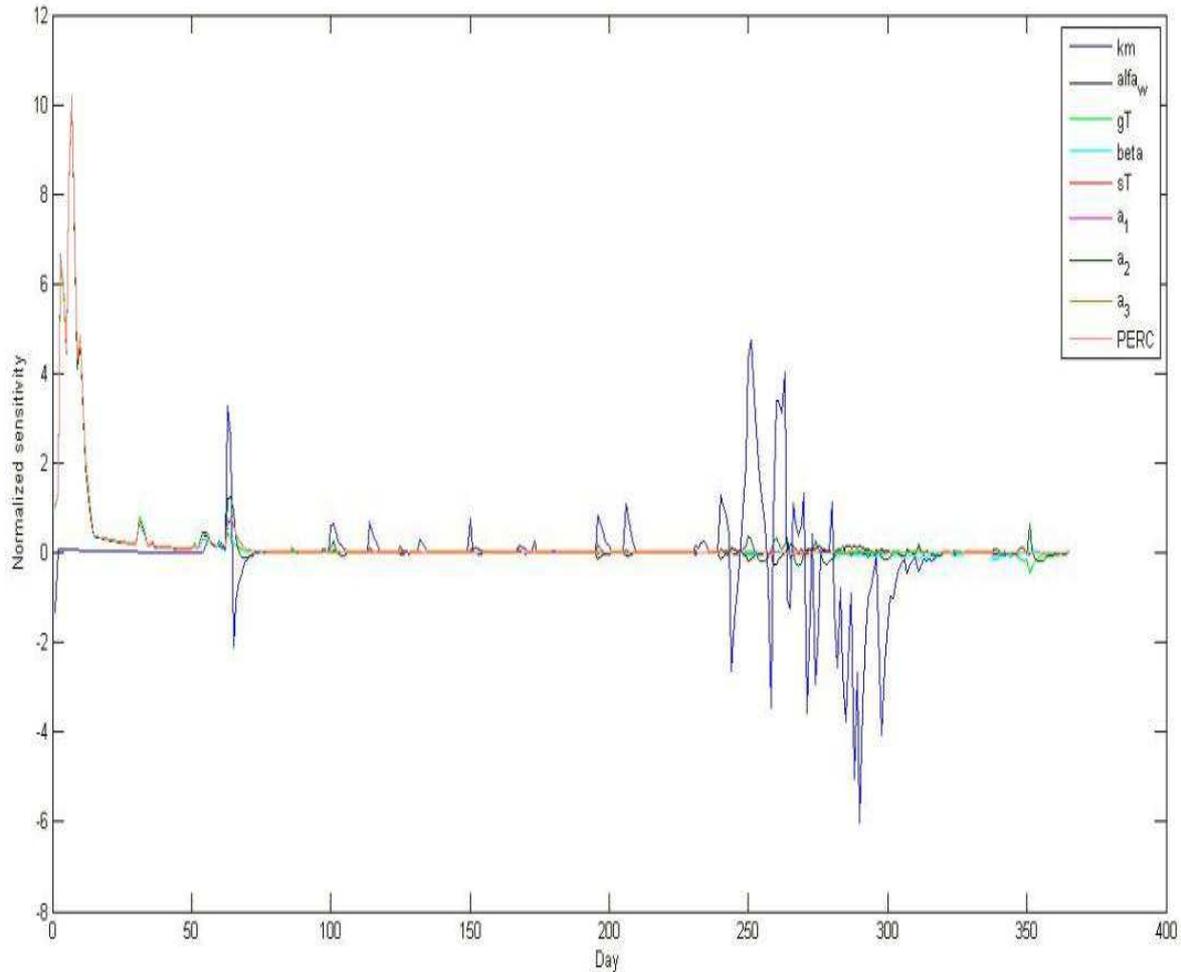


Figure 4: Normalized sensitivity index plot for 9 identifiable parameters of August in a year (01/08/1973 - 31/07/1974).

Table 2 shows parameters importance ranking with respect to time-varying characteristic (monthly analysis). Second and third columns indicate the most important parameters (from left to right) for each principal component.

Table 2: Parameter importance rankings.

Month	Parameters importance ranking during 01/08/1973 - 31/07/1974 (1st year)	Parameters importance ranking during 01/08/1973 - 31/07/1974 (2nd year)
August	PC1 (89%): $\alpha_\omega, g_T, \beta, s_T, \alpha_1, \alpha_2, \alpha_3, \text{PERC}, k_m$	PC1 (89%): $\alpha_\omega, g_T, s_T, \alpha_1, \alpha_3, \text{PERC}, \beta, \alpha_2, k_m$
September	PC1 (89%): $s_T, \alpha_1, \alpha_3, \alpha_\omega, \text{PERC}, \alpha_2, g_T, \beta, k_m$	PC1 (82%): $\text{PERC}, \alpha_\omega, \alpha_3, \beta, \alpha_1, g_T, \alpha_2, s_T, k_m$
October	PC1 (89%): $\text{PERC}, s_T, \alpha_1, \alpha_3, \beta, \alpha_\omega, g_T, \alpha_2, k_m$	PC1 (66%): $s_T, \alpha_3, \alpha_1, \beta, g_T, \alpha_\omega, \text{PERC}, \alpha_2, k_m$ PC2 (14%): $k_m, \alpha_2, g_T, \text{PERC}, \beta, \alpha_1, \alpha_\omega, s_T, \alpha_3$
November	PC1 (78%): $s_T, \alpha_1, \alpha_3, \alpha_\omega, \beta, g_T, \alpha_2, k_m, \text{PERC}$	PC1 (50%): $\alpha_\omega, k_m, \beta, g_T, \text{PERC}, \alpha_2, s_T, \alpha_1, \alpha_3$ PC2 (32%): $\alpha_1, s_T, \alpha_3, g_T, \text{PERC}, \beta, \alpha_2, \alpha_\omega, k_m$
December	PC1 (71%): $\beta, \alpha_1, s_T, \alpha_3, g_T, \alpha_\omega, \alpha_2, k_m, \text{PERC}$	PC1 (48%): $g_T, \beta, \alpha_\omega, k_m, \text{PERC}, \alpha_2, \alpha_3, \alpha_1, s_T$ PC2 (33%): $s_T, \alpha_1, \alpha_3, \alpha_2, \text{PERC}, \beta, \alpha_\omega, k_m, g_T$
January	PC1 (70%): $g_T, \beta, s_T, \alpha_1, \alpha_3, k_m, \text{PERC}, \alpha_2, \alpha_\omega$	PC1 (59%): $g_T, s_T, \alpha_1, \alpha_3, \beta, k_m, \alpha_2, \alpha_\omega, \text{PERC}$ PC2 (28%): $\alpha_\omega, \text{PERC}, k_m, \beta, \alpha_2, \alpha_3, \alpha_1, s_T, g_T$

February	PC1 (75%): $g_T, \alpha_\omega, s_T, \alpha_1, \alpha_2, \alpha_3, k_m, \beta, PERC$	PC1 (99%): $k_m, \alpha_\omega, g_T, \beta, s_T, \alpha_1, \alpha_2, PERC, \alpha_3$
March	PC1 (85%): $\alpha_1, s_T, \beta, g_T, k_m, \alpha_2, \alpha_3, PERC, \alpha_\omega$	PC1 (89%): $k_m, \alpha_\omega, g_T, \beta, s_T, \alpha_1, \alpha_2, PERC, \alpha_3, \alpha_\omega$
April	PC1 (55%): $\beta, s_T, \alpha_3, \alpha_\omega, \alpha_1, \alpha_2, k_m, PERC, g_T$	PC1 (56%): $g_T, \beta, s_T, PERC, \alpha_1, \alpha_\omega, \alpha_2, k_m, \alpha_3$
	PC2 (28%): $PERC, g_T, k_m, \alpha_3, s_T, \alpha_1, \alpha_\omega, \beta, \alpha_2$	PC2 (19%): $\alpha_3, \alpha_1, \alpha_2, \alpha_\omega, s_T, \beta, k_m, PERC, g_T$
May	PC1 (31%): $PERC, \beta, k_m, \alpha_\omega, \alpha_3, \alpha_2, g_T, \alpha_1, s_T$	PC1 (42%): $\alpha_3, \alpha_1, k_m, \alpha_2, \beta, \alpha_\omega, s_T, PERC, g_T$
	PC2 (28%): $g_T, s_T, \alpha_1, PERC, \alpha_2, \beta, \alpha_\omega, k_m, \alpha_3$	PC2 (24%): $PERC, s_T, \beta, g_T, \alpha_2, \alpha_1, \alpha_3, \alpha_\omega, k_m$
	PC3 (16%): $\alpha_3, \alpha_1, \beta, \alpha_2, \alpha_\omega, s_T, k_m, g_T, PERC$	PC3 (18%): $\alpha_\omega, g_T, \beta, s_T, \alpha_2, PERC, k_m, \alpha_3, \alpha_1$
June	PC1 (45%): $\beta, k_m, g_T, PERC, \alpha_2, \alpha_3, s_T, \alpha_\omega, \alpha_1$	PC1 (68%): $s_T, \alpha_1, \alpha_3, g_T, \beta, \alpha_\omega, PERC, k_m, \alpha_2$
	PC2 (32%): $\alpha_1, \alpha_\omega, s_T, \alpha_3, \alpha_2, PERC, g_T, \beta, k_m$	PC2 (23%): $\alpha_2, k_m, PERC, \alpha_\omega, \beta, g_T, \alpha_3, \alpha_1, s_T$
July	PC1 (55%): $s_T, k_m, \alpha_\omega, \alpha_1, \alpha_2, g_T, \alpha_3, PERC, \beta$	PC1 (79%): $k_m, \alpha_\omega, s_T, \alpha_1, \beta, g_T, PERC, \alpha_3, \alpha_2$
	PC2 (27%): $\beta, \alpha_3, g_T, PERC, \alpha_1, k_m, \alpha_\omega, s_T, \alpha_2$	

Note: The percentage value in front of each principal component explains the corresponding variances of the original variance along each principle component.

4.2. Collinearity analysis based on Principle Component Analysis

Interpretation of the loading plot indicates the importance and also the correlation between model variables (parameters). Those variables that lie far away from the origin yet close to the principle component axis, are the significant variables for that PC; those variables which lie near the origin are the less important ones. If the variables lie on the same side of the origin, they have positive correlation, but if they lie on opposite sides of the origin (more or less along a straight line through the origin), they are negatively correlated. It can be inferred that those variables which are located at 90 degrees to each other through the origin of the loading plot are independent. [5]

According to the explained method, those parameters which satisfy the following three conditions should be selected as desirable identifiable parameters:

1. Not negatively correlated with other parameters
2. Lie far from the origin
3. Located close to the Principle Component axis

4.3. Identifiability analysis based on Principle Component Analysis

By considering both the parameter importance ranking in Table 2 and parameter collinearity, the identifiable parameters are found as listed in Table 3.

Table 3: Identifiable parameters.

Month	Identifiable parameters (1st year)	Identifiable parameters (2nd year)
August	$\alpha_\omega, g_T, \beta, s_T, \alpha_1, \alpha_2, \alpha_3, PERC$	$\alpha_\omega, g_T, s_T, \alpha_1, \alpha_2, PERC, \beta, \alpha_2$
September	$s_T, \alpha_1, \alpha_3, \alpha_\omega, PERC, \alpha_2, g_T, \beta$	$PERC, \alpha_\omega, \alpha_3, \beta, \alpha_1, g_T, \alpha_2, s_T$
October	$PERC, s_T, \alpha_1, \alpha_3, \beta, \alpha_\omega, g_T, \alpha_2$	$s_T, \alpha_3, \alpha_1, \beta, \alpha_\omega, g_T, k_m$
November	$s_T, \alpha_1, \alpha_3, \alpha_\omega, \beta$	$\beta, PERC, s_T, \alpha_3$
December	α_1, s_T, α_3	g_T, k_m, α_1, s_T
January	$\beta, s_T, \alpha_1, k_m, PERC$	$g_T, s_T, \alpha_1, \alpha_3, PERC$
February	$\alpha_\omega, s_T, \alpha_3$	$k_m, \alpha_\omega, g_T, \beta, s_T, \alpha_1, \alpha_2, PERC$
March	α_1, s_T, α_2	$k_m, \alpha_\omega, g_T, \beta, s_T, \alpha_1, \alpha_2, PERC$

April	$\beta, \alpha_2, g_T, PERC$	$g_T, S_T, \alpha_1, \alpha_3$
May	$k_m, g_T, \alpha_3, S_T, \alpha_1$	$\alpha_3, \alpha_1, k_m, \alpha_\omega, PERC, S_T$
June	$k_m, \alpha_1, \alpha_\omega$	$S_T, \alpha_1, \alpha_3, \alpha_2$
July	$S_T, k_m, \alpha_\omega, \alpha_1, \beta$	$k_m, \alpha_\omega, S_T, \alpha_1, \beta, g_T, \alpha_3$

It is evident that identifiable parameters for the same month in two consecutive years are different. This result is in line with Abebe's (2010) study which has confirmed that optimum parameters are not constant in time, and the range of optimum parameter values varies with time. Consequently, a dynamic identifiability analysis should be conducted based on forecasted temperature and precipitation to find better identifiable parameters for future days/forecasts.

5. Concluding remarks

In order to improve the production planning in a hydropower system based on forecast of runoff/inflow of water to a reservoir, an assessment of a mechanistic hydrological model has been introduced. The model was obtained by reformulating the HBV-3 model based on mass balance and physical laws for single altitude. The model was then extended by considering the hypsography concept and thus including all relevant altitudes. More accurate results would have been reached by increasing the number of altitudes to get a smoother function between the snow surface fraction and the altitude; the drawback of this would be that more available data lead to a heavier computational load.

The MATLAB programming language was used to solve the model equations, and model verification was based on information from the Eggedal catchment in Norway, as provided by Skagerak Kraft AS and NVE. The model displayed some sensitivity to long sample times in temperature and precipitation measurements.

To find the most important parameters for model calibration, sensitivity analysis and identifiability analysis were done based on principal component analysis. Identifiability analysis indicated that identifiable parameters were not constant in time and which parameters are identifiable also varies with time. For simplicity and for getting reasonable results, parameters were assumed to be constant in a month.

References

- [1] Magne Fjeld, Sverre Aam, "An Implementation of Estimation Techniques to a Hydrological Model For Prediction of Runoff to a Hydroelectric Power Station," IEEE Transactions on Automatic Control, vol. AC.-25, pp. 151-163, 1980.
- [2] Skagerak Kraft AS. (2013). Fakta om Skagerak Energi. Available: http://www.skagerakerenergi.no/eway/default.aspx?pid=300&trg=MainRight_9194&MainArea_8872=9194:0:&MainRight_9194=9216:0:10,3419
- [3] Roland Brun, Martin Kühni, Hansruedi Siegrist, Willi Gujer, P. Reichert "Practical identifiability of ASM2d parameters—Systematic selection and tuning of parameter subsets," Water Research, vol. 36, pp. 4113-4127, 2002.
- [4] Nibret A. Abebe, Fred L. Ogden, Nawa R. Pradhan, "Sensitivity and uncertainty analysis of the conceptual HBV rainfall–runoff model: Implications for parameter estimation," Journal of Hydrology, vol. 389, pp. 301-310, 2010.
- [5] Kim. H. Esbensen, Multivariate data analysis - in practice: Camo, 2001.
- [6] Sobhan Shafiee, "Automatic updating of hydrological models for runoff/inflow forecasting to hydropower system," MSc thesis, Telemark University College, Porsgrunn, Norway, 2013.

Fuel Efficient Speed Profiles for Finite Time Gear Shift with Multi-Phase Optimization

Xavier Llamas*, Lars Eriksson* and Christofer Sundström*

* *Vehicular Systems, Dept. of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, {xavier.llamas.comellas, lars.eriksson, christofer.sundstrom}@liu.se*

Abstract: A method that finds fuel optimal speed profiles for traveling a predefined distance is presented. The vehicle is modeled using a quasistatic formulation and an optimal control problem is defined. In addition, the solving method is based on a multi-phase optimization algorithm based on dynamic programming. This approach results in lower computational time than solving the problem directly with dynamic programming, it also makes the computational time independent of the travel distance. In addition, the simulation generated data can be used to get the solution to several optimal control problems in parallel that have additional constraints. Further a finite time gear shift model is presented to include the gear selection in the optimization problem. The problem also considers speed losses and fuel consumption during the maneuver. The results presented show the optimal speed and gear profiles to cover a distance, making special emphasis at the acceleration phase, where it is optimal to perform a fast acceleration to engage the highest gear as soon as possible. Finally a proposed application is to use the simulation data to provide eco-driving tips to the driver.

Keywords: Dynamic Programming, Simulation, Eco-driving.

1. INTRODUCTION

Driving more efficiently has become an important issue since the fuel cost has increased significantly during the last decades. One way to reduce the fuel consumption, and thereby reduce the CO_2 emissions, is to improve the efficiency of the powertrain by means of technical advances, e.g. hybridization. Another way to reduce the fuel consumption is optimizing the way how the driver operates the vehicle. Giving advice to the driver, or directly controlling the vehicle speed and gear shifting, can be useful to use the powertrain at the most efficient operating points, and thus reduce the fuel consumption.

The problem studied is to find vehicle optimal speed profiles for covering a predefined distance. Previous research has been done in this area. The first analytic approach to solve the problem was done by Schwarzkopf and Leipnik (1977), which was revised later by Chang and Morlok (2005). These two papers solve the problem using the Pontryagin maximum principle.

Optimal speed profiles have also been computed using numerical methods, mostly using Dynamic Programming (DP). Starting with the paper by Hooker (1988), or more recent papers like the ones by Saerens et al. (2009), Gausemeier et al. (2010), Luu et al. (2010), Jorge et al. (2011) and Mensing et al. (2011), where optimal speed profiles for several situations are presented.

The optimal control problem is solved by means of DP, see Bellman and Dreyfus (1962) and Bertsekas (1995). One difference with previous research is that the problem is solved as three separate phases that are joined optimally

together. This formulation decreases computational time and is able to compute several speed profiles with the same output data from the optimization. Moreover it also adds the possibility to apply additional constraints once the simulation is computed.

Another contribution consists in presenting a model that optimizes the engaged gear together with the vehicle speed. It also takes into account the gearshift speed losses, due to no traction torque during the maneuver, and fuel consumption in the optimization.

Previous authors that studied the influence of gear shift as an instantaneous maneuver are Hooker (1988) and Mensing et al. (2011). Other studies used an automatic gearbox with a predefined gear shift strategy, see Gausemeier et al. (2010) and Luu et al. (2010). In Saerens et al. (2010), a speed penalty factor is taken into account during gear shift, however no fuel consumption is considered. On the other hand, in Hellström et al. (2010) both speed loss and fuel consumption are considered, however it requires to interpolate between state values to compute the cost.

A criterion for choosing a right discretization to ensure accurate results is also described in the study. Finally several optimal speed profiles are presented and analyzed, making specially emphasis at the acceleration phase.

2. VEHICLE MODEL

A quasistatic approach is used for modeling the vehicle. This approach allows to compute the vehicle fuel consumption as function of vehicle speed and gear. Thus, it is a well suited modeling approach for the DP algorithm, because the gear engaged and the speed at the beginning

Table 1. Vehicle, driveline and engine parameters used in the study.

Parameter	Symbol	Value	Unit
Vehicle mass	m_v	1500	kg
Frontal area	A_f	2	m^2
Air drag coeff.	c_d	0.3	-
Rolling res. coeff.	c_r	0.01	-
Wheel radius	r_w	0.3	m
Wheel inertia	J_w	0.6	$kg \cdot m^2$
Gearbox efficiency	η_{gb}	0.98	-
Gear ratio 1	i_1	13.0529	-
Gear ratio 2	i_2	8.1595	-
Gear ratio 3	i_3	5.6651	-
Gear ratio 4	i_4	4.2555	-
Gear ratio 5	i_5	3.2623	-
Displacement	V_d	$2.3 \cdot 10^{-3}$	m^3
Indicated efficiency	e	0.35	-
Idling speed	$\omega_{e,idle}$	95	rad/s
Stroke	S	$79.5 \cdot 10^{-3}$	m
Bore	B	$96 \cdot 10^{-3}$	m
Cylinders	-	4	-
Engine inertia	J_e	0.2	$kg \cdot m^2$

and at the end of the arc are given by the optimization algorithm and the fuel arc cost can be obtained easily. Further information about the quasistatic approach can be found in Guzzella and Sciarretta (2013).

In the quasistatic formulation some assumptions are considered in order to compute the vehicle speed and acceleration for a given arc. The first assumption is that the acceleration remains constant over a step in distance. The second assumption is that the mean speed, computed between the initial and the final speed values of the current arc, is used in order to calculate the arc cost. These are reasonably good approximations for a sufficiently dense speed and distance grids.

2.1 Driveline Model

Following the previous mentioned quasistatic approach, the torque required at the wheels is computed using a longitudinal vehicle model (1), from Guzzella and Sciarretta (2013)

$$T_w = r_w \left(\underbrace{\frac{1}{2} A_f \rho_a c_d v^2}_{\text{air drag}} + \underbrace{c_r m_v g}_{\text{roll resist.}} + \underbrace{\dot{v} \left(m_v + \frac{J_w}{r_w^2} \right)}_{\text{inertia}} \right) \quad (1)$$

and a manual gearbox with a constant efficiency is used. Table 1 contains the main parameters used in this study. More details regarding the model are given in Llamas (2012).

2.2 Engine model

The Willans line approximation is used to compute the fuel consumption. Despite being a simple model, it gives reasonably good fuel consumption values at a low computational cost. The approximation is based on normalized engine variables, that do not depend on the engine size, like the mean effective pressure, p_{me} , and the fuel mean pressure, p_{mf} . The output power of the engine, represented by the mean effective pressure, is computed by an affine function (Guzzella and Sciarretta, 2013)

$$p_{me} = e \cdot p_{mf} - p_{me0} \quad (2)$$

where e is the indicated engine efficiency and p_{me0} represents the friction and pumping losses in the engine. The term p_{me0} is modeled using the ETH friction model from Guzzella and Onder (2004). Table 1 contains the engine parameters used.

The maximum torque available, $T_{e,max}(\omega_e)$, is determined by interpolating the maximum torque data from the QSS toolbox from Guzzella and Amstutz (1999). This data is scaled in order to match the engine dimensions.

3. PROBLEM FORMULATION

The problem goal is to minimize fuel consumption, thus the optimal control problem is formulated as follows:

$$\min \int_0^T \dot{m}_f dt \quad (3)$$

$$\dot{X} = f(X, U)$$

where $X = (x, v, g)^T$ is the state vector and consists of position, speed and engaged gear and $U = (T_e, u_g)^T$ is the control vector that consists of engine torque and gear shift command. This minimization is restricted to the following constraints:

$$\begin{aligned} T_e &\leq T_{e,max} & a &\leq a_{max} \\ \omega_e &\geq \omega_{e,idle} \end{aligned} \quad (4)$$

$$X(0) = (0, 0, 0)^T \quad X(T) = (x_{final}, 0, 0)^T$$

that are interpreted as functional limits in engine torque and engine speed as well as a limit in vehicle acceleration. Initial and final conditions of the optimal control problem are also defined, as: drive a certain distance, x_{final} and starting and finishing at stand still with neutral gear.

4. MULTI-PHASE DYNAMIC PROGRAMMING ALGORITHM

As DP is used to find the optimal speed profile, a discrete state-space model is required. The variables to be optimized through a certain distance are speed and engaged gear. The gear engaged is already a discrete variable, thus speed has to be discretized as well as distance, which is used as a tracking variable.

A three dimensional DP algorithm is required to compute the optimal speed profile. More dimensions in the DP algorithm means that the computational time becomes much more influenced by the discretization size, due to the "curse of dimensionality" (Bellman and Dreyfus, 1962). Hence, it is interesting to adapt the DP algorithm to the particularities of this specific problem to save computational time. It is known from previous research that keeping constant speed is a fuel optimal policy (Chang and Morlok, 2005; Schwarzkopf and Leipnik, 1977). It is also pointed out in Llamas (2012) that the optimal policy is to accelerate fast to reach a constant cruising speed and then do coasting until stand still.

Thus, there is a long phase of the optimization that consists of keeping constant speed and therefore having a fine grid in this phase is a waste of resources. The key idea is to split the optimization algorithm into three phases that later on are optimally joined together. These three phases are: acceleration, constant cruising and deceleration, see Figure 1.

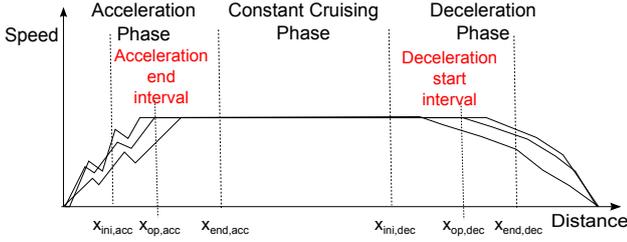


Fig. 1. Multi-Phase DP algorithm diagram.

With this approach the acceleration and deceleration phases are still optimized using DP with a dense grid to obtain accurate results. Note that the deceleration phase consists in doing coasting and thereby the speed profile could be computed analytically. However, this is not possible because the gearshift profile needs to be optimized. The two phases do not require from data of each other, and they can thus be solved in parallel to save computational time. Using the results from both DP phases, optimal profiles for different constant cruising speeds can be computed.

The algorithm that optimally merges the three phases requires a predefined constant cruising speed value and gear. Then, as illustrated in Figure 1, it generates the cost to go matrices over the distance intervals at the acceleration ($x_{ini,acc}, x_{end,acc}$) and deceleration phases ($x_{ini,dec}, x_{end,dec}$) to find the joining points ($x_{op,acc}, x_{op,dec}$) that result in the lowest fuel consumption. A reference value for setting suitable intervals can be found by looking at the optimal profiles in Llamas (2012). Note that the maximum length of these intervals is given by the distance grid used in the DP simulations and also that the required distance grid is proportional to the desired cruising speed.

In addition, obtaining longer distance optimal profiles do not affect the computational time because it only requires a longer constant cruising speed phase. It has also the possibility to add other constraints to the problem, e.g. maximum distance to accelerate or decelerate the vehicle.

4.1 Acceleration Phase

The acceleration profile is optimized using Forward DP. The difference with the well known Backward DP is that the arc costs are computed forwards, from the initial state to the end state, while Backward DP computes them from the final state to the initial state. Once the arc costs are known, the Forward algorithm computes the optimal profile from a given final speed, gear and distance to the initial condition point $X(0)$. Thus with one simulation all optimal acceleration profiles from the initial condition point $X(0)$, to every speed, gear and distance grid points are known. In this phase the vehicle is only allowed to accelerate if no gear shift is performed.

4.2 Deceleration Phase

The deceleration profile is computed using two strategies that are later compared in Section 7. The first one is to compute the optimal profiles using a Backward DP algorithm. The Backward formulation is used here because it gives all optimal profiles from a certain grid point to the final condition point $X(T)$. For this strategy the engine is either running or in fuel cut.

The second strategy is coasting with the engine shut down. That means that the engine is disengaged from the powertrain and thus the vehicle speed loss is only due to air drag and rolling resistance. The speed profiles are computed analytically from an initial distance and speed state.

4.3 Constant Cruising Phase

This phase joins the acceleration and deceleration phases with a constant speed and gear profile. Given the cruising speed and gear, the algorithm finds the optimal joining points, $x_{op,acc}$ and $x_{op,dec}$, that result in the global lowest fuel consumption, see Figure 1.

5. FINITE TIME GEAR SHIFT MODEL

A new way to approach a finite time gear shift maneuver is developed. The model computes the fuel consumption and the speed loss during the maneuver time and then these values are taken into account by the optimization algorithm. This approach has been inspired by the gear shift model from Hellström et al. (2010), however when it is implemented for the DP algorithm it does not require to interpolate between state values if there is a gearshift.

The maneuver time is defined as the time elapsed since the clutch is pressed until the engine is engaged again. In this study it is set to be 1 s. The assumption made to compute the fuel consumption and the speed loss is that during the maneuver time the engine cannot provide traction torque. Thus the vehicle speed is computed by

$$\frac{dv}{dt} = \frac{-1}{m_v + \frac{J_w}{r_w^2}} \left(\frac{1}{2} \rho_a A f c_d v^2 + m_v g c_r \right) \quad (5)$$

since T_ω in (1) is set to zero. The cost of the maneuver is assumed to be the fuel consumed during the maneuver and that the engine is idling.

The distance traveled during the gear shift maneuver can be computed using the maneuver time and the final speed computed with (5). Hence, with that information the point in the state-space grid can be set (x_l, v_l) . If $x_l \ll x_{k+1}$ the cost to go from that point to every next grid point, in x_{k+1} , can be computed. A gear shift computation is illustrated in Figure 2.

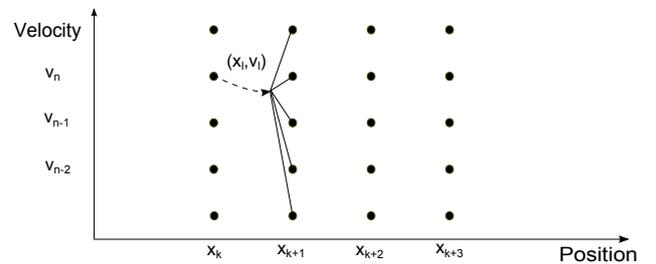


Fig. 2. Computation of the cost to go from (x_k, v_n) to each state of x_{k+1} with a gear shift maneuver.

However, some problems arise when the distance traveled during the gear shift is bigger, or nearly equal, to the step length used in the algorithm. These two situations have two related problems. The first one is that if the maneuver

finishes close to the next distance step, the next speed points might not be reachable due to a high acceleration or braking torque is requested in the small distance that exists from the maneuver end point to the next grid step. The second issue is that if the gear shift maneuver ends after the next distance grid step, the cost to go cannot be computed following the standard DP algorithm.

In order to solve these problems, the standard DP algorithm is modified. If the gear shift maneuver ends after or too close to the next distance grid step, the algorithm computes the costs to the next available grid point and saves to which distance step the costs are computed. The model makes a difference if the next step has to be reached by accelerating or braking because if braking is needed, the model allows to do a gradual braking through the distance step even while performing the gearshift maneuver. The strategy is illustrated in Figure 3.

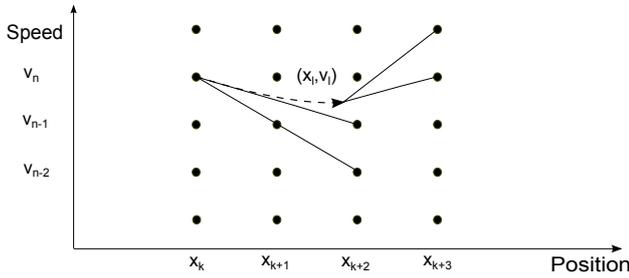


Fig. 3. Skipping grid points strategy from (x_k, v_n) during a gear shift maneuver.

6. DISCRETIZATION

The discretization of the continuous problem plays an important role in the accuracy of the results. Hence, how the discretization is defined on the acceleration phase is investigated because it is the phase with highest fuel consumption per distance. Then the discretization used in the deceleration phase is the same that is considered sufficiently accurate for the acceleration phase.

The acceleration profile to reach 8.75 m/s with fifth gear in 50 m is optimized with several grid choices and the results are presented in Figure 4. It can be observed that speed and gear profiles change especially at the beginning if a smaller distance step is selected. First gear is engaged instead of second for a certain step distance smaller than 5 m , in Figure 4 for values of 1 m and 0.5 m .

In addition, one also must ensure that the assumptions made to compute the arc costs are reasonably fulfilled with the grid choice. If the grid is too sparse, there will be less distance points where the profile is computed, making the computations less accurate. E.g. for $\Delta x = 5 \text{ m}$, with only one distance step and thus one computation, the speed profile reaches 6 m/s which is more than the half of the final speed value.

A grid is sufficiently accurate if further decrease of the grid size do not change the optimal profile, thus the optimal profile converges. Looking again at Figure 4, once first gear is selected (for $\Delta x \leq 1 \text{ m}$), decreasing the distance step more does not change the optimal profile and thereby the distance step is sufficiently small.

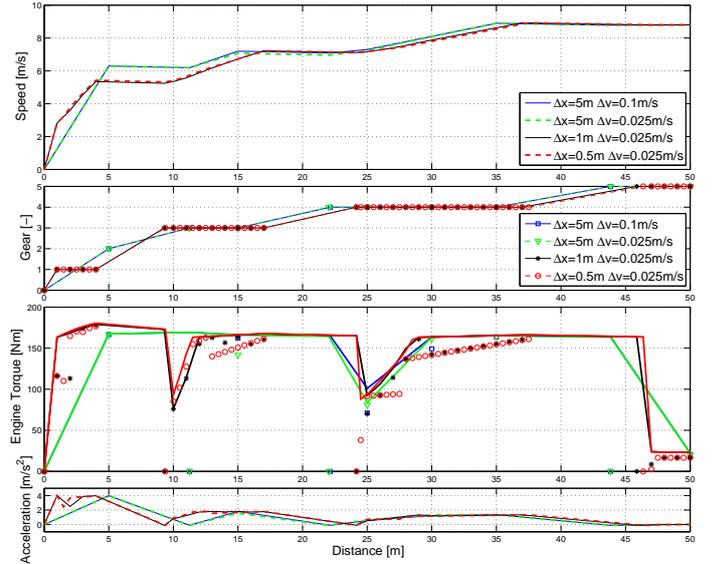


Fig. 4. Acceleration Phase with several grid step sizes. The thick lines in the Engine Torque plot represent the maximum engine torque.

Moreover, looking at the acceleration values for the different grid choices, it can be seen that for the first step the algorithm choice is always at the maximum acceleration boundary (which for this test is set to $a_{max} = 4 \text{ m/s}^2$). Hence the maximum acceleration value defines the profile shape at the very beginning. In this case, in order to ensure convergence, one must select a distance step that engages first gear at the beginning. This means that the optimal speed value at the first distance step must be lower than the minimum speed needed to engage second gear.

The speed at one step is computed as the mean value between the start and end points

$$v_m = \frac{v_{k+1} + v_k}{2} \quad (6)$$

thereby for the first step, using (6) with $v_1 = 0$ and kinetic energy, the speed after the first distance step is

$$v_2 = \sqrt{2a_{max}\Delta x} \quad (7)$$

thus the relation that must hold to ensure that the first gear can be selected at the first distance step is

$$v_2 < \frac{\omega_{idle} r_w}{i_2} \quad (8)$$

that can be satisfied with a small enough value of Δx . With the parameters used, the relation is fulfilled with $\Delta x < 1.5 \text{ m}$, and the selected value is $\Delta x = 1 \text{ m}$. In addition, the selected speed step value is $\Delta v = 0.025 \text{ m/s}$.

7. RESULTS

First of all, an optimization of a speed profile simulation is carried out without additional constraints. Figure 5 presents the optimal speed profiles for several cruising speed values. It can be seen that the acceleration and deceleration phases are really similar for the each cruising speed.

The cruising speed of 8.75 m/s is selected because this is the one that leads to the lowest fuel consumption, see

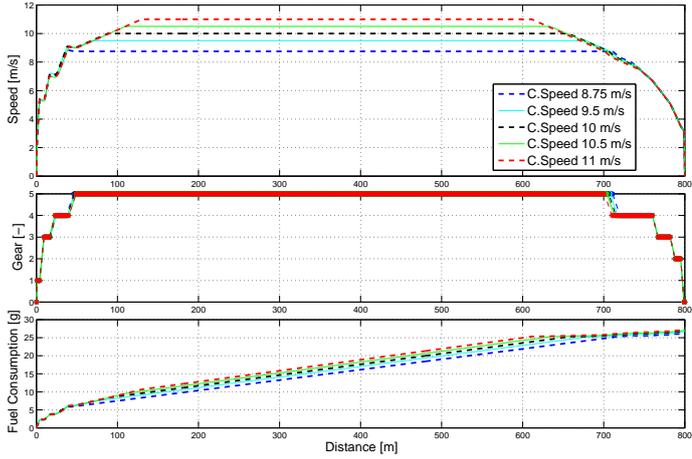


Fig. 5. Optimal speed profiles for several cruising speeds without constraints.

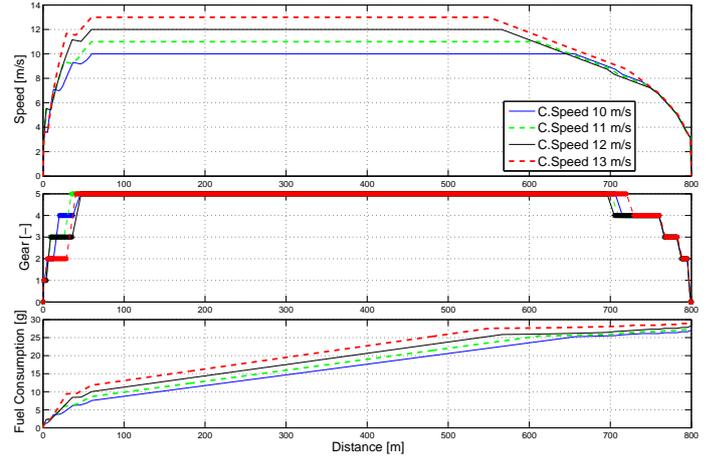


Fig. 7. Optimal speed profiles for several cruising speeds a maximum acceleration distance of 60 m

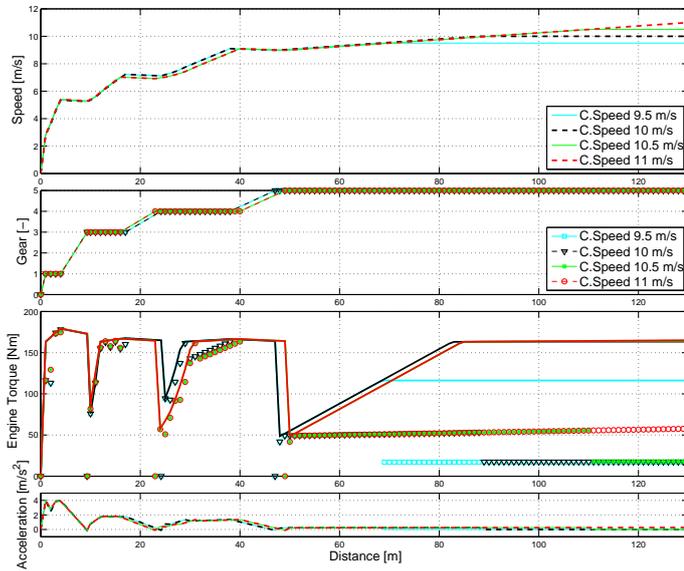


Fig. 6. Optimal acceleration profiles extracted from Figure 5. The thick lines in the Engine Torque plot represent the maximum torque available.

Table 2. In Llamas (2012) the whole drive mission is optimized by a one phase DP algorithm, and the results show that the optimal policy is to cruise at the lowest speed possible with the highest gear engaged. This optimal cruising speed is defined by the engine idling speed, the wheel radius and the gear ratio at the highest gear. The other cruising speeds are selected near that value to provide a comparison.

Figure 6 presents a closer look at the acceleration phase. It can be seen that the torque values are always close to the maximum, thus the optimal policy is to have high acceleration and engage higher gears as soon as they are available in order to reach fifth gear as soon as possible (see also Figure 4). As there is no time constraint, once the acceleration profile reaches the optimal cruising speed, 8.75 m/s, it continues accelerating as slow as possible until the fixed cruising speed. This slow acceleration depends on the grid choice (for each Δx , the speed increases Δv), due to that the model is not allowed to keep constant speed.

Table 2. Fuel consumption and trip time for both strategies and several cruising speeds.

C. Speed	Fuel DP	Fuel E. off	Time DP	Time E. off
8.75m/s	4.48l/100km	2.95l/100km	104.65s	145.68s
9.5m/s	4.54l/100km	2.82l/100km	98.96s	140.83s
10m/s	4.58l/100km	2.72l/100km	95.88s	141.18s
10.5m/s	4.61l/100km	2.61l/100km	93.43s	140.93s
11m/s	4.65l/100km	2.50l/100km	91.37s	140.03s

If constant speed is allowed, then the acceleration profile does cruising at 8.75 m/s during a certain distance before accelerating to reach the imposed cruising speed.

In order to see the effects of applying additional constraints into the original problem, a limit is set into the acceleration distance, e.g. the cruising speed must be reached within 60 m. Figure 7 displays the results with the new constraint. Note that these new profiles can be obtained with the same DP simulation data.

Looking again at the acceleration phase, presented in Figure 8. With the applied constraint, the acceleration profiles now differ depending on the cruising speed. The gearshift strategy depends on this as well. However, the acceleration limit is reached at the very beginning by all speed profiles, the same acceleration is reached in Figures 4 and 6. Moreover the engine torque values are again close to its maximum values. Hence, in general the optimal speed and gear choice is the one that places the engine torque as close to the maximum as possible, which is due to that these are the zones of the engine map with high efficiency and thus lead to the lowest fuel consumption.

The results for the second deceleration strategy are presented in Figure 9. The optimal profiles have much longer coasting profiles because the speed losses are lower without the engine engaged to the driveline. In addition, the acceleration to the cruising speed is exactly the same as for the optimized deceleration using DP.

Table 2 presents the total fuel consumption as well as the trip time for the different cruising speed values with the two deceleration strategies. As one might think beforehand, it is true that shutting down the engine leads to a lower fuel consumption.

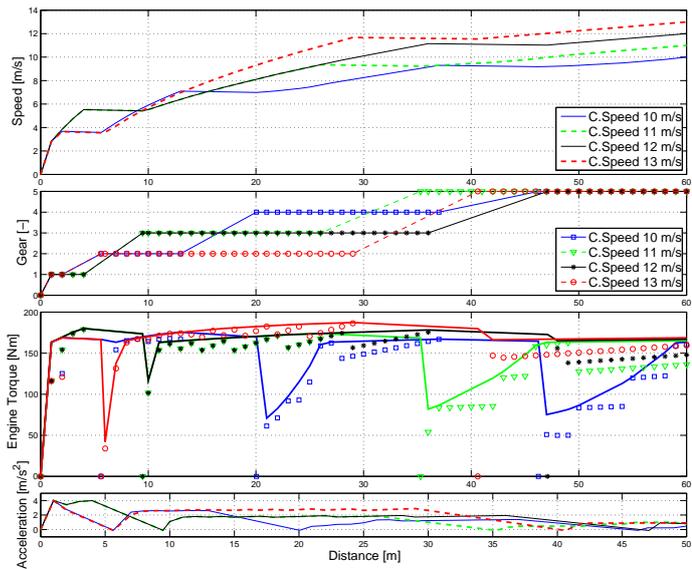


Fig. 8. Optimal acceleration profiles extracted from Figure 7. The thick lines in the Engine Torque plot represent the maximum torque available.

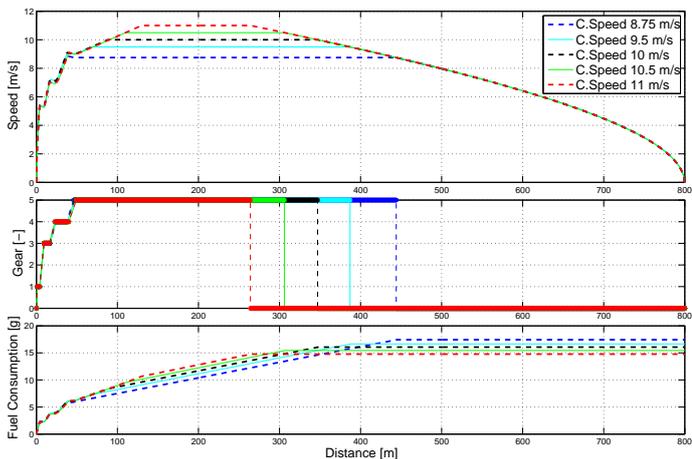


Fig. 9. Optimal speed profiles for several cruising speeds with engine shut down strategy.

8. CONCLUSION

A new approach to obtain speed profiles has been proposed. The algorithm is able to provide results applying different constraints with one single simulation data, while a normal DP algorithm would need one simulation for each optimal profile. This new algorithm has been built by taking assumptions about the shape of the optimal speed profile, i.e. constant cruising speed and coasting.

An advantage to use this method is to store simulation data for acceleration and deceleration conditions and consult them to obtain multiple optimal profiles without much more computational effort. It can be easily implemented into a vehicle computer to provide eco-driving tips to the driver.

Also a new way to handle gearshifts has been proposed, that takes into consideration the speed losses and the fuel consumption during the time that the gearshift maneuver

takes place. This model obtains more realistic results than without consider gearshift losses, and takes into consideration the vehicle engaged gear.

Results regarding optimal speed policies have been presented. Performing fast gearshifts until the highest gear is engaged appoints as optimal, as well as keeping constant cruise speed with the highest gear engaged until decelerate using the fuel cut-off feature.

REFERENCES

- Bellman, R. and Dreyfus, S. (1962). *Applied Dynamic Programming*. Princeton University Press.
- Bertsekas, D. (1995). *Dynamic Programming and Optimal Control*, volume I. Athena Scientific.
- Chang, D.J. and Morlok, E.K. (2005). Vehicle speed profiles to minimize work and fuel consumption. *Journal of Transportation Engineering*, 131(3), 173–182.
- Gausemeier, S., Jäker, K.P., and Trächtler, A. (2010). Multi-objective optimization of a vehicle velocity profile by means of dynamic programming. In *Prep. IFAC symp. Advances in Automotive Control*. AAC2010.
- Guzzella, L. and Amstutz, A. (1999). CAE-tools for quasistatic modeling and optimization of hybrid powertrains. *IEEE Transactions on Vehicular Technology*, 48(6), 1762–1769.
- Guzzella, L. and Onder, C.H. (2004). *Introduction to Modeling and Control of Internal Combustion Engine Systems*. Springer Verlag.
- Guzzella, L. and Sciarretta, A. (2013). *Vehicle Propulsion Systems - Introduction to Modeling and Optimization*. Springer Verlag, 3 edition.
- Hellström, E., Åslund, J., and Nielsen, L. (2010). Design of an efficient algorithm for fuel-optimal look-ahead control. *Control Engineering Practise*, 18(11), 1318–1327.
- Hooker, J.N. (1988). Optimal driving for single-vehicle fuel economy. *Transportation Research, Part A*(22A(3)), 183–201.
- Jorge, T.R., Lemos, J.M., Barao, M.S., and Bemporad, A. (2011). Hybrid dynamic optimization for cruise speed control. In *Prep. 18th IFAC World Congress*.
- Llamas, X. (2012). *Optimal Speed Profiles for Passenger Cars with Gear Shifting*. Master's thesis, Linköping University.
- Luu, H.T., Nouvelière, L., and Mammar, S. (2010). Dynamic programming for fuel consumption optimization on light vehicle. In *Prep. IFAC symp. Advances in Automotive Control*. AAC2010.
- Mensing, F., Trigui, R., and Bideaux, E. (2011). Vehicle trajectory optimization for application in eco-driving. In *Prep. 18th IFAC World Congress*.
- Saerens, B., Vandersteen, J., Persoons, T., Swevers, J., Diehl, M., and Van den Bulck, E. (2009). Minimization of the fuel consumption of a gasoline engine using dynamic optimization. *Applied Energy*, 86, 1582–1588.
- Saerens, B., Diehl, M., and Van den Bulck, E. (2010). Optimal control using pontryagin's maximum principle and dynamic programming. In *Automotive Model Predictive Control*, chapter 8. Springer Verlag.
- Schwarzkopf, A.B. and Leipnik, R.B. (1977). Control of highway vehicles for minimum fuel consumption over varying terrain. *Transportation Research*, 11(4), 279–286.

Factors Influencing the Estimation of Downhole Pressure far Away From Measurement Points During Drilling Operations

Eric Cayeux*
Hans Petter Lande**

*International Research Institute of Stavanger, Stavanger, N-4068,
Norway (Tel: +47 51 87 50 07; e-mail: eric.cayeux@iris.no).

**University of Stavanger, Stavanger, N-4036, Norway (e-mail: hp.land@outlook.com)

Abstract: In any drilling operation, it is important to maintain the pressure in the well within the geo-pressure margins (above collapse and pore pressure, and below fracturing pressure). The downhole pressure management consists primarily of selecting the operational drilling parameters (flow-rate, pump acceleration, rotational and axial velocities and accelerations of the drill-string) in such a way that the well pressure stays within the geo-pressure margins in the part of the well that is open to the formation. Alternatively, the pressure may be actively controlled by adjusting one of the parameters that influences the hydrostatic pressure, like the well head pressure in a back-pressure MPD (Managed Pressure Drilling) method or the level of the interface between the blanket and drilling fluid in a DG (Dual Gradient) method.

Primary indicators of pressure-related drilling problems, such as lost circulation or formation fluid influx, include surface measurements of active pit volume, unexpected pump pressure or hook loads in combination with downhole pressure measurements transmitted to the surface via telemetry systems. However, due to limited sampling rate and resolution, modern downhole telemetry systems are often incapable of capturing transient pressure pulses that may lead to drilling problems. Mathematical models can to some extent compensate for sparse downhole data measurements, but the reliability and accuracy of such models are sensitive to variations in wellbore position and diameter, including the presence of cuttings beds and hole enlargements, annulus temperature, and drilling fluid rheology. Deviation in any of these parameters influences the mathematical models cumulatively over distance, which can significantly limit their accuracy far from the measurement location.

This paper features an introduction to the drilling process, pressure-related drilling problems and a comprehensive review of sources of uncertainty in pressure predictions. Two case studies based on recent drilling operations are presented in detail and used to quantitatively estimate the impact of uncertainty on pressure predictions at the open hole sections of the wells. We show for both case studies how downhole pressure readings can substantially reduce the uncertainty in pressure predictions.

Keywords: drilling, wellbore position uncertainty, systematic error, downhole pressure uncertainty.

1. INTRODUCTION

In order to drill a well, one needs to rotate and press a drill-bit against the formation rock to cut into the rock materials and thereby create a hole.

To remove the produced cuttings from the hole, one uses a flow of drilling fluid (see **Fig. 1**) which transports the formation rocks up to the surface where they are separated from the drilling fluid using a system of screens with different meshes (so-called shale shakers). The clean drilling fluid flows back into the mud pits and can be reused. To allow for the flow of fluid down to the bit, a hollow drill-string is used. Mud pumps are used to circulate the fluid into the drill-string and after exiting at the bit, the mud continues its path to the surface within the annulus (*i.e.* the inter-space between the borehole and the drill-string). The same drill-string is used to exert a force from the bit upon the formation. The drill-string is hanging in a hoisting system (draw-works

or ram rig) that controls the position and velocity of the drill-string inside the borehole. The bit can be rotated in two different ways: either using a downhole motor (*i.e.* a positive displacement motor or a turbine) whose shaft is directly connected to the drill bit or by rotating the whole drill-string. In some cases, both means of rotation are applied simultaneously. When rotating the whole drill-string, two methods can be used: either a rotary table which is rotating a special hexagonal pipe connected to the top of the drill-string (a so-called kelly), or a top-drive which consists of a motor whose shaft is directly connected to the top of the drill-string.

The role of the drilling fluid is not limited to the transport of the cuttings back to the surface. It also ensures that the pressure in the borehole is:

- sufficient to stabilize the wellbore walls,

- larger than the pressure of the fluid contained in the porous rock so that formation fluid does not flow uncontrolled into the well while drilling,
- not too excessive so as to avoid fracturing the rock and thereby risking uncontrolled losses of drilling fluid to the formation.

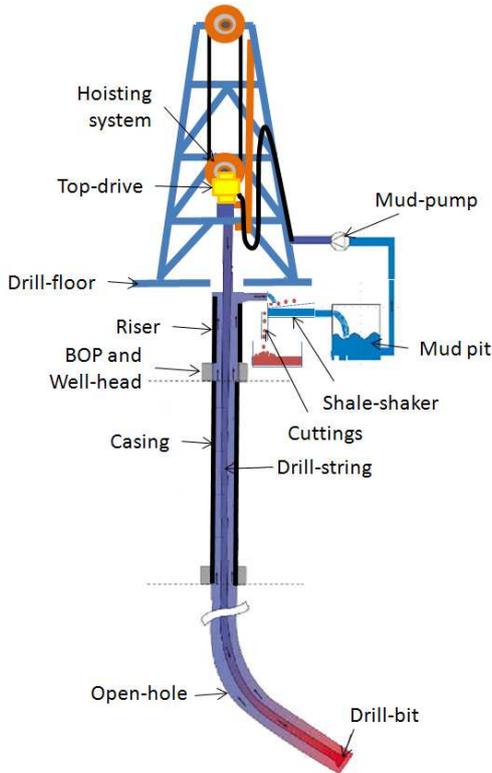


Fig. 1: Schematic depicting the main components used during a drilling operation

Hence, good estimation and control of the downhole pressure is essential for conducting safe drilling operations. However, downhole pressure measurements are sparse in both time and space. It is therefore necessary to consider how accurately we can estimate the downhole pressure at any depth and time during a drilling operation and if the accuracy of those calculations are compatible with the open hole pressure tolerances.

2. GEO-PRESSURE MARGINS

In this section, we will briefly describe the genesis of hydrocarbon reservoirs and the depth dependence of geo-mechanical properties of formation rocks in order to explain how the drilling process can be performed in a safe manner by ensuring borehole stability and controlling downhole fluid migration.

2.1 Genesis of Hydrocarbon Reservoirs

In aqueous environments with a low oxygen content, organic matter can deposit without being completely decomposed into mineral substances (such as water and CO_2). Especially in fine sediments (*e.g.*, shale, fine

limestone) deposited in water, the environment tends to be anaerobic because the sediments are impermeable and oxygen is quickly consumed by micro-organisms without being renewed. These sediments, filled with organic matter, constitute the source rock. In a subsiding basin, these sediments are buried deeper and deeper. At burial depths greater than 20 meters, the organic matter is transformed into kerogen, an organic substance with a high molecular weight. When exposed to the right temperature and pressure conditions, kerogen transforms into petroleum substances. The hydrocarbon material is thereafter expelled from the source rock by compaction and migrates towards the surface due to buoyancy. If it encounters an impermeable barrier during this vertical migration, it accumulates in the formation rock below the trap (*e.g.*, a cap rock is an impermeable seal that can form a trap) and thereby creates a hydrocarbon reservoir. Provided that the formation rock containing the hydrocarbons is porous and of sufficient permeability, it can constitute an economically viable source for hydrocarbon production.

2.2 Pore Pressure

The pressure of the fluid contained in the formation rock matrix is the result of the equilibrium between the weight exerted by the overlying sediments (lithostatic stress $\sigma_v(z) = p_{atm} + \int_0^z \rho_b g dm$, where p_{atm} is the atmospheric pressure, ρ_b is the bulk density of the rock, g is the gravitational acceleration and z is the vertical depth) and the capacity of the formation to expel the fluid such that it reaches a normal state of compaction (Charlez 1991).

With good permeability conditions, the formation rock is continuously in a perfectly drained condition and the sediments are said to be normally consolidated. A typical relationship for the porosity of normally consolidated material is:

$$\frac{\Phi}{1 - \Phi} = \frac{\Phi_0}{1 - \Phi_0} - \lambda \ln(gz(\rho_b - \rho_f)), \quad (1)$$

where Φ is the porosity at the vertical depth z , Φ_0 is the porosity of the material at the deposition time (typically 70% to 80%, Charlez 1991), λ is the compressibility coefficient of the material (assumed to be constant) and ρ_f is the density of the pore fluid.

However, if the formation has a low permeability, the drainage and compaction of the rock matrix is not synchronised resulting in the development of abnormal pore pressure. The consolidation process happens in four principle steps:

- In the initial conditions, *i.e.* at the initial porosity Φ_0 , the lithostatic stress (σ_v) and the pore pressure (p) are equal to the hydrostatic pressure at the top of the depositional environment (in most cases the sea bottom).
- When the sediment gets buried deeper, the fluid cannot flow out of the rock matrix because of the low permeability, and if the fluid is only slightly compressible, it can almost completely balance

the overburden pressure ($\sigma_v = p$) while the porosity does not change ($\Phi = \Phi_o$). This corresponds to undrained conditions.

- With time, the over-pressured fluid is expelled from the rock matrix: the effective stress increases while both the pore pressure and the porosity decrease.
- At the end of the transient phase, the pore pressure corresponds to the drained conditions and the porosity is equal to the normally consolidated material.

Impermeable rock materials are fairly common above the reservoir (for instance the cap rock has necessarily a very low permeability) and therefore formation pore pressure may vary greatly depending on the age of the sediment, its permeability and the variations in burial depth through the geological history of the basin.

2.3 Formation Fracturing

Formation rock is usually anisotropically loaded. There is obviously a vertical component (the lithostatic stress: σ_v) and two horizontal stress components: a minimum (σ_h) and a maximum (σ_H) horizontal stress that results from geological conditions. The orientation of the horizontal stress is also important. The magnitude and direction of the horizontal stress components varies at different depths and locations and can be related to rifting, subduction and other tectonic activities or the proximity to faults (Charlez 1991).

Drilling creates a hole in the formation (see Fig. 2). The stability of the borehole is related to the stress within the formation in the proximity of the wellbore considering that the hole is filled with a fluid at pressure p_w .

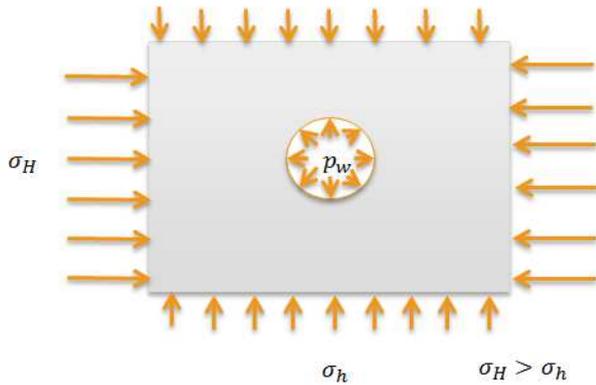


Fig. 2: The borehole stability of a vertical well corresponds to the classic Kirsh's problem (a hole in a plate with a bi-axial stress state).

As an illustration and for the sake of simplicity, we will consider drilling vertically through a brittle formation. Schmitt and Zoback (1989) derived the breakdown pressure (p_b), *i.e.* the pressure at which fractures are initiated, in a low permeability formation rock as:

$$p_b = T_0 + 3\sigma_h - \sigma_H - \alpha p_o \quad (2)$$

where T_0 is the tensile strength of the formation rock, α is the Biot's effective stress coefficient and p_o is the pore pressure. For more permeable rocks, the breakdown pressure is defined by:

$$p_b = \frac{T_0 + 3\sigma_h - \sigma_H - 2\zeta p_o}{1 + \alpha - 2\zeta}, \zeta = \alpha \frac{1 - 2\nu}{2(1 - \nu)} \quad (3)$$

where ν is the Poisson's ratio.

When a fracture is initiated and as long as the borehole pressure is large enough, the fracture remains open and may continue to propagate, giving the possibility for large quantities of drilling fluids to be lost to the formation.

The determination of the minimum horizontal stress can be obtained during a leak-off test (LOT). With an extended leak-off test (XLOT), it is also possible to determine the tensile strength and the maximum horizontal stress. During such tests, a short portion (typically 10 m) of the formation is exposed to the borehole pressure. The well is shut-in and the downhole pressure is increased by pumping drilling fluid in the closed well until the formation breaks down. As long as pumping continues, the fracture propagates. When pumping stops, with the well still shut-in, the pressure drops to the fracture closure pressure which is equal to the minimum horizontal stress. Fig. 3 illustrates schematically the expected pressure variations over the duration of an XLOT. Following the first cycle, in which breakdown and closure pressures are recorded, the pressure is bled off, then a second pressure cycle is initiated and one finds the re-opening fracture pressure, which is in fact lower than the initial breakdown pressure because of the destruction of the rock tensile strength, T_0 .

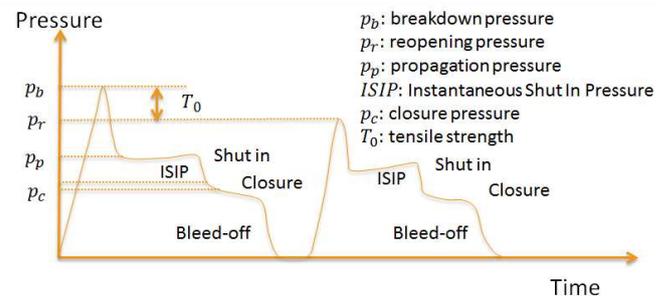


Fig. 3: Pressure variations during an XLOT

As it is necessary to perform a LOT or an XLOT with a very short open hole section, in practice it is possible to measure the minimum horizontal stress at only one depth per drilled section. To estimate the minimum horizontal stress at other depths, correlations based on various LOTs made in the same region are used (Breckels and van Eekelen 1982). For instance in the North Sea, the correlation for the minimum horizontal stress is:

$$\begin{cases} \sigma_h = 0.0053z^{1.145} + 0.46(p_f - p_{fn}), z < 3500m \\ \sigma_h = 0.0264TV D_{ss} - 31.7 + 0.46(p_f - p_{fn}), z \geq 3500m \end{cases} \quad (4)$$

where z is the vertical depth measured from MSL (Mean Sea Level), p_f is the pore pressure and p_{fn} is the normal pore pressure (based on water gradient).

2.4 Collapse Pressure

As seen above, the creation of a hole in an anisotropically loaded formation results in a redistribution of stress around the wellbore. If the formation strength cannot withstand the new stress distribution, then the formation rock breaks out resulting in pieces of rock called cavings falling into the borehole. Several shear failure criteria (σ_{cl}) are commonly used in the industry. The simplest of them is the Mohr-Coulomb criterion:

$$\sigma_{cl} = \frac{\sigma_1 + \sigma_3}{2} \sin \varphi + S_0 \cos \varphi - \frac{\sigma_1 - \sigma_3}{2} \leq 0 \quad (5)$$

where S_0 is the inherent shear strength and φ is the internal friction angle (Yi *et al.* 2005).

Another collapse failure index is the circumscribed Drucker-Prager criterion (Yu *et al.* 2001):

$$\sigma_{cl} = -\sqrt{J_2} + AJ_1^{ef} + B \leq 0 \quad (6)$$

with $J_1^{ef} = \frac{\sigma_{\rho\rho} + \sigma_{\theta\theta} + \sigma_{ss}}{3} - p$ and $J_2 = \frac{1}{9} \left((\sigma_{\rho\rho} - \sigma_{\theta\theta})^2 + (\sigma_{\theta\theta} - \sigma_{ss})^2 + (\sigma_{ss} - \sigma_{\rho\rho})^2 + 6\sigma_{\theta s}^2 \right)$ where $A = \frac{2\sqrt{2} \sin \varphi}{3 - \sin \varphi}$, $B = \frac{2\sqrt{2} S_0 \cos \varphi}{3 - \sin \varphi}$.

Yet another commonly used criterion is the modified Lade criterion (Ewy 1999):

$$\sigma_{cl} = 27 + \xi - \frac{I_1^{n_3}}{I_3} \leq 0 \quad (7)$$

where $I_1'' = (\sigma_1 + S_1) + (\sigma_2 + S_1) + (\sigma_3 + S_1)$, $I_3'' = (\sigma_1 + S_1)(\sigma_2 + S_1)(\sigma_3 + S_1)$, with $S_1 = \frac{S_0}{\tan \varphi}$ and $\xi = \frac{4 \tan^2 \varphi (9 - 7 \sin \varphi)}{1 - \sin \varphi}$.

Based on the studies conducted so far, no one criterion model can be singled out as superior to the others when it comes to predicting shear failure (Colmenares and Zoback 2001, Yi *et al.* 2005, Nawrocki 2010).

Note that calliper logs or wellbore imaging can provide information about the orientation of the horizontal stresses by observing the orientation of breakouts or breakdowns, since breakouts are oriented by the minimum horizontal stress and breakdowns by the maximum horizontal stress.

2.5 Geo-pressure Window

If the borehole pressure is lower than the pore pressure in an open hole section of the well, formation fluid starts to enter the annulus. During drilling, this is a situation that in most cases is undesirable as it can lead to contamination of the drilling fluid with external components, but mostly because hydrocarbons may be present in an uncontrolled manner at the rig site with potential risks of fire and explosion. Thus, unless during an under-balanced drilling operation where adequate surface equipment is used to manage hydrocarbons while drilling, it is crucial to maintain the downhole pressure above the pore pressure of open hole sections at all times to avoid a formation fluid influx.

Similarly, the downhole pressure shall be larger than the collapse pressure, in order to avoid cavings falling into the

well which, because of their large size cause difficulties in cleaning the hole with the potential of degenerating into a stuck pipe situation.

At the opposite extreme, the downhole pressure shall remain below the fracture pressure in order to avoid creating fractures in the formation and thereby risking mud losses which in the worst case scenario, can result in a loss circulation incident.

To sum up, the borehole pressure is bounded on the low side by the maximum of the collapse pressure and the pore pressure and on the high side by the fracturing pressure (see Fig. 4). The interval defined by these limits is called the safe downhole pressure window.

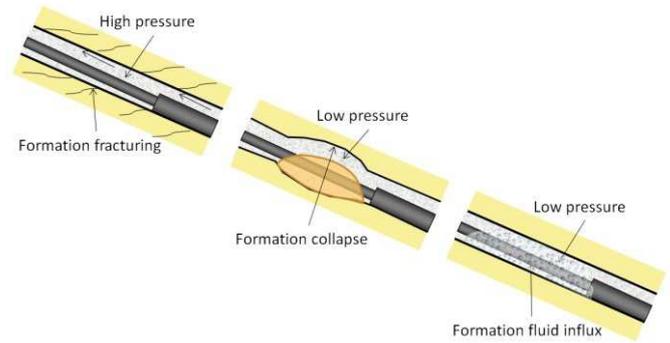


Fig. 4: Downhole pressure related geo-mechanical and well control problems.

Since pressure increases with depth, it is relatively difficult to work directly with a pressure window. Instead, it is more convenient to work with the gradient of pressure compared to the vertical depth. For convenience, the pressure gradient is transformed into a density by dividing it by the earth gravitational acceleration. This value is referred to as an equivalent mud weight (EMW) because it is the density of the fluid that would have caused the same pressure at the same vertical depth:

$$\rho_{emw} = \frac{p - p_{atm}}{gh} \quad (8)$$

where ρ_{emw} is the equivalent mud density and h is the vertical depth.

It is quite common that the fracturing pressure gradient is at its minimum at the start of the section and that the pore pressure gradient is at its maximum toward the end of the section. This corresponds to an overall limited geo-pressure window, as illustrated in Fig. 5.

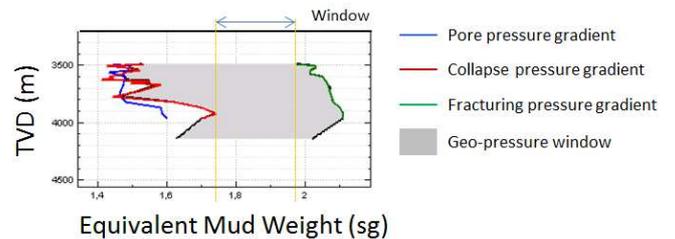


Fig. 5: Typical geo-pressure window

In the case of over-pressured formations, it is also possible that a narrow geo-pressure window occurs in the

same range of depths. **Fig. 6** shows an example where the geo-pressure margin at 1730 m true vertical depth (TVD) is no larger than 12 bars.

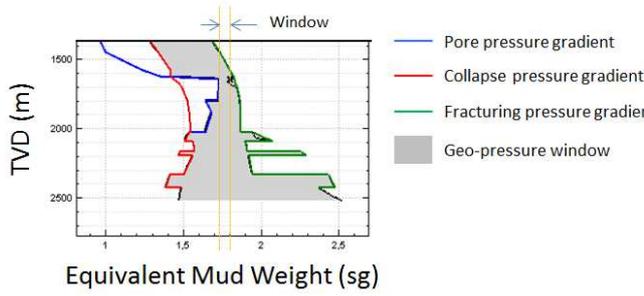


Fig. 6: Example of a narrow geo-pressure window

3. DOWNHOLE PRESSURE

In this section we will review how downhole pressure is generated and controlled. We will also analyse the source of deviations of downhole pressure.

3.1 Drilling Methods

When a well is filled with a drilling fluid, the hydrostatic downhole pressure at a given measured depth (*i.e.* a curvilinear abscissa s) is:

$$p(s) = p_0 + \int_{s_0}^s \rho_m g \cos I \, dm \quad (9)$$

where s_0 and p_0 are, respectively, the curvilinear abscissa and the pressure at the start of the borehole, ρ_m is the density of the drilling fluid at a given depth, I is the wellbore inclination angle relative to the downward vertical at the same depth.

In conventional drilling, the top of the well is not sealed and therefore $p_0 = p_{atm}$. In addition, only one single drilling fluid is used. Usually the drilling fluid has relatively low compressibility and thermal expansion (see a in **Fig. 7**).

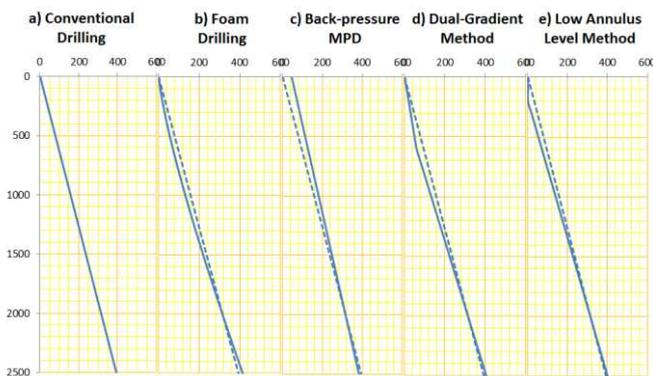


Fig. 7: hydrostatic downhole pressure with different drilling methods in a 2500 m vertical well.

In circumstances where the water table is significantly deeper than the ground level, it is necessary to use a drilling fluid with a lower density than water to obtain pressure gradients that are compatible with the geo-pressure window. This has led to the use of foams as drilling fluids. As a consequence, the density of the drilling fluid is sensitive to

both pressure and temperature because of its large compressibility and thermal expansion (see b in **Fig. 7**).

It is also possible to seal the top of the wellbore and to apply a back pressure on the well. By varying the back pressure applied at the top of the well, it is possible to control the downhole pressure. This is a so-called back-pressure managed pressure drilling (MPD) method (see c in **Fig. 7**).

In deep water environments, the formation pore pressure and geo-mechanical characteristics are strongly influenced by the water pressure gradient and the lithostatic stress gradient below the sea bottom. In order to obtain downhole pressure gradients that are compatible with those conditions, dual-gradient methods can be used (see d in **Fig. 7**) where two different fluids fill the wellbore. In this drilling method, the top of the well is filled with a blanket fluid (typically seawater) while the bottom of the well is filled with a drilling fluid. A lift pump, placed just below the interface between the blanket fluid and the drilling fluid, ensures the return of the drilling fluid without disturbing the blanket fluid.

An alternative dual-gradient method consists of using air as a blanket fluid. In this case the liquid level in the wellbore is adjusted by controlling the lift pump rate to produce the dual-gradient effect (see e in **Fig. 7**).

See, for instance, Cayeux *et al.* (2013a) for more detail concerning the arrangement of equipment used for these different drilling methods.

3.2 Downhole Pressure Measurements

It is possible to measure the downhole pressure using an annular-pressure-while-drilling tool. The tool is located in the BHA (Bottom Hole Assembly) about 10 to 20 m above the bit. The sensor has a typical accuracy of 0.6 bars over the pressure range 0 to 1500 bar (Ward and Andreassen 1998). In drilling today, the most frequently used data transmission technology is mud pulse telemetry. With this technology, downhole pressure information is available with refresh intervals ranging from 30 s to a few minutes and only when the flow-rate is above a certain threshold fixed by the MWD (measurement while drilling) transmission tool. As a consequence, the downhole pressure is not transmitted in real-time when the mud pumps are turned off or when the flow-rate is low. However, when the mud pumps are cycled, the MWD tool transmits information about the average downhole pressure recorded during the period when the mud pumps were off, therefore providing some information about the hydrostatic pressure. This information may be of limited value if the drill-string has moved during that period of non-data transmission, because swab and surge pressures resulting from the axial movement of the drill-string are superposed on the hydrostatic pressure.

New high-speed telemetry systems based on induction couplings in the drill-string, acoustic waves along the drill-pipes or electro-magnetic waves through the formation have lately become available. These modes of transmission allow for a higher bandwidth and also for distributed sensors along the drill-string. Coley and Edwards (2013) reported how distributed pressure sensors have been used to monitor the

circulation of high density sweep, locating downhole events such as pack-offs and monitor the actual efficiency of transporting cuttings. The nominal accuracy of the pressure sensors used during those drilling operations was no better than 6.9 bar but the investigators reported that the pressure sensors did react to pressure variations as low as 0.1 bar and that the observed inaccuracy could be compensated for by performing local sensor calibration.

Running a downhole pressure tool in the BHA, or distributed pressure sensors along the drill-string increases cost. In addition increasing the number of complex elements used in a drilling operation is often associated with increasing the risk that one of them can fail. So the question is to evaluate with which accuracy we can estimate the annulus pressure far away from downhole sensor and check whether that precision is compatible with the geo-pressure margins.

3.3 Effect of Temperature and Pressure on Drilling Fluid Density

There are several types of drilling fluids. The most commonly used are the water-based mud (WBM), the oil-based mud (OBM) and the synthetic-based mud (SBM). For these types of fluid, the density of the mud is obtained by changing the proportion of high gravity solid (e.g. barite) in the mix. Another way of controlling the density of the drilling fluid consists of diluting a heavy liquid such as cesium formate with another lighter liquid such as potassium formate. The drilling fluid contains other additives in order to manage its chemical, viscous, gelling and fluid loss control properties.

Due to its compressibility and thermal expansion, the density of each of the components changes as a function of pressure and temperature. Because drilling fluids are used in a wide range of pressures and temperatures, it is seldom acceptable to consider the compressibility and thermal expansion of the fluid components as constant. For instance, Isambourg *et al.* (1996) recommend a biquadratic relationship for the density of the components with respect to temperature and pressure (see Fig. 8).

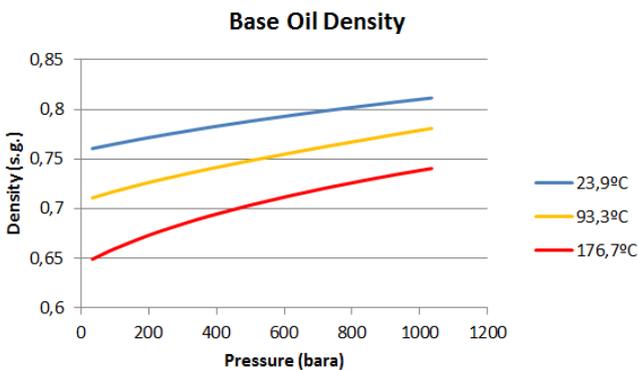


Fig. 8: Example of the density variations of a typical base oil as a function of pressure and temperature.

The pressure-volume-temperature (PVT) properties of the base constituents of the drilling fluid may be quite different from one type of mud to another (see Fig. 9).

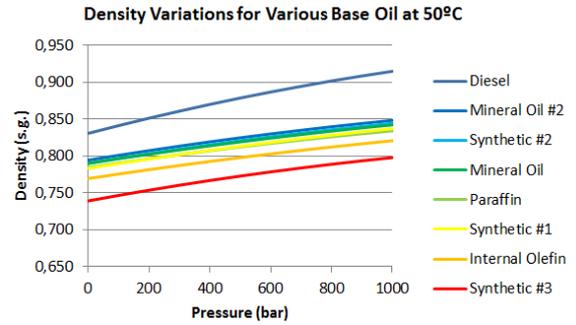


Fig. 9: Example of density dependence on pressure at 50°C for different base oils

Unfortunately, there are instances where the PVT properties of the base constituent are not known. In that case, it is necessary to consider the possible statistical variations of the PVT properties of the constituent type, in order to evaluate its influence on the downhole pressure estimation.

For instance, with the different base oils used to construct Fig. 9 (but excluding diesel since it is seldom used nowadays), we can fit a generic base oil model based on a polynomial of degree 2 in pressure and degree 1 in temperature, as recommended by the API Recommended Practices 13D (2006):

$$\rho_o = (A_o + B_o T) + (C_o + D_o T)p + (E_o + F_o T)p^2 \quad (10)$$

where ρ_o is the oil density, T is the temperature and p is the pressure. Using p in units of psi, T in °F, ρ_o in kg/m³ then $A_o=824.52$ ($\sigma^2 = 35.27$ where σ is the standard deviation), $B_o=-0.375$ ($\sigma^2 = 0.000906324$), $C_o=0.003021$ ($\sigma^2 = 0.00000362931$), $D_o = 0.000012877$ ($\sigma^2 = 0.00000000009326$), $E_o=-2.31E-8$ ($\sigma^2 = 1.6E - 14$), $F_o=-3.23E-10$ ($\sigma^2 = 4.11E - 19$) (see Fig. 10).

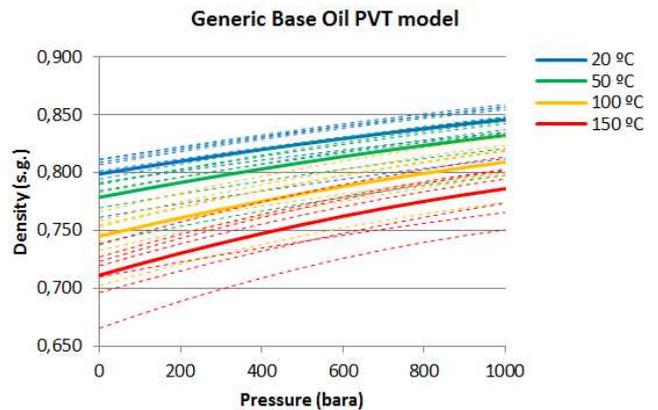


Fig. 10: Generic Base Oil PVT model generated from 7 specific base oil density models.

A similar treatment can be made for the brine phase. However, in that case the lack of information usually relates to the type of salts and their concentrations in the electrolyte. Kemp *et al.* (1989) give a precise description on how to calculate the density of a brine containing different salts at various concentrations. Using this model, it is possible to generate individual PVT models for the most common

solutes used in drilling fluids, *i.e.* NaCl, KCl and CaCl₂ (see Fig. 11).

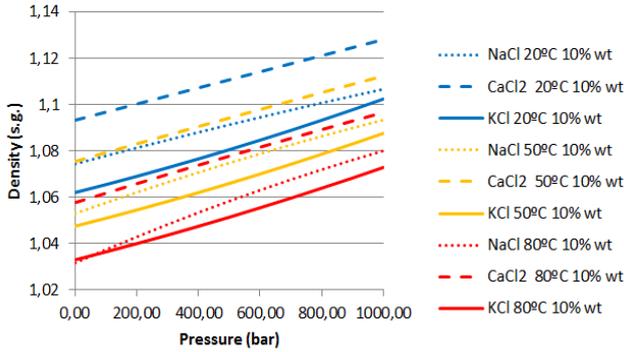


Fig. 11: Brine PVT for different dissolved salts for a concentration of 10% in weight.

As for the oil phase, it is possible to derive a generic brine density model with associated statistical characteristics that can further be used to produce statistically meaningful brine density models when the type of solute is not defined:

$$\begin{aligned} \rho_w & \\ &= (S_0 + S_1 w_t + S_2 w_t^2 + S_3 w_t^3 + B_w T) \\ &+ (C_w + D_w T)p + (E_w + F_w T)p^2 \end{aligned} \quad (11)$$

where w_t is the weight fraction of the salt at normal condition of temperature and pressure (atmospheric pressure and 297 K). With w_t in %, p in units of psi, T in °F, ρ_w in kg/m³, then the fitted parameters based on NaCl, KCl and CaCl₂ are $S_0 = 1024.17$ ($\sigma^2 = 1.41E - 6$), $S_1 = 7.28$ ($\sigma^2 = 3.65E - 7$), $S_2 = 2.25E - 2$ ($\sigma^2 = 5.90E - 9$), $S_3 = -6.00E - 5$ ($\sigma^2 = 6.38E - 12$), $B_w = -3.31E - 1$ ($\sigma^2 = 1.26E - 2$), $C_w = 1.96E - 3$ ($\sigma^2 = 2.17E - 5$), $D_w = 6.31E - 6$ ($\sigma^2 = 1.29E - 9$), $E_w = 1.57E - 8$ ($\sigma^2 = 9.58E - 14$), $F_w = -1.45E - 10$ ($\sigma^2 = 5.69E - 18$).

If the weight fraction of the salt is not fully known or is biased by contamination from downhole formation fluids, then the salt weight fraction shall also be statistically varied.

For an OBM or an SBM, the oil/water ratio K_1 is known at surface conditions (temperature T_1 and pressure p_1). Because the compressibility and thermal expansion of the oil phase and the brine phase are not identical, the oil/water ratio does not remain constant when the drilling fluid is exposed to different temperatures and pressures. At new conditions of temperature T and pressure p , the oil/water ratio is defined by:

$$K(p, T) = K_1 \frac{\rho_{o1} \rho_w(p, T)}{\rho_{w1} \rho_o(p, T)} \quad (12)$$

where ρ_{o1} and ρ_{w1} are the densities of the oil and brine phases (respectively) at the surface conditions, and $\rho_o(p, T)$ and $\rho_w(p, T)$ are the densities of the oil and brine phases at the temperature T and pressure p . In view of the above discussion on the uncertainty of the oil and brine densities, one shall also consider that the oil/water ratio is subject to uncertainty as well. Using the above relation, we can deduce the density of the liquid emulsion:

$$\begin{aligned} \rho_l(p, T) & \\ &= \frac{K(p, T)}{1 + K(p, T)} \rho_o(p, T) + \frac{1}{1 + K(p, T)} \rho_w(p, T) \end{aligned} \quad (13)$$

where ρ_l is the density of the liquid mix.

Similarly, if we have a volume ratio (Λ) between the low gravity solids (V_{lgs}) and the high gravity solids (V_{hgs}):

$$\Lambda = \frac{V_{lgs}}{V_{hgs}} \quad (14)$$

then we can express the combined solid density (ρ_s) as:

$$\rho_s = \frac{\Lambda}{1 + \Lambda} \rho_{lgs} + \frac{1}{1 + \Lambda} \rho_{hgs} \quad (15)$$

where ρ_{lgs} is the density of the low gravity solids and ρ_{hgs} is the density of the high gravity solids. Here, we will consider that the density of solids is not changing with pressure and temperature even though some other investigators (Isambourg *et al.* 1996) do account for those tiny variations.

The volume fraction of solids in the drilling fluid changes with the conditions of pressure and temperature because of the compressibility and thermal expansion of the liquid phase. The estimation of the volume fraction of solid is obtained from the initial density of the drilling fluid (ρ_{m2}) at surface conditions of temperature (T_2) and pressure (p_2). In those conditions the reference solid volume fraction (f_{s2}) is:

$$f_{s2} = \frac{\rho_{m2} - \rho_{l2}}{\rho_s - \rho_{l2}} \quad (16)$$

where ρ_{l2} is the density of the liquid in those conditions of temperature (T_2) and pressure (p_2). The solid volume fraction at different conditions of the temperature and pressure is then:

$$f_s(p, T) = \frac{f_{s2} \rho_l(p, T)}{(1 - f_{s2}) \rho_{l2} + f_{s2} \rho_l(p, T)} \quad (17)$$

So finally, we can calculate the density of the drilling fluid (ρ_m) at any temperature T and pressure p :

$$\rho_m(p, T) = f_s(p, T) \rho_s + (1 - f_s(p, T)) \rho_l(p, T) \quad (18)$$

It should be noted that the exact densities of the low gravity solids and high gravity solids are usually not documented in drilling fluid or daily drilling reports. So the uncertainty on the exact density of these components has, in turn, a consequence on the calculation of the volume fractions and therefore on the way the density of the drilling fluid varies as a function of pressure and temperature. For instance, the density of barite (a typical high gravity solid) is typically between 4.3 sg to 5 sg and the density of bentonite (a low gravity solid) ranges from 2.4 sg to 2.7 sg.

To illustrate the effect of these uncertainties on the PVT properties of a drilling fluid, let us consider an oil based mud with the following characteristics at 50°C and at atmospheric pressure: density 1.7 sg, oil/water ratio 70/30. We do not

have the PVT characteristics of the base oil and we do not know the salinity nor the amount of low gravity solids. However, since we do know the density of the fluid at one given temperature and pressure, we can estimate the quantities of each component by assuming some values for the missing information:

- density of each component,
- salinity,
- fraction of low gravity solids.

By performing Monte Carlo simulations on the unknown parameters of the model, we can obtain the most likely density variations as a function of temperature and pressure as well as their associated standard deviations. Note that for all the following examples, the stochastic simulations are based on 150000 trials.

Assuming that the probabilistic simulations are following a Gaussian law, the variations of densities within 1 time the standard deviation corresponds to a 68% confidence that the density is within that interval and the ones within 2 times the standard deviation are inside a confidence interval of 95% (see Fig. 12).

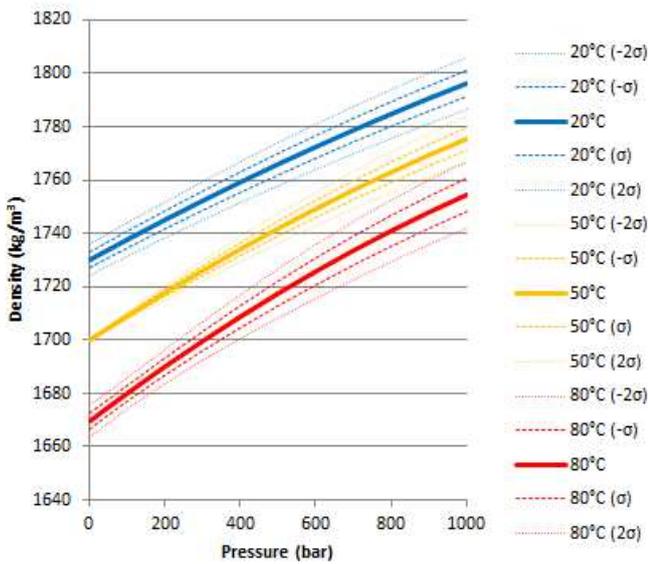


Fig. 12: PVT properties of an unspecified OBM with a nominal density of 1.7 sg at 50°C (atmospheric pressure) having an oil/water ratio of 70/30. The 1σ margins correspond to a 68% confidence for the density and the 2σ margins represent a 95% probability of presence of the drilling fluid density.

Eq. 9 shows that the hydrostatic pressure at a given depth is directly related to the density of the drilling fluid above that depth. At the same time the density of the drilling fluid at any depths depend on the local pressure and temperature conditions. Therefore the solution closure depends on the estimation of the downhole temperature at any depth of the well.

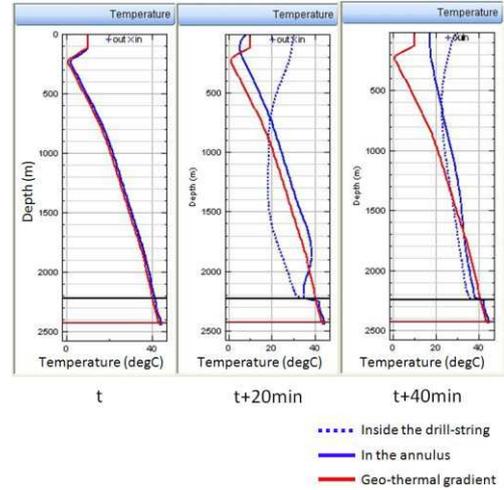


Fig. 13: These three graphs show the evolution of the temperature of the fluid inside the drill-string and in the annulus when circulating at a constant flow-rate of 2500l/min and starting at geothermal conditions.

As for downhole pressure, downhole temperature measurements are very sparse, if available at all. It is possible to calculate the temperature of the drilling fluid at any depth as a function of time by using a transient heat transfer model (Corre *et al.* 1984). It should be noted that the heat transfer process never reaches any steady state conditions (see Fig. 13) because the mud pumps are stopped and started each time a new stand is added, and because the depth of the well continuously changes due to the drilling process, hence exposing deeper formations at higher temperature. It is therefore important to use a transient model connected to the real-time drilling parameters in order to more accurately estimate the downhole temperature.

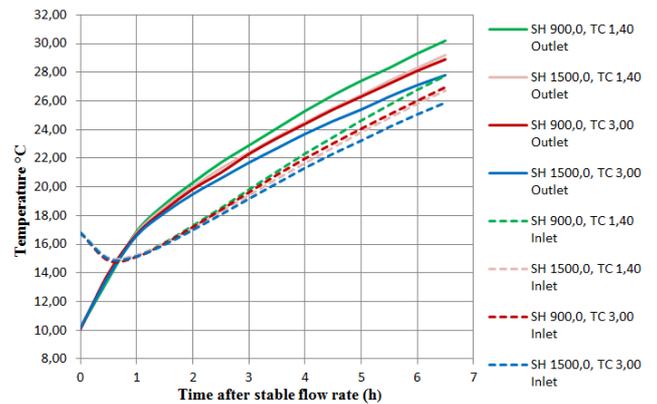


Fig. 14: Inlet and outlet temperatures when circulating at a constant flow-rate of 2500 l/min with a typical OBM (SH stands for Specific Heat Capacity, TC stands for Thermal Conductivity) (Courtesy Toft 2011).

However such models depend on parameters that are seldom known such as the specific heat capacity and thermal conductivity of the various formation rocks penetrated by the well. Considering that the specific heat capacity of sedimentary rock can easily vary between 500 J/kg.K and 1500 J/kg.K, and that the thermal conductivity of such rocks can be found between 1.4 W/m.K to 3 W/m.K, the predicted

temperature along the borehole may deviate from reality by a few degrees (see Fig. 14).

The specific heat capacity and thermal conductivity of the drilling fluid depend on the composition of the drilling fluid. There are relatively large differences for the specific heat capacity and thermal conductivities of different variants of the mud constituents, so if the exact thermal properties of the individual constituents are not known then we shall account for the possible variations. For instance, the specific heat capacity of standard oil is 1550 J/kg.K while the one of mineral oil is 400 J/kg.K. As another example, the thermal conductivity of bentonite can be found between 0.34 and 0.58 W/m.K.

Using the same mud description as above, Monte Carlo simulations, based on the uncertainty of the mud component characteristics, give a specific heat capacity of 1126 J/kg.K with a standard deviation of 144 J/kg.K and a thermal conductivity of 0.43 W/m.K with a standard deviation of 0.16 W/m.K.

The thermo-physical properties of the drilling fluid also depend on the concentration of cuttings in suspension (Cayeux *et al.* 2013b). Erroneous estimations of the cuttings concentration at various depth as a function of time is yet another cause of discrepancies between the actual downhole temperature and the estimated one.

Furthermore, the estimation of the amount of heat generated by the work of the bit, mechanical friction and hydraulic turbulence is somewhat debatable. Unless there are reliable downhole temperature measurements available for calibration, altogether these sources of uncertainty in the heat transfer calculations may cause deviations in the estimation of downhole temperature compared to actual values in the magnitude of a dozen degrees Celsius.

3.4 Effect of Cuttings Concentration on Drilling Fluid Density

The drilling process produces cuttings that need to be removed from the well: this is one of the reasons why drilling fluids are circulated through the well. With an adequate flow-rate and drill-string rotational velocity, the cuttings are kept in suspension, even when the borehole is deviated, hence they can be transported out of the hole together with the drilling fluid (Cayeux *et al.* 2013b).

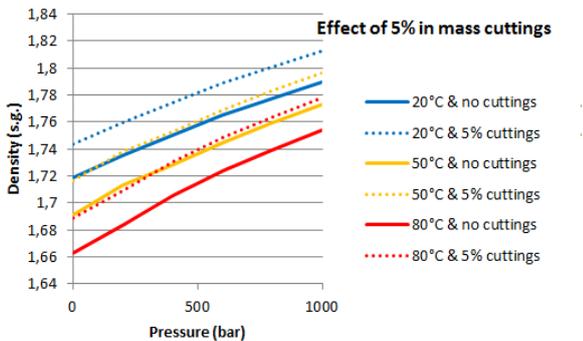


Fig. 15: Change of density of a typical drilling fluid when there is a 5% concentration of cuttings in suspension.

The cuttings in suspension affect the local density of the drilling fluid as a function of their concentration and density (see Fig. 15).

The determination of the cuttings concentration at a certain depth depends on what has happened to the cuttings particle during its transport. For instance, the size of the particle changes during transport due to the grinding action of the drill-string rotation against the wall of the borehole. Since the size of the particle is quite important in the calculation of its slip velocity, it obviously has an impact on the cuttings concentration at different depths. The cuttings size and shape also influences the probability that a particle will settle on the low side of inclined sections of the well when the drilling parameters (*i.e.* flow-rate and drill-string rotational speed) are not sufficient to keep it moving on the high side of the annulus. Furthermore, the cuttings dimensions play a role in determining cuttings bed erosion. As a consequence, there is some degree of uncertainty on the actual concentration of cuttings particles at any depth along the wellbore (see Fig. 16).

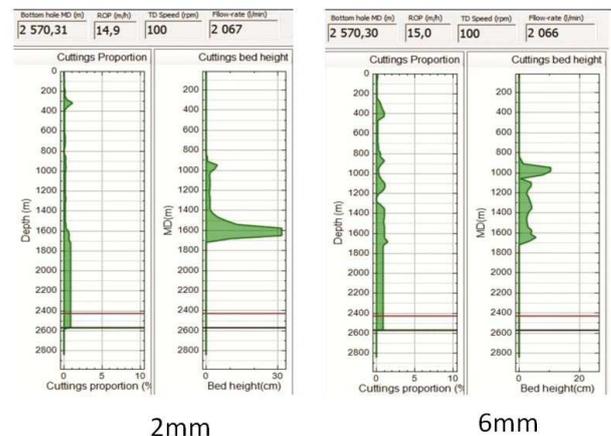


Fig. 16: Cuttings concentration and cuttings bed height for two different cuttings particle size and cuttings bed erosion conditions.

3.5 Effect of Gas Concentration on Drilling Fluid Density

Formation rock may contain gases such as CO₂, H₂S or hydrocarbon gases in different proportions. When drilling through these gas bearing formations, gas invades the drilling fluid simply because it is released from the cuttings.

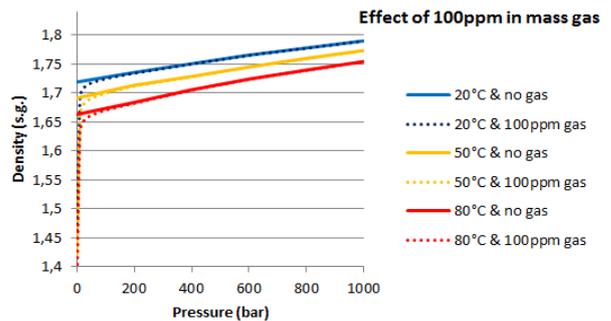


Fig. 17: Effect of the presence of gas on drilling fluid density.

Gas may dissolve in the drilling fluid, especially if the drilling fluid is an OBM or a SBM. When dissolved, the density of the drilling fluid is slightly changed (O'Bryan *et al.* 1988, Monteiro 2005). But when the pressure drops below the bubble point pressure, dissolved gases return to gas form with bubble sizes that increase greatly as the pressure decreases, causing drastic changes in the drilling fluid density (see Fig. 17).

The concentration of the various gases is measured when the drilling fluid returns to the surface installation, but this is too late if one wants to account for the effect of gas on the *in situ* drilling fluid density. This leads to an additional source of uncertainty for the actual mud density at different depths of the borehole.

3.6 Effect of Fluid and Drill-String Movements on Downhole Pressure

So far, we have concentrated on the estimation of downhole pressure for hydrostatic conditions, but in practice the downhole pressure is also influenced by pressure losses along the annulus when the drilling fluid is circulated and the drilling-string is moved axially and rotationally. The downhole pressure under such conditions can be described by differential equations for mass and momentum balance, (Fjelde *et al.* 2003):

$$\begin{cases} \frac{\partial}{\partial t} (A\rho) + \frac{\partial}{\partial s} (A\rho v) = q \\ \frac{\partial}{\partial t} (A\rho v) + \frac{\partial}{\partial s} (A\rho v^2) + A \frac{\partial}{\partial s} p = -A(K - \rho g \cos(I)) \end{cases} \quad (19)$$

where t is time, s is the curvilinear abscissa, A is the cross-sectional area of a fluid element, ρ is the averaged density, v is the average velocity, q is the source term (a mass per length per time through the fluid element side walls), p is the pressure and K is the friction pressure-loss term.

The friction pressure loss depends on the fluid velocity, its viscosity and whether the flow regime is laminar, intermediate or turbulent. Drilling fluids are non-Newtonian and their rheology is relatively well described by a Robertson-Stiff constitutive law (Robertson and Stiff 1976):

$$\tau = A (\dot{\gamma} + C)^B \quad (20)$$

where τ is the shear stress, $\dot{\gamma}$ is the shear rate, A , B and C are the coefficients of the model. As for any fluid, the viscosity of drilling mud depends on the temperature but it can also depend on the pressure as well (Houwen and Geehan 1986). In practice, this means that the coefficients A , B and C of the Robertson-Stiff model are functions of temperature and pressure.

The rheology of the drilling fluid is measured at regular intervals (typically four times a day) and usually this measurement is done at atmospheric pressure and at a given temperature (see Fig. 18).

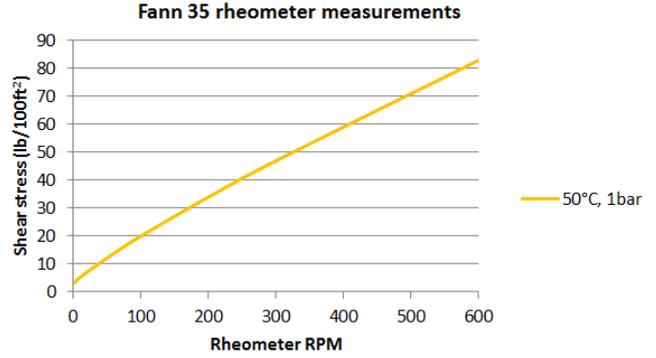


Fig. 18: Rheometer measurements for a typical drilling fluid made at 50°C and atmospheric pressure.

To retrieve the pressure and temperature dependence of the mud rheology, one relies on empirical models for the particular drilling fluid that is in use. If the rheology dependence on pressure and temperature of the drilling fluid is not known, or if the formulation of the mud currently used differs in some ways from the reference model, the estimation of the effective viscosity ($\mu = \frac{\tau}{\dot{\gamma}}$) at *in situ* conditions will be uncertain. Fig. 19 shows the variation in the estimation of the effective viscosity at various pressures and temperatures when using three different reference rheological models for the drilling mud (example based on the rheometer measurements from Fig. 18).

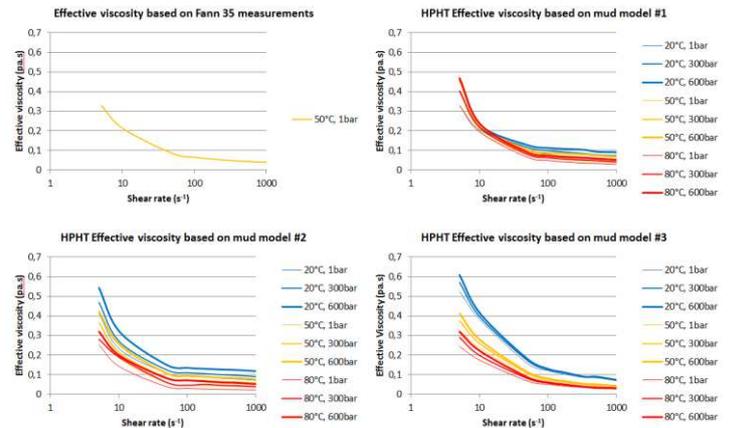


Fig. 19: Estimation of the effective viscosity based on a single rheometer measurement with 3 different mud rheological models.

Furthermore, the presence of cuttings and gas also changes the rheological properties of the drilling fluid (see Fig. 20) and since we have some uncertainty on the actual concentration of cuttings particles and gas bubbles, this adds to the uncertainty on pressure loss calculations.

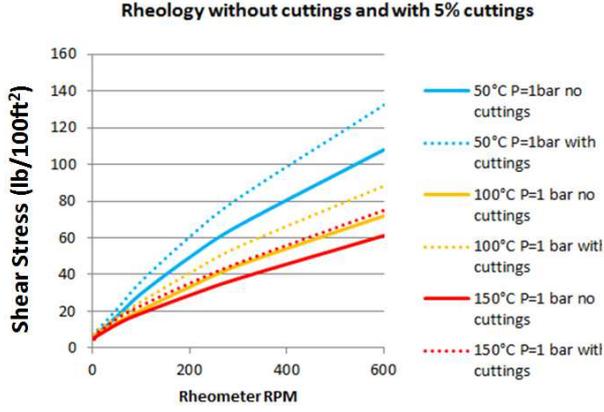


Fig. 20: Effect of cuttings in suspension on the rheology of a drilling fluid.

Another source of uncertainty is the actual flow-rate at which the drilling fluid is pumped into the drill-string. It is extremely seldom that a precise flow-meter is available at the inlet of the system. Instead the flow-rate is estimated based on the stroke rate of the mud pumps ($Q = fV_{st}\eta$ where Q is the flow-rate, f is the stroke-rate, V_{st} is the volume of one stroke and η is the pump efficiency). Traditionally the pump efficiency is considered to be constant and is calibrated once in a while, typically during a cement job. However, Shafer *et al.* (1992) reported that in their test setup where an accurate flow-meter (electro-magnetic) was connected to the inflow of the mud pumps, the pump efficiency had varied from 86% to 96% during the experiment with daily variations as high as 5%. The pump efficiency reflects the leakage between the pistons and the liners as well as the fluid flow back of the inlet and outlet valves, dependent on the valve design, closing speed and condition of the valve springs. Therefore the pump efficiency is in fact not constant and depends on the flow-rate, pressure and properties of the drilling fluid.

Finally, the basic pressure loss calculations in the annulus are usually based on assumptions of a concentrically oriented drill-string and no drill-string rotation. However, in deviated wells the drill-string axis will be decentred compared to the wellbore axis and, unless drilling with a downhole motor (PDM) in oriented mode, the drill-string will rotate. Empirical models to correct for these effects have been developed (Pilehvari and Serth 2009, Hansen and Sterri 1995), but experience shows that these are limited in accuracy. So, in deviated wells with drill-string rotation, one must expect to obtain slightly incorrect estimations of the pressure losses in the annulus.

3.7 Effect of Apparent Hole Diameter on Pressure Losses

Pressure loss calculations depend very much on the bulk velocity of the drilling fluid. However, in the presence of cuttings beds or hole enlargements due to washout or hole collapse, the actual cross sectional area of the wellbore may be quite different from the expected one. Even though a downhole calliper may be run in some circumstances while drilling, the position of this instrument in the BHA, and therefore close to the bit, limits the availability of the actual

hole size at shallower depths. Those restrictions and enlargements may play an important role in the pressure losses occurring through the annulus.

4. WELLBORE POSITION UNCERTAINTY AND VERTICAL DEPTH ACCURACY

Any downhole pressure estimation in the open hole section of a well has to be related to a corresponding formation layer in order to retrieve the correct geo-pressure boundaries. In practical terms, that means a true vertical depth. In this section we will review how the wellbore position can be measured and the methods that have been developed to estimate its accuracy. We will thereafter utilize the wellbore position accuracy estimation methods to derive trajectories that most likely represent extreme situations in terms of vertical bias, so as to evaluate the impact of wellbore position uncertainty on pressure calculations.

4.1 Wellbore Position Surveying

The position of a well cannot be measured directly because it is difficult to make direct spatial measurements through the earth crust (both electro-magnetic and pressure waves are quickly attenuated and dispersed when traveling through rock materials). In fact it is derived from the integration of tangential measurements made along the borehole. Indeed, at a given location, it is relatively simple to measure the inclination (I) of the wellbore axis compared to the gravitational field and to measure the direction (azimuth: A) of the well axis relative to the geomagnetic field or by recording direction changes using a gyroscope. Furthermore, the length of the borehole is relatively well approximated by the length of the drill-string, thereby providing a curvilinear abscissa (s) for the measurement.

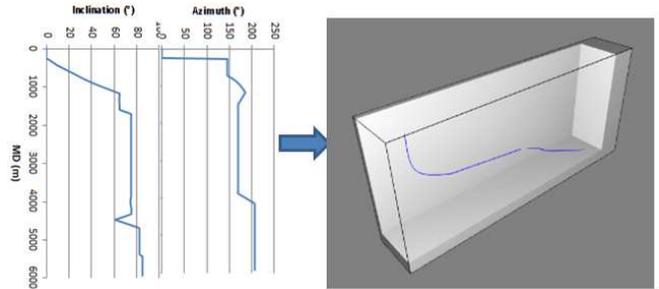


Fig. 21: An example of a well trajectory generated from measured depth, inclination and azimuth survey stations.

In order to calculate the position, some assumptions must also be made about the curvature between two measurement points (survey stations). The industry standard method assumes a circular arc. The result, the so-called minimum curvature method (Sawaryn and Thorogood, 2005), is used to calculate the position of a well by recursive integration of the following equations (see Fig. 21):

$$\begin{cases} x_{i+1} = x_i + r_f(s_{i+1} - s_i)(\sin I_i \cos A_i + \sin I_{i+1} \cos A_{i+1}) \\ y_{i+1} = y_i + r_f(s_{i+1} - s_i)(\sin I_i \sin A_i + \sin I_{i+1} \sin A_{i+1}) \\ z_{i+1} = z_i + r_f(s_{i+1} - s_i)(\cos I_i + \cos I_{i+1}) \end{cases} \quad (21)$$

where i is the index of the previous survey station, $i + 1$ is the index of the current survey station, x is the coordinate in

the north direction, y is the coordinate the east direction, z is the coordinate in the downward vertical direction, $r_f = \frac{\tan^{D_l/2}}{D_l} \forall D_l \neq 0, r_f = 0.5$ if $D_l = 0$ and $D_l = \cos^{-1}(\cos(I_{i+1} - I_i) - (1 - \cos(A_{i+1} - A_i)) \sin I_{i+1} \sin I_i)$.

4.2 Wellbore Position Uncertainty Models

The measurement of the inclination, azimuth and measured depth may be slightly biased for each survey station. If that bias is systematic, then the errors accumulate and, after integration, result in a potentially large discrepancy between the estimated borehole position and the true position of the well. The importance of systematic errors in the calculation of wellbore trajectories has been reported and analysed by Wolff and de Wardt (1981). These authors recognized six sources causing systematic errors on the measurement of inclination, azimuth and measured depth. The authors also developed a method for calculating the region of space that, with a given probability, will contain a survey station. This region of space is defined by a 3x3 covariance matrix and corresponds to an ellipsoid (see Fig. 22):

$$\Sigma = \begin{bmatrix} \text{var}(x,x) & \text{cov}(x,y) & \text{cov}(x,z) \\ \text{cov}(x,y) & \text{var}(y,y) & \text{cov}(y,z) \\ \text{cov}(x,z) & \text{cov}(y,z) & \text{var}(z,z) \end{bmatrix}, \quad (22)$$

Further analysis of the problem of wellbore position uncertainty estimations has led to more advanced models. One dedicated to magnetic measurements is described by Williamson (2000), which is now considered as today's industry standard. An extension to this model to encompass gyroscopic measurements has been described by Torkildsen *et al.* (2008). Note that the new formalism still uses a covariance matrix to describe the wellbore position uncertainty at a given survey station.

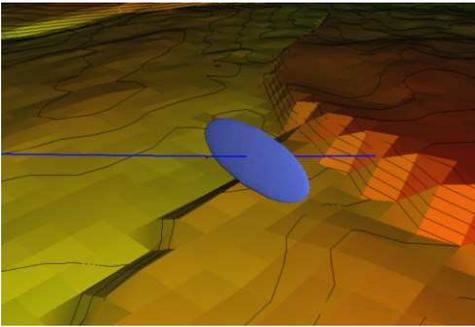


Fig. 22: Ellipsoid of uncertainty at certain depth of a deviated well.

Based on these calculation methods of the wellbore position uncertainty, we will now estimate the most probable trajectories that pass furthest away, in the vertical direction, to the depth at which a downhole pressure measurement is taken.

4.3 Trajectory Estimations at Shallowest and Deepest Vertical Depths

The prediction interval for the multivariate normal distribution yields a region defined by:

$$\Delta \vec{r}^T \cdot \Sigma^{-1} \cdot \Delta \vec{r} \leq \chi_3^2(p), \quad (23)$$

where Σ^{-1} represents the inverse covariance matrix, $\chi_3^2(p)$ is the quantile function for probability p of the chi-squared distribution with 3 degrees of freedom (see Fig. 23), $\Delta \vec{r} = \vec{r} - \vec{r}_0$ with \vec{r}_0 being the mean position of the well, and \vec{r} the borehole position when accounting for the uncertainty. This is the definition of an ellipsoid in a 3 dimensional space.

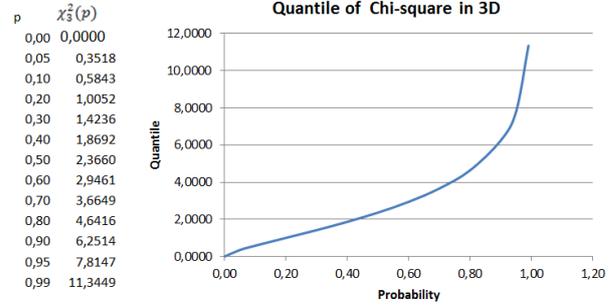


Fig. 23: Quantile function for probability p of the chi-squared function with 3 degrees of freedom.

So if we choose $\chi_3^2(p) = 1$ then we obtain an ellipsoid that represents the region with about a 20% chance of containing the actual trajectory. We need to take $\chi_3^2(p) = 7.8147$ to calculate the ellipsoid that has a 95% chance of contain the real position of the well. Similarly, to obtain a 99% confidence of containing the trajectory, we need to calculate the ellipsoid based on $\chi_3^2(p) = 11.3449$.

For simplicity, the elements of the inverse covariance matrix will be denoted as $H_{ij}, \forall i, j \in \{1,2,3\}$. The covariance matrix is a symmetric matrix and therefore $\Sigma = \Sigma^T$. This property implies that the inverse matrix will also be symmetric. In practice this gives, $H_{12} = H_{21}, H_{13} = H_{31}$ and $H_{23} = H_{32}$. This property was utilized to derive the Cartesian equation of the ellipsoid:

$$f(x, y, z) = H_{11}x^2 + H_{22}y^2 + H_{33}z^2 + 2H_{12}xy + 2H_{13}xz + 2H_{23}yz = \chi_3^2(p), \quad (24)$$

The maximum and minimum points in the z direction are characterized by a normal to the surface of the ellipsoid that is oriented in the z -direction and is described by:

$$\begin{cases} \frac{\partial f}{\partial x} = 0 \\ \frac{\partial f}{\partial y} = 0 \\ \frac{\partial f}{\partial z} = \alpha \end{cases} \quad (25)$$

where α is a scalar value. This can be re-written as:

$$\begin{cases} H_{11}x + H_{12}y + H_{13}z = 0 \\ H_{22}y + H_{12}x + H_{23}z = 0 \\ H_{13}x + H_{23}y + H_{33}z = \alpha \end{cases} \quad (26)$$

By solving this set of three equations, expressions for the x, y and z coordinates are given as follows:

$$\begin{cases} x = \frac{\alpha(H_{12}H_{23} - H_{13}H_{22})}{\Delta} \\ y = -\frac{\alpha(H_{11}H_{23} - H_{12}H_{13})}{\Delta} \\ z = \frac{\alpha(H_{11}H_{22} - H_{12}^2)}{\Delta} \end{cases} \quad (27)$$

where:

$$\Delta = H_{33}(H_{11}H_{22} - H_{12}^2) - H_{11}H_{23}^2 + 2H_{12}H_{13}H_{23} - H_{13}^2H_{22} \quad (28)$$

By inserting the three expressions for x , y and z in the Cartesian equation of the ellipsoid, an expression for α can be derived:

$$\alpha = \pm \sqrt{\frac{\chi_3^2(\varphi)(H_{11}H_{22}H_{33} - H_{12}^2H_{33} - H_{11}H_{23}^2 + 2H_{12}H_{13}H_{23} - H_{13}^2H_{22})}{H_{11}H_{22} - H_{12}^2}} \quad (29)$$

Due to α having a positive and a negative value, two solutions $\vec{r}_h = (x_h, y_h, z_h)$ and $\vec{r}_l = (x_l, y_l, z_l)$ are obtained from this set of equations. These are the points of interest, denoted \vec{r}_h for the highest point and \vec{r}_l for the lowest point.

The size and direction of an ellipsoid at a given depth of the trajectory are the results of the accumulation of the systematic errors all along the trajectory prior to that depth. When using a single survey instrument, those systematic errors are unknown but do not change through the integration.

For instance, if the magnitude of the systematic error in the azimuth measurement is much larger than those for inclination and measured depth, the projection of the ellipsoid of uncertainties on the horizontal plane is laterally elongated. However, if the borehole changes direction substantially, the ellipsoid continues to carry the lateral uncertainty from before the change of direction and will stay mostly in the same orientation (Brooks and Wilson 1996) even though the borehole changes direction (see **Fig. 24**). Of course, this principle applies in all directions.

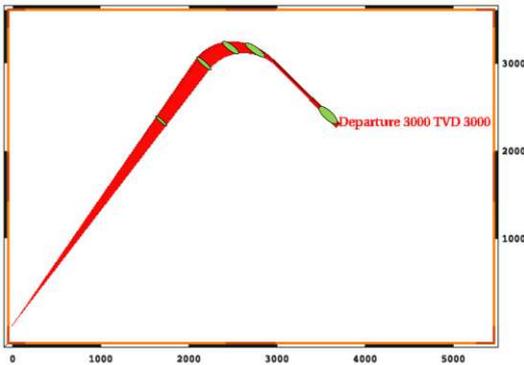


Fig. 24: The ellipsoids of uncertainty never decrease in size, but they accumulate the systematic error from the start of measurement series.

We can therefore consider that the location at the surface of the ellipsoid is a characteristic of the effect of the systematic errors on the inclination, azimuth and measured depth. By using the same relative position at each ellipsoid in the measurement series, we can reconstruct a trajectory that bears the particular effect of those systematic errors. To find

that characteristic location, we will use the parametric coordinates of a point at the surface of the ellipsoid.

Let us consider the parametric equations of an ellipsoid in a coordinate system (X, Y, Z) that is oriented by the three axes (a, b, c) of the ellipsoid:

$$\forall \phi \in [0, \pi[, \forall \theta \in [0, 2\pi[, \begin{cases} X = a \sin \phi \cos \theta \\ Y = b \sin \phi \sin \theta \\ Z = c \cos \phi \end{cases} \quad (30)$$

The angles ϕ and θ characterize the position of a point at the surface of that ellipsoid and are specific to the systematic errors that cause the trajectory to end up at that position of the ellipsoid. Using the same parameters ϕ and θ at each ellipsoid along the wellbore, we can re-construct the trajectory that embeds the same systematic errors all along the measurements made in that wellbore.

To obtain the coordinate system oriented by the axes of the ellipsoid, we need to diagonalize the covariance matrix and find the transfer matrix that transforms the covariance matrix into its diagonalized version:

$$\Sigma = PDP^{-1}, \quad (31)$$

where P is the transfer matrix and D is the diagonal matrix. Note that the transfer matrix is in fact the multiplication of three rotations necessary to transform the global Cartesian system into a local Cartesian system defined by the axes of the ellipsoid. The following equation shows how to transform a point from the global coordinate system into the local coordinate system attached to the axes of the ellipsoid:

$$P \times \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad (32)$$

The diagonal values of the matrix D are actually the eigenvalues of the covariance matrix. They can be found by solving the equation:

$$|\Sigma - \lambda I| = 0, \quad (33)$$

where I is the identity matrix.

So to reconstruct a trajectory that corresponds to one of the extreme vertical depths associated with a particular downhole pressure measurement, one shall apply the following procedure (see **Fig. 25**):

- Calculate the global coordinates of the extreme vertical point at the depth of interest using eq. 29 and 27
- Transform that point into the local coordinate system of the ellipsoid using the eq. 32
- Calculate the corresponding ellipsoid parameters (ϕ and θ) using eq. 30
- Apply the same ellipsoid parameters (ϕ and θ) at all other survey stations of the wellbore and convert the local coordinates into the global coordinates.

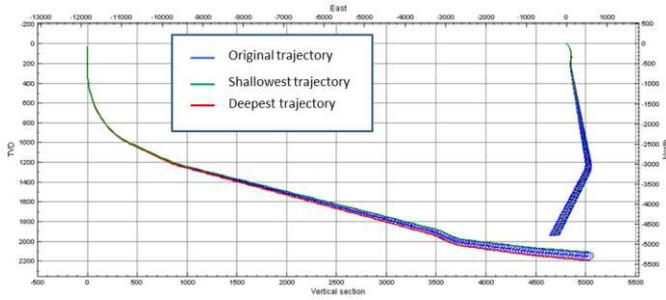


Fig. 25: Reconstruction of shallowest and deepest trajectories

5. COMBINATION OF UNCERTAINTIES AND EFFECTS ON DOWNHOLE PRESSURE ESTIMATION

As seen in the above sections, lack of information on the characteristics of the drilling fluid (including the side effects of contaminants such as cuttings or formation fluids), the formation rock thermal properties, the real position of the wellbore and the actual borehole size (due to cuttings beds and hole enlargements) are sources of uncertainty in the evaluation of downhole pressures when utilizing hydraulic models. It is possible to estimate the most probable hydrostatic and circulating pressure with their associated standard deviations by performing stochastic simulations on the various parameters that are sources of uncertainty. The statistical variations shall be adjusted to account for additional knowledge that may restrain their domain of variability. The pressure estimates and standard deviations can help in determining drilling parameters that ensure a safe drilling operation.

For the examples of section 6, we have used the following variations:

- Wellbore position uncertainty (Gaussian distribution)
- Specific heat capacity of formation rock (Uniform distribution)
- Thermal conductivity of formation rock (uniform distribution)
- Specific heat capacity of base oil (Gaussian distribution)
- Thermal conductivity of base oil (Gaussian distribution)
- Specific heat capacity of brine (Gaussian distribution)
- Thermal conductivity of brine (Gaussian distribution)
- Low gravity solid specific heat capacity (Gaussian distribution)
- Low gravity solid thermal conductivity (Gaussian distribution)
- High gravity solid specific heat capacity (Gaussian distribution)

- High gravity solid thermal conductivity (Gaussian distribution)
- Cuttings specific heat capacity (Gaussian distribution)
- Cuttings thermal conductivity (Gaussian distribution)
- Factor controlling the heat generation from the bit (Gaussian distribution)
- Factor controlling the heat generation from mechanical friction (Gaussian distribution)
- Factor controlling the heat generation from turbulence (Gaussian distribution)
- Cuttings proportion (Uniform distribution)
- Cuttings density (Uniform distribution)
- Cuttings position in annulus (Uniform distribution)
- Cuttings bed position, length and height in annulus (Gaussian distribution)
- Cuttings concentration around the BHA (Uniform distribution)
- Hole collapse position, length and size (Gaussian distribution)
- High gravity solid density (Gaussian)
- Low gravity solid density (Gaussian)
- Type of PVT for base oil (Uniform or fixed if known)
- Type of PVT for brine (Uniform or fixed if known)
- Oil/water ratio (Gaussian if known or Uniform otherwise)
- Salinity (Gaussian if known or uniform otherwise)
- Low gravity solid proportion (Gaussian if known or uniform otherwise)
- Mud pump efficiency (Gaussian)
- Surface roughness (Gaussian)

If downhole measurements are available, it is possible to further restrain the estimations. From all the downhole pressure solutions generated using the stochastically varied parameters of the problem, we keep only those that respect the measurement reading within its tolerance. An envelope of those depth-based downhole pressure curves can be generated to characterize the remaining possible variations of the downhole pressure at depths distant from the measurement. It should be noted that there exist many possible ways to reach the downhole pressure measurement and therefore the variability of the estimated downhole pressure may be larger

at depths farther from the measurement than the one at the measurement depth (see Fig. 26).

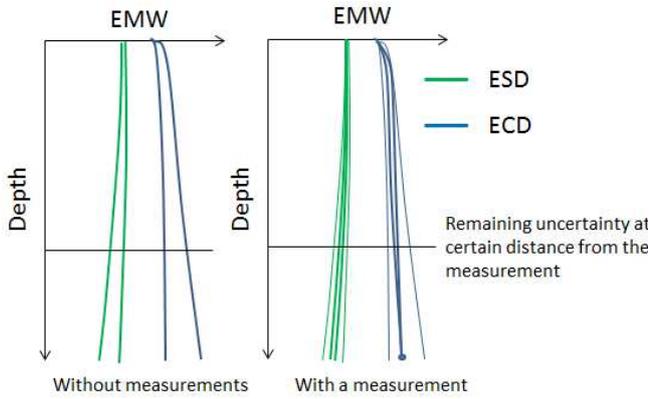


Fig. 26: Variability of the downhole pressure estimations with and without downhole pressure measurements.

6. EXAMPLES

In this section we will analyse how the reviewed source of errors during downhole pressure estimation can influence the results of calculations for two different cases.

5.1 8 1/2 × 9 1/2 in. Section of a Long Horizontal Well

In this first example, we will consider a well (from now on referred as Well A) with a total length of 5934 m and a final TVD of 2151 m (see Fig. 25). The wellbore architecture is given in Table 1.

Table 1: wellbore architecture of well A

Section	MD at sect. TD [m]	TVD at sect. TD [m]	Mud Weight [s.g.]	Incl. at sect. TD [°]	Az. at sect. TD [°]
32" (26" conductor)	396	395,2	1,03 – 1,07 WBM	7,5	146
17 1/2" (13 3/8" casing)	1561	1175	1,06 – 1,18 WBM	65,0	169,57
12 1/4" (9 5/8" casing)	3384	1654	1,60 – 1,62 OBM	75,27	169,57
8 1/2" x 9 1/2" (7 5/8" liner)	4150	1851	1,32 – 1,34 OBM	75,41	207
6 1/2" Open-hole	5934	2151	1,12 – 1,14 OBM	85	207

Well A has a long horizontal departure which incorporates a horizontal turn, changing the azimuth from 169° to 207° and reaching an inclination of 85° at TD.

Fig. 27 shows the geo-pressure margins and the geo-thermal gradient for this well.

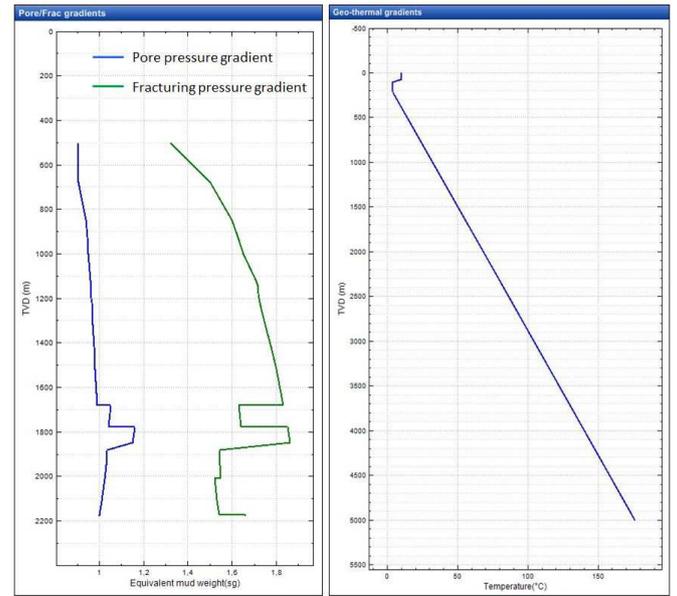


Fig. 27: Geo-pressure window and geo-thermal gradient for well A.

We will in particular focus on the 8 1/2 × 9 1/2 in. section of this well. A synthetic based mud of nominal density 1.32 sg at 22°C with an oil/water ratio of 75/25 has been used for the drilling of that section. The brine is based on CaCl₂ at a weight fraction of 13.5%. The volume fraction of low gravity solid to high gravity solid is 35% (see Fig. 28 for a description of the PVT behaviour and the rheology of that mud).

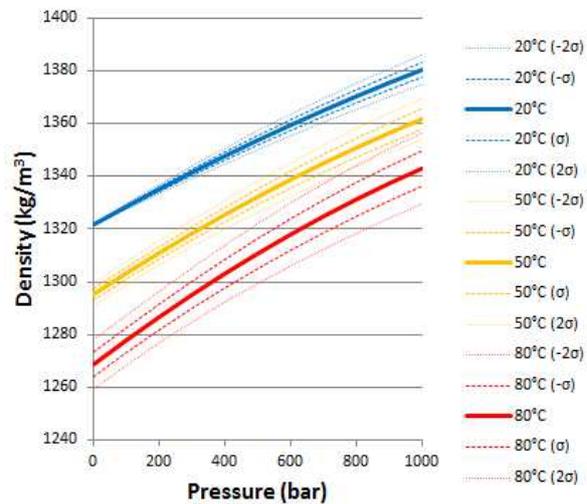


Fig. 28: PVT of the mud used in well A.

The estimated specific heat capacity, accounting for the uncertainties of the mud component characteristics, is 1265 J/kg.K with a standard deviation of 148 J/kg.K. Similarly the estimated thermal conductivity is 0.318 W/m.K with a standard deviation of 0.067 W/m.K.

For rate of penetrations ranging from 5 m/h to 40 m/h and with flow-rate in between 2000 l/min and 2500 l/min, the cuttings mass concentration during transport varies from 0.3% to 2.8%. When accounting for the presence of cuttings,

the PVT of the mud shifts upwards and the interval of confidence increases. This can be seen by comparing the PVT curves of the drilling fluid in Fig. 28 with those for the fluid with cuttings to the right in Fig. 29. The effect of the presence of cuttings on the thermal properties is as follow:

- specific heat capacity: 1283 J/kg.K with a standard deviation of 152 J/kg.K
- thermal conductivity: 0.322 W/m.K with a standard deviation of 0.068 W/m.K

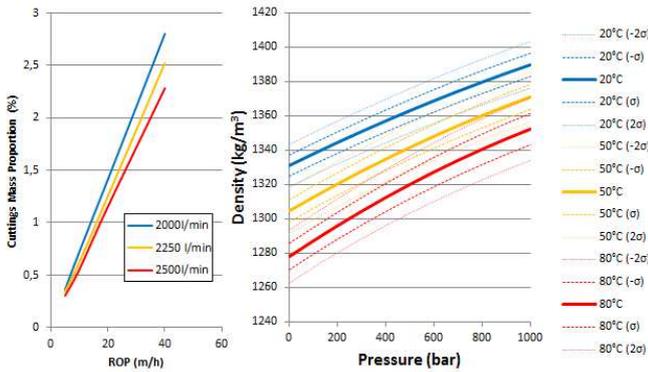


Fig. 29: Impact of the presence of cuttings on the PVT of the drilling fluid used in well A.

Stochastic variations of the thermo-physical parameters used for heat transfer calculations provide possible variations of the annulus temperatures as show in Fig. 30.

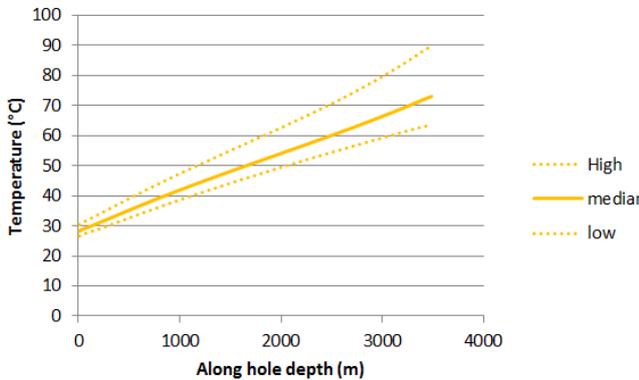


Fig. 30: Possible range of variations of the annulus temperature for well A.

Using all the possible variations on the uncertain parameters of this particular case, it is possible to estimate the hydrostatic pressure at the bit depth and its standard deviation. For the well A at the bit depth 4100 m MD, the estimated hydrostatic pressure is 240 bar with a standard deviation of 3.7 bar, *i.e.* the 95% confidence interval for the hydrostatic pressure at that depth is [232.6 bar, 247.4 bar], or in equivalent mud weight [1.305 sg, 1.318 sg] and the corresponding ESD interval at the casing shoe is [1.309 sg, 1.363 sg] (see Fig. 31).

When circulating at 2000 l/min, the estimated downhole pressure at the bit is 263 bar with a standard deviation of 4.1 bar. Note that the standard deviation of the estimated pressure while circulating is not much increased compared to the

hydrostatic conditions. This is due to the relatively large open hole size (9 ½ in) that limits the impact of pressure losses in the annulus. In equivalent mud weight, the 95% confidence interval for the ECD at bit is [1.405 sg, 1.478 sg] and at the casing shoe it is [1.410 sg, 1.489 sg] (see Fig. 32).

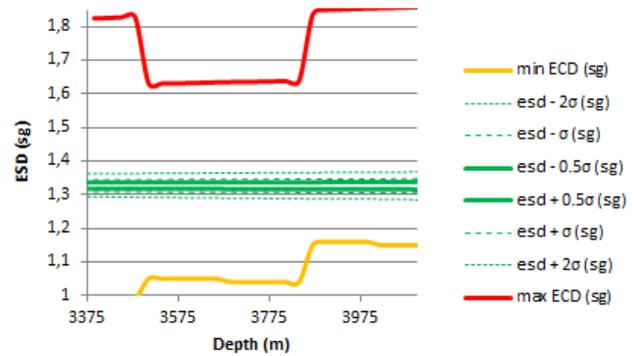


Fig. 31: Equivalent static density plot in the open hole section of well A when there are no constraints on the downhole pressure estimation at the bit depth.

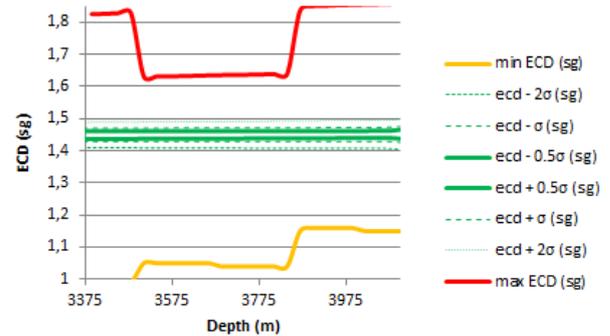


Fig. 32: Equivalent circulating density plot in the open hole section of well A when there are no constraints on the downhole pressure estimation at the bit depth.

If we have a downhole pressure reading close to the bit depth of 262 bar \pm 0.5 bar, we can constrain the stochastic simulations with this particular information. The 95% confidence intervals for the ECD at the casing shoe reduces to [1.452 sg, 1.469 sg] (see Fig. 33) and the ESD confidence interval reduces as well (see Fig. 34).

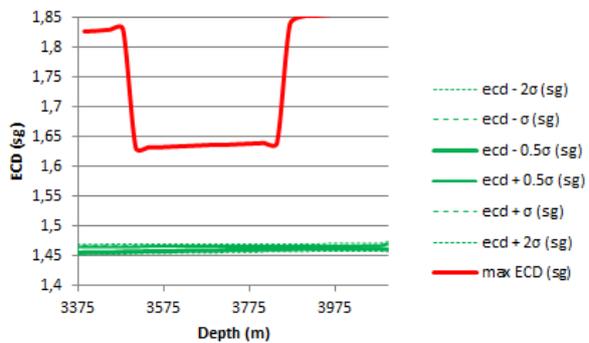


Fig. 33: Equivalent circulating density plot in the open hole section of well A with a downhole pressure constraint at the bit depth.

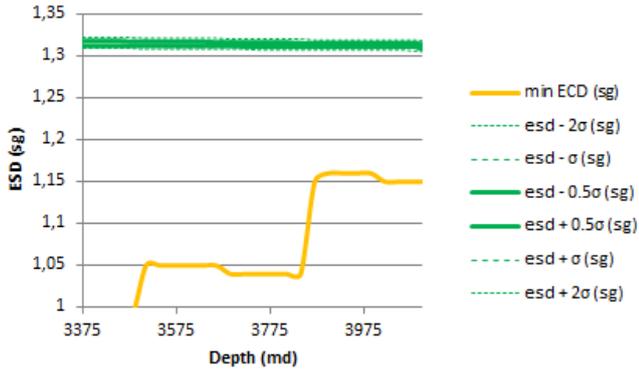


Fig. 34: Equivalent static density plot in the open hole section of well A with a downhole pressure constraint at the bit depth.

5.2 8 ½ in. Section of an Ultra ERD Well

In this second case, we will consider a well (referred to as well B) with a total length of 8467 m reaching a final TVD of 4146 m (see Fig. 35).

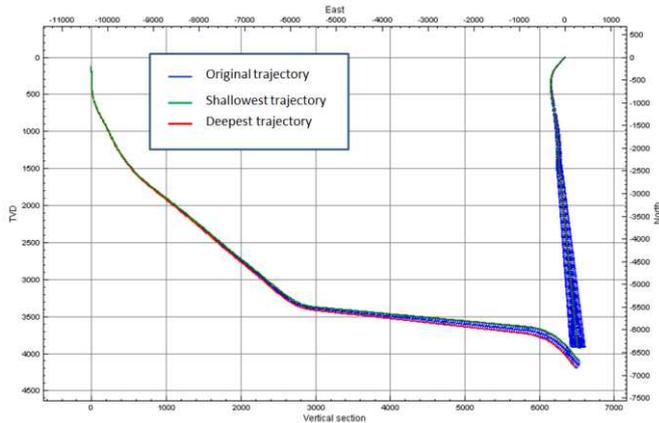


Fig. 35: Trajectory and wellbore position uncertainty of well B.

The wellbore architecture is given in Table 2. Due to its length and vertical depth, this well is particularly challenging in terms of pressure losses inside the drill-string. Hence large drill-pipes are necessary to ensure that an acceptable flow-rate can be used without exceeding the maximum pump pressure capability. As a result of using heavy drill-pipes with large connections, the mechanical forces (tension and torque) are also pushing the top-drive to its limit.

Table 2: Wellbore architecture of well B

Section	MD at sect. TD [m]	TVD at sect. TD [m]	Mud Weight [s.g.]	Incl. at sect. TD [°]	Az. At sect. TD [°]
32" (27" conductor)	309	309		2,7	115,6
24" (18% casing)	1192	1130	1,03 – 1,30 WBM	25,4	220,5
17½" (13% casing)	4277	3200	1,65 OBM	51,9	174,6
12¼" (9% casing)	6500	3571	1,56-1,60 OBM	84,1	174,3
8½" (7" liner)	8250	3980	1,60-1,70 OBM	42,6	174,3
6" (4½" liner)	8467	4146	1,5-1,62 OBM	40	174,3

In terms of geo-pressures, the end of the 8 ½ in. section (3900 m TVD) is the most challenging because of a zone with a high collapse pressure gradient (see Fig. 36).

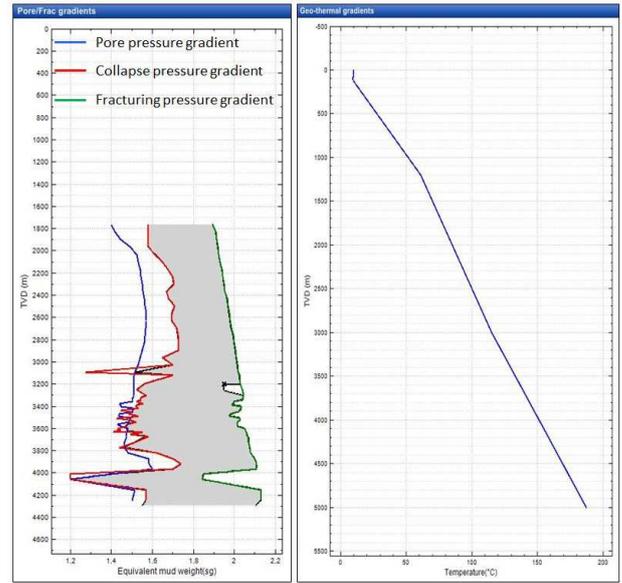


Fig. 36: Geo-pressure window and geo-thermal gradient for well B.

For this example, we will analyse the 8 ½ in. section. The drilling fluid used for that section is a low viscosity SBM of nominal density of 1.64 sg at 45°C and an oil/water ratio of 84/16. The base oil is of type “synthetic #1” (ref. Fig. 9). We do not know which types of salts are used in the brine but the volume ratio of low gravity to high gravity solids is 2.8. (see Fig. 37).

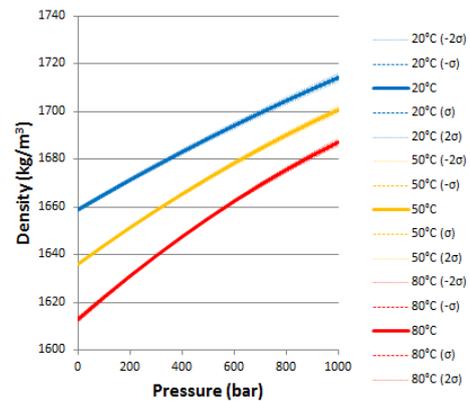


Fig. 37: PVT of the mud used in well B.

The estimated specific heat capacity is 1117 J/kg.K with a standard deviation of 134 J/kg.K and the estimated thermal conductivity is 0.320 W/m.K with a standard deviation of 0.061 W/m.K.

For rates of penetration ranging from 5 m/h to 40 m/h and with flow-rates between 1750 l/min and 2250 l/min, the cuttings mass concentration during transport varies from 0.2% to 2.2%. When accounting for the presence of cuttings, the PVT of the mud shifts upwards and the interval of confidence increases (see Fig. 38), as was also the case for

the fluid used in well A. The effects of the presence of cuttings on the thermal properties are now as follow:

- specific heat capacity: 1128 J/kg.K with a standard deviation of 134 J/kg.K
- thermal conductivity: 0.325 W/m.K with a standard deviation of 0.063 W/m.K

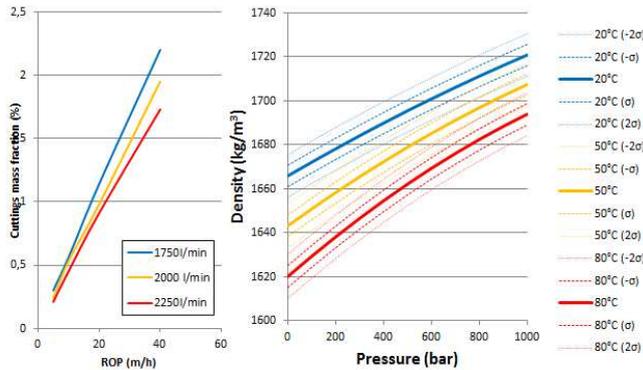


Fig. 38: Impact of the presence of cuttings on the PVT of the drilling fluid used in well A.

The influence of the uncertainty on the thermo-physical properties results in the temperature variations shown on Fig. 39.

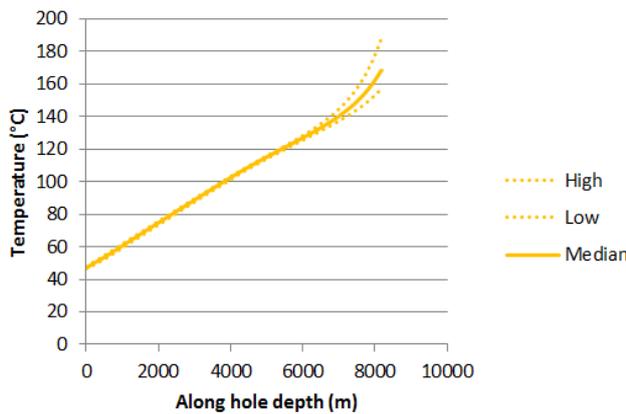


Fig. 39: Possible range of variations of the annulus temperature for well B.

Similar stochastic simulations as those performed for the previous example give the following results when the bit is at 7650 m MD:

- Estimated hydrostatic pressure at bit: 601.5 bar with a standard deviation of 8.9 bar.
- 95% confidence interval for the ESD at bit without downhole pressure constraints: [1.611 sg, 1.710 sg].
- 95% confidence interval for the ESD at the casing shoe without downhole pressure constraints: [1.621 sg, 1.703 sg] (see Fig. 40).
- When circulating at 1500 l/min, the pressure at bit is 711 bar with a standard deviation of 34.5 bar.

- 95% confidence interval for the ECD at bit without downhole pressure constraints: [1.817 sg, 2.154 sg]
- 95% confidence interval for the ECD at the casing shoe when circulating without downhole pressure constraints: [1.718 sg, 2.072 sg] (see Fig. 41).
- 95% confidence interval for the ECD at the casing shoe when circulating with a downhole pressure constraints of 715 bar \pm 1 bar: [1.928 sg, 1.940 sg] (see Fig. 42).
- 95% confidence interval for the ESD at the bit depth using a circulating downhole pressure of 730 bar \pm 1 bar: [1.657 sg, 1.665 sg] (see Fig. 43).

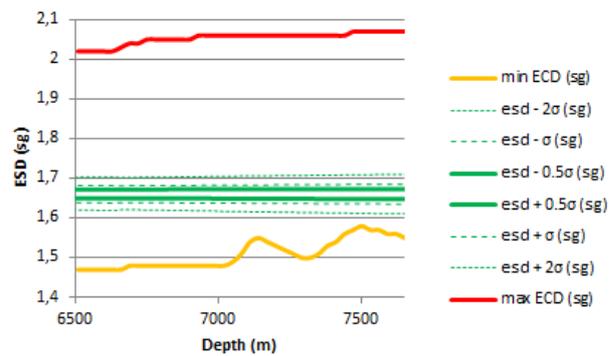


Fig. 40: Equivalent static density plot in the open hole section of well B when there are no constraints on the downhole pressure estimation at the bit depth.

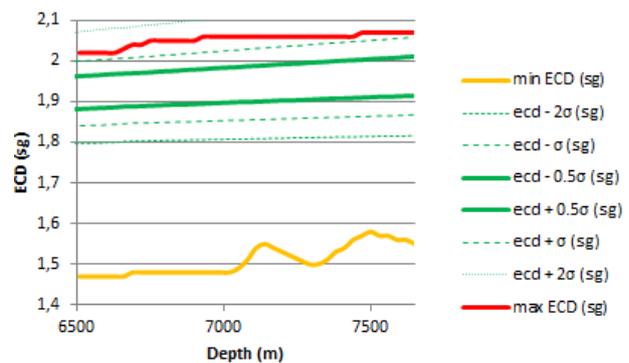


Fig. 41: Equivalent circulating density plot in the open hole section of well B when there are no constraints on the downhole pressure estimation at the bit depth.

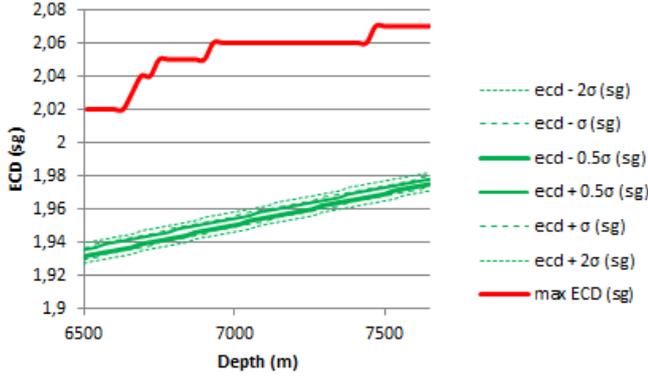


Fig. 42: Equivalent circulating density plot in the open hole section of well B with a downhole pressure constraint at the bit depth.

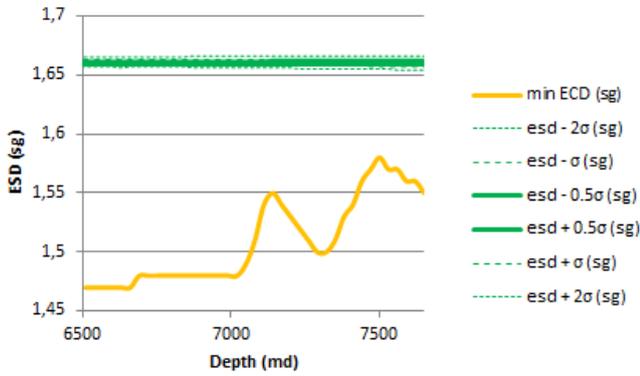


Fig. 43: Equivalent static density plot in the open hole section of well B with a downhole pressure constraint at the bit depth.

7. CONCLUSIONS

Lack of detailed information about the drilling fluid characteristics or the thermo-physical properties of the formation rocks, combined with inherent well bore position uncertainty and uncertain evaluations on the cuttings transport or the eventual presence of small quantity of gas, can be the source of substantial uncertainty in the estimation of downhole pressures along the whole length of the open hole section. This paper has presented how systematic analysis of the possible sources of uncertainties can be used to quantify the uncertainty on the downhole pressure estimations. Additional information such as downhole pressure measurements can help reduce these uncertainties.

Through two example cases, we have shown that the downhole pressure inaccuracy can add up to relatively large values. When drilling with small geo-pressure windows, it is therefore of paramount importance to conduct this type of analysis to ensure that the drilling operation can be performed in a safe manner.

This systematic approach opens the way to more extensive analyses of the minimum requirements for the number of pressure sensors and their accuracy in order to obtain a trustworthy estimation of downhole pressures at different places along the wellbore.

8. ACKNOWLEDGEMENTS

The authors acknowledge the Research Council of Norway, ConocoPhillips, Det norske oljeselskap, Statoil, Talisman, TOTAL and Wintershall for financing the work through the research centre SBBU (Centre for drilling and wells for improved recovery) at IRIS.

ABBREVIATIONS

BHA	Bottom Hole Assembly
ECD	Equivalent Circulating Density
EMW	Equivalent Mud Weight
ERD	Extended Reach Drilling
ESD	Equivalent Static Density
LOT	Leak-Off Test
MD	Measured Depth
MSL	Mean Sea Level
MWD	Measurement While Drilling
OBM	Oil-Based Mud
PDM	Positive Displacement Motor
PVT	Pressure Volume Temperature
SBM	Synthetic-Based Mud
TD	Total Depth
TVD	True Vertical Depth
WBM	Water-Based Mud
XLOT	eXtended Leak-Off Test

NOMENCLATURE

A	cross-sectional area [L^2](m^2)
f_s	solid volume fraction
f_{s2}	solid volume fraction at reference conditions
g	gravitational acceleration [LT^{-2}] (m/s^2)
H_{ij}	element of the inverse covariance matrix [L^{-2}](m^{-2})
h	vertical depth [L](m)
I	wellbore inclination (radian)
p	pressure [$ML^{-1}T^{-2}$] (pa)
p_b	breakdown pressure [$ML^{-1}T^{-2}$] (pa)
$p_{fn}^{1T^{-2}}$	normal pore pressure based on water gradient [$ML^{-1}T^{-2}$] (pa)
p_0	pressure at the start of the borehole [$ML^{-1}T^{-2}$] (pa)
p_{atm}	atmospheric pressure [$ML^{-1}T^{-2}$] (pa)
p_o	pore pressure [$ML^{-1}T^{-2}$] (pa)

p_w	wellbore pressure [ML ⁻¹ T ⁻²] (pa)	ρ_w	water density [ML ⁻³](kg/m ³)
Q	volumetric flow-rate [M ³ T ⁻¹](m ³ /s)	ρ_{w1}	water density at reference conditions [ML ⁻³](kg/m ³)
\vec{r}	position of a point on an ellipsoid [L] (m)	Σ	covariance matrix [L ²](m ²)
\vec{r}_h (m)	highest point in the z-direction on an ellipsoid [L]	σ_h	minimum horizontal stress [ML ⁻¹ T ⁻²] (pa)
\vec{r}_l (m)	lowest point in the z-direction on an ellipsoid [L]	σ_H	maximum horizontal stress [ML ⁻¹ T ⁻²] (pa)
\vec{r}_0	centre of an ellipsoid [L] (m)	σ_v	lithostatic stress [ML ⁻¹ T ⁻²] (pa)
s	curvilinear abscissa [L] (m)	σ_1	maximum principal stress [ML ⁻¹ T ⁻²] (pa)
S_0 (pa)	formation rock inherent shear strength [ML ⁻¹ T ⁻²]	σ_2	intermediate principal stress [ML ⁻¹ T ⁻²] (pa)
T_0	formation tensile strength [ML ⁻¹ T ⁻²] (pa)	σ_3	minimum principal stress [ML ⁻¹ T ⁻²] (pa)
T	temperature (K)	$\sigma_{\rho\rho}$	stress in the radial direction [ML ⁻¹ T ⁻²] (pa)
t	time [T](s)	σ_{ss}	stress in the axial direction [ML ⁻¹ T ⁻²] (pa)
V_{lgs}	volume of low gravity solid [L ³](m ³)	$\sigma_{\theta s}$	shear stress [ML ⁻¹ T ⁻²] (pa)
V_{hgs}	volume of high gravity solid [L ³](m ³)	τ	shear stress [ML ⁻¹ T ⁻²] (pa)
V_{st}	pump stroke volume [L ³](m ³)	ν	Poisson's ratio
v	bulk fluid velocity [LT ⁻¹](m/s)	Φ	porosity (fraction)
w_t	salt weight fraction	Φ_0	initial porosity (fraction)
x	north coordinate [L](m)	φ	formation rock internal friction angle (radian)
y	east coordinate [L](m)	χ_3^2	quantile function for probability p of the chi-squared distribution with 3 degrees of freedom
z	vertical depth [L] (m)		

Greek letters:

α	Biot's effective stress coefficient
$\dot{\gamma}$	shear rate [T ⁻¹](s ⁻¹)
η	pump efficiency
K	oil water ratio
K_1	oil water ratio at reference conditions
Λ	solid volume ratio
λ	rock compressibility
μ	effective viscosity [ML ⁻¹ T ⁻¹] (pa.s)
ρ_b	bulk density of the rock [ML ⁻³](kg/m ³)
ρ_{emw}	equivalent mud weight [ML ⁻³](kg/m ³)
ρ_f	density of the pore fluid [ML ⁻³](kg/m ³)
ρ_l	density of liquid mix [ML ⁻³](kg/m ³)
ρ_{l2} [ML ⁻³](kg/m ³)	density of liquid mix at reference conditions [ML ⁻³](kg/m ³)
ρ_m	density of the drilling fluid [ML ⁻³](kg/m ³)
ρ_o	oil density [ML ⁻³](kg/m ³)
ρ_{o1}	oil density at reference conditions [ML ⁻³](kg/m ³)

REFERENCES

- API Recommended Practice 13D (2006). Rheology and Hydraulics of Oil-well Drilling Fluids. Fifth Edition, June 2006
- Breckels, I.M., van Eekelen, H.A.M. (1982). Relationship Between Horizontal Stress and Depth in Sedimentary Basins. *Journal of Petroleum Technology*, vol. 34, No 9, pp. 2191-2199, Sep. 1982.
- Brooks, A., Wilson, H. (1996). An Improved Method for Computing Wellbore Position Uncertainty and its Application to Collision and Target Intersection Probability Analysis. *SPE European Petroleum Conference*, Milan, Italy, Oct. 22-24, 1996.
- Cayeux, E., Daireaux, B., Dvergsnes, E.W., Florence, F. (2013a). Toward Drilling Automation: On the Necessity of Using Sensors That Relate to Physical Models. *SPE/IADC Drilling Conference and Exhibition*, Amsterdam, The Netherlands, 5-7 March, 2013.
- Cayeux, E., Mesagan, T., Tanripada, S., Zidan, M., Fjelde, K.K. (2013b). Real-Time Evaluation of Hole Cleaning Using a Transient Cuttings Transport Model. *SPE/IADC Drilling Conference and Exhibition*, Amsterdam, The Netherlands, 5-7 March, 2013.
- Charlez, P. A. (1991) Rock mechanics (II: petroleum applications). *Technip, Paris* (1991).
- Coley, C.J., Edwards, S.T. (2013). The Use of Along String Annular Pressure Measurements to Monitor Solids Transport and Hole Cleaning. *SPE/IADC Drilling*

- Conference and Exhibition*, Amsterdam, The Netherlands, Mar 5-7, 2013.
- Colmenares, L.B., Zoback, M.D. (2001). Statistical Evaluation of six Rock Failure Criteria Constrained by Polyaxial Test Data. US Symposium on Rock Mechanics, Washington D.C., USA, Jul 7-10, 2001.
- Corre B., Eymard R., Guenot A. (1984). Numerical Computation of Temperature Distribution in a Wellbore While Drilling. *SPE Annual Technical Conference and Exhibition*, Houston, Texas, USA, 16-19 September 1984.
- Ewy, R.T. (1999). Wellbore-Stability Predictions by Use of a Modified Lade Criterion. *SPE Drilling Completion Journal*, vol. 14, No 2, pp. 85-91, June 1999.
- Fjelde K.K., Rommetveit R., Merlo A., Lage A. (2003). Improvements in Dynamic Modeling of Underbalanced Drilling. *SPE/IADC Underbalanced Technology Conference and Exhibition*, Houston, Texas, USA, March 25-26, 2003.
- Hansen, S.A., Sterri, N. (1995). Drill Pipe Rotation Effects on Frictional Pressure Losses in Slim Annuli. *SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA, 22-25 Oct., 1995.
- Houwen, O.H., Geehan T. (1986). Rheology of Oil-Based Muds. *SPE Annual Technical Conference*, New Orleans, Louisiana, USA, October 5-8, 1986.
- Isambourg, P., Anfinsen, B.T., Marken, C. (1996). Volumetric Behavior of Drilling Muds at High Pressure and High Temperature. *SPE European Petroleum Conference*, Milan, Italy, 22-24 Oct, 1996.
- Kemp, N.P., Thomas, D.C., Atkinson, G., Atkinson, B. (1989). Density Modeling for Brines as a Function of Composition, Temperature, and Pressure. *SPE Production Engineering*, vol. 4, No 4, pp. 394-400, Nov. 1989.
- Monteiro, E.N. (2005). Study of Methane Solubility in Organic Emulsions Applied to Drilling Fluid Formulation and Well Control. *SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA, 9-12 Oct., 2005.
- Nawrocki, P.A. (2010). Critical Wellbore Pressures Using Different Rock Failure Criteria. *ISRM Internationals Symposium 2010 and 6th Asian Rock Mechanics Symposium – Advances in Rock Engineering*, New Delhi, India, 23-27 Oct, 2010.
- O'Bryan, P.L., Bourgoyne, A.T., Monger, T.G., Kopcsó, D.P. (1998). An Experimental Study of Gas Solubility in Oil-Based Drilling Fluids. *SPE Drilling Engineering*, vol. 3, No 1, pp. 33-42, March 1988.
- Pilehvari, A., Serth, R. (2009). Generalized Hydraulic Calculation Method for Axial Flow of Non-Newtonian Fluids in Eccentric Annuli. *SPE Drilling & Completion Journal*, vol. 24, No 4, pp. 553-563, Dec. 2009.
- Robertson R.E., Stiff H.A. (1976). An Improved Mathematical Model for Relating Shear Stress to Shear Rate in Drilling Fluids. *SPE Journal*, February 1976, Volume 16 (1), pp. 31-36.
- Sawaryn, S.J., Thorogood, J.L., 2005. A Compendium of Directional Calculations Based on the Minimum Curvature Method. *SPE Drilling & Completion Journal*, vol. 20, No 1, pp. 24-36, Mar 2005.
- Schmitt, D.R., Zoback, M.D. (1989). Poroelastic effects in the determination of the maximum horizontal principal stress in hydraulic fracturing tests – A proposed breakdown equation employing a modified effective stress relation for tensile failure. *International Journal of Rock Mechanics and Mining Science*, vol. 26, pp. 499-506.
- Shafer, D.M., Loeppke, G.E., Glowka, D.A., Scott, D.D., Wright, E.K. (1992). An Evaluation of Flowmeters for the Detection of Kicks and Lost Circulation During Drilling. *IADC/SPE Drilling Conference*, New Orleans, Louisiana, USA, Feb 18-21, 1992.
- Toft, R.E. (2011). Influence of the Thermo-Physical Parameters of Formation Rock on Temperature Modelling, and Comparison With Recorded Data From a Drilling Operation. B. Sc. Thesis, University of Stavanger, 2011.
- Torkildsen, T., Håvardstein, S. Weston, J., Ekseth, R. (2008). Prediction of Wellbore Position Accuracy When Surveyed With Gyroscopic Tools. *SPE Drilling & Completion*, vol. 23, No 1, pp. 5-12, March 2008.
- Ward, C., Andreassen, E. (1998). Pressure-While-Drilling Data Improve Reservoir Drilling Performance. *SPE Drilling & Completion Journal*, vol. 13, No 1, pp. 19-24, March 1998.
- Williamson, H.S. (2000). Accuracy Prediction for Directional Measurement While Drilling. *SPE Drilling & Completion Journal*, vol. 15, No. 4, pp. 221-233, Dec. 2000.
- Wolff, C.J.M., de Wardt, J.P. (1981). Borehole Position Uncertainty- Analysis of Measuring Methods and Derivation of Systematic Error Model. *Journal of Petroleum Technology*, vol. 33, No. 12, pp. 2338-2350, Dec. 1981.
- Yi, X., Ong, S.H., Russel, J.E. (2005). Improving Borehole Stability Analysis by Quantifying the Effects of Intermediate Principal Stress Using Polyaxial Rock Strength Test Data. US Symposium on Rock Mechanics, Anchorage, Alaska, USA, June 25-29, 2005.
- Yu, M., Chen, G., Chenevert, M.E., Sharma, M.M. (2001). Chemical and Thermal Effects on Wellbore Stability of Shale Formations. *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, 30 Sep-3Oct, 2001.

Evaluation of various methods for estimating downhole pressure during drilling operations

Hans Petter Lohne^{1,3}, Gerhard Nygaard^{1,4}, Rune Time¹ and Michael Nikolaou²

¹University of Stavanger, Department of Petroleum Engineering, NO-4036 Stavanger, Norway

²University of Houston, Chemical Engineering, Houston, TX-77004, Texas

³International Research Institute of Stavanger, NO-4068 Stavanger, Norway

⁴Cybernetic Drilling Technologies AS, NO-5008, Bergen, Norway

Abstract:

The formation pressures in challenging areas such as deepwater and depleted reservoirs, requires an accurate determination of the downhole pressures. This paper presents several different approaches in determining these pressures while a drilling operation is ongoing.

The downhole pressure can be measured directly, using a downhole pressure sensor, and may be transmitted to the surface using mud-pulse telemetry or wired pipe technologies. The downhole pressure may also be obtained by using wellbore geometry information and surface measurements combined with flow modeling techniques. The accuracy, robustness and cost of the obtained dynamic downhole pressure vary depending on method selected.

This paper presents a comparison between the approaches based on accuracy, robustness and cost. The best method is not necessarily the most accurate one, but the one which satisfies its purpose at a reasonable cost.

1. INTRODUCTION

As the simpler and easier accessible reservoirs have been produced, the oil industry is moving towards drilling more challenging wells. Increasing recovery from older fields require drilling through depleted reservoirs, while new fields are often in deep water. Such wells have a smaller pressure margin between pore and fracturing pressures. Accurate information of the bottomhole pressure (BHP) is then crucial to manage to control the pressure in these wells. The drilling fluid is one of the important factors influencing the downhole pressure.

Today, the common practice during drilling operations is to identify the various drilling fluid parameters using a sample of the drilling fluid. The density is found by weighing a known volume at atmospheric pressure. A rheology value is typically determined by measuring the

time a mud sample takes to run out of the Marsh funnel. The dynamic viscosity of the sample would then be assessed by using a parameter table that correlates the measured flow duration with dynamic viscosity. More accurate rheology measurements would be found by using a standard Fann viscometer.

However, drilling fluids are typically non-Newtonian and other methods of defining the rheology should be found. Direct measurement of rheology of non-Newtonian drilling fluids based in drilling fluid samples is challenging. Another method that has been discussed in literature the recent years, is utilizing differential pressure transducers for providing rheology estimates. The method would benefit from the robustness, the reliability, and the high accuracy of differential pressure sensors that have been significantly improved over the last decade.

Estimates of the BHP can be found several ways. The pressure can be measured by providing tools in the bottomhole assembly (BHA), and sent to the surface by mud pulse telemetry or wired drillpipe. Another method is using a numerical flow model to calculate the BHP based on information related to the wellbore geometry and surface measurements. A third method is to use simple “rule-of-thumb” methods which only require a few input parameters. A fourth method is to utilize differential pressure measurements at the surface and calculate BHP based on drillpipe length and diameters.

In this paper these methods are compared in terms of accuracy, robustness and cost. The accuracy is evaluated based on simulated results, and the robustness and costs comparison is based on an coarse discussion relative to high and low robustness and high and low cost.

The following section presents these four methods on how to determining the downhole pressures. The third section presents a case description in where the methods are evaluated, followed by a section presenting results and discussions from the comparison. Finally the conclusions and suggestions for further work are described in the last section.

2. METHODS

Four methods for estimating the bottomhole pressure (BHP) is described in this section. The first method involves direct downhole pressure measurements, the second using a numerical flow model, the third utilizing a “rule-of-thumb”, and the last method uses surface differential pressure sensors. Each has their own advantages and drawbacks, which will be commented on in the description of each method in the results sections.

Downhole pressure measurements

A measurement-while-drilling (MWD) tool is part of the bottomhole assembly (BHA) of a drillstring, and takes pressure measurements in the annulus just above the drill bit. These are taken at frequent intervals, usually once every second. The measurements are then sent to surface using mud pulse or wired pipe telemetry.

Mud pulse telemetry sends information to surface using pressure waves in the mud, and cannot be used during hydrostatic conditions. Mud pulse has a fairly low bandwidth limiting the amount of information sent to the surface, thus using mud pulse the measurements are usually only sent every 5 minutes. Another problem with such down-hole tools is that when they break down or run out of power, the entire drill string must be pulled to surface in order to repair the tool.

An option to mud pulse telemetry is using wired pipe telemetry. Wired pipe is quite costly, as the special wired drill pipe has to be rented for the entire length of the well. The benefit is a, comparatively speaking, nearly endless bandwidth and nearly instant transportation of the measurements. It can also incorporate measurements along the drill string, but that is not considered here. All parts of the drillstring must also be working for the telemetry to work, thus there is a reliability issue should a stand of drillpipe become malfunctioning.

Hydarulic flow model

The BHP can also be estimated using basic fluid mechanics, based on conservation of mass

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho v}{\partial x} = 0$$

and conservation of momentum

$$\frac{\partial \rho v}{\partial t} + \frac{\partial \rho v^2}{\partial x} + \frac{\partial p}{\partial x} = G$$

Here ρ is the density of the drilling fluid, v is the velocity of the drilling fluid, p is the pressure and G is work performed by gravity and wall friction. Due to the length of an oil well compared to the diameter the flow is treated as one-dimensional pipe and annular flow. The above equations are for single phase flow, which is used in this work, although the flow model used is a two-phase model. All equations are presented for single phase flow due to notation simplicity.

The numerical scheme to solve the above set of equations is the AUSM scheme, [Evje and Fjelde, 2003]. Some additional equations are needed to solve the equations. First, the density is given by a linear equation

$$\rho = \rho_0 + \frac{p - p_0}{c^2}$$

where ρ_0 is the density at the pressure p_0 , and c is the sound speed in the drilling fluid. Secondly the source terms G needs to be defined. The hydrostatic pressure loss caused by gravity is given by

$$DP_{hyd} = \rho g \sin(\theta)$$

where g is the gravitational constant and θ is the inclination of the well. The frictional pressure loss is based on the generalized friction loss equations for flow through pipes and concentric annuli, [Reed and Pilehvari, 1993]. It is given by

$$DP_{fric} = -\frac{2f\rho v^2}{D_{hy}}$$

where D_{hy} is the hydraulic diameter (outer diameter minus inner diameter for annuli), and f is the Fanning friction coefficient. For fully developed laminar flow

$$f = \frac{16}{Re_G} = \frac{16\mu_{app}}{\rho v D_{eff}}$$

where Re is the generalized Reynolds number, D_{eff} is the effective diameter and μ_{app} is the apparent viscosity. For Newtonian flow $D_{eff} = D_{hy}$ and μ_{app} is the Newtonian viscosity. When the generalized Reynolds number is high enough, turbulent flow sets in and the Colebrook-White equation is used to solve for the turbulent f :

$$\frac{1}{\sqrt{\frac{f}{4}}} = -2 \log_{10} \left(\frac{\epsilon}{3.7D_{hy}} + \frac{2.51}{Re_G \sqrt{\frac{f}{4}}} \right)$$

Here ϵ is the relative pipe roughness. The change between laminar and turbulent flow occurs, [Reed and Pilehvari, 1993].

Such a transient fluid model can be run in real-time using the data of ongoing drilling operations. In advance the diameters of the equipment used in the well must be known, along with the properties of the drilling fluid. The model can then be used to simulate the well using the flow rate provided by the mud pumps. This provides a continuous estimate of the BHP,

the accuracy depending on the accuracy of the input data and the model itself. It only follows the physical models, thus if the downhole environment change in unexpected ways, the model will not be able to follow the actual BHP.

“Rule-of-thumb”

Drilling wells have been performed successfully using much the same methods for a very long time before the advent of fast computers and electronic tools. In the Wild Well Control handbook [Wild Well Control, 2013] there is described some simple equations for estimating what pressure you will achieve for a typical well using very few input parameters.

The equations are based on oilfield units. The required input in addition to the mud weight in pounds per US gallon, vertical depth in feet and annular dimensions in inches are the annular frictional pressure loss in psi, the yield point of the fluid in $\frac{lb}{100ft^2}$ or the yield point, plastic viscosity and fluid velocity in $\frac{lb}{100ft^2}$, cP and feet per minute respectively.

The simplest formula used for equivalent circulation density (ECD – mud weight a hydrostatic column would have)

$$ECD_{ppg} = MW_{ppg} + \frac{AnnularFrictionPressureLoss_{psi}}{0.052 * TVD_{ft}}$$

where MW is the mud weight, $AnnularFrictionPressureLoss$ is the pressure loss in the annulus usually taken to be about 10% of the pump pressure for normal hole geometries, and TVD is the vertical depth of the hole.

There are also another model using the yield point (YP) of the drilling fluid

$$ECD = MW + \frac{0.1YP}{OD - ID}$$

when $MW \leq 13 ppg$, and where OD is the hold diameter and ID is the outer diameter of the drill pipe. When $MW > 13 ppg$ another formula is used

$$ECD = MW + \frac{0.1}{OD - ID} \left(YP + \frac{PVv}{300(OD - ID)} \right)$$

where PV is the plastic viscosity and v is the flow velocity in $\frac{ft}{minute}$ given by

$$v = 24.51 * \frac{Q_{gpm}}{OD^2 - ID^2}$$

The simplest formula is the one used here to estimate the bottomhole pressure. In SI units the equation for bottomhole pressure become

$$BHP = \rho g TVD + 0.1 \cdot SPP$$

Where ρ is the mud weight, $g = 9.81 \frac{m}{s^2}$, TVD is the vertical depth of the well and SPP is the pressure at the pump.

Differential pressure sensors

The method of using two differential pressure transducers to obtain information of the drilling fluid is based on mounting one differential transducer along a horizontal segment of the drilling fluid flow path at the drillfloor, and another differential pressure transducer at a (upwards) vertical segment of the drilling rig.

Thus, instead of calculating the friction factor of the mud from measurements of the mud viscosity through manual conventional methods and then calculating the resultant frictional pressure, it is easier and more accurate to measure directly by automation means the mud frictional pressure drop over a certain horizontal section.

By using the horizontal differential pressure measurements P_H and the vertical pressure readings P_V in addition to the measurement pipe length h , an estimate for the bottomhole pressure can be found by multiplying the pressure differentials with the length of the well.

$$BHP = SPP + \frac{P_H}{h} \cdot MD - \frac{(P_V - P_H)}{h} \cdot TVD$$

It should be noted that such measurements could also be used indirectly to measure physical quantities. The density of the drilling fluid mixture could be obtained using

$$\rho = \frac{P_V - P_H}{gh}$$

Since the volume flow rate $Q = Av$, where A is the cross-sectional area of the pipe, the apparent viscosity could also be expressed as

$$\mu = \frac{\rho Q D}{Re A}$$

When calculating the apparent viscosity, the flow regime should be identified to obtain the correct values.

Some initial simulations and experiments utilizing these quantities have been performed, [Carlsen et al, 2012a], [Carlsen et al, 2012b].

3. CASE DESCRIPTION

The hydraulic flow model simulates a 2000 meters deep vertical well, where the flow rate is varied over a period of 2200 seconds. A 5 inch diameter drillpipe in a 8.097 inch hole is used. There is an additional pressure loss over the bottom-hole-assembly (BHA - a collection of tools including drill bit at the end of the drillstring). The fluid used has a mud density of 1000 kg/m³ at standard conditions, and a low viscosity of 5 cP.

In order to compare the methods for estimating the bottomhole pressure, i.e. to determine the benefits and value of each method the described hydraulic flow model is used as "truth". The method of comparison assumes the volumetric flow rate into the well as well as the pump pressure (SPP) is available for use in all the methods.

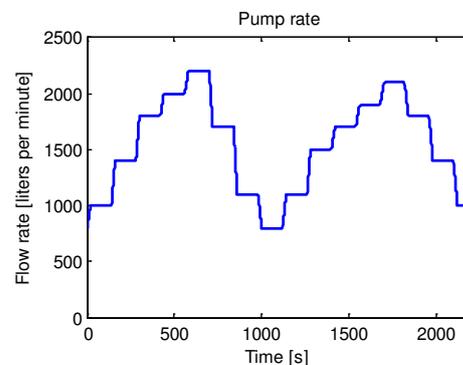


Figure 1 Fluid pump rate into the well used as input to the models

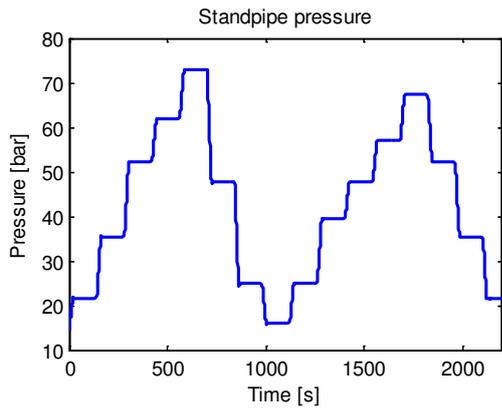


Figure 2 Simulated standpipe pressure used as input to some of the models

The bottomhole pressure calculations are then performed:

- The downhole pressure measurement provides the “true” pressure every 5 minutes
- The hydraulic model uses input that are slightly different causing an error assumed the roughness has increased by a factor of 10 due to regular wear, the inner diameter of the pipe has increased by 1% and the viscosity is guessed 10% higher than it actually was.
- The simple rule-of-thumb method is used, based on an annular friction loss of 10% of the measured SPP.
- The differential pressure sensors provide accurate pressure losses over a short segment of pipe which is then multiplied to represent the pressure loss in the entire drillstring.

The resulting bottomhole pressure calculations are compared with the simulator results. However, the method of determining the downhole pressure directly is only discussed from robustness and cost viewpoints.

4. Results

The accuracy of the direct measurement of downhole pressure is assumed to be sufficiently accurate. The “truth” simulation with the measurements plotted in is shown in Figure 2.

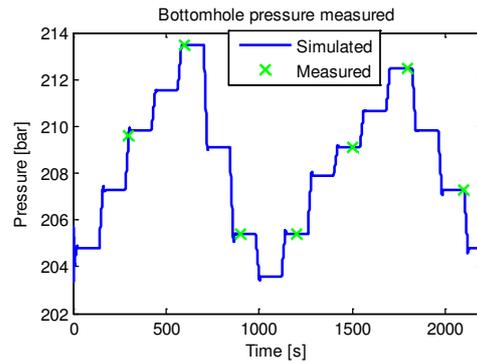


Figure 3 Downhole pressure measurements.

In Figure 1 a comparison between the simulated model and the “Rule of thumb” is shown.

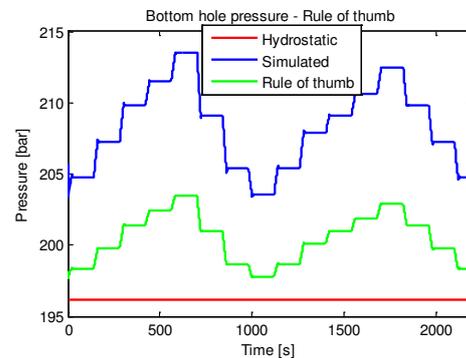


Figure 4 Downhole pressure estimated using “Rule of thumb”

The results show that the “rule of thumb” is not giving sufficient accuracy needed in estimating the downhole pressure.

In Figure 2 the downhole pressure estimated for the advanced flow model is presented. The advanced flow model gives a fairly good estimate of the downhole pressure.

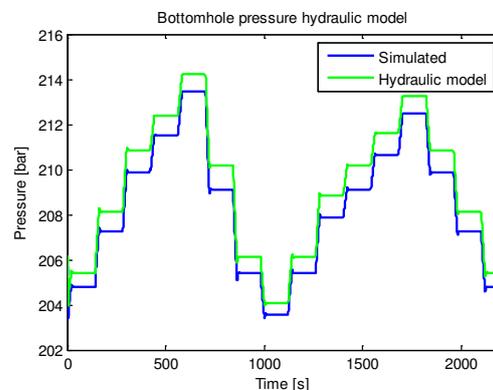


Figure 5 Downhole pressure determined using advanced model

The corresponding pump pressures are shown in Figure 3.

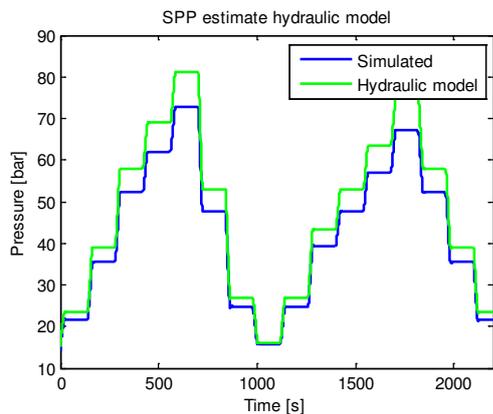


Figure 6 Standpipe pressure determined using advanced model

The pressure losses in a pipe segment at surface is used to estimate the bottomhole pressure in Figure 5

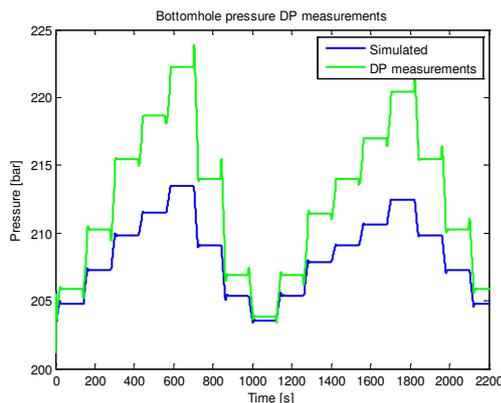


Figure 7. Downhole pressure determined using DP measurements

5. DISCUSSIONS

This section discusses the various methods.

Measurement-while-drilling

A fairly good estimate of the BHP is available, and is kept above the pore pressure and below the fracture pressure. It gives little indication of the pressure in other parts of the well.

As drillings stops, so does the measurements. Thus when starting the pumps again it will take some time of increasing pressure until the measurements will be available again. In addition the measurement can become

unavailable for other reasons such as tool failure, and will then either need to be pulled out (which takes time) or continue drilling without the measurements.

Numerical model

The numerical model estimates the pressure along the entire wellbore. The estimate may not be very good on the other hand. As the SPP is available, the estimates can be somewhat improved using the measurement to calibrate the model, or be an indication of problems while drilling. It should be noted that poor input relevant for flow in the pipe may cancel out poor input given for the annulus. Pump pressure can be kept high, as the pressure is known at the shoe, and the safety factor of friction loss in the cased annulus is replaced by a safety factor based on the uncertainty of the model.

“Rule of thumb”

The rule of thumb method has a rather poor estimate, but provides an estimate of the pressure at the shoe, or at the bottom of the well. The rule is likely to perform better for larger hole sizes, as the annular pressure loss would be a smaller portion of the SPP. Due to the poor pressure estimate, an additional pressure safety margin is needed.

Surface differential pressure sensors

Surface differential pressure sensors give an early measurement of the frictional and hydrostatic pressure loss that can be expected within the drillpipe. Unknown discrepancies between the measurement pipe and the drillpipe can of course exist, which could reduce the efficiency of the measurements. Only frictional pressure losses inside the drillstring is estimated here, thus neglecting the pressure loss through the bit and BHA, which can explain most of the difference seen in

Due to the low cost, this method could be well used together with other methods such as a hydraulic flow model to get improved accuracy.

A summary of the discussion is given in Table 1.

Table 1 Comparison of methods for estimating downhole pressure

	Accuracy	Robustness	Cost
Downhole pressure measurement	High	Low	High
Numerical model	Medium	High	Low
Rule of thumb	Low	Low	Low
Surface differential pressure sensors	Medium	High	Low

6. CONCLUSIONS

The results indicate that the various methods have both benefits and drawbacks. Further works should be performed where focus should be on combining the various methods.

By combining the differential pressure transducer method with an advanced hydraulic flow model, a reasonably accurate system could be developed, regarding accuracy, cost and robustness.

7. REFERENCES

- S. Evje, K.K. Fjelde, "On a rough AUSM scheme for a one-dimensional two-phase model", *Computers and Fluids*, vol. 32 (2003)
- T.D. Reed, A.A. Pilehvari, "A New Model for Laminar, Transitional and Turbulent Flow of Drilling Muds", Paper SPE 25456 presented at the Productions Operations Symposium, Oklahoma City, OK, 21-23 March (1993)
- Wild Well Control, "Technical Data Handbook", <http://wildwell.com/literature-on-demand/literature/technical-data-book.pdf> (2013)
- G. Nygaard, "Evaluation of automated drilling technologies developed for petroleum drilling and their potential when drilling geothermal wells", presented at GeoEnergy2011, Bergen, 7.-8. September, 2011.

L.A. Carlsen, G. Nygaard, R. Time, "Utilizing Instrumented Stand Pipe for Monitoring Drilling Fluid Dynamics for Improving Automated Drilling Operations", Proceedings of the 2012 IFAC Workshop on Automatic Control in Offshore Oil and Gas Production, Norwegian University of Science and Technology, Trondheim, Norway, May 31 - June 1, 2012.

L.A. Carlsen, N. L. Rolland, G. Nygaard, R. Time, "Simultaneous Continuous Monitoring of the Drilling-Fluid Friction Factor and Density", *Journal SPE Drilling & Completion*, Vol. 28, No. 1 pp. 34-44, March, 2013.

Using Probabilistic Analysis with Finite Element Modeling of High Temperature Geothermal Well Casings

Gunnar Skúlason Kaldal*, Magnús Þór Jónsson, Halldór Pálsson, and Sigrún Nanna Karlsdóttir

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science,
University of Iceland, Hjarðarhagi 2-6, 107 Reykjavík, Iceland

Abstract

High temperature geothermal wells which are drilled in geothermal areas are constructed of several concentric steel casings that are cemented together. The structural integrity of such well casings is essential for the utilization of high temperature geothermal wells. The temperature change in high temperature geothermal wells is large and much larger than commonly seen in oil wells. This large temperature change can cause problems in the casing due to thermal expansion of materials. The wellhead rises during discharge due to thermal expansion of the steel in the casing and the large temperature change can also lead to casing collapse due to expanding annular fluids. With recent increasing interest in drilling deeper geothermal wells the strength of the casing becomes one of the most limiting factor. A nonlinear structural finite element model of the cased well is presented and discussed here. The purpose of the model is to evaluate the structural integrity of the casing when it is subjected to thermo-mechanical loads. The outcome of the model depends highly on the accuracy of the input parameters, i.e. geometrical sizes and material properties. The accuracy of the results are evaluated with the use of probabilistic design analysis where selected input parameters of the model are assumed to contain a reasonable amount of scatter. The uncertainties of the model can thus be quantified.

1. Introduction

Geothermal wells are constructed of several concentric steel casings which are fully cemented together and cemented to the rock formation. Usually, three casings are used; the production casing, anchor casing and surface casing. The wellhead consists of a casing head flange, expansion spool and a master valve. The wellhead is attached to the top of the anchor casing and the production casing movements relative to the anchor casing is accommodated below the master valve inside the expansion spool [5].

During the discharge of high temperature geothermal wells, the temperature difference in the well from non-flowing to flowing conditions is large. To take an example, reservoir temperatures in the Krafla high temperature geothermal area in Iceland typically range from 210 °C to 350 °C [11]. The large temperature change generates thermal stress in the casing which is partially constrained by the concrete. While the well warms up the wellhead rises as a result of thermal expansion of the casings and concrete. Cas-

ing failures can lead to a reduced energy output from the well, render it inoperative and in worst cases cause unsafe conditions above the surface. Thus the structural integrity of well casings is essential for the utilization of high temperature geothermal wells.

The casings and the wellhead form a structural system which is unpractical to solve analytically mainly due to the nonlinear behavior of the contacting surfaces. Therefore, the structural system is analyzed numerically with the use of the nonlinear finite element method (FEM). A thermal and nonlinear structural model of the cased well is constructed where nonlinearities, e.g. friction, plasticity and large non-uniform deformations are accounted for. The nonlinear axi-symmetric model described here is a continuation of the work described by Kaldal [7]. Other models of geothermal wells have been created, e.g. an elastic 2D FEM model presented by Gretarsdottir [4] and a nonlinear FEM model by Magnúsdottir where the bonding characteristics between the production casing and its outer concrete were analyzed [10]. The collapse of

*gunnarsk@hi.is

the production casing is a non-symmetrical phenomenon which cannot be analyzed with this model but has been modeled with nonlinear 3D FEM models by Kaldal [6][8].

As opposed to deterministic analysis, where input parameters are treated as constants which results in a one possible solution, a probabilistic approach is used, where the input parameters are assumed to contain a reasonable amount of scatter, which then gives an estimate of the model uncertainties. Here, selected results from the FEM model are used as outputs for the probabilistic analysis. Scatter plots of the input parameters versus the output results reveal which input parameters are significant to the results of the FEM model. A cumulative distribution function of the wellhead movement and the maximum von Mises stress in the casings and surrounding concrete is obtained here and are used to review the uncertainty of the model. In this paper, probabilistic analysis of the FEM model is presented where several input parameters are evaluated.

2. Probabilistic design

Probabilistic design is an analysis technique for assessing the effect of uncertain input parameters and model assumptions [1]. Using this method the uncertainties of the model can be quantified by acknowledging that the input parameters are not constants but rather parameters that follow statistical distribution functions such as Gaussian or normal distribution. By this assumption the limited outcome of deterministic results is avoided and the uncertainties in the model and probability distribution of the results are analyzed. Of course, the modeling error which refers to the difference between the physical system and its mathematical model [2] can be larger due to numerous approximations made to the geometry, material properties, load, etc. But all in all this method gives a map of the results rather than one point.

Intertwined in the method is the determination of the sensitivity of individual parameters to the results. For the model described below, each input parameter is assumed to follow normal distribution given by assumed mean and standard deviation. The probabilistic analysis employs the Monte Carlo Simulation method with Latin Hypercube Sampling, which avoids repeated samples [1]. A given number of simulation loops are performed and before each loop, individual input parameter is randomly given a value within its normal distribution domain. When the simulation loops are finished the sensitivity of the in-

put parameters to the results can be visualized with scatter plots.

For the probabilistic analysis used on the FEM model described below, 400 simulation loops were used. The selected input parameters that were used for the probabilistic analysis are listed in Table 1. Their assumed means and standard deviation are listed as well. The standard deviation σ provides the sample range for the parameter and 99.7% of the samples should fall within 3σ from the mean provided that the number of simulation loops is sufficient.

Table 1: Probabilistic design input parameters (mean and assumed standard deviation).

Parameter	Units	Mean	Std
μ	-	0.45	0.15
τ_{max}	MPa	0.46	0.13
E_{co}	GPa	2.4	0.6
E_{gr}	GPa	80	20
f_c	MPa	30	7.5
ρ_{st}	kg/m ³	6125	150
ρ_{co}	kg/m ³	1600	200
α_{st}	1/°C	12e-6	1e-6
α_{co}	1/°C	10e-6	1.5e-6
$\sigma-\epsilon_{sc}$	-	1	0.1

3. FEM model description

A nonlinear thermal and structural model of a high temperature geothermal well which reaches from the wellhead to the bottom of the production casing is constructed with the use of the finite-element method (FEM). The focus is to analyze the structural system which consists of a wellhead and several concentric casings connected together and to the formation with concrete. The model is two dimensional and axisymmetric around the center of the well. It includes nonlinearities which are found in large geometrical displacements, in material properties and in connectivity between contacting surfaces.

The model, shown in Figure 1, is parametrically designed so geometrical sizes and material properties are adjustable by the user. Simplified couplings with no threads are included in the production casing to account for the anchoring effect of the couplings in the concrete. A simplified wellhead based on an actual design is also included to account for pressure loads and the interaction between the casing and the wellhead. Material properties and reference values that are used in the model are listed in Table 2. Additionally, the reference value for the coefficient of friction between steel and concrete is $\mu = 0.5$ and

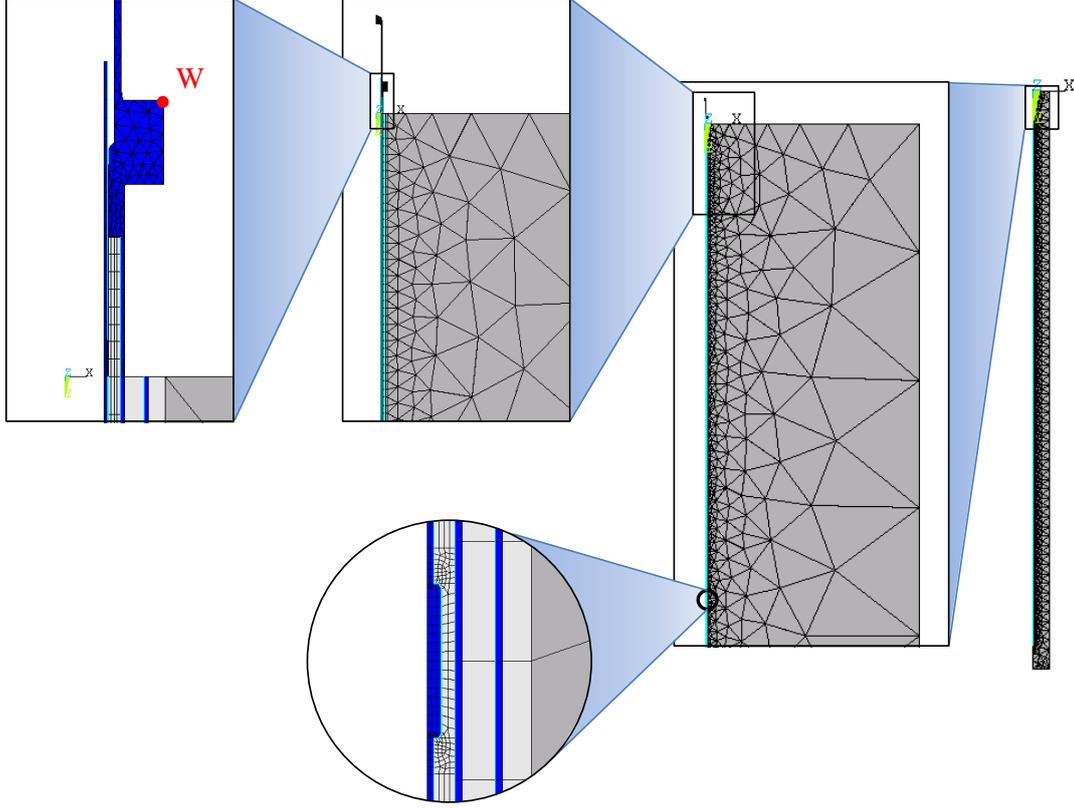


Figure 1: The geometry of the model.

the maximum shear stress when sliding initiates is $\tau_{max} = 0.46$ MPa. Nonlinear material properties of steel grades K55, L80, T95 and X56 are implemented in the model with the use of stress-strain curves which were obtained from tensile tests [9]. The concrete is assumed to yield plastically above its maximum compressive strength and the formation is assumed to be solid basaltic rock.

Table 2: Material properties and numerical values used in the model.

Material property	Units	Steel	Concrete
Young's modulus (E)	GPa	210	2.4
Poisson's ratio (ν)	-	0.3	0.15
Density (ρ)	kg/m ³	7850	1600
Th. conductivity (K)	W/m ² C	50	0.81
Specific heat (c)	J/kg ² C	400	880
Th. expansion (α)	1/ ^o C	12e-6	10e-6
Compressive strength (f_c)	MPa	-	25e6

The frictional connection between surfaces in particular makes the model computationally complex. Contact element pairs are used between contacting surfaces. Their main purpose is to prevent surfaces to intersect each other, while still allowing gap formation and tangential movement between casings and concrete. The Coulomb friction model is used to describe fric-

tion between contacting surfaces, where it can withstand shear stresses up to a certain magnitude across its interface before they start sliding relative to each other [1]. Once the equivalent shear stress exceeds τ_{max} relative sliding begins. The Coulomb friction model is defined as:

$$\tau = \begin{cases} \mu P + b & \text{if } \tau < \tau_{max} \\ \tau_{max} & \text{if } \tau \geq \tau_{max} \end{cases} \quad (1)$$

where τ is the equivalent shear stress, τ_{max} is the maximum shear stress, μ is the isotropic coefficient of friction, b is the contact cohesion and P is the contact normal pressure, see Figure 2 for the graphical interpretation of the Coulomb friction model.

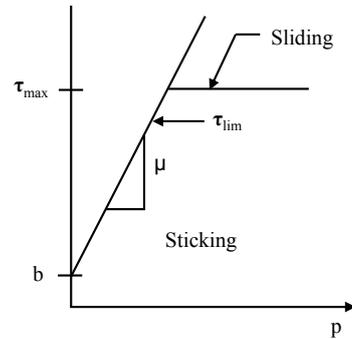


Figure 2: The Coulomb friction model in Ansys [1].

The boundary conditions of the model are defined such that no displacements are allowed in the lower and outer boundary of the model. The lower boundary is located 20 m deeper than the production casing shoe and the outer (radial) boundary of the rock formation is 20 m outward from the well, which is sufficient for both the thermal and the structural parts of the model. The wellhead movement is observed at the nodal point denoted with *W* on Figure 1.

4. Results

4.1. FEM results from a single simulation run

FEM results from a single simulation run, with the material properties values listed in Table 2, are presented here.

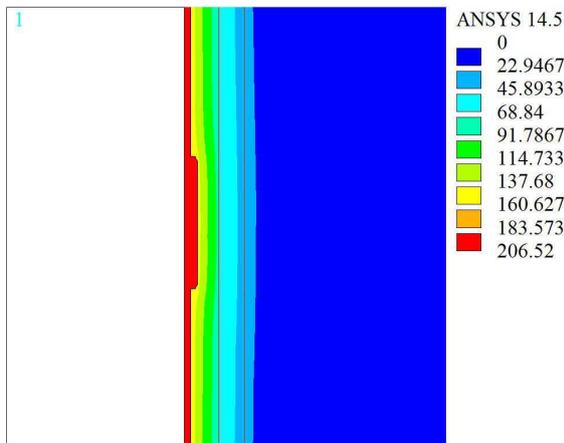


Figure 3: Temperature distribution 2.5 hours after discharge (°C).

The calculated temperature distribution at the top of the well 2.5 hours after discharge, which will be the reference time of the subsequent results, is displayed in Figure 3. A temperature change of 200 °C is assumed. The displacement at the top of the well is displayed in Figure 4. In this run the displacement of the production casing is 35.8 mm and the wellhead displacement is 11.7 mm. Stress concentration near the couplings of the casing is illustrated in Figure 5 where the maximum stress in the steel is produced near the couplings and in Figure 6 where the maximum stress in the concrete forms at the top of the couplings. In this case the maximum stress for both the concrete and the casing is formed at the second highest coupling.

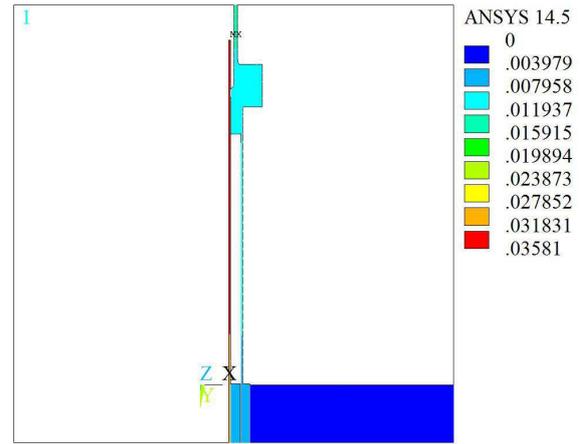


Figure 4: Displacement of the wellhead 2.5 hours after discharge (m). The displacement of the production casing is 35.8 mm and the wellhead displacement is 11.7 mm.

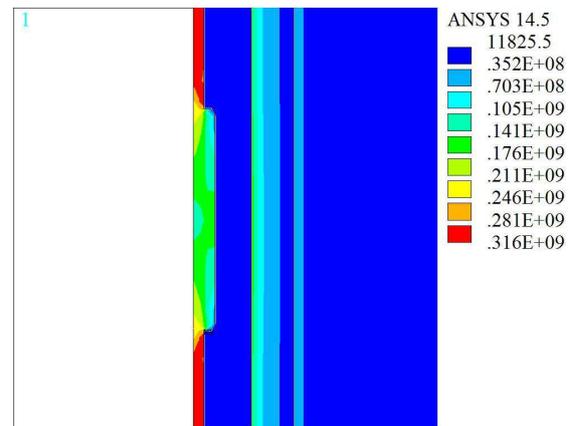


Figure 5: Von Mises stress at the second highest coupling (Pa).

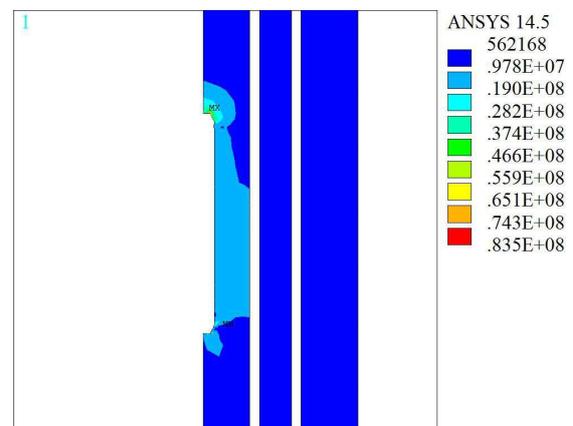


Figure 6: Von Mises stress of concrete at the second highest coupling (Pa).

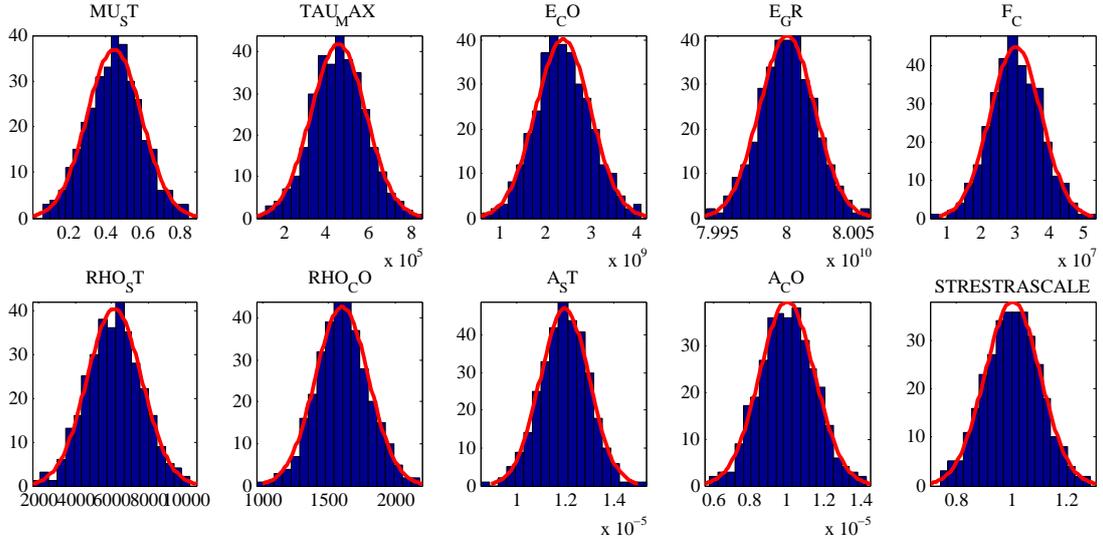


Figure 7: Histograms of normally distributed input parameters.

4.2. Distribution of inputs and outputs

The material properties for the FEM model that were selected as input parameters for the probabilistic analysis are listed in Table 1. Histograms of the input parameters which were assumed to follow normal distribution are seen in Figure 7. The samples of the input parameters follow normal distribution which confirms that for the probabilistic analysis, 400 Monte Carlo simulation loops are sufficient.

The results from the FEM model that were selected as outputs for the probabilistic analysis, i.e. wellhead displacement and maximum von Mises stress in the production casing, the anchor casing and the surrounding concrete for each casing, are displayed in the histograms in Figure 8 and discussed in the discussion section below.

4.3. Correlation between inputs and outputs

Scatter plots showing the correlation between the selected input parameters and the selected outputs of the model are illustrated in Figure 14. A significance level of the correlation between input and outputs is chosen to be 2.5%, so that $R^2 < 0.025$ is dismissed as insignificant. For the first column, the wellhead displacement, the significant parameters are; μ_{st} , τ_{max} and α_{st} with correlations of $R^2 = 0.02528$, $R^2 = 0.32987$ and $R^2 = 0.35066$. All other input parameters are of no significance to the wellhead movement. Summary of the significant parameters and correlations is listed in Table 3.

Table 3: Significant input parameters and correlation with the results.

Output	Input	R^2
Wellhead displacement	μ_{st}	0.02528
	τ_{max}	0.32987
	α_{st}	0.35066
Max. von Mises stress of the production casing	α_{st}	0.36826
	$\sigma-\epsilon_{sc}$	0.73963
	τ_{max}	0.62363
Max. von Mises stress of the production concrete	E_{co}	0.15270
	f_c	0.03274
	α_{st}	0.09283
	$\sigma-\epsilon_{sc}$	0.02940
Max. von Mises stress of the anchor casing	ρ_{st}	0.02701
	ρ_{co}	0.21588
	α_{st}	0.73663
Max. von Mises stress of the anchor concrete	E_{co}	0.83435
	α_{st}	0.11622

4.4. Cumulative distribution function

If the results from all the Monte Carlo simulations are sorted and plotted against the proportion of the result values an empirical cumulative distribution function (CDF) of the results is obtained. These plotted curves can then be used to visualize the results. From the wellhead displacement results for example, seen in Figure 9, it can be stated with 90% certainty that the wellhead displacement is less than or equal to

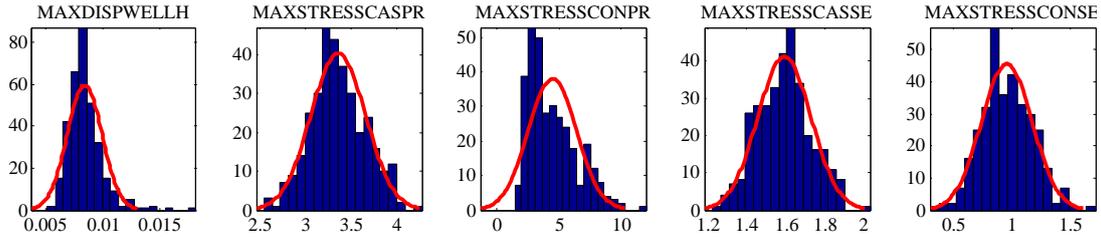


Figure 8: Histograms of the selected output results from the probabilistic analysis.

10 mm, with 95% certainty that the displacement is below 11 mm and with 99% certainty that the maximum rise of the wellhead during discharge will be 15 mm. This of course depends on the premises of all the input parameters, the model assumptions and limitations, and the assumed load the well is subjected to.

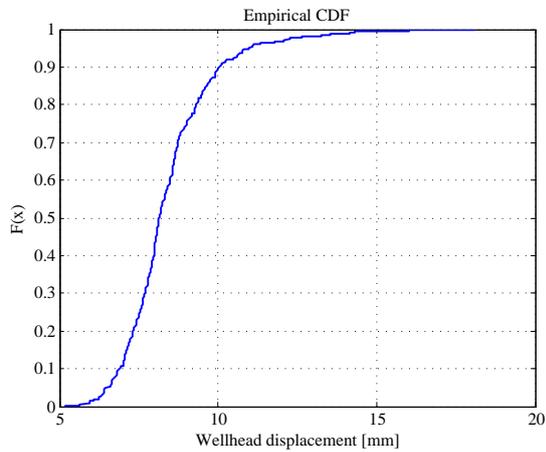


Figure 9: Empirical cumulative distribution function (CDF) plot of the wellhead displacement

Similar statements, as were made to the wellhead movement, can be made to the cumulative distribution function plot of the maximum von Mises stress for the production casing, its surrounding concrete, the anchor casing and its surrounding concrete which are displayed in Figures 10-13. Statistical summary of the output results from the probabilistic analysis is listed in table 4.

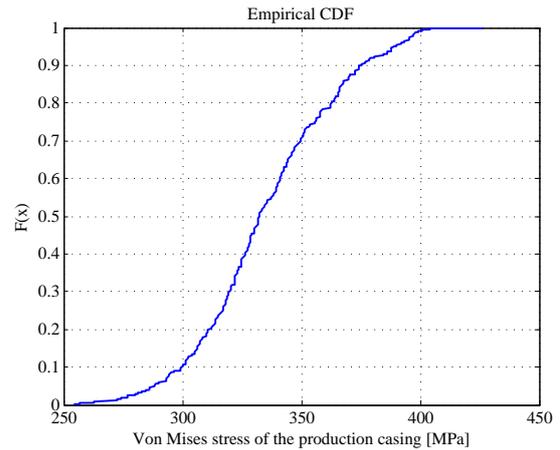


Figure 10: Empirical CDF plot of the maximum von Mises stress in the production casing

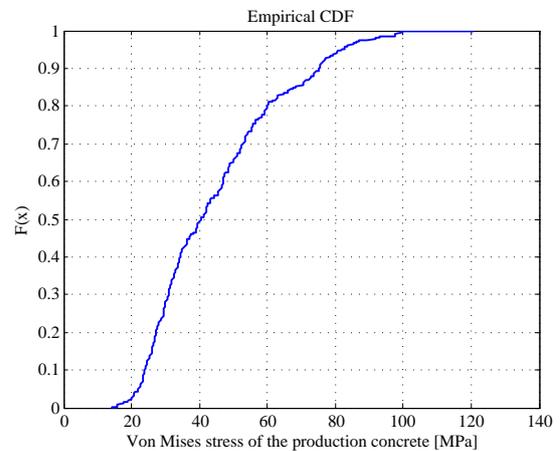


Figure 11: Empirical CDF plot of the maximum von Mises stress in the production concrete

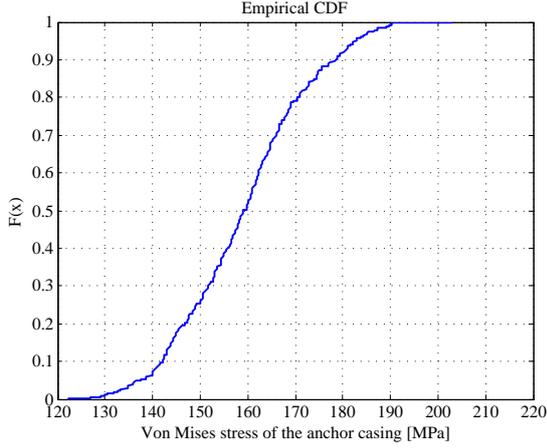


Figure 12: Empirical CDF plot of the maximum von Mises stress in the anchor casing

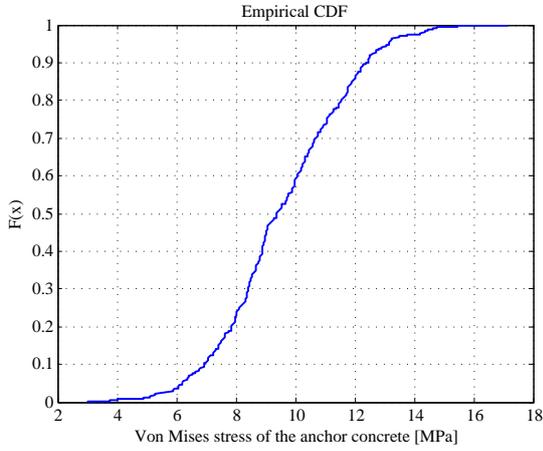


Figure 13: Empirical CDF plot of the maximum von Mises stress in the anchor concrete

Table 4: Statistical summary of the output results from the probabilistic analysis.

	Mean	Std	Min	Max
i.	8.43	1.52	5.19	18.1
ii.	335.6	29.73	254.6	426.5
iii.	44.86	19.55	14.11	120.6
iv.	159.2	13.69	122.3	203.1
v.	9.589	2.163	3.010	17.16

- i: Wellhead displacement.
- ii: Max. von Mises stress of the production casing.
- iii: Max. von Mises stress of the production concrete.
- iv: Max. von Mises stress of the anchor casing.
- v: Max. von Mises stress of the anchor concrete.

5. Discussion

In the single run of the FEM model it was illustrated how the results appear after 2.5 hours which was the reference time for the probabilistic analysis. The temperature distribution shown in Figure 3 illustrates how shallow the temperature front has reached during this time. Figure 4 illustrates the wellhead displacement and how the production casing slides inside the wellhead. The stress concentration region which is located at the couplings is illustrated in Figures 5 and 6.

The selected outputs of the FEM model for the probabilistic analysis are not as distinctively normally distributed as the input parameters which were randomly given a value within the normal distribution domain. The histograms of the results seen in Figure 8 do however reveal that, apart from the maximum von Mises stress in the concrete surrounding the production casing, the results follow normal distribution nevertheless. The standard deviation given for the input parameters in the probabilistic analysis are intentionally large but might be a bit too spacious. Excluding the insignificant input parameters and narrowing the standard deviation of the significant input parameters should remove some of the noise and improve the results from the probabilistic analysis.

The correlations between input and output parameters reveal which input parameters are significant for each of the output results, see Table 3. Thermal expansion of the casings α_{st} and the parameters for friction, μ_{st} and τ_{max} , proved to be significant to the wellhead displacement, μ_{st} the least significant of the three with a correlation of $R^2 = 0.02528$. Thermal expansion of steel α_{st} is significant for all the selected output results. Specially for the maximum von Mises stress of the production casing and the anchor casing. Scaling the stress-strain curve for steel appears only to be significant for the maximum stress in the production casing and its surrounding concrete sliding freely inside the anchor casing and the wellhead. The anchor casing on the other hand is connected to the wellhead which results in less degree of freedom.

The maximum von Mises stress might not be a good output from the probabilistic analysis since it is a local peak stress which does not resemble the whole casing. Instead or rather additionally, because the maximum is surely of interest, mean stress with standard deviation could be a better option of outputs for comparison.

The cumulative distribution functions (CDF) of the output results, Figures 9 to 13, illustrate the uncertainties of the model. For ex-

ample it can be stated with 95% certainty and with these premises that the wellhead displacement is below 11 mm, the maximum von Mises stress in the production casing will be lower than 390 MPa, its surrounding concrete 82 MPa, the anchor casing 182 MPa and its surrounding concrete 13 MPa. This means that the production casing has reached beyond the proportional limit of the stress-strain curve for K55 steel which has a minimum yield strength of 379 MPa[3], its surrounding concrete is partially broken, but the security casing and its surrounding concrete are still intact, again with 95% certainty.

6. Conclusion

In this paper, probabilistic analysis of a structural FEM model of a high temperature geothermal well was presented. Using probabilistic methods on FEM models provide a broader understanding of the problem and the model itself and produce a topography of the results as well as enabling the uncertainties of the model to be quantified.

7. Acknowledgment

This work was supported by the University of Iceland research fund, the Technology Development Fund at RANNIS - The Icelandic Center for Research, the Innovation Center Iceland, Landsvirkjun Energy Research Fund, and GEORG - Geothermal Research Group. Their support is much appreciated.

Nomenclature

E	Young's modulus
E_{st}	Young's modulus of steel
E_{co}	Young's modulus of concrete
E_{gr}	Young's modulus of formation
ν	Possion's ratio
ρ	Density
ρ_{st}	Density of steel
ρ_{co}	Density of concrete
f_c	Compressive strenght of concrete
K	Thermal conductivity

c	Specific heat
α	Thermal expansion
α_{st}	Thermal expansion of steel
α_{co}	Thermal expansion of concrete
μ	Coefficient of friction
τ_{max}	Maximum shear stress
τ_{lim}	Limit shear stress
τ	Equivalent shear stress
P	Contact normal pressure
b	Contact cohesion
R^2	Coefficient of determination
σ	Standard deviation
$\sigma-\epsilon_{sc}$	Stress-strain curve scaling factor

References

- [1] *Release 14.5 documentation for ANSYS*, 2012.
- [2] Robert D. Cook, David S. Malkus, Michael E. Plesha, and Robert J. Witt. *Concepts and Applications of Finite Element Analysis*. John Wiley & Sons, Inc, 111 River Street, Hoboken, NJ 07030, 2002.
- [3] Gilles Gabolde and Jean-Paul Nguyen. *Drilling Data Handbook*. IFP publications, Éditions Technip, 27 rue Ginoux, 75737 Paris Cedex 15, France, 2006.
- [4] S. Gretarsdottir. Mechanical and thermal properties of well cement. Master's thesis, University of Iceland, 2007.
- [5] Hagen Hole. Geothermal well design - casing and wellhead. *Petroleum Engineering Summer School*, June 2008. Workshop 26.
- [6] Gunnar Skúlason Kaldal. Load history and buckling of the production casing in a high temperature geothermal well. *Thirty-Sixth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California*, february 2011.
- [7] Gunnar Skúlason Kaldal. Thermal and structural analysis of the casing in a high temperature geothermal well during discharge. *Thirty-Seventh Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California*, february 2012.

- [8] Gunnar Skúlason Kaldal. Collapse analysis of the casing in high temperature geothermal wells. *Thirty-Eighth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California*, february 2013.
- [9] S. N. Karlsdottir and I. O. Thorbjornsson. High Temperature Geothermal Wells - Center of Excellence in Iceland - Phase I: Corrosion Testing of Steel in High Temperature Geothermal Wells in Iceland. Technical Report for RANNIS (The Icelandic Centre for Research), October 2009.
- [10] L. Magnúsdóttir. Nonlinear finite element model of a geothermal well. Master's thesis, University of Iceland, 2009.
- [11] A. Ragnarsson. Utilization of Geothermal Energy in Iceland. *International Geothermal Conference, Reykjavik, Iceland*, pages 39–45, September 2003.

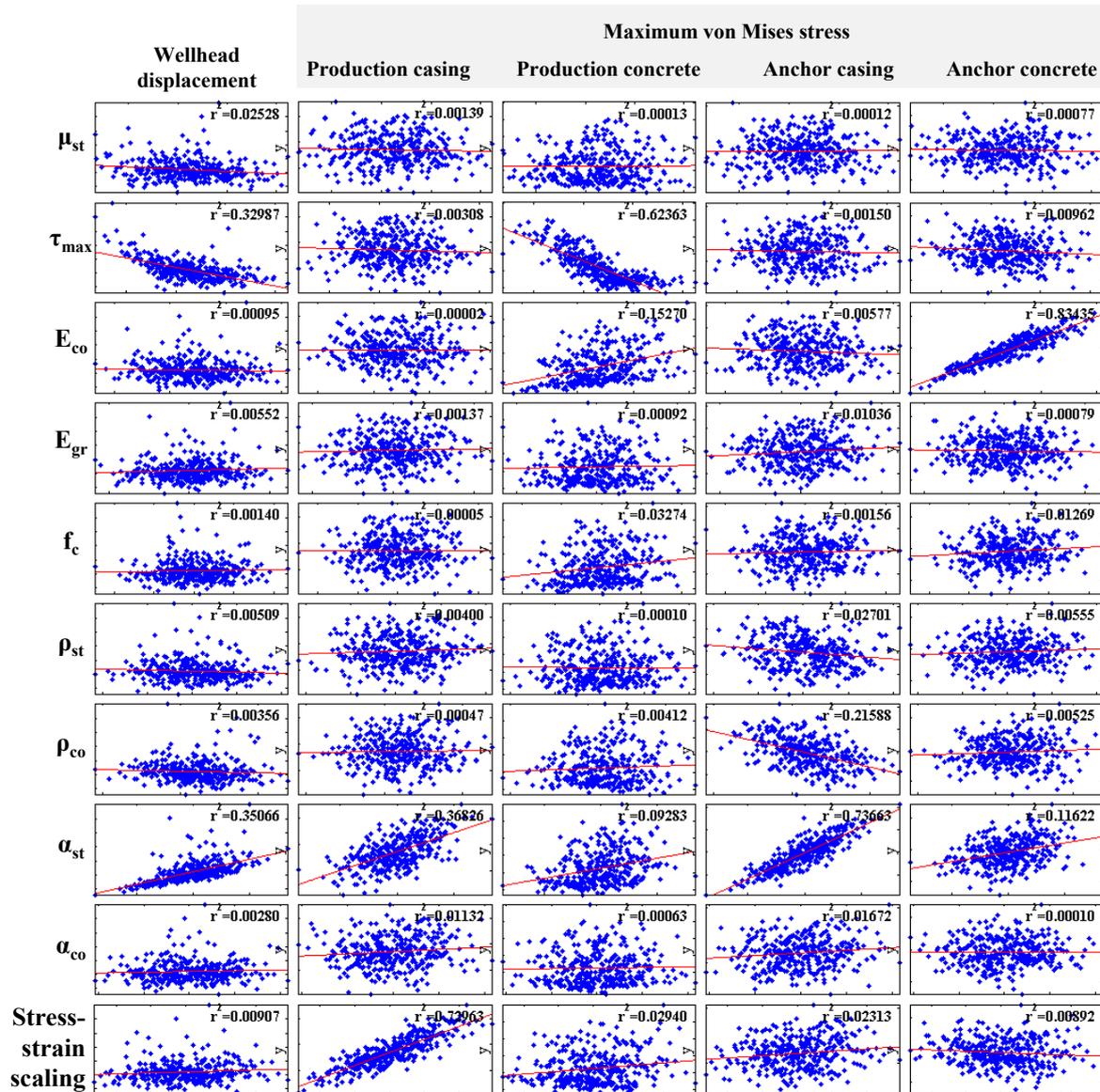


Figure 14: Correlation between selected input parameters and results.

A Numerical Simulation of a Boiling Front Moving Through Porous Medium

Larus Thorvaldsson Halldor Palsson
University of Iceland

August 15, 2013

Abstract

A boiling front of water moving through porous medium is simulated using a numerical finite volume method (FVM) which is then compared to an analytical solution. The FVM is constructed on top of the OpenFOAM framework, which is a highly customizable set of C++ libraries and tools for the solution of problems in continuum mechanics. The solution is stabilized using Fréchet derivatives which enables the use of appropriately sized time steps. The physical properties of the water are determined from the IAPWS-IF97 thermodynamic formulation. The results show that the numerical model is able to simulate the overall behavior of the boiling front sufficiently well. This increases the confidence in the numerical solutions outside the limits of the analytical solution.

1 Introduction

Numerical simulation of hydrothermal systems has played an important role in their modeling for the past decades. For researchers it has been used to test competing hypothesis in these complex environments, where data is often scarce, and in industry numerical simulation has become standard practice in the planning and management of the development of geothermal fields [O’Sullivan, 2001].

The earliest efforts to apply numerical models to geothermal reservoirs were made in the early 1970’s, but the usefulness of numerical modelling did not begin to gain acceptance by the geothermal industry until after the 1980 Code Comparison Study [Stanford Geothermal Program, 1980]. Since that study was performed, the experiences

gained in carrying out site-specific studies as well as generic reservoir modelling studies have led to a constant improvement in the capabilities of numerical reservoir models.

Numerical modeling of hydrothermal systems is often defined by which components of the system are taken into account. Traditionally they have been divided into hydrological (H), thermal (T), mechanical (M) and chemical (C) components. Those components are coupled together in a way that is inherently multiscale in nature, such that their temporal and spatial scales vary by several orders of magnitude [Ingebritsen et al., 2010]. Because of the complex nature of those couplings, models involving all four components are rare.

The equations that describe hydrothermal systems are relatively complex but they can nevertheless be solved analytically for a highly idealized set of initial and boundary conditions. Such cases usually only involve one of the four (HTMC) components, where the Theis problem is an example thereof [Theis, 1935]. Some analytical solutions also exist for two components, such as the description of a boiling front moving through a porous medium [Pruess and Celatis, 1987] and the advance of a diffused salt water wedge in a confined aquifer [Henry, 1964]. These analytical solutions are very important in validating numerical models that are supposed to handle more complicated problems.

The current generation of numerical simulators is in most cases able to account for multi-phase, multi-component flow. The most versatile ones are software packages such as Finite Element Heat and Mass Transfer (FEHM) [Keating et al., 2002] and the Transport of Unsaturated Groundwater

and Heat (TOUGH) family of codes [Pruess, 1991]. These solvers have been applied to a wide variety of problems, such as CO₂ sequestration, geothermal studies and other environmental issues [Ingebritsen et al., 2010].

Other solvers are more specialized, such as the Complex Systems Modeling Platform (CSMP++) [Matthäi et al., 2007] and Fully Implicit Seafloor Hydrothermal Event Simulator (FISHES) [Lewis and Lowell, 2009]. They have been developed specifically to allow simulation of high-temperature multiphase flow of NaCl–H₂O fluids. Other codes such as FALCON [Podgorney et al., 2011], developed at Idaho National Labs have focused on the tightly coupled process of fluid rock interaction.

In this paper the applicability of using a free and open source package named OpenFOAM for modeling hydrothermal systems is examined. OpenFOAM is a highly customizable set of C++ libraries and tools for the solution of problems in continuum mechanics. It is also gaining considerable popularity in academic research and among industrial users, both as a research platform and a black-box CFD and structural analysis solver [Jasak et al., 2007].

The object orientation and operator overloading of C++ has enabled the developers of OpenFOAM to build a framework for computational fluid dynamics that enables modelers to work at a very high level of abstraction [Weller et al., 1998]. This makes it possible to manipulate the set of partial differential equations that describe the problem and customize the solver itself for each class of cases that need to be solved. This will enable researchers with sufficient knowledge about the relevant dynamics of each problem to construct efficient and accurate solvers for it. This is the main motivation for using OpenFOAM, rather than currently existing models.

As well as showing how easily an equation of state can be implemented in the previously existing framework, this paper also shows how the underlying equations can be modified in an easy way. This is demonstrated by linearizing the partial differential equations describing the hydrology of the problem using Fréchet derivatives. In this manner it is possible to stabilize the solution for superheated steam, which makes it possible to treat the pressure equation as steady state, allowing for a more

computationally efficient solution.

2 Methods and materials

2.1 Governing equations

For mass conservation the continuity equation must be satisfied

$$\frac{\partial \phi \rho}{\partial t} + \nabla \cdot (\rho \vec{u}) = 0 \quad (1)$$

where ϕ is porosity and ρ is density. In this equation \vec{u} denotes the superficial velocity, which is a hypothetical velocity calculated as if the fluid were the only thing present in a given cross sectional area. For pressure-velocity coupling, Darcy’s law can be applied

$$\vec{u} = -\frac{\kappa}{\mu} (\nabla p - \rho \vec{g}) \quad (2)$$

where κ is permeability, μ is viscosity and \vec{g} is gravitational acceleration.

This gives following equation for groundwater flow

$$\frac{\partial \phi \rho}{\partial t} - \nabla \cdot \left(\frac{\rho \kappa}{\mu} \nabla p \right) + \nabla \cdot \left(\frac{\rho^2 \kappa}{\mu} \vec{g} \right) = 0 \quad (3)$$

The energy equation includes both effects from the fluid and the soil, and is given as

$$\frac{\partial}{\partial t} (\phi \rho h + (1 - \phi) \rho_s c_s T) + \nabla \cdot (\rho \vec{u} h) = \nabla \cdot (\Gamma \nabla T) \quad (4)$$

where ϕ is porosity, ρ_s is the density of rock, c_s is the heat capacity of soil and Γ is the combined conductivity of fluid and soil. If the properties of the soil are assumed to be constant the laplacian of temperature can be broken up in terms of enthalpy and pressure. In this case, conduction effects in the water are also neglected, which gives the following equation for energy

$$\phi \frac{\partial \rho h}{\partial t} + (1 - \phi) \rho_s c_s \left(\frac{\partial T}{\partial h} \frac{\partial h}{\partial t} + \frac{\partial T}{\partial p} \frac{\partial p}{\partial t} \right) + \nabla \cdot (\rho \vec{u} h) = 0 \quad (5)$$

Since the density is a strong function of pressure in the liquid-vapor phase, it can be accounted for in the time derivative by using a first order Taylor expansion. If it is also assumed that the porosity is

not a function of time, the Taylor expansion of the time derivative becomes

$$\frac{\partial \phi \rho}{\partial t} = \phi \frac{\partial \rho}{\partial t} + \phi \frac{\partial}{\partial t} \left. \frac{\partial \rho}{\partial p} \right|_{p=p_0} (p - p_0) \quad (6)$$

The stability of the pressure equation can be increased by linearizing the laplacian term. This is possible by applying a Fréchet derivative operator on the term, such that

$$\frac{\delta F}{\delta p} = \frac{\delta}{\delta p} \nabla \cdot \left(\frac{\rho \kappa}{\mu} \nabla p \right) = \nabla \cdot \left(\frac{\kappa}{\mu} \frac{\partial \rho}{\partial p} \nabla p \right) + \nabla \cdot \left(\frac{\kappa}{\mu} \rho \nabla \right) \quad (7)$$

this can then be applied to a first order Taylor expansion of the function

$$\begin{aligned} p &= p_0 + \left. \frac{\delta F}{\delta p} \right|_{p=p_0} (p - p_0) \\ &= \nabla \cdot \left(\frac{\kappa}{\mu} \rho \nabla p \right) + (p - p_0) \nabla \cdot \left(\frac{\kappa}{\mu} \frac{\partial \rho}{\partial p} \nabla p_0 \right) \end{aligned} \quad (8)$$

Having applied both the Taylor expansion of the time derivative and the linearization of the laplacian term, the pressure equation finally becomes

$$\begin{aligned} \phi \frac{\partial \rho}{\partial t} + \phi \frac{\partial}{\partial t} \left. \frac{\partial \rho}{\partial p} \right|_{p=p_0} (p - p_0) - \nabla \cdot \left(\frac{\kappa}{\mu} \rho \nabla p \right) \\ - (p - p_0) \nabla \cdot \left(\frac{\kappa}{\mu} \frac{\partial \rho}{\partial p} \nabla p_0 \right) + \nabla \cdot \left(\frac{\rho^2 \kappa}{\mu} \vec{g} \right) = 0 \end{aligned} \quad (9)$$

The equation of state is implemented from the IAPWS-IF97 thermodynamic formulation [IAPWS, 2007]. The primary variables for the equation of state are taken to be pressure and enthalpy, since pressure-temperature formulation has been shown to have more difficulties close to the critical point [Ingebritsen et al., 2010]. Given those two state variables the algorithm returns the steam quality x , the density ρ , the temperature T and the partial derivatives of all those variables both with respect to pressure and enthalpy. Those values are then used in the system equation, which makes it possible to solve for each timestep.

2.2 Analytical solution

The problem which Pruess derived an analytical solution to [Pruess and Celatis, 1987] concerns cold water injection into a reservoir initially filled with superheated steam.

If the reservoir is assumed to be infinite and the initial conditions to be uniform, this problem can be shown only to depend on the similarity variable $\eta = r^2/t$, where r is the radial distance from the well, and t is the time. Therefore the boiling front must occur at a fixed value $\eta = \eta_f$ and by extension, the front temperature and pressure must be time-independent.

Pruess gives the solution in terms of pressure at the front p_f and the boiling fraction $b = q_{vf}/q_l$, where q_l is the mass flow rate of liquid water into the vapor zone and q_{vf} is the mass flow rate of vapor at the boiling front. For the vapor flow Pruess applies mass balance, Darcy's law and the real gas law, which gives the pressure at the front in the following terms [Pruess and Celatis, 1987]

$$p_f^2 = p_0^2 - \frac{ZR_s T_0 \mu_v q_{vf}}{2\pi \kappa H} \exp\left(\frac{r_f^2}{4\alpha t}\right) \text{Ei}\left(\frac{-r_f^2}{4\alpha t}\right) \quad (10)$$

where Ei is the exponential integral and H is the depth of the reservoir. The compressibility factor Z is evaluated at the initial steam temperature T_0 and the front pressure p_f . The viscosity μ_v is also evaluated at T_0 but an average pressure between the initial and front pressure is used, such that

$$\bar{p} = \frac{1}{2}(p_f + p_0). \quad (11)$$

The diffusivity parameter α in equation 10 is given as

$$\alpha = \frac{\bar{p} \kappa}{\phi \mu_v}. \quad (12)$$

Using mass conservation it is possible to derive a relation for the location of the front in terms of the similarity variable $\eta_f = r_f^2/t$. Looking at the total liquid mass present in the injection plume gives the following

$$M_l(t) = (q_l - q_{vf})t = \pi r_f^2 H \phi \bar{\rho}_l \quad (13)$$

$$\Rightarrow \eta_f = \frac{r_f^2}{t} = \frac{(1-b)q_l}{\pi H \phi \bar{\rho}_l} \quad (14)$$

where $\bar{\rho}_l$ is the average liquid density in the single phase liquid plume. Neglecting small pressure effects it is given as follows

$$\bar{\rho}_l = \omega \rho_l(T_{inj}) + (1-\omega) \rho_l(T_f) \quad (15)$$

where ω is the retardation factor which is the ratio of cold volume to total swept volume. The

relation for the retardation factor is given by [Bodvarsson, 1972] as follows

$$\omega = \frac{\phi c_{l,h} \rho_{l,h}}{(1 - \phi) \rho_s c_s + \phi c_{l,h} \rho_{l,h}} \quad (16)$$

where $c_{l,h}$ and $\rho_{l,h}$ are respectively the heat capacity and density of the liquid water that has been heated by the superheated steam.

Pruess then applies heat balance at the front, which is derived from the fact that temperature and pressure at the front must be related by the vapor pressure relationship for water

$$p_f = p_{\text{sat}}(T_f) = p_{\text{sat}} \left(T_0 - \frac{h_{vl} \phi \bar{\rho}_l}{(1 - \phi) \rho_s c_s (b^{-1} - 1)} \right). \quad (17)$$

This set of non-linear equations can be solved in various ways, Pruess uses a Newton-Raphson method, while this paper applies a trust region dog-leg method [Powell, 1968]. In order for the solution method to function robustly, the front pressure (p_f) has to be scaled, such that it is on the same order of magnitude as the boiling fraction b . This is done by defining the non-dimensional pressure in the following way

$$p_{nd} = \frac{p_f - p_0}{p_{\text{max}} - p_0} \quad (18)$$

where p_{max} is an estimation of the maximum pressure in the reservoir.

The physical properties are retrieved from a C++ implementation of the IAPWS-IF97 thermodynamic formulation [IAPWS, 2007]. The primary variables are taken to be pressure and temperature which return saturation pressure $p_{\text{sat}}(T)$, density ρ , heat capacity c , viscosity μ_v and the compressibility factor Z .

2.3 Discretization and boundary conditions

The numerical solution is calculated on a axisymmetrical grid with logarithmically spaced cells, with the boundaries of the domain at r_0 and r_N . The logarithmic spacing is constructed by making an equally spaced vector

$$x_i = \log_{10}(r_0) + \frac{i}{N} (\log_{10}(r_N) - \log_{10}(r_0)) \quad (19)$$

$$i = 0, \dots, N$$

Table 1: Numerical values for problem parameters

param.	value
κ	$5 \cdot 10^{-14} \text{ m}^2$
H	200 m
r_0	1 m
r_N	2500 m
ρ_s	2600 kg/m ³
c_s	920 J/kg/K
Γ	5 W/m/K
h_{vl}	2260 kJ/kg
q_l	27.8 kg/s
ϕ	8 %
p_0	600 kPa

which then gives a logarithmically spaced r_i vector

$$r_i = 10^{x_i}, \quad i = 0, \dots, N. \quad (20)$$

In the numerical solution, the injection well has to have some finite radius, r_0 . If the injection rate is given as the mass flow rate q_l the appropriate Neumann pressure boundary condition can be found from Darcy's law, such that

$$\vec{u} = \frac{\kappa}{\mu_{inj}} \nabla p \quad (21)$$

$$\Rightarrow A \rho_{inj} \vec{u} = \frac{A \rho_{inj} \kappa}{\mu_{inj}} \nabla p \quad (22)$$

$$\Rightarrow q_l = 2\pi H r_0 \rho_{inj} \kappa \mu_{inj}^{-1} \nabla p \quad (23)$$

$$\Rightarrow \frac{\partial p}{\partial r} = \frac{q_l \mu_{inj}}{2\pi H r_0 \rho_{inj} \kappa} \quad (24)$$

In order to be able to compare the numerical solution to the analytical one, the outer radius is made sufficiently large, as to approximate boundary conditions at infinity. A Dirichlet boundary condition for pressure is applied there with $p = p_0$. Dirichlet boundary conditions are assumed at both sides for the enthalpy, where a temperature of $T_{inj} = 302.55 \text{ K}$ at the well gives $h_{inj} = 1.258 \cdot 10^2 \text{ kJ/kg}$. At the boundary the temperature is $T_0 = 513.15 \text{ K}$ which corresponds to enthalpy of $h_0 = 2.937 \cdot 10^3 \text{ kJ/kg}$.

3 Results

In order to validate the OpenFOAM model, a case where cold water was injected into a reservoir with

superheated steam, was set up and then compared to the analytical solution. The parameters which were chosen for the problem are given in table 1.

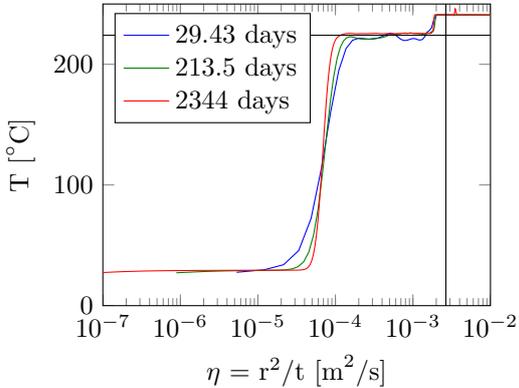


Figure 1: Temperature as a function of the similarity variable $\eta = r^2/t$, along with the analytical solution for the boiling front.

Figure 1 shows the temperature of the reservoir as a function of the similarity variable $\eta = r^2/t$. The figure shows the temperature profile after approximately 29.5 days, 213.5 days and 2344 days. The cross shows the location of the analytical solution as given by equations 10 and 17.

The numerical solution confirms that the location of the front in terms of the similarity variable $\eta_f = r_f^2/t$ is in fact constant. Even though that the time differs by a factor of 10, η_f stays approximately the same. Even though that conduction effects are not taken into account in the numerical solution, it shows some clear signs of numerical diffusion at the interface between hot and cold single phase liquid. However it seems to get sharper as time progresses, probably due to the fact the width of the interface stays constant in terms of r .

The analytical solution gives the location of the front at $\eta_f = 2.663 \cdot 10^{-3} \text{ m}^2/\text{s}$ with temperature of $T_f = 224.1 \text{ }^\circ\text{C}$. When compared to the numerical solution at at time $t = 2344$ days the temperature of the front is approximately $T_f = 225.6 \text{ }^\circ\text{C}$, which gives a $\Delta T_f = 1.5 \text{ }^\circ\text{C}$. For the location of the front there is slightly more difference, where the numerical solution at the same time gives $\eta_f = 1.953 \cdot 10^{-3} \text{ m}^2/\text{s}$ which results in a difference of $\eta_f = 7.1 \cdot 10^{-4} \text{ m}^2/\text{s}$. This could be due to the fact that in the analytical case, the injection well

is infinitely small and the reservoir infinitely large. Numerical constraints result in a finite value for both those quantities, which might result in some small inaccuracies when comparing them to the analytical solution.

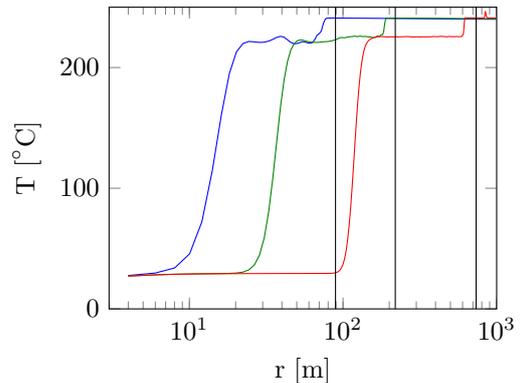


Figure 2: The temperature of the reservoir as a function of distance from the injection well. The temperature profile is given at three different times, after 29.43 days, 213.5 days and 2344 days.

Figure 2 gives a more physical representation of the solution by plotting it as a function of radial distance from the well. In that case it can clearly be seen how the front propagates in time. The vertical lines show the location of the analytical solution for each respective time and seem to show a rather good agreement between the analytical solution and the numerical one.

Figure 3 gives again a more physical representation of figure 1, but now in terms of time. The phase change from superheated steam to liquid water occurs after 12, 23 and 58 days for the respective distances from the well. After that the hot water slowly cools down as the injection into the well is continued and finally reaches the injection temperature of $29.4 \text{ }^\circ\text{C}$.

4 Discussion

This paper has described the applicability of the OpenFOAM framework to take on problems involving phase change in hydrothermal systems. In particular the results show that the numerical model is able to simulate the overall behavior of

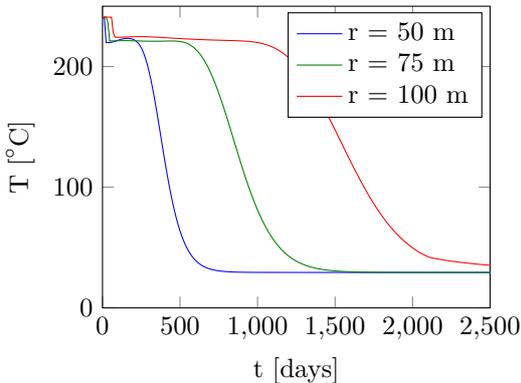


Figure 3: The temperature of the reservoir as a function of time for three different distances from the well.

the benchmark problem of a boiling front, propagating through porous medium, sufficiently well. This is confirmed by comparing the numerical solution $(\eta_f, T_f)_{\text{num}} = (1.953 \cdot 10^{-3} \text{ m}^2/\text{s}, 225.6 \text{ }^\circ\text{C})$ to an analytical solution $(\eta_f, T_f)_{\text{an}} = (2.663 \cdot 10^{-3} \text{ m}^2/\text{s}, 224.1 \text{ }^\circ\text{C})$, which seem to be in rather good agreement.

However both the numerical and analytical solution provided in this paper seems to differ from the solution provided in [Pruess and Celatis, 1987] which uses the same parameters. One explanation for that difference could be that the papers use different thermodynamic formulations for the physical properties. While this paper uses IAPWS-IF97 for the physical properties of steam and water, [Pruess and Celatis, 1987] uses the IFC-67 standard.

The ability of the OpenFOAM solver to handle phase change in porous media has also been demonstrated. Phase change can often result in numerical instabilities as displayed by [Thorvaldsson and Palsson, 2013]. These instabilities however do not seem to pose a problem in this paper. This is largely due to the fact that by applying Fréchet derivatives a more stable solution can be reached in the two phase region, and in the single phase region an unconditionally stable solution is made possible.

5 Conclusion

This paper has described the applicability of the OpenFOAM framework to take on problems involving phase change in hydrothermal systems. A benchmark case involving a boiling front moving through porous medium, which there exists an analytical solution of, was used to validate the results. The numerical model was also able to simulate the overall behavior of the boiling front sufficiently well. This increases our confidence in the numerical solver outside the limits of the analytical solution, and encourages continued usage of the OpenFOAM framework in geothermal reservoir modeling.

6 Acknowledgements

The authors would like to thank Eimskip University Fund, Landsvirkjun Energy Research Fund and Geothermal Research Group (GEORG) for their financial support of the project.

References

- [Bodvarsson, 1972] Bodvarsson, G. (1972). Thermal problems in the siting of reinjection wells. *Geothermics*, 1(2):63–66.
- [Henry, 1964] Henry, H. R. (1964). Effects of dispersion on salt encroachment in coastal aquifers. *Sea water in coastal aquifers*, 70.
- [IAPWS, 2007] IAPWS (2007). Revised Release on the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam. Technical Report August 2007, Lucerne, Switzerland.
- [Ingebritsen et al., 2010] Ingebritsen, S., Geiger, S., Hurwitz, S., and Driesner, T. (2010). Numerical simulation of magmatic hydrothermal systems. *Reviews of Geophysics*, (2009):1–33.
- [Jasak et al., 2007] Jasak, H., Jemcov, A., and Tukovic, Z. (2007). OpenFOAM : A C++ Library for Complex Physics Simulations. In *International Workshop on Coupled Methods in Numerical Dynamics*, volume m, pages 1–20.

- [Keating et al., 2002] Keating, G. N., Geissman, J. W., and Zyvoloski, G. A. (2002). Multiphase modeling of contact metamorphic systems and application to transitional geomagnetic fields. *198:429–448*.
- [Lewis and Lowell, 2009] Lewis, K. C. and Lowell, R. P. (2009). Numerical modeling of two-phase flow in the NaCl-H₂O system: Introduction of a numerical method and benchmarking. *Journal of Geophysical Research: Solid Earth*, 114(B5).
- [Matthäi et al., 2007] Matthäi, S. K., Geiger, S., Roberts, S. G., Paluszny, A., Belayneh, M., Burri, A., Mezentsev, A., Lu, H., Coumou, D., Driesner, T., and Heinrich, C. A. (2007). Numerical simulation of multi-phase fluid flow in structurally complex reservoirs. *Geological Society, London, Special Publications*, 292(1):405–429.
- [O’Sullivan, 2001] O’Sullivan, M. J. (2001). State of the art of geothermal reservoir simulation. *Geothermics*, 30(4):395–429.
- [Podgorney et al., 2011] Podgorney, R., Huang, H. C., Lu, C., Gaston, D., Permann, C., Guo, L., and Andrs, D. (2011). FALCON: A Physics Based, Massively Parallel, Fully Coupled, Finite Element Model for Simultaneously Solving Multiphase Fluid Flow, Heat Transport, and Rock Deformation for Geothermal Reservoir Simulation. Technical report, Idaho National Laboratory.
- [Powell, 1968] Powell, M. J. D. (1968). A FORTRAN subroutine for solving systems of non-linear algebraic equations. Technical report, Atomic Energy Research Establishment, Harwell (England).
- [Pruess, 1991] Pruess, K. (1991). TOUGH2 — A General Purpose Numerical Simulator for Multiphase Fluid and Heat Flow. Technical Report May, Lawrence Berkeley Laboratory, Berkeley, California.
- [Pruess and Celatis, 1987] Pruess, K. and Celatis, R. (1987). An analytical solution for heat transfer at a boiling front moving through a porous medium. *30(12):2595–2602*.
- [Stanford Geothermal Program, 1980] Stanford Geothermal Program (1980). Proceedings of the Special Panel on Geothermal Model Intercomparison Study. In *Report SGP-TR-42*, Stanford, CA.
- [Theis, 1935] Theis, C. (1935). The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage. *Transactions - American Geophysical Union*, 16:519–524.
- [Thorvaldsson and Palsson, 2013] Thorvaldsson, L. and Palsson, H. (2013). A numerical analysis on flow in hydrothermal systems. In *Stanford Geothermal Workshop*.
- [Weller et al., 1998] Weller, H. G., Tabor, G., Jasak, H., and Fureby, C. (1998). A tensorial approach to computational continuum mechanics using object-oriented techniques. *Computers in Physics*, 12(6).

Electric Transmission and Distribution Loss Reduction Strategies for Iran's Electricity Energy Sector

Kouros Mousavi Takami Hassan Gjolinejad¹ Amin Mousavi Takami Erik Dahlquist¹
k.takami@aol.com Hassan.gholinejad@mdh.se erik.dahlquist@mdh.se

¹IST Department, Malardalen University, Box 721 23, Vasteras, Sweden

Abstract

Although Iran's energy sector has undergone substantial changes within the last 10 years there remain many issues related to the development of an economically cycle combined power plants, sustainable energy sector and electricity losses reduction in distribution and transmission systems.

Low losses in world-class utilities are the result of economic system design, as well as hard work on metering, rate-making, billing, collection and energy theft processes.

Overall losses in Iran are around 27%, roughly evenly split between technical and commercial losses. Average losses in the world are about 8.3%. Author believes that a reduction in energy theft is likely to result in reduced electricity consumption, rather than increased revenue, due to the fact that electricity- bills represent a substantial share of typical family and business income.

Then Iranian electric system has very high rates of technical and commercial losses in transmission and distribution networks. Energy consumption in Iran's steam power plant is around 7% and needs to save more. It is studied in the presented paper.

This paper discusses both technical and commercial losses. It presents a mathematical model for losses. The model is versus load factor, power factor, peak load, etc. Model can be used to find losses in each time and can predict future losses.

Finally a proposal package is presented to reduce such losses in Iranian electricity network. Package consists of rehabilitations, optimization, modernization and automation.

For future work, multivariable data analysis is used to find the model. Unscrambler software together with MATLAB are used to data processing. A simulation by COMSOL software is done to show how connections and conductors get higher resistance by overload.

1. Introduction

Iran is among the top 14 countries with higher capacity of installed electricity power plants in the world. Iran's electricity generation exceeds from 220000 million KWh in 2013. More than 30 million customers in Iran consume electricity with:

Domestic: 31.5%

General: 8.9%

Agricultural: 16.8%

Industrial: 34.4%

Commercial etc.: 6.5%

Lighting: 2%.

Average statistics per person for electricity generation in Iran at 2012 were:

Average consumption for domestic customers: 2622 KWh

Per capita consumption: 2552 KWh

Per capita generation: 3345 KWh

Iranian electricity network was connected to the most of neighbor countries.



Fig1. Connection of Iranian electricity network to neighbors countries.

Iran exported about 11.054 billion KWh electricity (about 800 million USD) in 2012 and its exchanged power reached to 1432 MW. Iran imported 3.9 billion KWh that is equal to 216 MW. Electricity exchanging had 8.61% rise rather than to 2011.

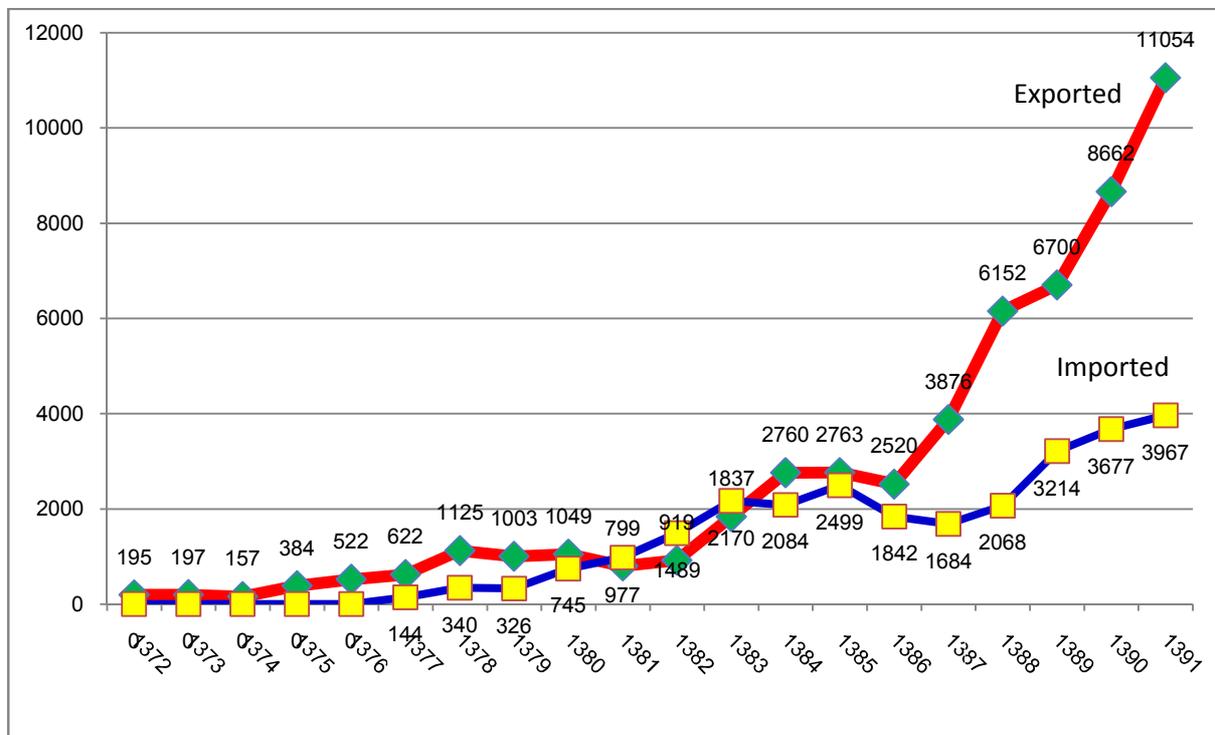


Fig 2. Exchanged electricity by neighbor countries since 1992 (1372) until 2012 (1391).

Iran's installed power plants exceed from 73000 MW. For 2012, it was about: 26000 MW Gas power plants, 16000 MW Combined Cycle, 16000 MW Steam, 11000 Hydro and others are Nuclear, wind, photovoltaic etc. In 2011, about 45% of generated power was produced by steam power plants.

Iranian fossil fuel power plants consumed about 142000 million liters in cold days and about 214000 million liters in hot days, in 2012. Totally power plants consumed equivalently 60000 million liters that about 75% was gas and other was different oils, in 2012. Sixty billion liter oil means 1033859 equivalent oil barrels per day. Each barrel is 159 liters.

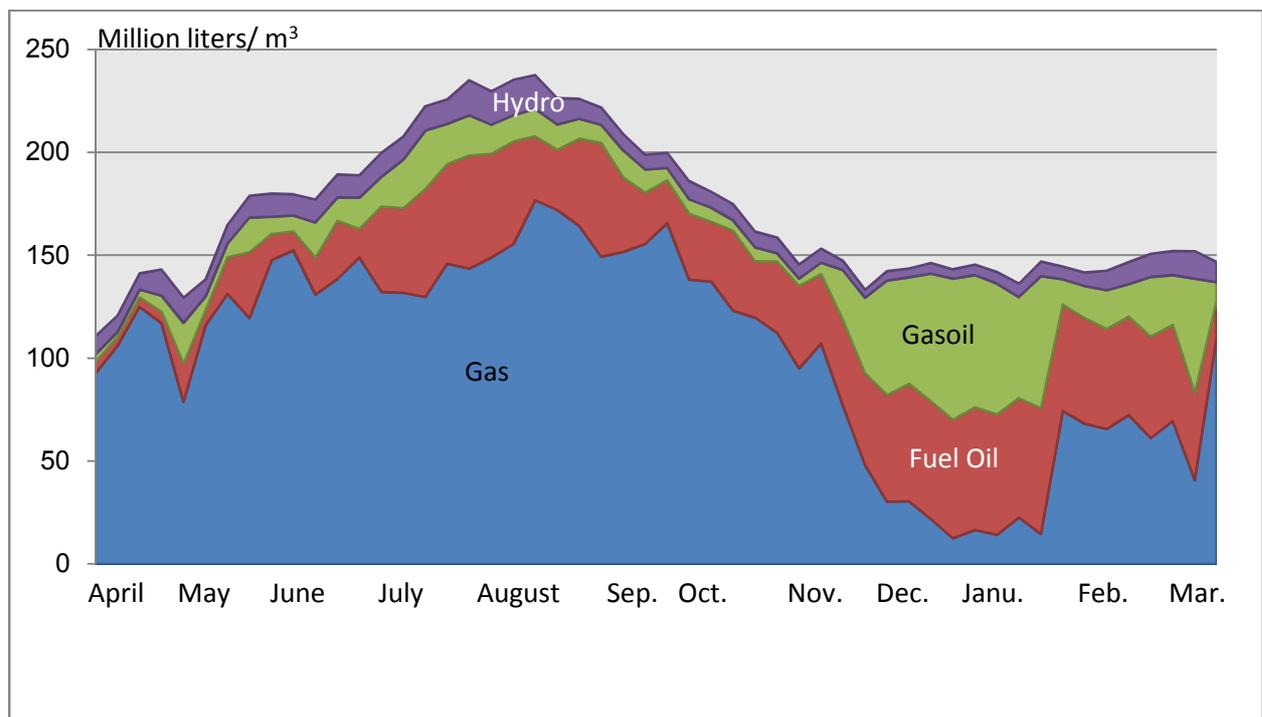


Fig 3. Type of consumed fuels in Iran's power plants, since April 2012 to March 2013.

Average electricity energy consumption (auxiliary) in Iranian steam power plants are about 7%, it is 0.5-2% for gas power plants.

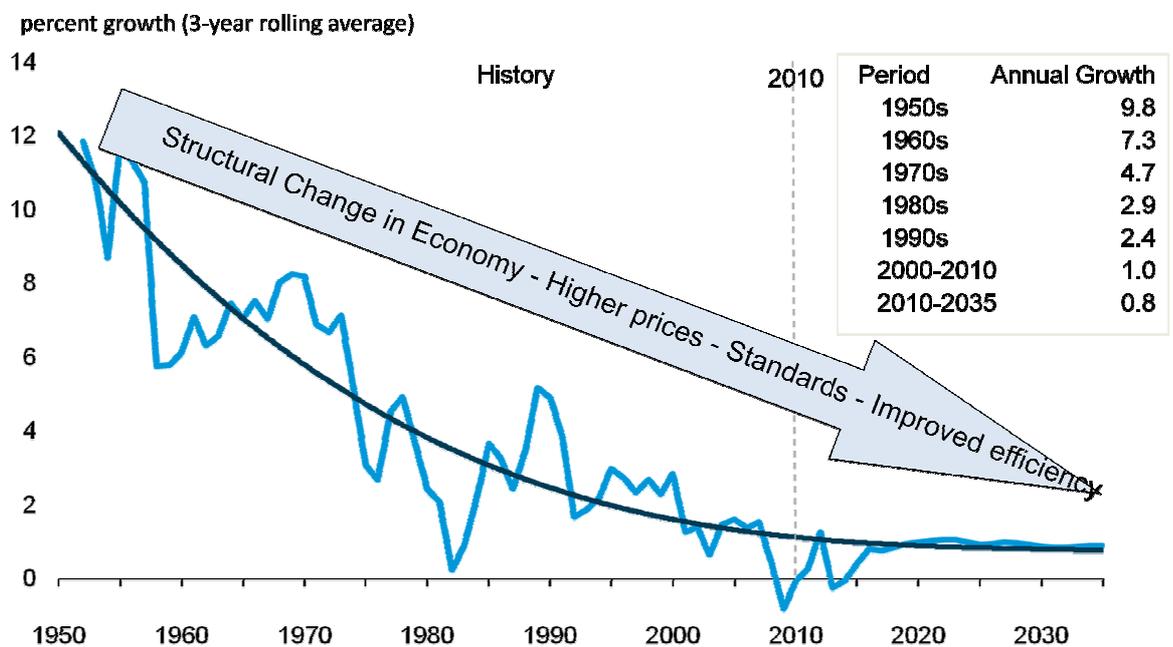
Paper topics included: possible causes of losses; loss estimation methods; meter testing procedures; and procedures for improving internal financial controls to reduce commercial losses.

The paper objectives were:

- To provide theoretical studying on technical and organizational measures and techniques to reduce both technical and commercial losses in Iran's electric transmission and distribution networks.
- To study on the effect of short time frequency controller and voltage stabilizers on the losses reduction.
- To present a method to reduce of energy consumption in power plants.
- To define a mathematical model for Iranian electricity losses, monthly or daily that needs to be taken to reduce losses in line with Government of Iran targets and sound utility practice.
- To present a proposal package to reduce of losses in Iranian electricity networks and to reduce green house gas generation in power plants as well.

1.1. Importance of attendance to electricity losses

According to the following graph, average percentage of annually growth of electricity plants in the world is decreasing. It means most of the world concerns are related to optimization, automation, distributed generation and losses reduction. Iranian electricity network is one of the high loss networks in the world. Iranian electricity demand growth is about 5-10% but networks losses is more than 27%. Iran's electricity consumption in 2012 was 195219 million KWh. Its total price with 7 cent/KWh is about 13.67 Billion USD. In other scenario, if Iran's electricity sector sold 60 Billion liter refined oil (that is used for fossil fuel power plants) with 108 USD/ barrel, it must be 40.8 Billion USD. About 60% of final cost for electricity is related to fuel, then for electricity sector total cost is 68 Billion USD. It means if Iran had imported electricity in 2012 must paid 13.67 million USD while expensed 68 Billion USD. If consider to the world average loss that is about 8.3, so about 19% of electricity is dissipated in Iran's networks. It is equal to 37091 million KWh. The mentioned value of extra loss produces 14 million tons CO₂ and lost energy by firing is more than 72 million oil barrels.



Source: EIA, Annual Energy Outlook 2012

Fig 4. Annual energy world outlook.

1.2. Material Covered

The paper was divided into three main areas.

The first part of the paper focused on reducing technical losses, including measuring and estimating energy and demand losses on both an annual and an hourly basis, and allocating their electric system components, using the Vattenfall or malarenergy or Tavanir as an example. This segment also provided a methodology for valuation of both energy and demand losses by voltage level, including methods for forecasting the value of future losses. Energy consumption in power plant is discussed and is proposed a method to save energy. Equipping every large electro motors such as FD fan, circulating water pump, feed water pump, condensing pump, etc. to variable speed drives are discussed.

The second part of the paper covered: commercial loss mitigation, including meter characteristics, accuracy, calibration and testing; billing systems and processes; electric rates and bills, using examples from Malarenergy or vattenfall or Tavanir (Iranian generation, transmission and distribution of electricity management company); revenue collection and non-payment issues; and energy theft and revenue protection including investigation, past due bill calculation, payment arrangements and criminal prosecution.

The final part of paper that will be as a future work included a demonstration of the hardware provided through the author's metering, billing and collection system project; a discussion of commercial losses in Iran. A mathematical model is presented for losses. It will be monthly, daily and hourly for each season.

1.3. Anticipated Outcomes

It is anticipated that some of the approaches presented will be implemented in Iran's distribution utilities, which should result in improved losses and reduced expenses in the Iranian power system.

The following model is an ideal model for Iranian network.

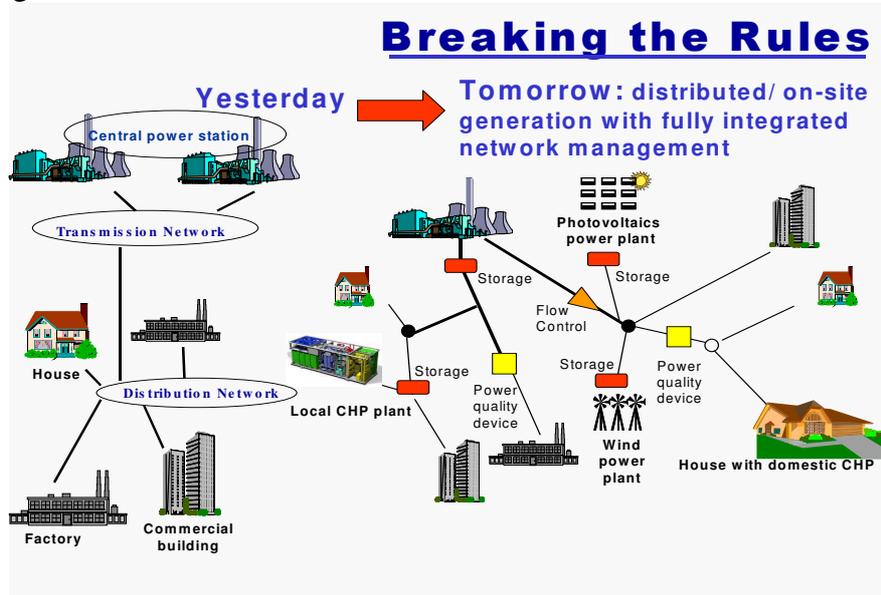


Fig. 5. Future configuration for distributed generation in Iranian network to reduce losses.

1.3.1. Achieve Goals of the Enterprise

Improve System Efficiency

- Reduce Technical Losses
- Improve Revenue Collection

Improve System Reliability

- Reduce Power Outages
- Improve Power Quality

Reduce Cost to Customers

- Improved System Efficiency
- Reduce Operation & Maintenance Cost

Achieve Financial Stability

- Improved Collections
- Reduced Debt Burden

1.4. Iran's Network Losses

Technical losses of power system components in Iran:

- Generator step-up banks 0.5-1%
- Transmission loss 1-4%
- Power transformers 1-2%
- Distribution lines 1-20%
- Distribution transformers 1-3%
- Secondary services 0.5-4%
- Metering 0.1-1%
- Energy Theft 0-15%

Total electricity losses for Iranian networks, according to claim of Iranian National grid (Tavanir), is shown in the following figure. Years are Iranian calendar, 1391 means 2012, 1384 means 2005 in Christian calendar.

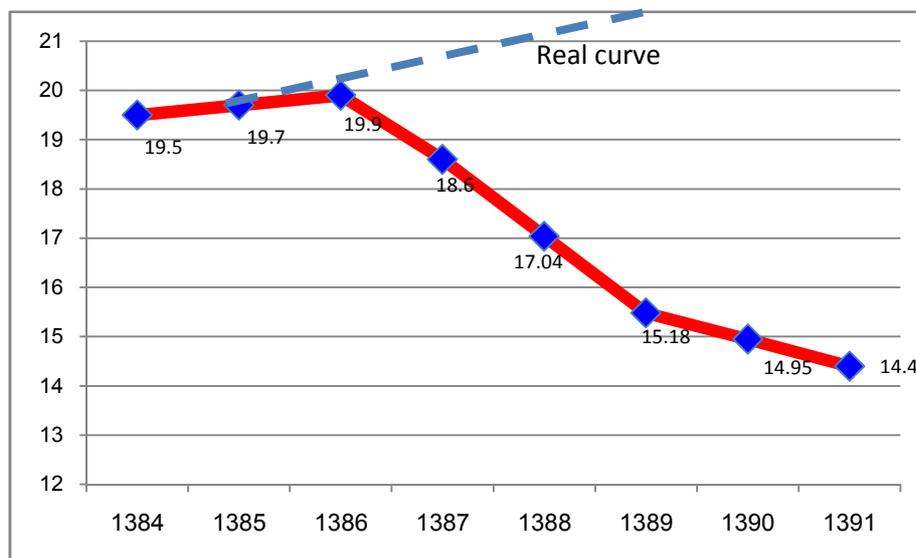


Fig. 6. Iranian governmental Tavanir presented losses diagram since 2005 (1384) - 2012 (1391).

In authors opinion, it is not real and it effected by mangers competition (to save their chair) in each period. The real losses in 2012 were more than 27%.

Authors are trying to find a method to reduce the losses until average international loss, 8.3%.

The following sections investigate the methods to reduce of auxiliary energy in power plants, losses reduction in transmission networks and losses management on distribution networks.

2. Power plants energy saving

As mentioned in the last section, most of steam power plants consume high electricity energy. Gas power plants often need to energy saving in mechanical environments. For example installing of cooling system in input air system before or after to compressor using Fog, Media, Swirl flash, chilling systems etc.

In general, air cooling methods for gas turbines are:

- Chiller cooling systems
 - Absorption Cooling Systems
 - Compression Cooling Systems
 - Thermal Storage Systems
- Evaporating systems

- (water injection)
 - Cooling over saturation,
 - Overspray
 - Hot Overspray
 - SwirlFlash
 - TOPHAT
 - 10-15% Water + Recuperation
 - Swirl Flash
 - 2% Water + Retrofit
 - Cold Overspray
 - Wet compression
- Cooling until saturation (input cooling)
 - Media
 - Fog

The following table shows an optional comparison between different types of cooling system for gas turbines in Abadan power plant.

Main features	Swirl flash (water injection with 40-150 bar and 150-250 °C)	Wet compression (cold water)	Evaporative cooling	Chilling
Power increase	10-15%	10-15%	5-10%	15-20%
Use at high humidity	Excellent	Excellent	Poor	Good
Use at low temperature	Excellent	Poor	Neutral	Neutral
Risk of blade corrosion	Low	High	Low	None
NOx emission reduction	High	High	Medium	Minor effect
Additional KWh/year	High	Medium	Low	Medium/high
Complexity	Medium	Medium/low	Low	High
Investment %	100	70	50-70	500-800
Increased fuel consumption for 7 hot months (m ³)	12153475	11465543	10258672	---
Increased water consumption for 7 hot months(m ³)	55576	52430	46911	--

The following curves are obtained with study on GE frame 9 Abadan gas power plant in south of Iran.

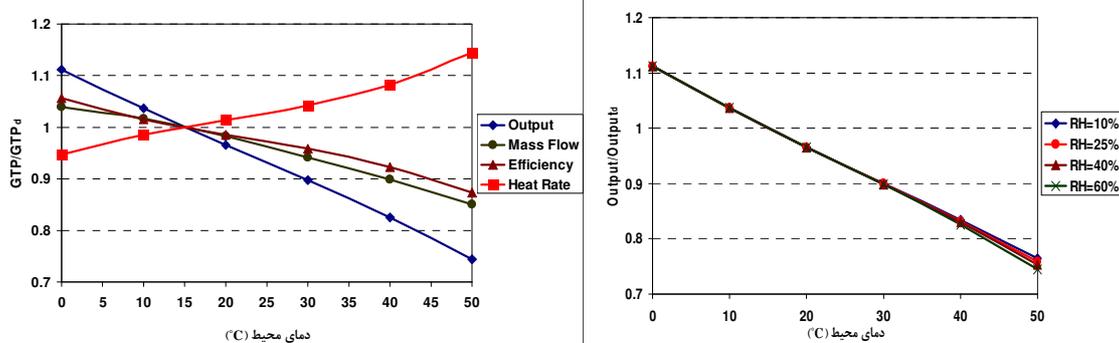


Figure 7. Temperature rise (left) and humidity (right) effect on gas turbine performance

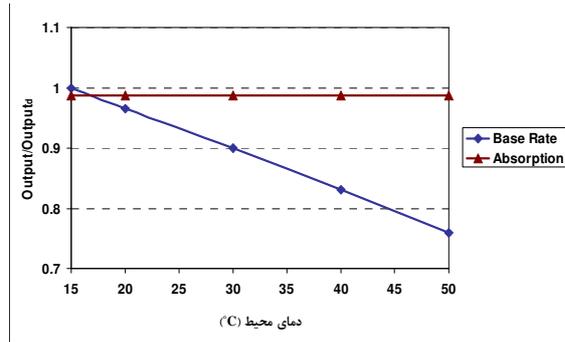


Figure 7. Effect of chiller cooling system on the gas turbine performance



Figure 8. Upstream fog system

swirl flash

In this study, mechanical method to increase of power plant efficiency is not authors concerns. But in electricity consumption management in steam power plants, several studies carried out on the energy saving. Because of poor proposed techniques, most of them had not been considerable to perform. However, some studies on forced draft fan (FD) showed that damper or vane are not good to control of input air volume to furnace, it needs to install drive on its electromotor. It is also need for CW pump, feed water pump, condensing pump etc. Also Jungstrom's (air pre heater) seal have considerable effects on the current of FD fan. In normal FD fans, about 8-14% air leakage between two different sides of jungstrom (hot and cold air) can be accepted, but in some power plants this leakage exceeds from 60%. It leads to over current about 40% and injected air to furnace reduces to 63%. Air volume reduction means 10% reduction to electricity generation. For example in 3*320 MW Bandar Abbas steam power plant, a 58% air leakage in one unit Jungstrom led to air reduction about 37% (air to furnace) and FD fan consumption 600 KW increased. Also just one unit electricity generation had 20-35 MW reduction.

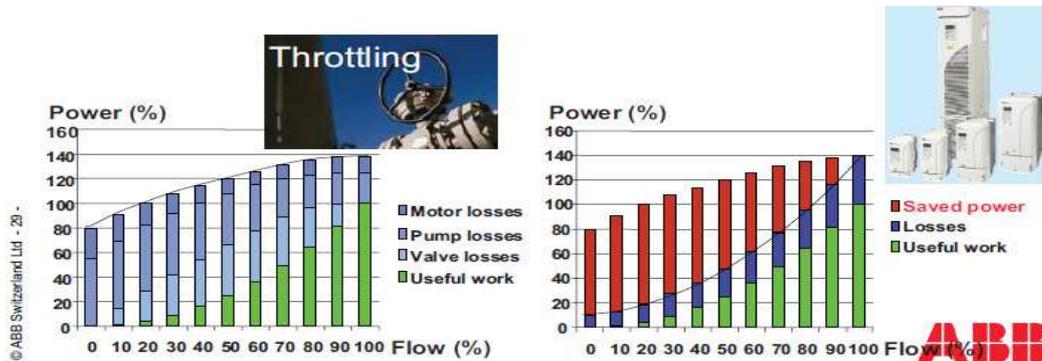


Figure 8. Energy saving with ABB drives (Curtsey by ABB)

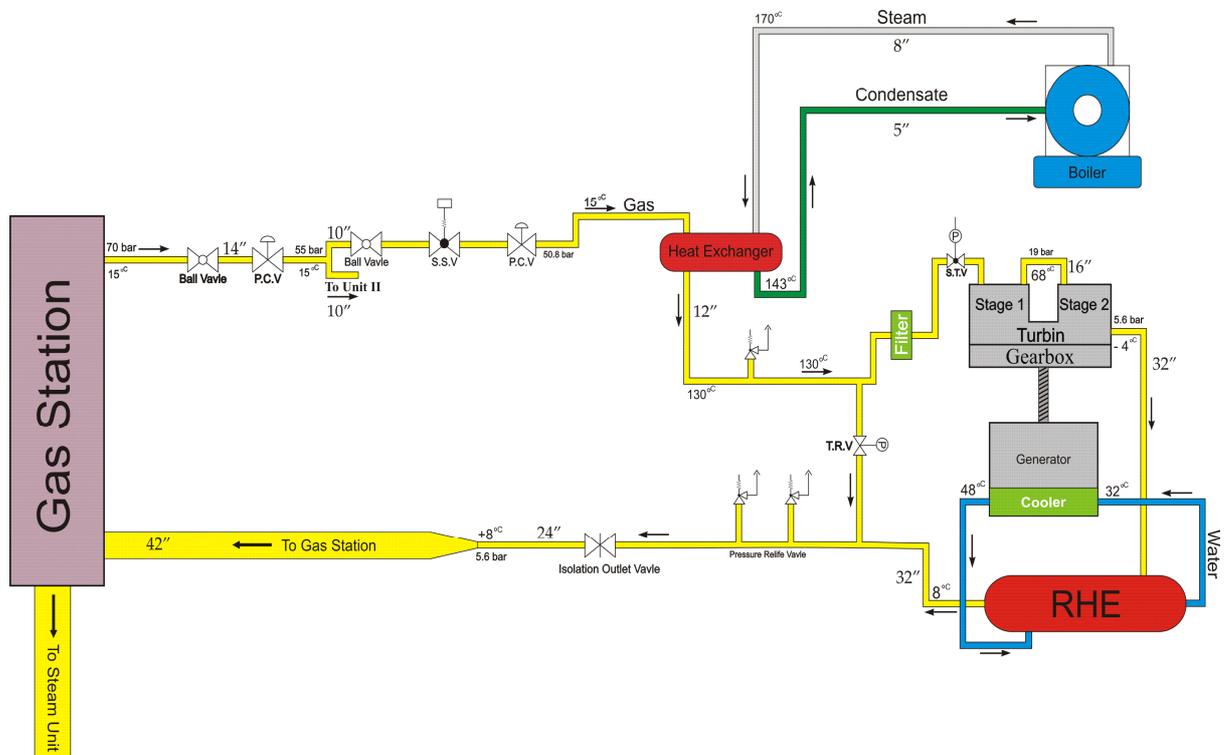
Leakage in feed water pump can increase electricity consumption in steam power plants. For example in Bea'sat power plant, to generate of 82.5 MW, feed water pump's electromotor nominal current in designing time was 465 A. At now, it reaches to 490 A. Pump's flow rate difference in to different sides were 39000 lb/h but it is 190000 lb/h, today.

Lighting in some power plants managed. For example, they used human sensor or timer for turbine generator site. Also lamps changed from tungsten or fluorescent type to LED lamps.

Most of electromechanical relays changed to digital types that have too much lower energy consumption.

One other method to increase of power plants efficiency and so self generation of auxiliariy electricity is to install of turbo expander in the path of gas high pressure lines to drive medium size generators. Gas is used as fuel for combustion and heat generation in boiler.

In Neka and two other power plants, two turbo expanders (for each power plant) in two paths of high pressure gas lines with 9.4MW capacity were installed. Most of in feed high pressure gas lines for Iranian power plants have Throttling Valves to reduce gas pressure. According to author's calculation, with installing of turbo expander in four Iranian power plants including of Rey (980MW in 29 Gas units), Bandar Abbas (4*320 steam), Tabriz (2*368 steam) and Toos (4*150MW steam), can generate 43 MW electricity with 22 millions USD capital investment. At now, Iranian high pressure gas lines network has a potential to run many turbo expanders. The capacity of this potential in cities or big industries entrances are about 800 MW electricity. It can produce electricity and 170 °c steam as well. See the following figure.



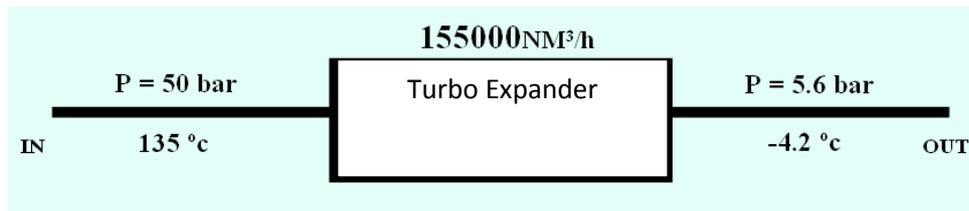


Figure 9. Turbo expander and its related circuits

Fuel consumption in Iranian Power plants in 2012 were 60 Billion Liters and it produced more than 24 million tones CO₂.

3. Energy saving in networks

Electrical network have three levels in Iran. Transmission (High voltage: 230 and 400 kV), sub transmission (63, 66, 132 kV) and medium & low voltage (33, 20, 11 and 0.4 kV). Transmission and sub transmission are discussed in a same way and medium and low voltage networks are studied in other form.

First of all have to formulate that How Are These Losses Determined? The answer:

1. Generation Input to the System

- a) Metered at Power Plants
- b.) Purchases from Other Utilities

2. Power Output to Customer

- a) Customer Revenue Meters
- b.) Sales to Other Utilities

3.1. Loss Reduction - Transmission and Substations

3.1.1. Transmission

- Correct Power Factor by installation of SVC and other FACT devices in the mid of lines.
- Reconductor Existing Circuits, rearrangement of towers cross arm. They must be like a triangle when 3 phases conductors are mounted under isolators. Transposition for longer lines and make bundle is well.
- Improve Voltage Plane. Using reactors, capacitors and thyristor switched devices it can be achieved.
- Balance Loads on Circuits

3.1.2. Substations

- Add Station VARs to Match Transformer
- Reactive Demand
- Replace Aging Transformers
- Retire Synchronous Condensers and Add Static Capacitor

3.1.3. Synchronous Condensers Study

Synchronous Condenser Status:

- There are no condensers in Iranian networks.

3.1.3.1. Problem

- High operation & maintenance costs, high losses

3.1.3.2. Study Objectives

- Reduce operation & maintenance costs and losses
- Maintain or exceed present quality of service to customers

For example, Rey gas power plant's efficiency is about 17%. This power plant can generate more than 900 MW, if all healthy units can be run. Authors propose to use the units of Rey power plant as synchronous condensers.

This plant has low efficiency and is not economic to generate electricity. Then a study procedure has to be defined.

3.1.3.3. Study Procedure

- Establish a criteria for synchronous condenser
- Identify function of each synchronous condenser
- Do economic analysis
- Evaluate operational effects of adding synchronous condensers
- Make recommendations.

3.2. Distribution network

3.2.1. Sources of Losses

Losses in distribution networks divided to two losses: technical and commercial losses.

A. Technical Losses

- Losses Which Vary with Electric Demand (I^2R)
- Losses Which are Constant (Magnetizing Currents)
- Losses Which Vary with Weather (Corona)
- Losses in connections, panels, customer tributary cables (because of low cross section), transformers, leakage in isolators and surge arresters, conductors, lighting etc.
 - Losses due to designing, long distribution networks (more than 400 meters distance from transformer), unbalanced loading, non symmetrical for system voltage, circulating current in network, low power factor, low voltage etc.
- Losses due to peak load, overload, demand.

B. Commercial Losses

- Uncollectible Accounts
- Energy Theft
- Measurement Errors
- Unmetered Accounts

3.2.2. Source of Errors in Measurement

A. Metering Inaccuracies

- At Power Plants
- Revenue Meters

B. Interchange Inaccuracies

- Allowance for Losses in Utility-to-Utility Transactions
- Allowance for Losses in Power Transmitted across Neighboring Power Systems

C. Timing Differences

- Between Power Transmitted and Meter Reading

D. Meter Reading Errors

3.3. Loss Reduction in Distribution System

Several methods are proposed to reduce losses in distribution network. Rehabilitation is the first option. Iranian medium and low voltage lines must be transferred to a proprietary tunnel. The using of self supporting cable with aluminum conductor should be stopped. At now all of overhead lines with low voltage copper conductors are replaced by 70 and 90 mm² self supporting cables. Those cross sections are not enough and increases losses. Each copper conductor must be replaced by an aluminum conductor with a 1.65 times cross section, because of lower conductivity of aluminum. Anyway, the following suggestions are proposed:

- Circuit Balancing
- Circuit Management
- Power Factor Correction
- Voltage Upgrade/cutovers
- Replace Oldest Transformers
- Reconductor
- Load Shaping/Demand-Side Management
- Maintain Voltage with Distribution
- Automation
- Add distributed generation
- Loose connection correction, it can find by a thermograph equipment. See the following figures.

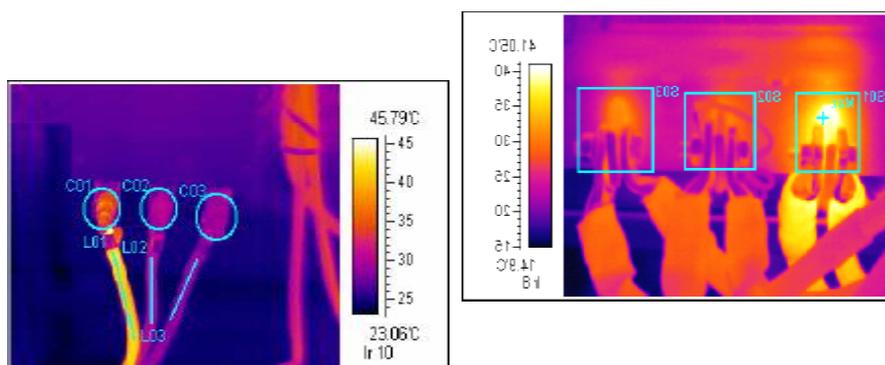


Fig 10. To find loose connection using thermograph methods.

Some devices are used to short time frequency and voltage compensation such as: flywheel, super capacitor, high energy batteries etc. but those have no priority in Iranian networks. Problems are deeper than to use of these devices.

3.4. Demand Side Management Choices

- Built-in incentives

- TOU rates
- Off-peak storage
- Swimming pool trippers
- Consumer education
 - Utility-actuated
- Heating and air conditioning cycles
- Interruptible load
- Demand subscription service
 - Conservation
- Price induced
- Programs
- Communication
 - Demand-Side Management Embraces, Several Load Shape Objectives
- Peak Clipping, Load Shifting
- Valley filling, conservation
- Growth, Flexible Load Shape
- Automatic Regulation of Customer Voltage and VAR Control

3.5. Load and loss factor

According to collected data for 2012, Load factor curve has been obtained.

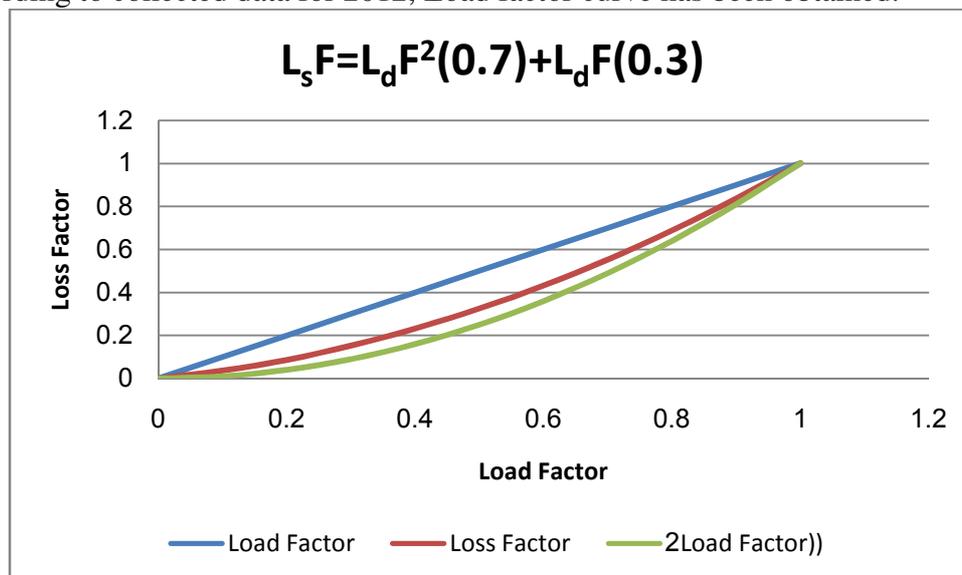


Fig 10. Load and loss factor

The calculation of load losses depends to the loss factor.

KWh of loss in a defined time= (Loss factor)*Hours* (Peak KW loss)/100

And loss factor is equal to $0.3 * \text{Load factor} + (1-0.3) * \text{Load factor}^2$.

Impact of customer load profile on cost of service and losses:

	Lower Cost per kwh	Higher Cost per kwh
Level of Service	Higher Voltage	Lower Voltage
Efficiency		
○ Peak Load	Off-Peak	On-Peak
○ Load Factor	High	Low
Amount of Usage	High	Low

3.6. Commercial Losses - Causes: Metering, Billing and Collections

Commercial losses are not a technical losses but those effect on energy revenues is very considerable. Metering devices are not calibrated for many years ago. Billing system is not modern and several mistakes exist in billing system. Energy theft has been considered by utilities but just self supporting cables are used instead of copper conductors. Self supporting cable are using aluminum conductors and when are used instead of copper conductors their cross section should increase 65% over same size of copper conductor. It is not noted by utilities because of several reasons.

- Metering and Meter Testing
- Billing and Billing Systems
- Collections and Non-payment
- Energy Theft

3.6.1. Sources of Metering Errors

As mentioned in last section, most of metering devices are installed in Iranian distribution networks at last 40 years. Those are electromechanical metering systems and are not calibrated. Those have high energy consumption. In other hands, many of current transformers are illegally changed by turn ration to show lower energy consumption. The following items have to be considered more:

- Current Transformers and potential transformers used for customers above 240 volt service
- Older jeweled-type meters
- Mechanical meter characteristics
- Calibration

3.6.2. Energy Theft Mitigation

Energy Theft Program

Most of energy is theft in border provinces of Iran. Khoozestan and Ilam provinces are much known for Iranian electricity. Tehran suburb has the same problem as well. Self supporting cable was the first action to prevent the electricity energy theft.

3.6.2.1.Prevention:

- Security Locking Rings, "Meter Sentries", Sealing, "Eagle Eye", Video "To Catch a Thief".
- Detection
Observation, Tip Cards, Computer Detection, Detection Bonus, Energy Theft Hotline, Tap Detector.
- Investigation:
Resealing Meters, Meter Testing, Service Investigations
- Prosecution
- Restitution
 - Revenue Recovery
 - Revenue Protection Group
 - Revenue Protection Manual
 - International Utilities Revenue
 - Protection Newsletter

4. Loss Estimation Methods with Incomplete Metering

Losses estimation is a necessity for the courtiers like Iran that have no a responsible metering and system. A method basis of Least Square is used to predict of losses. To build this system several data have to be acquired. A summarized procedure is explained in the following:

Make Estimates Using:

- Known Data to the Extent Possible

- Filling in the Blanks with Best Judgements

How to make "Best Judgements":

- Use Equipment Ratings where available
- Use typical Equipment Ratings
- Use computer simulations

**4.1. Determine and Allocate System Technical Losses
Using Incomplete Metering**

A. Determine Annual Energy Loss

- Determine/Estimate Energy Inputs and Outputs
- Metered and Estimated Data - Adjust for Billing Lag

B. Make Reasonable Estimates for

- Generator Step-up Banks
- Components where Field Measurement/Metering is not available

C. Transmission Network

- I. Use Power Flow Simulation and Load Factor/Loss Factor Approximations

D. Power Transformers

- II. Calculate Losses for a "Typical" Transformer, then Multiply by Number of Transformers

E. Distribution Circuits, Distribution Transformers, etc.

- III. Calculate Losses for a "Typical" Component by Voltage Class, then Multiply by Number of Components.

5. Cost of Energy for Use in Loss Analysis

Marginal Costs

- Based on Incremental Output of Marginal Generation
- Includes Price of Fuel, Generator Incremental Efficiency, Variable Operation and Maintenance

Expressed Hourly or Grouped in Time Periods of Similar Value

Includes Forecast of Future Values

- Determined from Computer Simulations

Values can be Used for Marginal Cost Ratemaking

5.1. Cost of Demand (or Capacity)

✓ **Marginal Costs of Capacity**

- Based on Plans for New Generation Capacity or Refurbishments
- Includes Installed Cost of New Capacity, Owners hip Costs (Financing, Depreciation, Taxes, Etc.), "Fixed" Operation and Maintenance, Inflation

✓ **Takes Into Account:**

- Probability That Losses Will Affect Timing Decisions on New Generation Capacity
- Reserve Margin

✓ **Expressed Annually or by Grouping in Time Periods with Similar Hourly Values.**

✓ **Calculated for Future Years**

✓ **Values Can Be Used for Marginal Cost Ratemaking**

5.2. Carrying Charges

Return of capital (i.e. depreciation)

+ Return of capital (i.e., interest and profit)

+ Taxes on return on capital

+ Operating expenses

6. Losses and System Design Criteria

Losses must consider in system design. The following items affected by losses.

- Economic Conductor Size
- Distribution Transformer
- Distribution System Voltage Level
- Placement and Sizing of Station Capacitors
- Placement and Sizing of Distribution Circuit
- Capacitors
- Author comment: Zero VAR flow at Substation
- **Voltage, Reactive Power and Losses**
- Minimum VAR Flow
- Minimum Voltage Fluctuation
- Minimum Losses

7. Conclusion;

Losses in Iranian electricity network shortly and summarily evaluated. Methods to identification, measuring and calculation of losses investigated. Electricity energy saving in power plant studied.

Future work:

Author's suggestion is to present a technical method to reduce greenhouse gases. Several optimizers such as BoilerMax[®] are developed by ABB to optimize of fuel in power plants. Author will study to suggest for installing in Iranian steam power plants.

Author will study on installation of distributed generation in Iranian distribution network in low and medium voltage levels. Smart grid and micro grid is studied and its effect on the losses, peak shaping and frequency control is evaluated. All of frequency control and voltage stabilization devices such as flywheel, high energy batteries, super capacitors, fuel cell and devices for sustainable generation like photovoltaic cell, small wind turbines and micro gas turbines are studied.

A simulation and multivariable data analysis is done for future work to find a well strategy to reduce Iranian electricity losses reduction.

8. References:

1. www.tavanir.org.ir, 2013.
2. K. Takami, optimization of energy consumption in Iranian steam power plants, 2012, workshop, Tehran, Tavanir, Bureau of energy consumption management.
3. EIA, annually energy outlook, 2012.
4. www.abb.com, 2013, August.