

**Proxy model-based closed loop reservoir management: A  
data driven approach**

by

Ajay Ganesh

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Process Control

Department of Chemical and Materials Engineering  
**University of Alberta**

©Ajay Ganesh, 2019

# *Abstract*

Steam assisted gravity drainage (SAGD) is a widely used thermally enhanced oil recovery process in western Canada. Petroleum reservoirs are large scale distributed parameter processes from a systems and control theoretic perspective. Physics-based first principles models pose severe computational challenges in the control and optimization studies; as a solution, proxy/surrogate models which are computationally light are preferred to design the optimization based control strategy. Caprock layer is the outer layer of the reservoir containing the dynamic steam chamber from erupting out to the surrounding. Maintaining the factor of safety (FoS) within the safety limit is crucial in operating the SAGD process. This thesis provides comprehensive analysis of steam chamber and FoS to enable the development of computationally efficient proxy models based closed loop reservoir management strategy. All of the techniques presented are developed from the data collected from a first principles-based commercial reservoir and geomechanical simulator CMG-STARs sequentially coupled with FLAC3D. The first proxy model predicts the caprock pressure and temperature fields using reduced-order dynamic modelling. The second performs dynamic analysis of FoS by modelling the evolution of caprock pressure clusters of high, medium and low pressure regions using graph theory and subspace modelling. System theoretic properties of these proxy models and their practical relevance is also analyzed. Next, a data-driven polynomial chaos expansion (PCE)-based proxy model is developed to provide quick and accurate estimation of caprock FoS along with the intention of propagating the

uncertainty in well bottom hole pressure inputs and reservoir petrophysical parameters to the caprock FoS. The next contribution develops a static and dynamic measures for the maximum operating pressure of the SAGD process respectively from the PCE-based FoS proxy model and a model predictive controller (MPC) to achieve FoS-constrained production optimized closed loop control strategy. The final contribution analyzes the system theoretic properties viz. controllability and observability with the intention of actuator and sensor placement and also provides a method of assessment of partial actuation and in-sensor ranges.

# Preface

This research work was carried out under the supervision of Dr. Vinay Prasad, CME and Dr. Richard J Chalaturnyk, CEE.

Chapter 2 of this thesis is a matured version of the work presented at CSChE 2015 and was accepted in November 2018 as A. Ganesh, J. Xiong, R. J. Chalaturnyk, V. Prasad. (2018) “Proxy models for caprock pressure and temperature dynamics during steam-assisted gravity drainage process” *Computers & Chemical Engineering Journal, Volume 121, 594-607*. I carried out the data analysis and proxy modelling over the SAGD simulation data; it was generated in collaboration with Dr. Jun Xiong in 2D-CMG-STARs. I wrote the first draft of the paper and Dr. Vinay Prasad and Dr. Richard Chalaturnyk reviewed and provided suggestions for improvements.

Chapter 3 of this thesis is a matured version of the work presented at CSChE 2017 and was submitted in August 2019 as A. Ganesh, B. Zhang, R. J. Chalaturnyk, V. Prasad. (2019) “Uncertainty quantification of the factor of safety in a steam-assisted gravity drainage process through polynomial chaos expansion” *Computers & Chemical Engineering Journal*. I carried out the data analysis and proxy modelling over the SAGD simulation data; it was generated by Bo Zhang in a 3D-CMG-STARs and FLAC3D sequential coupling platform. I wrote the first draft of the paper and Dr. Vinay Prasad reviewed and provided suggestions for improvements. The paper is still under review.

Chapter 4 uses the data and results from Chapter 3, I carried out the additional modelling and analysis over the data. I wrote the first draft of the chapter and was subsequently reviewed by Dr. Vinay Prasad.

Chapter 5 of this thesis is an improvised version of the work presented at IFAC

2018 and was submitted in August 2019 as A. Ganesh, Z. Li, R. J. Chalaturnyk, V. Prasad. (2019) “Graph-based determination of structural controllability and observability for pressure and temperature dynamics during steam-assisted gravity drainage operation” *Journal of Process Control*. This work is done in collaboration with the international undergraduate internship students, starting with Ankur Agarwal, Shivam Sharma, Pratyush Chakraborty, Kejie Lyu and Zixuan Li. I collaborated with each one of them and carried out the data analysis and proxy modelling over the SAGD simulation data generated by Dr. Jun Xiong in 2D-CMG-STARs. I wrote the first draft of the paper and it was edited by Dr. Vinay Prasad. The paper is still under review.

Chapters 1 and 6 were written by me and reviewed by Dr. Vinay Prasad.

# Gratitude

ॐ सह नाववतु । सह नौ भुनक्तु  
। सह वीर्यं करवावहै । तेजस्वि  
नावधीतमस्तु । मा विद्विषावहै  
ॐ शान्तिः शान्तिः शान्तिः ॥

- कैवल्य उपनिषद्

Let us all move together, let us work together and  
get nourished, and let us grow together without  
hatred, peace to one and all.

- Kaivalya Upanishad

ॐ असतो मा सद्गमय ।  
तमसो मा ज्योतिर्गमय ।  
मृत्योर्मा अमृतं गमय ।  
ॐ शान्तिः शान्तिः शान्तिः ॥

- बृहदारण्यक उपनिषद्

Oh Lord, drift me away from falsehood to the lap  
of truth, show me the path towards knowledge  
from the illusion, take me towards the blissful life  
from the shackles of death. Peace to one and all.

- Bruhadaranyaka Upanishad

ॐ नमः प्रणवाथोय  
शुद्धज्ञानैकमूर्तये ।  
निर्मलाय प्रशान्ताय  
दक्षिणामूर्तये नमः ॥

गुरवे सर्वलोकानां

भिषजे भवरोगिणाम् ।

निधये सर्वविद्यानां

दक्षिणामूर्तये नमः ॥

- दक्षिणामूर्ति स्तोत्रम्

Salutations to the divine Master for leading me  
towards the sphere of knowledge from the grasp  
of ignorance.

- Dakshinamurty Stotram

ಅಂದೊಡೆ ಶರಣ ಮೆರೆದೊಡೆ ಮಾನವ  
ಮರೆಯದೆ ಪೂಜಿಸು ಕೂಡಲಸಂಗನ !

- ಶರಣರ ವಚನ ೧೨ನೇ ಶತಮಾನ

*Dedicated to my wife, parents and to the fond memories  
of my late brother.*

# *Acknowledgements*

I wholeheartedly thank Prof. Vinay Prasad for his kind, friendly and effective guidance which led my way to success with flying colors. I am grateful to Prof. Richard Chalaturnyk for providing elegant guidance, generous funding and industrial exposure which made my PhD life sustainable and comprehensive. Starting from Day 1 to the last day Dr. Vinay has been wonderfully supportive and conducive for having numerous discussions and brainstorming, thanks to his open door policy, time management skills and patience in correcting my drafts. Dr. Vinay nurtured the qualities of instructor and researcher in me and provided with ample of opportunities to sharpen it. I also whole heartedly acknowledge the emotional support that Dr. Vinay offered me during tough times be it during the struggle with my courses or personal problems. I gratefully acknowledge the financial support from the Energi Simulation Industrial Research Chair in Reservoir Geomechanics for Unconventional Resources, NSERC Canada through a Discovery grant, and the University of Alberta International Internship program.

I thank my wife Kavya Suresh for her unrelenting support in being with me during all the times be it rain or shine, I thank her for the patience and reliance through thick and thin of her student husband. I thank my parents for their continued endurance, I also thank my in-laws for their care and affection throughout. I thank my extended family members for all the support I got during ups and downs.

I like to thank my peers in the RGRG team, Bo Zhang for his great support in working with the coupled simulation environment; Song and Dr. Jun with the CMG-STARS, Nathan Diesman, Amir Haghi and Alireza Khani for all the other technical stuffs in Dr. Rick's team and Hope walls for her zealous support during all the fabulous RGRG events, be it with Dr. Rick's potluck or the industrial consortia. I would like to thank Rajan Patel and Tarang Jain for their help at times with CMG-STARS and the PCE work. I thank my peers Anjana (special thanks to her for proof reading the thesis), Fereshteh, Maryam, Sushmitha, Kaushik, Dr. Rahul, Gokul, Arun Senthil, Dr. Kiru, Dr. Nabil, Dr. Hariprasad and Dr. Sarang Gumfekar for all the wonderful discussions and conversations I have had throughout. I specially thank Bharath KS and Dr. Sarang Pendharker for being great friends and for continuing their companionship from IIT Bombay, special mention to my high school friend Adarsh and his wife Chaithra in Calgary for the mature

and wonderful companionship, I (and Kavya) always relish the time we have spent with them as a family living nearby. I also like to thank my Edmonton Kannada Association friends for creating a small homeland far away from home. I thank all my dear friends and family whom I might have missed quoting here for all the support they offered. Last but not the least, my salutations to the Almighty for gracing me this opportunity to pursue my studies at U of A and bestowing on me enough strength to sail through.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Preface</b>	<b>iv</b>
<b>Gratitude</b>	<b>vi</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Complex systems, SAGD process and proxy modelling . . . . .	1
1.2 Thesis contribution . . . . .	5
1.3 Thesis outline . . . . .	6
References . . . . .	6
<b>2 Proxy models for caprock pressure and temperature dynamics during steam-assisted gravity drainage process</b>	<b>8</b>
2.1 Introduction . . . . .	9
2.2 Reservoir and caprock modelling . . . . .	11
2.2.1 Input Design . . . . .	13
2.3 Proxy modelling methods . . . . .	14
2.3.1 Proper orthogonal decomposition . . . . .	14
2.3.1.1 System Identification . . . . .	15
2.3.2 Clustering-based modelling . . . . .	17
2.3.2.1 Data Clustering . . . . .	17
2.3.2.2 State Space modelling . . . . .	18
2.3.2.3 State Space Parameter Estimation . . . . .	19
2.4 Results . . . . .	19
2.4.1 POD-based proxy models . . . . .	19
2.4.1.1 ARX modelling . . . . .	20
2.4.2 Cluster-based graph modelling . . . . .	24
2.4.2.1 Residual Error . . . . .	26
2.4.3 Controllability and Observability . . . . .	28
2.4.3.1 Observer with ground sensor measurement . . . . .	29
2.4.4 Temperature data analysis . . . . .	30

2.4.5	Practical considerations . . . . .	31
2.5	Comparison of proxy modelling methods . . . . .	33
2.6	Conclusions . . . . .	33
	References . . . . .	34
<b>3</b>	<b>Uncertainty quantification of the factor of safety in a steam-assisted gravity drainage process through polynomial chaos expansion</b>	<b>39</b>
3.1	Introduction . . . . .	40
3.2	Process modelling, data generation and visualization . . . . .	43
3.2.1	Data visualization . . . . .	44
3.3	Working procedure . . . . .	47
3.3.1	FoS computation . . . . .	47
3.3.2	Singular value decomposition: (SVD) . . . . .	48
3.3.3	Polynomial chaos expansion . . . . .	50
3.4	Results and discussion . . . . .	52
3.4.1	FoS computation . . . . .	53
3.4.2	SVD . . . . .	53
3.4.3	Polynomial chaos expansion . . . . .	57
3.4.3.1	UQ through moments of the FoS . . . . .	63
3.5	Conclusions . . . . .	66
	References . . . . .	66
<b>4</b>	<b>Factor of safety constrained model predictive controller design for closed loop reservoir management</b>	<b>71</b>
4.1	Introduction . . . . .	72
4.2	Model generation . . . . .	73
4.3	Proposed methodology . . . . .	73
4.3.1	Robust optimization based MPC formulation . . . . .	75
4.3.2	Stochastic optimization based MPC formulation . . . . .	76
4.4	Results and discussion . . . . .	77
4.4.1	Static measure based on FoS ensemble . . . . .	77
4.4.2	Dynamic measure based on MPC design . . . . .	78
4.4.2.1	Proxy models for FoS and OP . . . . .	78
4.4.2.2	Stochastic MPC design . . . . .	79
4.5	Conclusions . . . . .	83
	References . . . . .	83
<b>5</b>	<b>Graph-based determination of structural controllability and observability for pressure and temperature dynamics during steam-assisted gravity drainage operation</b>	<b>86</b>
5.1	Introduction . . . . .	87
5.2	Reservoir modelling and data generation . . . . .	89
5.3	Methods . . . . .	90
5.3.1	Regionalization-based clustering . . . . .	91

---

5.3.2	Granger causality-based graph construction . . . . .	94
5.3.3	Structural controllability and observability . . . . .	95
5.3.3.1	Driver nodes . . . . .	96
5.3.3.2	Root SCCs . . . . .	97
5.3.4	Evaluation of actuator and sensor reachability range . . . . .	98
5.3.4.1	Actuator reachability range . . . . .	98
5.3.4.2	Sensor reachability range . . . . .	99
5.4	Results and discussion . . . . .	100
5.4.1	Numbers of driver nodes and root SCCs and regionalization- based clustering . . . . .	101
5.4.2	Graph-theoretic analysis . . . . .	102
5.4.3	Complete structural controllability and observability . . . . .	105
5.4.4	Actuator and sensor range . . . . .	107
5.5	Conclusions . . . . .	108
	References . . . . .	109
<b>6</b>	<b>Conclusions</b> . . . . .	<b>114</b>
6.1	Conclusion . . . . .	115
6.2	Future work . . . . .	116
	Bibliography . . . . .	117
	Appendix . . . . .	128
.1	Appendix: Temperature data analysis . . . . .	128

# List of Figures

1.1	Reservoir modelling strategies . . . . .	3
1.2	SAGD process overview . . . . .	3
1.3	Proxy model based closed loop management of complex system . . . . .	5
2.1	System theoretic perspective of a reservoir . . . . .	9
2.2	Perspective view of reservoir structure with well bores and the viewing directions. The 2D data frame (cross-section in the reservoir) is shown and the inset shows a detailed view of the same with reference directions X and Z. . . . .	11
2.3	A detailed simulation layout of the 2D frame . . . . .	12
2.4	Shale distribution of heterogeneous reservoir . . . . .	13
2.5	Schematic of the POD based proxy modelling technique . . . . .	16
2.6	A directed and edge weighted line graph shows the interaction between the clusters. It governs the bin count dynamics where the edge weight is the index of the interaction. Only the incoming edge weight affects the node state. . . . .	18
2.7	The Image representation of the pressure data frames at various instants (180, 140, 110, 50 from top left to bottom right). The pressure values are scaled and mapped to colors and an image is drawn (refer color bar). Number of grids and the direction is also shown. . . . .	20
2.8	POD spatial basis visualization: Scales of all of the subfigures is same as of Figure (2.7) . . . . .	21
2.9	Logarithmic plot of the POD weights showing a steep decrease in the order of weights with their indices, the ratio of the 10 <sup>th</sup> singular value (marked on the curve) to the 1 <sup>st</sup> being order 10 <sup>-3</sup> and capturing 99.58% of the total energy can be seen. . . . .	22
2.10	Reconstruction with different number of POD basis; Colormap see Fig(2.7) . . . . .	22
2.11	System identification on training dataset result: Actual and modelled weight $w_1(l)$ , dotted line indicates identified model and solid line indicates the actual data. . . . .	23
2.12	System identification on validation dataset result: Actual and modelled weight $w_1(l)$ , dotted line indicates identified model and solid line indicates actual. . . . .	24
2.13	POD+ARX based reconstruction visualization . . . . .	24

2.14	A typical frame with clustered pressure grid points: four coloured regions corresponds to four regions having high pressure - brown, yellow clusters, medium pressure - cyan and low pressure -blue cluster (outermost cluster near to the ground) . . . . .	26
2.15	Variation of pressure centres of every cluster with frames. Colour map is same as Figure(2.14). Variation is centred about Blue - 803.23 KPa, Cyan - 1052.70 KPa, Yellow - 1302.2 KPa and Brown - 1551.63 KPa with a corresponding standard deviations of 63.79 KPa, 155.98 KPa, 248.74 KPa and 341.6 KPa . . . . .	27
2.16	Variation of normalized bin count of each cluster vs frames. Colour map is same as Figure(2.15) . . . . .	27
2.17	Outermost cluster bin count versus its pressure centres look-up function response . . . . .	29
2.18	Ground observer bin count error dynamics. The figure shows that the observer is able to catch up with the actual bin count over the course of time . . . . .	30
2.19	The image representation of different temperature data frames at different time instants (180, 140, 110, 50 from bottom right to top left).The temperature values are scaled and mapped to colors and an image is drawn. The reservoir has 5 injectors and 5 producers wells. Temperature is measured in degrees Celsius. . . . .	31
2.20	A typical frame with clustered temperature grid points: : Four colors corresponds to four clusters with high temperature 158.15 C - Brown, 116.38 C - Yellow, medium temperature 74.63 C - Cyan and low temperature 32.86 C - Blue with corresponding standard deviations of 56.86 C, 40.62 C, 24.39 C and 8.15 C. . . . .	32
3.1	Uncertainty quantification. . . . .	40
3.2	Schematic of a SAGD process showing two parallel wells (steam injector and the oil producer). The caprock layer shown at the top experiences stress due to the growth of the steam chamber. . . . .	41
3.3	Sequentially coupled simulation between the CMG-STARs and FLAC3D; the pore parameters change affects the stress strain which in turn affects the petrophysical parameters, which changes the pore parameters. . . . .	43
3.4	The working reservoir model in 3D with permeability I distribution in milli darcy. Well pairs 1-6 can also be seen. . . . .	45
3.5	Reservoir model schematic (XZ plane description) with grid size and numbers. Caprock: Clear water Formations; Reservoir: Wablskaw and McMurry formations. . . . .	45
3.6	Reservoir steam chamber visualization at different times. . . . .	46

3.7	Working data format: 200 realizations ( $\Xi_1 - \Xi_{200}$ ) of spatiotemporal FoS data taken as the XZ frames. Data at every instant $T$ has 200 grid points in $X$ , 200 grid points in $Y$ and 6 grid points in $Z$ taken over 29 time frames. Every realization $\{\Xi_k\}_1^{200}$ is of dimension $200 \times 6$ . The PCE model is computed over the realizations for each time instant. . . . .	47
3.8	Reservoir permeability field visualization at different times. . . . .	48
3.9	Overview of the work flow to develop a PCE model for the FoS field in a sequentially coupled reservoir simulation using SVD. . . . .	49
3.10	Representing each realization with a set of $r$ representatives obtained from SV decomposition of the reservoir permeability field. . . . .	50
3.11	PCE workflow algorithm. . . . .	51
3.12	Evolution of actual minimum FoS over different realizations and time. . . . .	54
3.13	Evolution of representative FoS (normalized highest SV) over different realizations and time. . . . .	54
3.14	Minimum FoS locations in X over realizations (Y) at Z=6 (the bottom layer of the caprock, i.e. the layer adjacent to the reservoir). These are the locations in the XY plane where the FoS is the least at different time instants. . . . .	55
3.15	Average permeability SVs over realizations vs their index and time. A sharp decrease at the knee point is clearly seen at the third SV, where $\frac{\sum_{i=1}^3 \xi_i}{\sum_{i=1}^{35} \xi_i} = 0.4912$ . Three SVs of each realization are used as its representative in PCE expansion. . . . .	56
3.16	Average FoS SVs over all realizations versus their index and time. A sharp decrease at the knee point is clearly seen at the second SV, where $\frac{\sum_{i=1}^2 S_i}{\sum_{i=1}^6 S_i} = 0.9995$ . Hence, at most two FoS SVs of each realization can capture 99.95% of the variance of the FoS field. . . . .	56
3.17	SVD-based FoS field reconstruction with two SVs; visualization at $t=1$ . . . . .	57
3.18	SVD-based FoS field reconstruction with two SVs; visualization at $t=10$ . . . . .	58
3.19	SVD-based FoS field reconstruction with two SVs; visualization at $t=25$ . . . . .	59
3.20	Performance of PCE order 1 regression and collocation models for $S_1$ at $t=10$ . . . . .	61
3.21	Performance of PCE order 2 regression and collocation models for $S_1$ at $t=10$ . . . . .	61
3.22	Performance of PCE order 3 regression model for $S_1$ at $t=10$ . . . . .	61
3.23	Performance of PCE order 1 regression and collocation models for $S_1$ at $t=25$ . . . . .	62
3.24	Performance of PCE order 2 regression and collocation models for $S_1$ at $t=25$ . . . . .	62
3.25	Performance of PCE order 3 regression model for $S_1$ at $t=25$ . . . . .	62

3.26	Performance of the PCE order 2 model in reconstructing the validation realizations of the representative FoS (minimum FoS) over all times. Only one SV is used in the reconstruction. . . . .	63
4.1	Process schematic for geomechanics based closed-loop reservoir control	73
4.2	Schematic of a MPC . . . . .	74
4.3	MPC control schematic . . . . .	75
4.4	Static MOP measure based on FoS ensemble generated from PCE .	78
4.5	Input/output data set to build proxy models. Variation of representative FoS $S_1$ (refer chapter 3), well BHP and OP over time. . .	79
4.6	MPC based MOP with the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5. .	80
4.7	Static MOP with the corresponding FoS and OP. The red line indicates the FoS setpoint at 1.5. . . . .	81
4.8	Fixed MOP at 1000KPa with the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5. . . . .	81
4.9	Fixed MOP at 2000KPa the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5. .	82
4.10	Fixed MOP at 1900KPa the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5. .	83
5.1	SAGD reservoir with schematic showing injector producer inputs, pore pressure/temperature and the measurement sensors on wells and at ground level. . . . .	88
5.2	The edges considered for regionalization . . . . .	92
5.3	Illustration of partitioning. . . . .	93
5.4	Partition algorithm flowchart . . . . .	94
5.5	Illustration of matched edges, matched nodes and driver nodes in an example directed graph. A set of edges (maximum) are said to be matched if they do not have a source (arrow going out) or sink (arrow coming in) node in more than one edge. Matched nodes are the ones pointed by the matched edges, remaining nodes are the driver nodes where the input is applied to achieve full controllability. 96	96
5.6	Illustration of an inference diagram with connected components, strongly connected components (SCCs) and root SCCs in an example directed graph. A graph can be split into largest connected components (sub-graphs) which are individually connected and having no edges amongst the connected components. In SCCs, all nodes are bidirectionally connected to each other including a self loop. SCC having no incoming edge is a root SCC. Sensors are placed in any one node of each of the root SCC to achieve full observability. .	97
5.7	Sensor/actuator locations in the neighbourhood of well pairs . . . .	98
5.8	Algorithm to find the control range. . . . .	99
5.9	Algorithm to find the observe range. . . . .	100

5.10	The variation of the spatial distribution of clusters with the number of clusters. Each color corresponds to a separate (spatially contiguous) cluster and the trend shows that the density of clusters increases in the central region, which is the region that exhibits large variations in pressure and temperature in the original data. . . . .	102
5.11	The graph model for the system (reservoir+caprock) based on the spatiotemporal pressure and temperature data, with nodes and edges identified. . . . .	103
5.12	(In- and out-) Degree distribution of the nodes of the graph as a function of the number of clusters (nodes) in the range 50-150. Red markers correspond to the top 20% nodes in terms of in- or out-degree, blue markers to the bottom 20%, and black markers to the middle 60% in terms of connectivity. . . . .	104
5.13	(In- and out-) Degree distribution of the nodes of the graph as a function of the number of clusters (nodes) in the range 200-300. Red markers correspond to the top 20% nodes in terms of in- or out- degree, blue markers to the bottom 20%, and black markers to the middle 60% in terms of connectivity. . . . .	105
5.14	Location of driver nodes for complete controllability as a function of the number of clusters. Each colour corresponds to a separate cluster identified as a driver node. Blank regions do not contain driver nodes. . . . .	106
5.15	Location of the root strongly connected components as a function of the number of clusters. Each colour corresponds to a separate cluster identified as a root SCC. . . . .	107
5.16	Actuation range for the actuator locations as shown in Figure 5.7. The graphs with higher numbers of nodes yield larger controllable regions. . . . .	108
5.17	In-sensor-range for the sensor locations as shown in Figure 5.7. The graphs with higher numbers of nodes yields larger observable regions. . . . .	109
1	Variation of temperature centres of every cluster with frames. Color map same as the temperature cluster . . . . .	128
2	The variation of normalized bin count of each cluster vs frames. Trend shows that, with time clusters with higher pressure value overtakes the lower ones . . . . .	128
3	Normalized bin count error (actual - estimated) for all frames . . . . .	129
4	POD+ARX based reconstruction visualization . . . . .	129
5	System identification on training dataset result: Actual and modeled weight $w_1(l)$ , dotted line indicates identified model and solid line indicates the actual data. . . . .	130
6	System identification on validation dataset result: Actual and modeled weight $w_1(l)$ , dotted line indicates identified model and solid line indicates actual. . . . .	130

# List of Tables

2.1	CMG-STARS simulation parameters $S_o$ : Oil saturation, $S_w$ : Water saturation . . . . .	12
2.2	System identification of POD weights. <sup>a</sup> Fit with 10 step ahead prediction, <sup>b</sup> Fit with one step ahead prediction. . . . .	23
2.3	POD verses Cluster Graph Comparison . . . . .	33
3.1	Hermite ( $H_i$ ) and Legendre ( $L_i$ ), $i^{th}$ order polynomials defined respectively over the intervals $[-\infty, \infty]$ and $[0, 1]$ alongwith their roots. . . . .	52
3.2	Regression and collocation PCE coefficients for order 1, 2 and 3 models with the corresponding basis at time t=10. . . . .	64
3.3	Regression and collocation PCE coefficients for order 1, 2 and 3 models with the corresponding basis at time t=25. . . . .	65
4.1	Comparison of different cases . . . . .	82
5.1	Number of driver nodes and root SCCs . . . . .	101

# Chapter 1

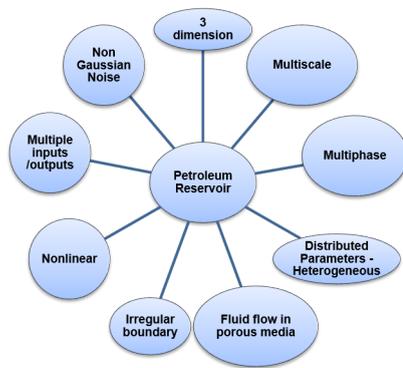
## Introduction

### 1.1 Complex systems, SAGD process and proxy modelling

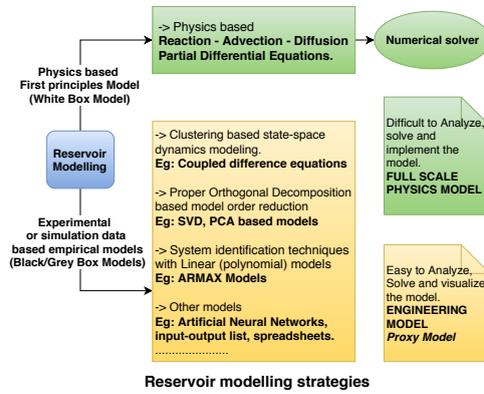
Proxy or surrogate modelling has been successfully used in many engineering disciplines, especially in control and optimization studies of complex processes; this requires tasks like gradient computation, multi-step ahead forecasting, soft sensor development and others[1]. Physics-based first principles models may not be feasible and/or may be computationally too costly to achieve these tasks. Weather forecasting models, ocean models and optimal organizational policy making processes are some of the examples of complex processes to use proxy modelling strategy [2, 3]. In this research work, we concentrate on petroleum reservoirs [4] which are a class of complex processes having a large geographical spread with three dimensional multiphase fluid flow through porous media parametrized by spatiotemporal variations in petrophysical parameters contained in an irregular boundary having multiple steam injection inputs and multiple oil production outputs; such processes can be termed as large scale distributed parameter processes. Figure 1.1(a) shows some of the important features which make the analysis challenging. Figure 1.1(b) shows the two major modelling strategies employed in practice; the physics-based modelling is carried out through the partial differential equations (PDEs) and the data-based proxy modelling is carried out using the concepts of statistical learning, design of experiments/simulations and system identification techniques. Physics-based models are difficult to analyze, solve and implement in practice but proxy models can alleviate these challenges. Figure 1.1(c) compares different types

of modelling techniques available in literature using the factors viz. complexity, validity, phenomena capture, scale and precision. Physics-based models on one end are highly complex but have global validity, captures micro/macro scale phenomena, solvable at any scale with great precision; on the other end, the crude models are highly simple but fail to meet all of these criterion. The proxy models have to find a 'sweet spot' between these two ends to provide optimal performance with a fair computational cost.

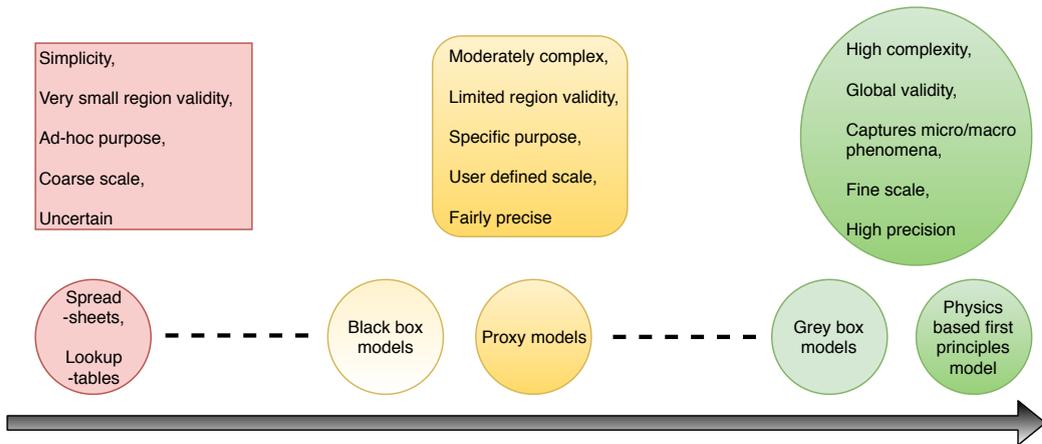
Figure 1.2 shows the schematic of a SAGD process [5] having two well-injector and producer running in parallel. Injector injects the steam which softens the bitumen to flow down due to gravity and gets collected by the producer. The injected steam creates a steam chamber which needs to be controlled. Out of many layers in the reservoir, the caprock layer, which is an impermeable layer holding the injected steam from erupting out to the surface, plays an important role in assessing the safety of the operation of the reservoir. Maintaining the caprock factor of safety (FoS) within the prescribed limits during the reservoir operation is crucial in adhering to safe operational standards. Deformations associated with the development of the steam chamber in the reservoir affect the FoS of the caprock significantly [6]. In this research work we present a proxy modelling based strategy to model and control the FoS of the caprock layer while optimizing the production parameters of the SAGD process completely from a data driven approach. The data is obtained from a first principles SAGD commercial simulator CMG-STARS [7] sequentially coupled with FLAC3D [8] for geomechanical analysis on the caprock layer [9]. To start with, we present two deterministic proxy models to capture the spatiotemporal pressure and temperature dynamics in the caprock layer. The first one is a proper orthogonal decomposition (POD)- based proxy model and the second one is a cluster graph-based proxy model. The first proxy model addresses the issue of reduced-order dynamic modelling of the caprock pressure and temperature fields based on POD and system identification using data from CMG-STARS. The second proxy model takes the first step towards dynamic analysis of FoS in reservoir management by modelling the evolution of clusters of high, medium and low pressure regions using graph theory and subspace modelling. System theoretic properties of these proxy models and their practical relevance is also analyzed.



(a) Physics based reservoir modelling challenges



(b) Reservoir modelling strategies, physics based vs. data based.



(c) Varieties of modelling techniques available in literature. Each model type is assessed over the factors of complexity, validity, phenomena capture, scale and precision. A desirable balance has to be found when working with proxy models w.r.t. these factors.

FIGURE 1.1: Reservoir modelling strategies

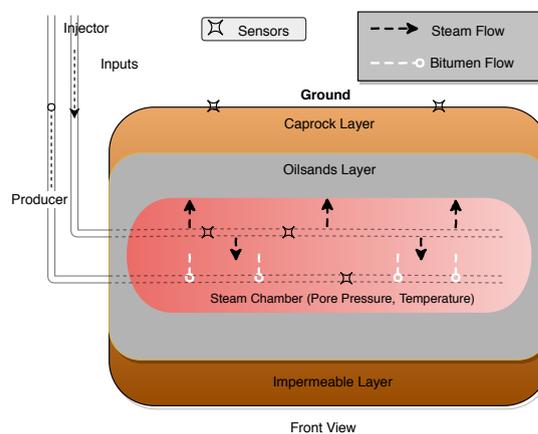


FIGURE 1.2: SAGD process overview

In the next part, we present an uncertainty quantification technique to quantify the uncertainty in the FoS with respect to the inputs and the petrophysical parameters. With a limited number of well-logs, precise quantification of heterogeneity in petrophysical and geomechanical parameters is a challenge in coupled reservoir-geomechanics modelling; this, along with nonlinearity in the process dynamics, gives rise to non-Gaussian uncertainties in the pore pressure/temperature, which poses severe challenges in reservoir control and optimization. A data-driven polynomial chaos expansion (PCE)-based proxy model is developed from sequentially coupled reservoir-geomechanics simulation. Proper orthogonal decomposition (POD) combined with the PCE yields a proxy model that can provide a quick and accurate estimation of caprock FoS along with quantifying its uncertainty. We desire the FoS to be greater than unity, but a very high FoS is uneconomical from the perspective of oil production rate. Hence, a FoS-constrained oil production rate (OP) maximized closed loop reservoir management is valuable in the oil and gas industry. A PCE based first order ARX model for the FoS is used as the proxy model to predict its dynamics and to quantify its uncertainty. A first order ARX model is also developed for the OP versus the well bottom hole pressure. Using these two proxy models, we build a stochastic model predictive controller (MPC) to decide the maximum operating pressure (MOP) of the SAGD process to balance the FoS limit and maximal OP. We also present a static measure for the MOP based on the PCE model along with the dynamic measure based on the stochastic MPC.

Apart from controller design, system theoretic analysis of complex systems is crucial in the framework of proxy modelling. The problem of sensor placement and the control of steam chamber growth and oil production, respectively, require analysis of the observability and controllability of the system. In SAGD, parametric sensitivity is traditionally used in lieu of observability and controllability has not been explored rigorously. In this work, we analyze the pressure and temperature fields of a SAGD model and present a data-driven technique to assess the structural controllability and observability of the system, with a view to determine optimal locations of sensors and actuators. An agglomerative hierarchical clustering technique is used to obtain a spanning tree of the clusters which is partitioned based on an objective function to arrive at a set of spatially contiguous clusters that display similar pressure/temperature dynamics. A Granger causality measure is used to create the linkage amongst the clusters to build a digraph model of the data. The

driver nodes of the graph identify locations for actuation which provide full control over the graph, and the root strongly connected components indicate sensor locations which ensure structural observability over the entire graph. We demonstrate the method using data generated from CMG-STARS simulator, identify the sensor and actuator locations required for complete structural observability and controllability of the system, and also provide a method of assessment of partial actuation and in-sensor ranges. We conclude our thesis highlighting the achieved results and presenting the future horizons in this area.

## 1.2 Thesis contribution

Figure 1.3 shows the proposed proxy model based closed loop control framework with an optimization objective for complex systems. The proxy models developed for the complex system are used to design a model predictive controller to decide on the control inputs that optimize a cost function while keeping the operation in the required zone of stability. As soon as the new set of process data is measured, the proxy models are calibrated/updated/recomputed to be used in the next step. This loop continues throughout the operation of the process. This thesis provide a similar proxy model-based framework to optimize the production profiles of the SAGD process while keeping the FoS within the prescribed limit.

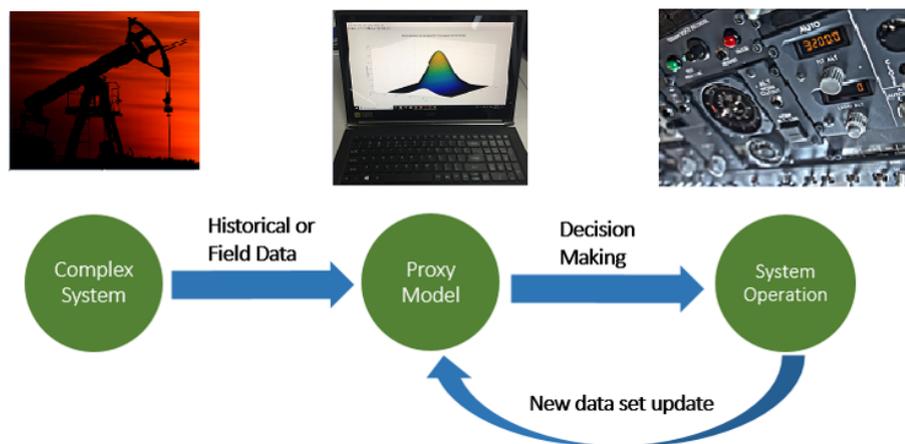


FIGURE 1.3: Proxy model based closed loop management of complex system

### 1.3 Thesis outline

The thesis has five chapters excluding the introduction. Chapter 2: *“Proxy models for caprock pressure and temperature dynamics during steam-assisted gravity drainage process”* proposes two proxy models; the first one estimates the caprock pressure and temperature fields using a reduced-order dynamic modelling. The second proxy model performs a dynamic analysis of FoS by modelling the evolution of caprock pressure clusters of high, medium and low pressure regions using graph theory and subspace modelling. In Chapter 3: *“Uncertainty quantification of the factor of safety in a steam-assisted gravity drainage process through polynomial chaos expansion”*, a data-driven polynomial chaos expansion (PCE)-based proxy model is developed to provide quick and accurate estimation of caprock FoS along with the intention of propagating the uncertainty in well bottom hole pressure inputs and reservoir petrophysical parameters to the caprock FoS. Chapter 4: *“Factor of Safety constrained Model Predictive Controller design for closed loop reservoir management”* combines the techniques from Chapters 2 and 3 to develop a model predictive controller (MPC) to design a FoS-constrained production-optimized closed loop control strategy. Chapter 5: *“Graph-based determination of structural controllability and observability for pressure and temperature dynamics during steam-assisted gravity drainage operation”* analyzes the system theoretic properties viz. controllability and observability of the SAGD process with the intention of actuator and sensor placement. It also provides a method of assessment of partial actuation and in-sensor ranges. The last chapter is the *Conclusions* which summarizes the thesis and discusses the scope for further development.

This is to re-emphasize that this thesis is a paper based thesis; hence, the reader is advised of the potential repetition of the main concepts surrounding the problem, especially in the introduction of each chapter.

# Bibliography

- [1] A. I. Forrester & A. J. Keane, (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3), 50-79.
- [2] A. Forrester & A. Keane, (2008). *Engineering design via surrogate modelling: A practical guide*. John Wiley & Sons.
- [3] A. C. Antoulas, *Approximation of large-scale dynamical systems*. Vol. 6. SIAM, 2005.
- [4] E. Gildin & T. Lopez, (2011, January). Closed-loop reservoir management: Do we need complex models?. In *SPE Digital Energy Conference and Exhibition*. Society of Petroleum Engineers.
- [5] R. M. Butler, (1994). Steam-assisted gravity drainage: concept, development, performance and future. *Journal of Canadian Petroleum Technology*, 33(02), 44-50.
- [6] M. R. Carlson, (2012, January). An analysis of the caprock failure at Joslyn. In *SPE Heavy Oil Conference Canada*. Society of Petroleum Engineers.
- [7] Computer Modeling Group, Calgary, Alberta, T2M 3Y7 Canada (2015 (10.5715.22942)).
- [8] *FLAC3D-Fast Lagrangian Analysis of Continua in 3Dimensions*, Minneapolis, USA (Version 7.0).
- [9] A. Azad, and R. J. Chalaturnyk, A mathematical improvement to SAGD using geomechanical modeling. *Journal of Canadian Petroleum Technology* 49.10 (2010): 53-64.

## Chapter 2

Proxy models for caprock  
pressure and temperature  
dynamics during steam-assisted  
gravity drainage process

## 2.1 Introduction

A continuous time dynamical system can be modelled through first principles using differential equations and modelled empirically (though with a limited region of validity) using input-output data. Reservoir modelling and analysis is typically carried out using commercial simulators which encode first-principles models and are computationally intensive. Parametric uncertainty is a significant challenge when these models are used for history matching; another challenge is the accurate modelling of near-well phenomena [1]. Proxy or surrogate models [2–4] are proposed as an alternative to address some of these shortcomings [5–7]. Ideally,

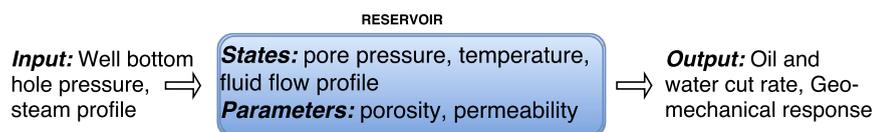


FIGURE 2.1: System theoretic perspective of a reservoir

proxy models are computationally efficient [8], simple and easily modifiable [9–12]. Depending on their nature, they may not capture all physically significant aspects of reservoir operation; however, they should at least be able to capture the relevant input-output dynamics, perhaps with periodic updating [15, 16]. A variety of proxy models are discussed in the literature related to reservoir modelling ranging from artificial neural networks (ANN) [17], multidimensional polynomials [18] to Monte Carlo sampling-based models [19] for uncertainty analysis, forecasting and history matching. Recent advancements in local model reduction techniques for accurately approximating highly nonlinear dynamics while considering spatiotemporal variations are discussed in [20–22].

The system theoretic perspective of a reservoir is shown in Figure (2.1). Steam Assisted Gravity Drainage (SAGD) [23] is a thermal recovery technique used for recovery of bitumen from oil sands. There are three important layers in the reservoir: the caprock layer, which is the top layer containing the oil and gas from coming out (approximately 300m from the ground surface); the impermeable layer, which is at the bottom, and the middle layer (the oil sands layer), which holds the bitumen. The SAGD process uses two parallel horizontal wells; the upper well, called the injector, injects steam into the reservoir. As a result, the bitumen captured in the oil sands becomes less viscous, flows down due to gravity and gets collected by the lower well (called the producer). Reservoirs typically have multiple injector

and producer wells. The injected steam creates a high pressure and temperature chamber in the reservoir. One of the implications of having high temperatures and pressures is that monitoring of the integrity of the caprock is essential for safe operation. Many caprock rupture incidents like Joslyn [25] in the past have proved the importance of monitoring the stress/strain over the caprock for a safer operation of the reservoir. The factor of safety (FoS) [26], defined as the ratio of the allowable stress to the actual stress in the caprock layer, specifies the constraints on reservoir operation. The pressure and temperature field variations in the chamber should be monitored and kept within the FoS limit. Caprock damage can result in great environmental damage, danger to humans and financial losses. Reservoir operational safety is therefore of prime concern in the oil and gas industry [27–29].

Closed loop optimization and control is extensively used in many engineering disciplines, and has been proposed for the management of petroleum reservoirs, too [30]. Control-relevant proxy models for the reservoir and the caprock can provide computationally inexpensive solutions for this task. In this work, we propose two such control relevant proxy modelling schemes. The first proxy model is developed for the pressure and temperature fields in the caprock based on proper orthogonal decomposition (POD) and system identification. This method enables visualization and forecasting of the entire pressure and temperature field variations using a set of basis vectors and time-varying coefficients. The second proxy model groups the pressure and temperature fields into different clusters (zones) and models interactions between the zones as a directed graph or network to capture the dynamic behaviour with their neighbouring interacting zones. This type of proxy model provides a qualitative and quantitative way to analyze the factor of safety in the reservoir based on the pressure and temperature fields and their effect on the stress and strain profiles in the caprock. Both of these techniques are designed to provide a computationally affordable approach considering the large scale nature of the problem.

This article is organized as follows: The first-principles models and data generated from them are explained in the next section, followed by a description of the proxy modelling techniques which is followed by results and discussion.

## 2.2 Reservoir and caprock modelling

The data used in the development of proxy models is obtained from a CMG-STARS reservoir simulator [33]. A model of a heterogeneous SAGD reservoir having 5 injectors and 5 producers is provided with the designed inputs (described later) and the data on the 2-dimensional cross section of the reservoir pressure and temperature fields at various time instants is gathered and analysed. Figure (2.2) explains the source of the data in the context of the reservoir geometry, showing the 5 pairs of injector and producer wells. The spatiotemporal data consists of the cross section of the pressure/temperature field at equally spaced time intervals, referred to as frames. The simulation domain is illustrated in Figure (2.3) showing

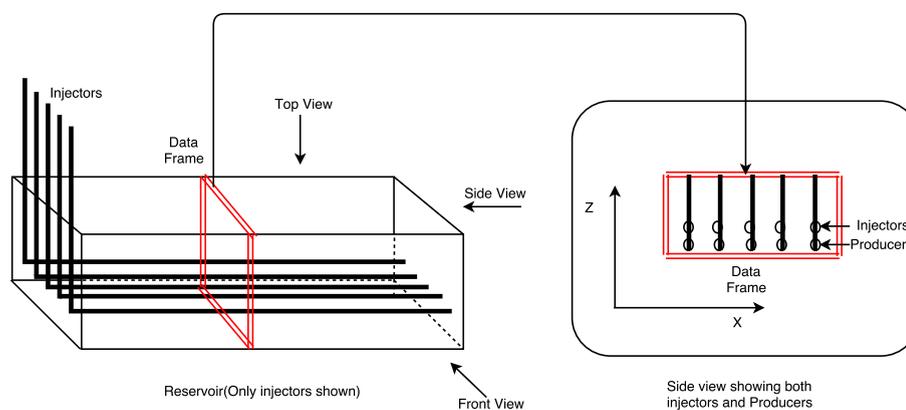


FIGURE 2.2: Perspective view of reservoir structure with well bores and the viewing directions. The 2D data frame (cross-section in the reservoir) is shown and the inset shows a detailed view of the same with reference directions  $X$  and  $Z$ .

different regions of interest. The thickness of the overburden was set to 120 m. Part of the overburden (60 m) is included in the CMG-STARS model (*STARS 2015.10*). In the vertical direction, the model grid comprised a total of 81 grid cells. The cell sizes were 10m per cell in the overburden (10 m  $\times$  6 cells); 1 m per cell in the caprock (1 m  $\times$  25 cells); 1 m per cell in the reservoir (1 m  $\times$  45 cells); and 4 m per cell in the underburden (4 m  $\times$  5 cells). In the horizontal direction, the model width was 3000 m, and the model grid contained 1200 cells for CMG-STARS. These cells were distributed from left to right as follows: 100 5-m cells; 1000 2-m cells; and 100 5-m cells. There are five wellpairs in the reservoir, with the horizontal wells in each well pair being 5 m apart, and the production well being located 5 m above the bottom of the reservoir. The horizontal coordinates of the 5 producer wells are  $X = 1340$  m, 1420 m, 1500 m, 1580 m and 1660 m. The

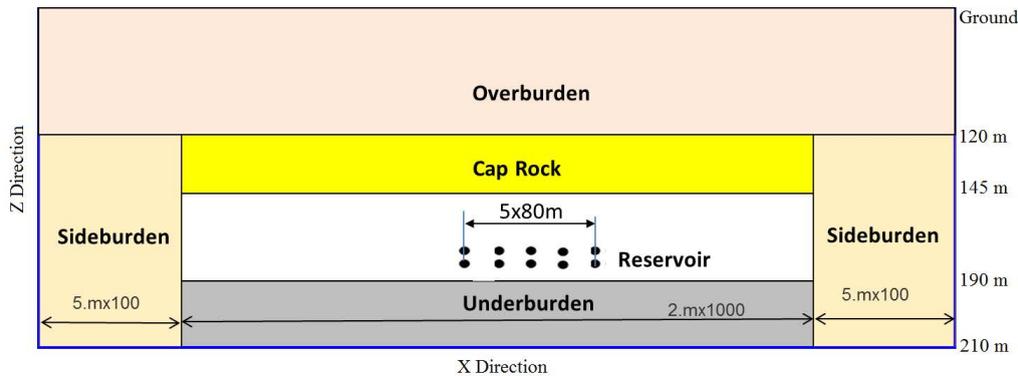


FIGURE 2.3: A detailed simulation layout of the 2D frame

shale distribution of the heterogeneous 2D reservoir model is illustrated in Figure (2.4), with 20% shale volume overall. The main parameters for the flow simulator are summarized in Table (2.1).

Each frame has  $1200 \times 81$  rectangular grid blocks (or points) in the X, Z direction corresponding to the width and height of the reservoir, and 184 such frames (representing monthly data) are available. The pressure and temperature values are reported for each grid block at each time sample. The inputs variables are the well bottom-hole pressures of the injector and producer wells. In this chapter, proxy models are developed for the caprock region. The dimension of the caprock (pressure/temperature) data is  $1000 \times 25 \times 184$  and that of the input data as  $10 \times 184$ .

TABLE 2.1: CMG-STARS simulation parameters  
 $S_o$ : Oil saturation,  $S_w$ : Water saturation

Particular	Porosity	Permeability	Permeability	Permeability	$S_o$	$S_w$
		J (md)	J (md)	K (md)		
Overburden	N/A	100	100	50	0.0	1.0
Caprock	0.1	1	1	0.001	0.0	1.0
Reservoir sand	0.3	5000	5000	3330	0.85	0.15
Reservoir shale	0.1	50	50	33	0.15	0.85
Underburden	0.1	1	1	0.001	0.00	1.0

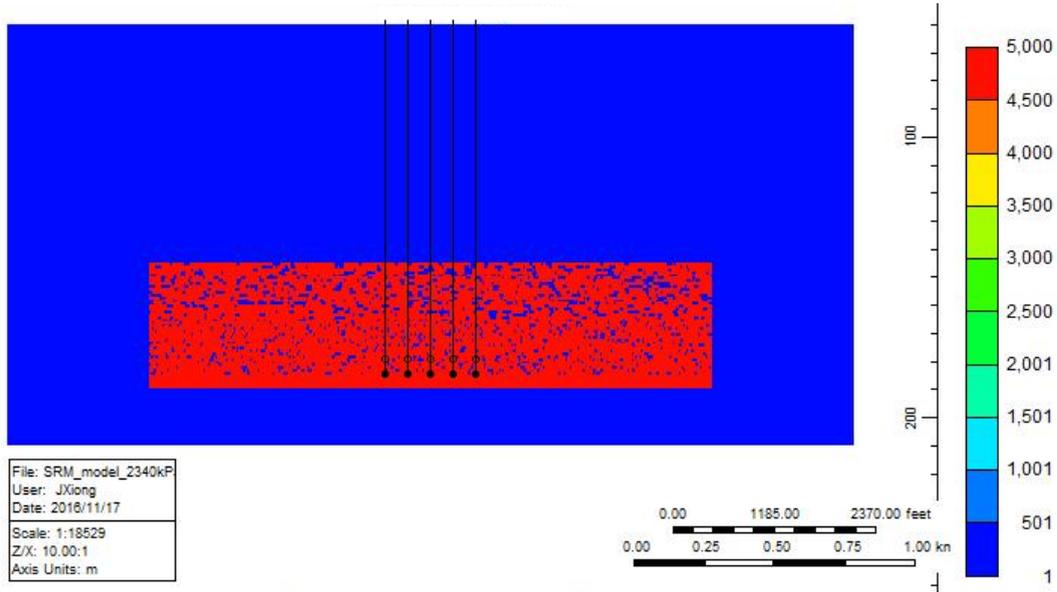


FIGURE 2.4: Shale distribution of heterogeneous reservoir

### 2.2.1 Input Design

Injector and producer well bottom hole pressures are the inputs to the SAGD model. Persistence of excitation [9, 10] is an essential consideration when designing inputs for the purpose of system identification. An input  $u(t)$  is said to be persistently exciting of order  $n$  if the following conditions hold true.

1. The limit

$$r_u(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u_{t+\tau} u'_t$$

exists.

2. The matrix  $R_u(n) = \begin{bmatrix} r_u(0) & r_u(1) & \dots & r_u(n-1) \\ r_u(-1) & r_u(0) & \dots & r_u(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ r_u(1-n) & \dots & \dots & r_u(0) \end{bmatrix}$  is positive definite.

A necessary condition for consistent estimation of an  $n^{\text{th}}$  order ARX model is that the  $u(t)$  should be of order  $2n$ . A pseudo random binary signal (PRBS) persistently excites a system of any order. In this work, two simulation cases are carried out with the CMG-STARS, one with MATLAB [31] designed PRBS for the injector and producer well bore pressures, and another with a step input to generate two data sets of the pore pressure and temperature. Both of the input

data are of dimension  $10 \times 184$ . The data set with PRBS input is used for building the POD based proxy model and the one with the step input for the clustering-based method. As the intended proxy models are linear, a PRBS signal approach was found to be sufficient for the identification.

## 2.3 Proxy modelling methods

### 2.3.1 Proper orthogonal decomposition

The  $3D$  matrix data  $D$  (dimension  $1000 \times 25 \times 184$ ) is folded into a  $2D$  matrix by stacking rows such that every column in the resultant matrix corresponds to a frame of the original data. Proper orthogonal decomposition [32] (POD) is a data/signal/feature decomposition technique based on orthogonal basis expansion which is used to obtain a low dimensional approximation of high dimensional spatiotemporal data. POD can be used to approximate a function  $z(x, t)$  as a finite sum in variables separated form as:

$$z(x, t) \approx \sum_{i=1}^N \psi_i(x) a_i(t) \quad (2.1)$$

where  $\psi_i(x)$  are called the basis functions or modes and the coefficients  $a_i(t)$  are called the basis weights. For a chosen  $N$  and  $\psi_i(x)$ , equation (2.1) is the POD expansion. In our finite dimensional discrete time case;  $z(x, t)$  is a  $D_{25K \times 184}$  matrix, and POD can be achieved through singular value decomposition (SVD) and is commonly referred as principal component analysis (PCA). We compute the spatial basis functions using the SVD method.

$$D_{25K \times 184} = U_{25K \times 184} \Sigma_{184 \times 184} V'_{184 \times 184} \quad (2.2)$$

$U$ ,  $V$  are orthogonal matrices and  $\Sigma$  is a diagonal matrix with the singular values of  $D$  arranged in descending order.

In the  $\mathbb{R}^{184}$  space, the principal component directions are the directions of the columns of  $U_{25K \times 184}$ . These principal directions obtained through SVD are by definition those basis vectors which capture the information in the least square sense and their corresponding singular values quantify the contribution of each of the basis function. The ratio  $\frac{\sum_{i=1}^r a_i}{\sum_{i=1}^{184} a_i}$  quantifies the percentage of energy captured

by the first  $r$  basis functions out of the set of 184 basis functions. The first  $r$  columns of  $U_{25K \times 184}$  are chosen as the truncated set of basis vectors, i.e.  $U_{25K \times r}$ . Expressing every column of  $D$  in terms of  $U_{25K \times r}$  gives

$$D_{25K \times 184} = U_{25K \times r} W_{r \times 184} \quad (2.3)$$

where  $W_{r \times 184}$  is the linear combination weights matrix, given as

$$W_{r \times 184} = \Sigma_{r \times r} V'_{r \times 184} \quad (2.4)$$

where  $\Sigma_{r \times r}$  is the  $r$  sized leading principal minor of  $\Sigma_{184 \times 184}$ . If full SVD computation is not possible due to size constraints, then the weights matrix can be computed as

$$W_{r \times 184} = U'_{25K \times r} D_{25K \times 184} \quad (2.5)$$

Here,  $U'_{25K \times r}$  is the transpose of  $U_{25K \times r}$ . Every column of  $W_{r \times 184}$  contains the  $r$  linear combination weights  $w_1 \dots w_r$ , and we have such 184 columns of weights corresponding to every frame.

### 2.3.1.1 System Identification

The pressure (or temperature) field at any point of time can be expressed as a function of its past values and the  $m$  inputs, which suggests the possibility of applying system identification techniques based on the proper orthogonal decomposition. We have previously demonstrated the applicability of system identification to proxy modelling [12, 13]. In the POD framework we see that every frame (column of  $D$ ) has  $r$  representative weights and these weights can be modelled as functions of their past values and the  $m$  process inputs. Modelling the frames translates to modelling the POD weights, described as follows. At any time frame  $t$ , we can express the  $k^{th}$  row of  $W = [w_1 \ w_2 \ \dots \ w_r]'$  with respect to input  $U(t)$  as

$$w_k(t+1) = f\{w_k(t), U(t)\} \\ \forall t \in \{1 \dots N-1\}, k \in \{1 \dots r\}, U(t) \in R^m \quad (2.6)$$

We use a linear auto-regressive with exogenous input (ARX) model ( $f$ ) to estimate the future frames. The ARX model is given by

$$A(z^{-1})w(t) = B(z^{-1})u(t - nk) + e(t) \quad (2.7)$$

$$A(z^{-1}) = 1 + a_1z^{-1} + \dots + a_{na}z^{-na}$$

$$B(z^{-1}) = b_1 + b_2z^{-1} + \dots + b_{nb}z^{-nb+1}$$

$A$  and  $B$  are polynomials of order  $na$  and  $nb - 1$ , respectively (to be estimated), and  $nk$  is the input delay. With  $m$  inputs, the ARX model (2.7) is given by (2.8)

$$A(z^{-1})w_k(t) = B_1(z^{-1})u_1(t - nk) + B_2(z^{-1})u_2(t - nk) + \dots + B_m(z^{-1})u_m(t - nk) + e(t) \quad (2.8)$$

Estimating a future frame that is  $l$  steps ahead translates to estimating the  $(N+l)^{th}$  column of  $D$ . So  $r$  ARX models (each for a POD weight) are fitted to find the  $(N+l)^{th}$  column of  $W$ .

$$W(:, N+l) = [w_1(N+l) \ w_2(N+l) \ \dots \ w_r(N+l)]'$$

The future frame can be estimated as:

$$D(:, N+l) = U_{25K \times r} W(:, N+l)$$

Figure (2.5) summarizes the POD based proxy modelling procedure.

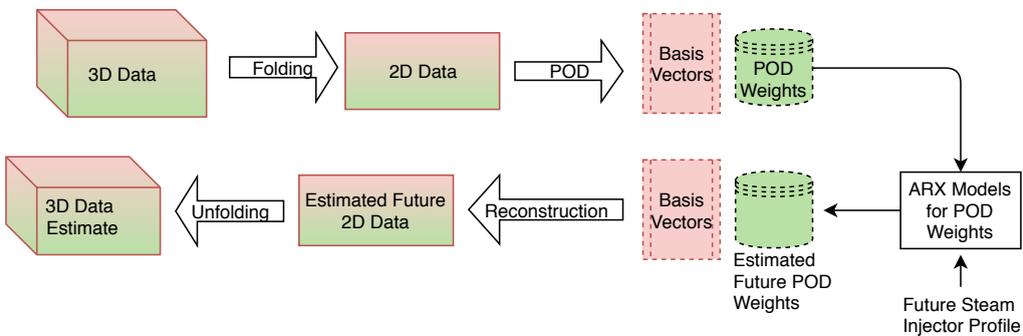


FIGURE 2.5: Schematic of the POD based proxy modelling technique

It should be noted that the approach presented here uses stacking (or vectorization) to fold the  $3D$  matrix data into a  $2D$  matrix for performing the POD analysis. A potentially more efficient method (in terms of the number of basis vectors needed to reproduce essential spatiotemporal features) is to use a tensor algebra-based

approach such as higher order singular value decomposition [14]. However, the approach used here is shown in the results to be adequate for developing a reduced order system identification-based framework for modelling.

## 2.3.2 Clustering-based modelling

The proxy model described in this section clusters the pressure (or temperature) field to identify characteristic regions and track their dynamics and the interactions between them.

### 2.3.2.1 Data Clustering

The pressure (or temperature) grid points in every frame are sorted into  $N_c$  bins (clusters) having equal spacings over the range of the pressure values (equivalent to constructing a  $N_c$  binned histogram for every frame), and the bin counts of the clusters (the number of grid points in that particular cluster) is computed for every frame. After binning every frame into  $N_c$  clusters, the average and standard deviation of pressure for each cluster over all the frames is calculated. Modelling the bin count variations over the frames serves as a proxy model to understand the dynamics of the clusters of different pressure ranges.

The interaction between the pressure clusters can be modelled using a graph, with each cluster a node and the interactions between the clusters represented by edges. Figure (2.6) shows how the clusters and interactions can be modelled by a directed weighted line graph. Each vertex of the graph corresponds to the bin count for that particular pressure cluster, every cluster interacts with its adjacent clusters and the bin count of every cluster at any instant depends on the interacting adjacent clusters' bin counts; this leads to a line graph. This can be generalized easily to any number of clusters  $N_c$ . For the specific case being considered here, the interactions only occur with neighbouring clusters, leading to a directed line graph. This leads to a tri-diagonal structure to the system matrix  $\phi$  of equation (2.9). The tri-diagonal structure with non-zero entries assures the controllability and observability of the model (explained in section 2.4.3) for any number of clusters. Reservoirs with a more heterogenous nature will require a larger number of clusters along with a more complex set of interactions between

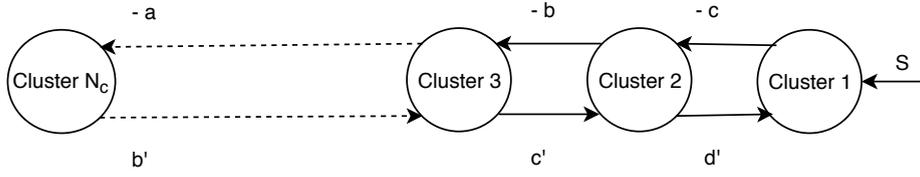


FIGURE 2.6: A directed and edge weighted line graph shows the interaction between the clusters. It governs the bin count dynamics where the edge weight is the index of the interaction. Only the incoming edge weight affects the node state.

the clusters; approaches such as those described in [34–36] can be used to develop the network descriptions for those cases.

### 2.3.2.2 State Space modelling

For the sake of demonstrating the technique, the number of clusters in each frame is chosen as 4, i.e.  $N_c = 4$ . The bin counts of the nodes at the  $(i + 1)^{th}$  instant can be represented in terms of their bin counts at the  $i^{th}$  instant:

$$N_4^{i+1} = N_4^i - aN_3^i + e_4^i$$

$$N_3^{i+1} = N_3^i - bN_2^i + b'N_4^i + e_3^i$$

$$N_2^{i+1} = N_2^i - cN_1^i + c'N_3^i + e_2^i$$

$$N_1^{i+1} = N_1^i + dN_2^i + e_1^i + S$$

As the pressures in the clusters are ordered as  $P_4 < P_3 < P_2 < P_1$ , the neighbouring higher pressure cluster serves to decrease a cluster's bin count, and the neighbour lower pressure cluster increases its bin count.  $e_k^i$  is the residual in the model for the bin count of the  $k^{th}$  state at the  $i^{th}$  instant; which is assumed to be Gaussian with zero mean. The structure of the interactions reported here is specific to the reservoir and operating conditions considered in this work, but the concept can be generalized easily.

The set of difference equations governing the cluster equations is described in matrix (state space) form as:

$$\begin{bmatrix} N_4^{i+1} \\ N_3^{i+1} \\ N_2^{i+1} \\ N_1^{i+1} \end{bmatrix} = \begin{bmatrix} 1 & -a & 0 & 0 \\ b' & 1 & -b & 0 \\ 0 & c' & 1 & -c \\ 0 & 0 & d & 1 \end{bmatrix} \begin{bmatrix} N_4^i \\ N_3^i \\ N_2^i \\ N_1^i \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} S + E^i$$

$$N^{i+1} = \phi N^i + \Gamma S + E^i \quad (2.9)$$

$E^i = [e_1 \ e_2 \ e_3 \ e_4]'$  is assumed to be an independently identically distributed noise vector, and here  $S$  is the source/input. The model development problem is to obtain estimates of the unknown coefficients  $a, b, b', c, c', d$  and  $S$ . In the simulations for this part of the work,  $S$  was chosen to be a step input.

### 2.3.2.3 State Space Parameter Estimation

The state space parameter vector  $\theta = [a \ b \ b' \ c \ c' \ d \ S]'$  can be estimated using the available bin count data reported in Figure (2.16) as follows:

$$\theta = \arg \min_{\theta} \|N_j^{i+1} - (\phi(\theta)N_j^i + \Gamma S)\|_2^2$$

and the estimates are obtained using a weighted least squares formulation.

## 2.4 Results

### 2.4.1 POD-based proxy models

The first set of results related to the POD and system identification-based proxy models are presented in this section. Figures (2.8(a)) to (2.8(i)) shows (9 out of 10) the image representation of the POD basis vectors  $U_{25K \times 10}$  of equation (2.3). Figure (2.9) shows the logarithmic variation of all of the POD weights of the data. Figures (2.10(a)) to (2.10(c)) illustrate the effect of the number of basis vectors  $r$  used in the POD reconstruction of the pressure field. A choice of 10 for  $r$  provided good reconstruction of the original pressure field in terms of capturing the variance in the data, with the ratio of the minimum and maximum singular values being of

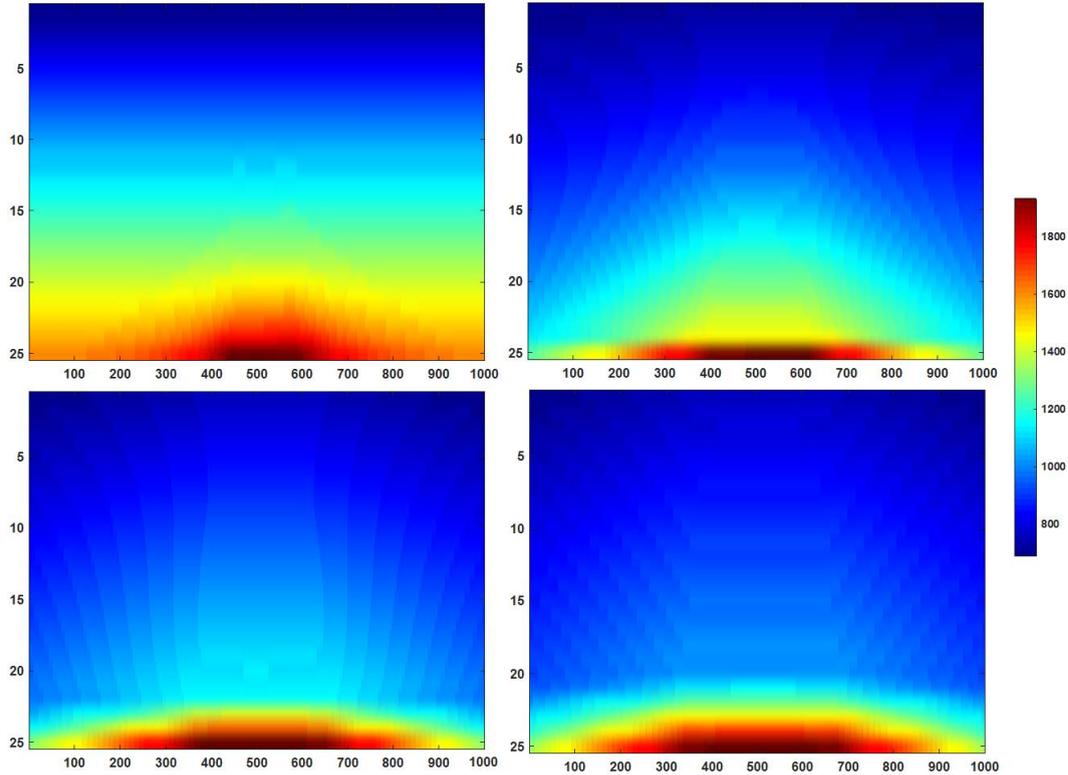


FIGURE 2.7: The Image representation of the pressure data frames at various instants (180, 140, 110, 50 from top left to bottom right). The pressure values are scaled and mapped to colors and an image is drawn (refer color bar). Number of grids and the direction is also shown.

the order of  $10^{-3}$  and the percentage of energy captured by the 10 out of 184 basis functions is 99.58%. For the system identification procedure, the first 170 of the 184 data points are used for training and the other 14 are reserved for validation. The parameter estimation was carried out using an autoregressive model with exogenous input (ARX model).

#### 2.4.1.1 ARX modelling

This section shows the modelling results of all the POD weights models (equation 2.6). Since there are 10 input variables, the structure of the ARX model used for each of the POD weights is given by equation (2.10)

$$A(p)w_k(t) = B_1(p)u_1(t-nk) + B_2(p)u_2(t-nk) + \dots + B_{10}(p)u_{10}(t-nk) + e(t); \quad (2.10)$$

Based on trial and error, polynomials of order 3 were used for  $A(p)$ ,  $B_1(p) - B_{10}(p)$  in the identification procedure, and the delays were set to one sample. The models

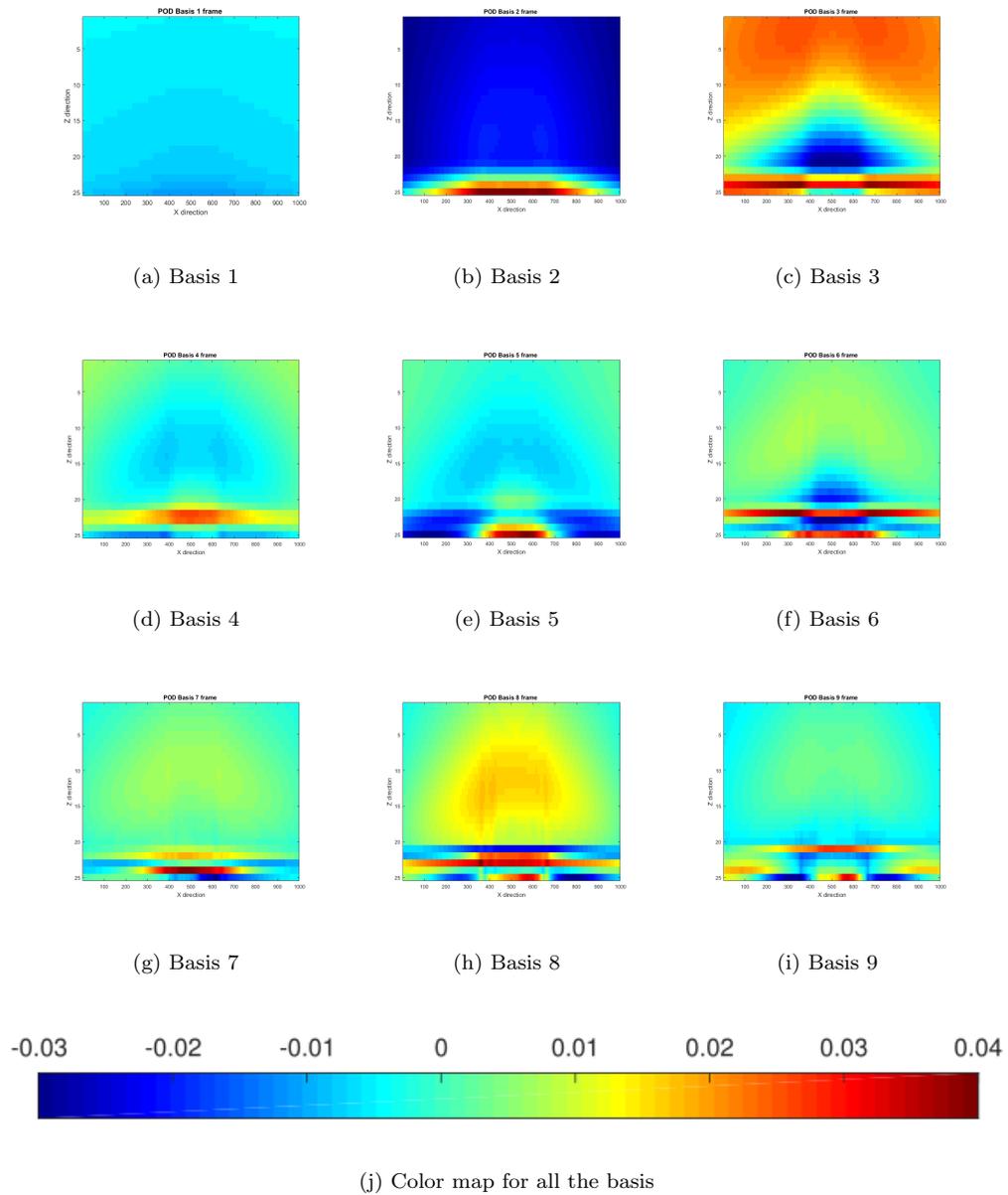


FIGURE 2.8: POD spatial basis visualization: Scales of all of the subfigures is same as of Figure (2.7)

were evaluated based on the normalized root mean squared error (NRMSE) encapsulated by the fit % reported by the system identification toolbox in MATLAB given by equation (2.11), and on the autocorrelations of the residuals and their cross-correlation with the inputs.

$$Fit \% = \left(1 - \frac{\|y - \hat{y}\|}{\|y - \text{mean}(y)\|}\right) \times 100 \quad (2.11)$$

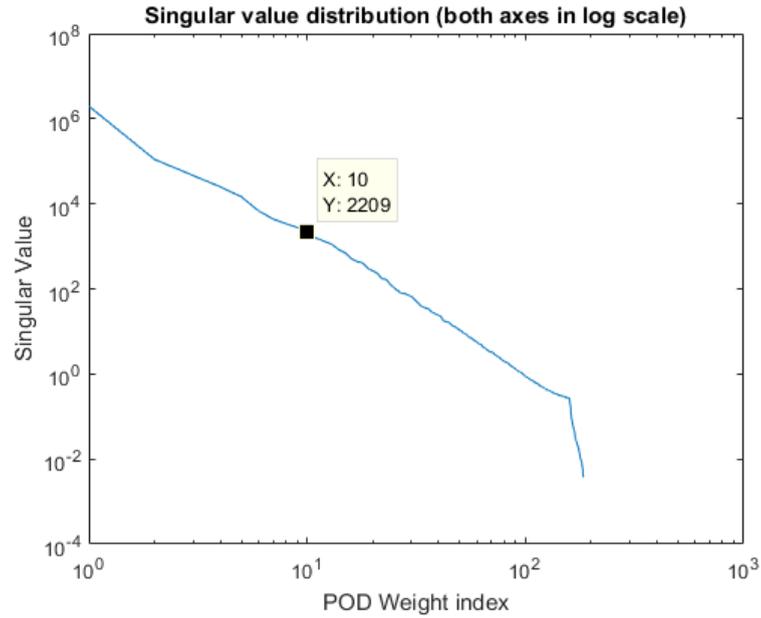


FIGURE 2.9: Logarithmic plot of the POD weights showing a steep decrease in the order of weights with their indices, the ratio of the 10<sup>th</sup> singular value (marked on the curve) to the 1<sup>st</sup> being order  $10^{-3}$  and capturing 99.58% of the total energy can be seen.

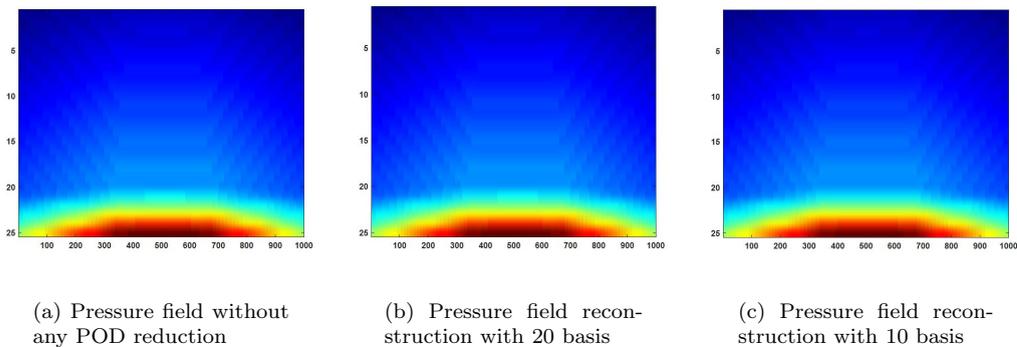


FIGURE 2.10: Reconstruction with different number of POD basis; Colormap see Fig(2.7)

$y$  is the actual data and  $\hat{y}$  represents the model predicted output. Figure (2.11) and Figure (2.12) show results for the modelling of POD weight 1 for the modelling and validation data sets, respectively, and Table (2.2) summarizes the model fit percentages for all the ten weights. Percentage fit is shown in both Figures (2.11), (2.12) and the Table (2.2). Figures (2.13(a), 2.13(b) and 2.13(c)) show the actual 174<sup>th</sup> frame in the validation set, the same frame reconstructed with the discussed POD+ARX technique and the error (actual-reconstructed) frame with the error Frobenius norm respectively. We reconstructed all the 14 frames in the validation set and compared them with the actual data. The error (Frobenius norm) in the

frames ranged from 85 to 115 and in the individual frame the error at every grid ranged between -5 and 5.

Weight index	% modelling Fit <sup>a</sup>	% Validation Fit <sup>b</sup>
1	91.91	94.36
2	82.17	97.51
3	87.63	54.48
4	62.34	96.44
5	46.98	63.59
6	24.86	68.65
7	-6.215	75.54
8	9.078	81.7
9	5.599	66.3
10	4.048	79.12

TABLE 2.2: System identification of POD weights. <sup>a</sup> Fit with 10 step ahead prediction, <sup>b</sup> Fit with one step ahead prediction.

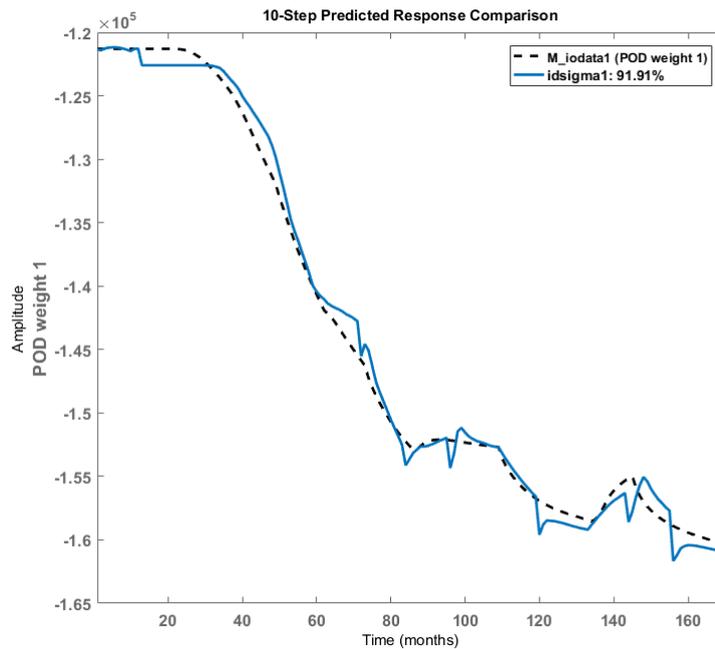


FIGURE 2.11: System identification on training dataset result: Actual and modelled weight  $w_1(l)$ , dotted line indicates identified model and solid line indicates the actual data.

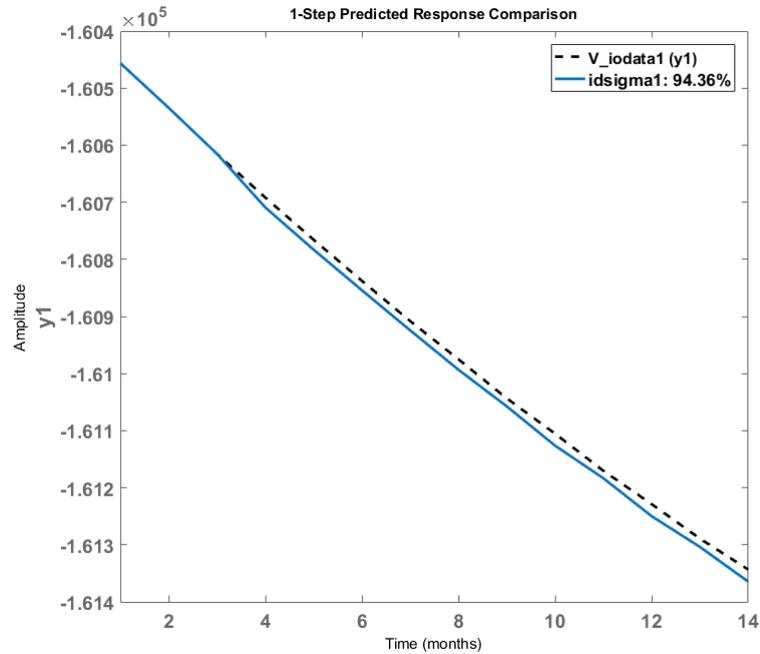
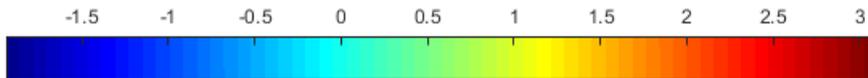
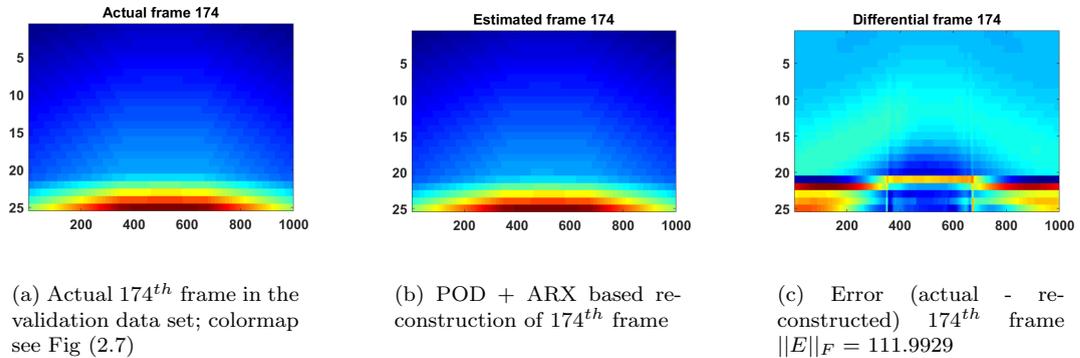


FIGURE 2.12: System identification on validation dataset result: Actual and modelled weight  $w_1(l)$ , dotted line indicates identified model and solid line indicates actual.



(d) Color map for the differential frame (2.13(c))

FIGURE 2.13: POD+ARX based reconstruction visualization

## 2.4.2 Cluster-based graph modelling

The computed average pressures of the four clusters are  $803.23\text{KPa}$ ,  $1052.70\text{KPa}$ ,  $1302.2\text{KPa}$  and  $1551.63\text{KPa}$ , with standard deviations of  $63.79\text{KPa}$ ,  $155.98\text{KPa}$ ,

248.74*KPa* and 341.6*KPa*, respectively. Figure (2.15) shows the frame-wise variation of the pressure centres, with the trends showing that the pressure centres are increasing with time. The framewise variation of the bin counts of all the four clusters are shown in Figure (2.16). The parameter estimation for the cluster-based modelling (equation 2.9) seeks to minimize the objective function

$$\|N_j^{i+1} - (\phi N_j^i + \Gamma S)\|_2^2$$

which can be portrayed as the follows

$$\begin{bmatrix} -N_3^i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -N_2^i & N_4^i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -N_1^i & N_3^i & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & N_2^i & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ b' \\ c \\ c' \\ d \\ S \end{bmatrix} + E^i = \begin{bmatrix} N_4^{i+1} - N_4^i \\ N_3^{i+1} - N_3^i \\ N_2^{i+1} - N_2^i \\ N_1^{i+1} - N_1^i \end{bmatrix}$$

Building this matrix for all the 183 instants we get a matrix of the following form

$$X_{468 \times 7} \theta_{7 \times 1} + E^i = Y_{468 \times 1}$$

The vector of parameter to be estimated,  $\theta$ , is obtained using an error weighted least square solution using the normal equation

$$\theta = (X' * W * X)^{-1} X' * W * Y \quad (2.12)$$

The estimated model parameters  $\theta$  and the corresponding state space matrix  $\phi$  (of equation (2.9)) are presented below.

$$\begin{bmatrix} a = -0.0107 \\ b = -0.0260 \\ b' = -0.0057 \\ c = -0.0211 \\ c' = -0.0096 \\ d = -0.0415 \\ S = 75.79 \end{bmatrix} ; \phi = \begin{bmatrix} 1 & 0.0107 & 0 & 0 \\ -0.0057 & 1 & 0.0260 & 0 \\ 0 & -0.0096 & 1 & 0.0211 \\ 0 & 0 & -0.0415 & 1 \end{bmatrix}$$

The bin count dynamics are governed by the weighted adjacency matrix ( $A_w$ ) of the graph, which is a tri-diagonal system matrix  $\phi = A_w + I$ . The spectrum of  $\phi$  is represented by the eigenvalues  $1 \pm 0.0068i$  and  $1 \pm 0.0338i$ . In the z-domain, this is equivalent to all the poles lying outside the unit circle, giving rise to slowly growing bin counts.

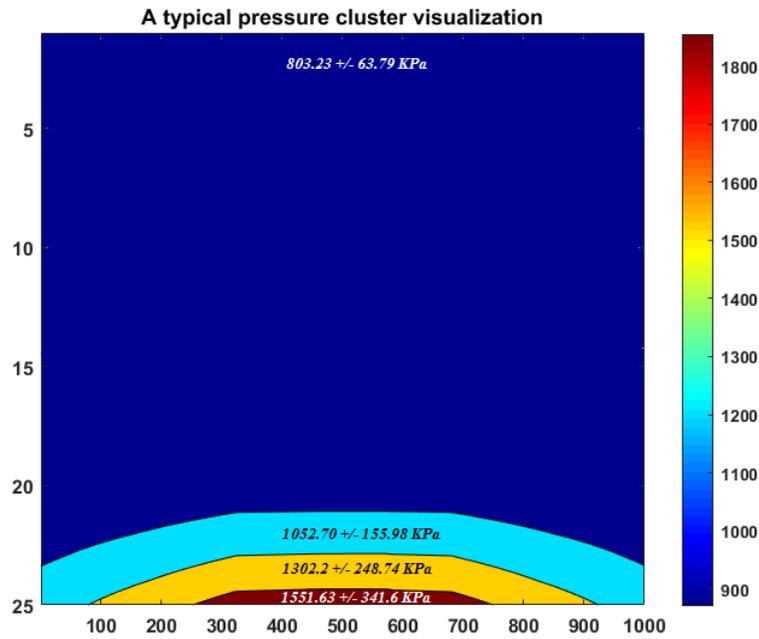


FIGURE 2.14: A typical frame with clustered pressure grid points: four coloured regions corresponds to four regions having high pressure - brown, yellow clusters, medium pressure - cyan and low pressure -blue cluster (outermost cluster near to the ground)

#### 2.4.2.1 Residual Error

The residual bin count error 2-norm  $\|E^i\|$  between the modelled and the original data is computed as

$$\|E^i\|_2^2 = \|N^{i+1} - (\phi N^i + \Gamma S)\|_2^2$$

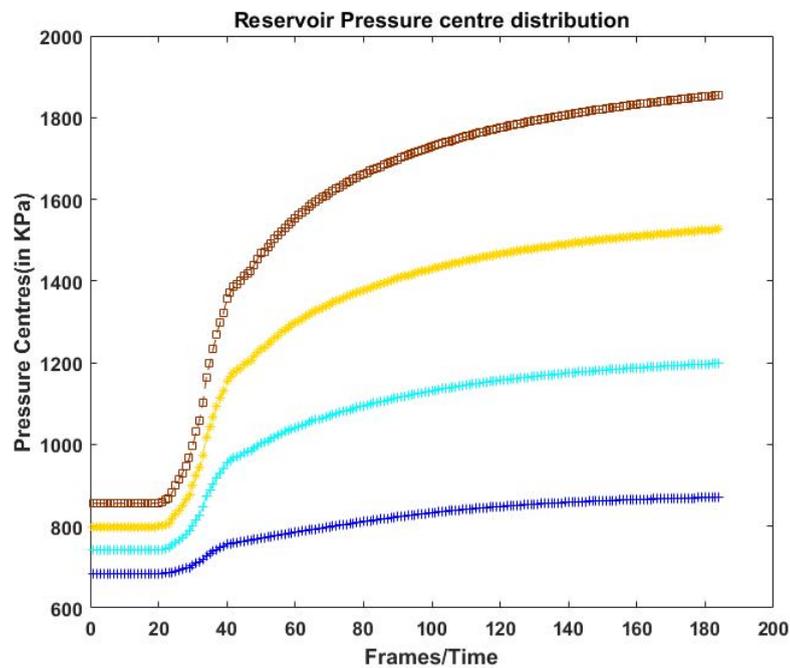


FIGURE 2.15: Variation of pressure centres of every cluster with frames. Colour map is same as Figure(2.14). Variation is centred about Blue - 803.23 KPa, Cyan - 1052.70 KPa, Yellow - 1302.2 KPa and Brown - 1551.63 KPa with a corresponding standard deviations of 63.79 KPa, 155.98 KPa, 248.74 KPa and 341.6 KPa

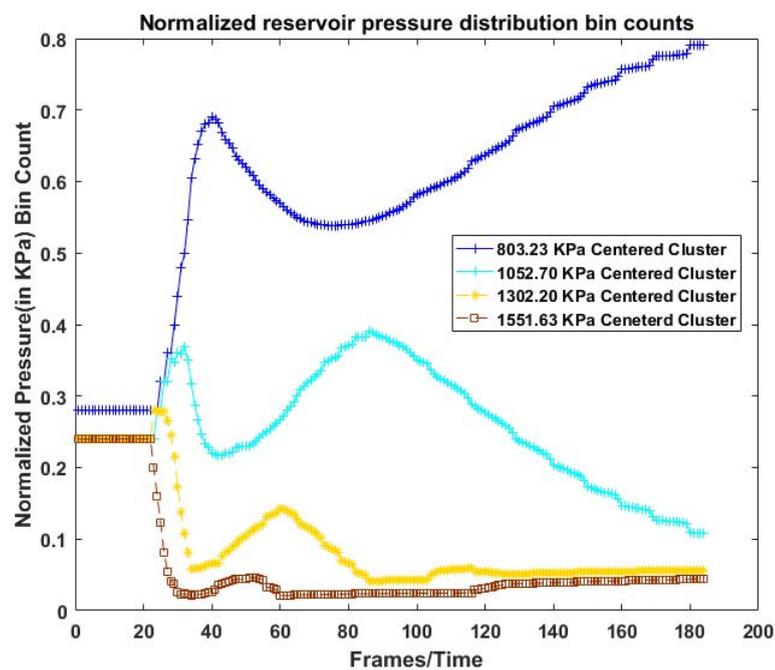


FIGURE 2.16: Variation of normalized bin count of each cluster vs frames. Colour map is same as Figure(2.15)

The norm is found to be  $[0.1252 \ 0.0936 \ 0.0930 \ 0.0752]'$ . The cross correlation coefficient matrix for the bin count error is:

$$Corr(E^i) = \begin{bmatrix} 1.0000 & -0.5115 & -0.6330 & -0.2264 \\ -0.5115 & 1.0000 & 0.0023 & -0.3431 \\ -0.6330 & 0.0023 & 1.0000 & -0.1904 \\ -0.2264 & -0.3431 & -0.1904 & 1.0000 \end{bmatrix}$$

Since most of the cross-correlation coefficients are relatively small (less than 0.35), this indicates that residuals in the estimation are not correlated strongly, giving confidence in the parameter estimates.

### 2.4.3 Controllability and Observability

The strong interaction between the clusters in the process assures non-zero parameters in  $\phi$  for the model described by equation (2.9). Consequently, the controllability matrix  $[\Gamma \ \phi\Gamma \ \phi^2\Gamma \ \phi^3\Gamma]$  is full rank, and the system is completely controllable by construction. This ensures that the bin count of any cluster can be controlled by the step input  $S$ . The feasibility of tracking the bin counts of all the clusters by knowing the bin count of one of the clusters may be referred as observability. As the bin count of a cluster is not directly measurable, we need to build a lookup function which maps the pore pressure measured through a ground sensor to the respective bin counts of that cluster.

Observability is evaluated using the model described by equation (2.9) along with the measurement equation given below

$$Y = CN^i \tag{2.13}$$

With  $C = [1 \ 0 \ 0 \ 0]$  (discussed in the subsequent section), the model (2.9) is fully observable for the same non-zero parameters reason in  $\phi$ . This assures us that an asymptotic observer can be designed to track the bin counts.

### 2.4.3.1 Observer with ground sensor measurement

With  $C = [1 \ 0 \ 0 \ 0]$  the sensor (perhaps placed on the ground surface) can measure the average pore pressure around the top of the caprock which corresponds to the outermost cluster, in Figure (2.14). A look-up function  $f_{GL}$  (refer Figure(2.17)) can be constructed to relate the bin count of the outermost cluster to this measured pressure, thereby acting as an indirect measurement for  $N_4^i$ .

*Ground lookup function:* The outermost bin count variation (Figure 2.16) can be modelled against the corresponding average pressure centres variation (Figure 2.15) to build a lookup function. An ARX model can be developed for the same.

$$\text{Cluster bincount} = f_L \{ \text{Pressure Centre} \}$$

An observer for the bin count developed on the basis of this look-up function is

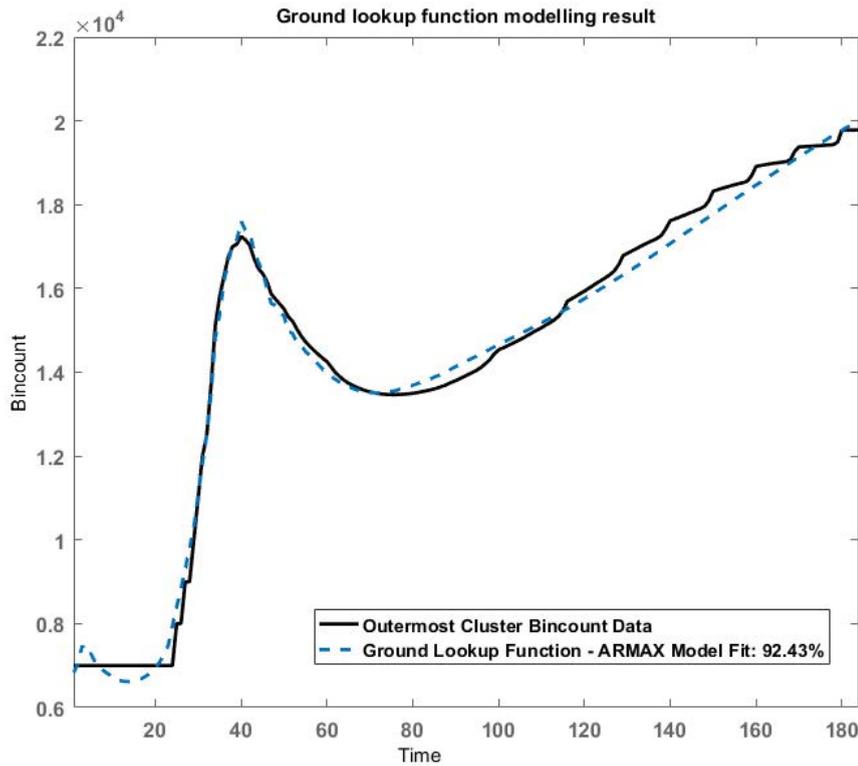


FIGURE 2.17: Outermost cluster bin count versus its pressure centres look-up function response

given below:

$$\hat{N}_{i+1} = \phi \hat{N}_i + L[y_i - C \hat{N}_i] + \Gamma S$$

$$y_i = f_{GL}(\text{Ground Sensor Measurement})$$

Figure (2.18) shows the observer bin count error dynamics versus time for a simple

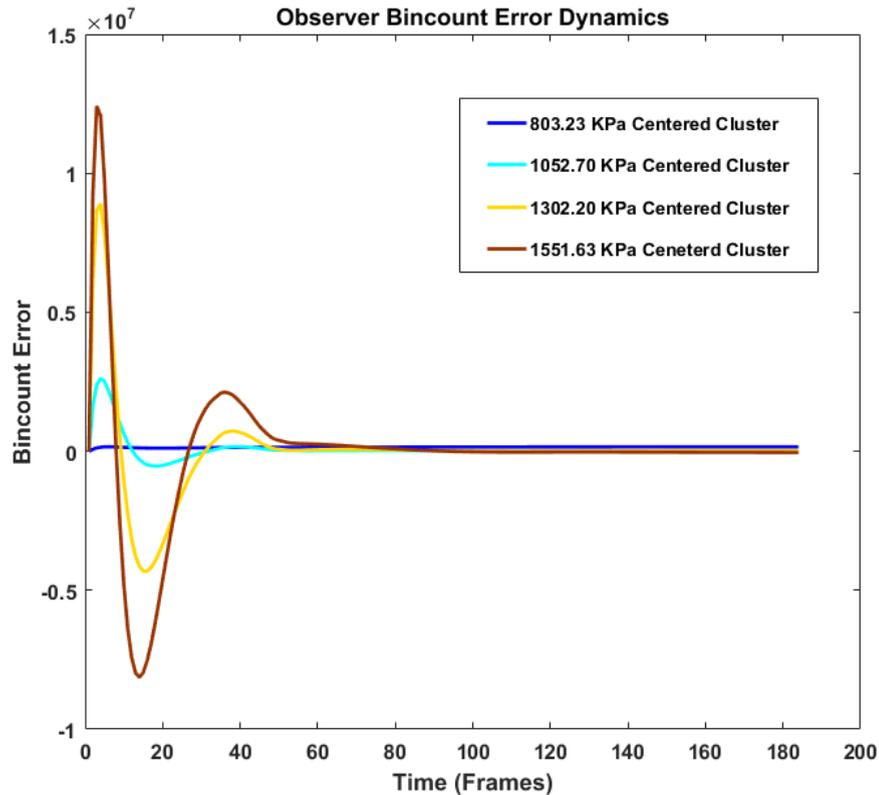


FIGURE 2.18: Ground observer bin count error dynamics. The figure shows that the observer is able to catch up with the actual bin count over the course of time

Luenberger observer developed in MATLAB with gain  $L = [0.62 \ 13.34 \ 50.68 \ 75.77]'$  so that the poles  $[0.85 \ 0.86 \ 0.87 \ 0.88]'$  of  $\phi - LC$  lie within the unit circle.

#### 2.4.4 Temperature data analysis

The previous sections explained the analysis over the pressure field of the caprock. Along the same lines Temperature or the steam chamber data can also be analysed. Figure (2.19) shows the evolution of the temperature field in the caprock. For brevity we do not provide detailed analysis here, but report the results in Appendix 6.2.

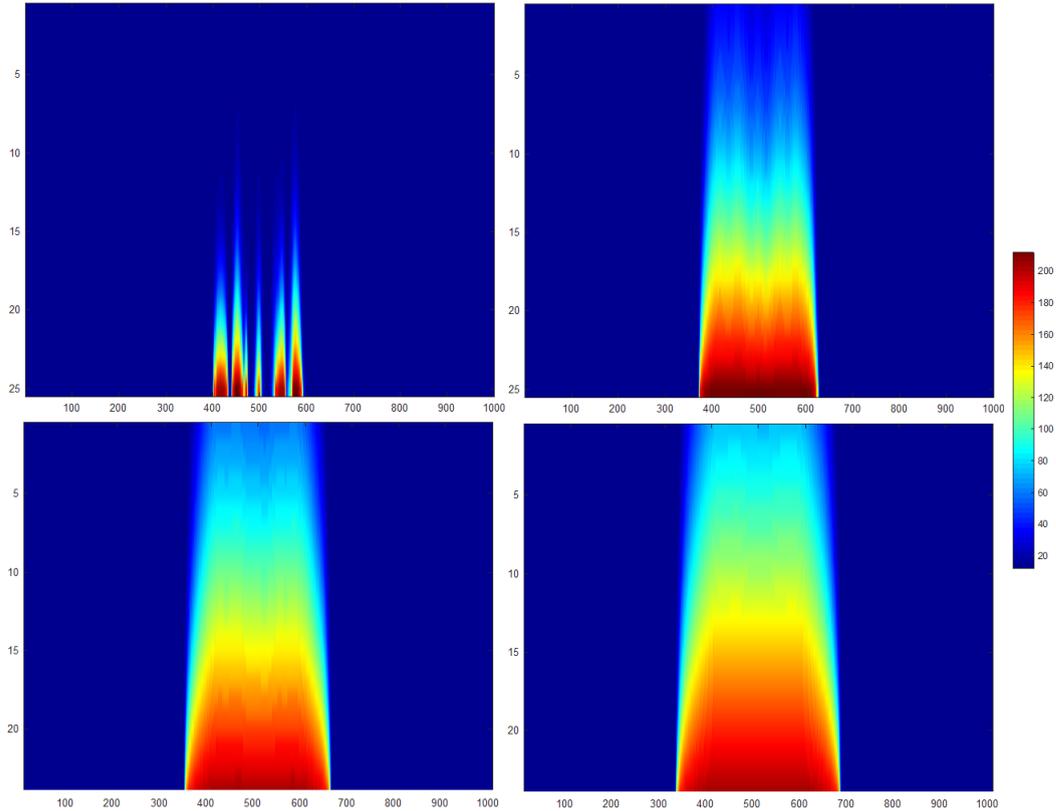


FIGURE 2.19: The image representation of different temperature data frames at different time instants (180, 140, 110, 50 from bottom right to top left). The temperature values are scaled and mapped to colors and an image is drawn. The reservoir has 5 injectors and 5 producers wells. Temperature is measured in degrees Celsius.

### 2.4.5 Practical considerations

The proxy models discussed so far are fairly computationally inexpensive and can be easily recomputed as soon as we get a new set of data, indicating that this can be used in monitoring and prediction. A single simulation for the reservoir and caprock over 184 time frames takes 24 hours, while each of the proxy models provide predictions of the behaviour over the same timeframe in seconds on the same computer. After discussing the different approaches to model the pore pressure and temperature, it is worth discussing the impact of pore pressure and temperature on the induced stress on the caprock hence on the FoS. Let us consider a simple model of a linear elastic rock reservoir which is assumed to be a very long cylinder with an elliptical cross section with an aspect ratio (thickness:width) of  $e$ . Pore pressure ( $\Delta P$ ) and temperature ( $\Delta T$ ) are uniformly changing within the reservoir, and the change in vertical stress in caprock and reservoir is directly proportional to the change in pore pressure and temperature as described by the

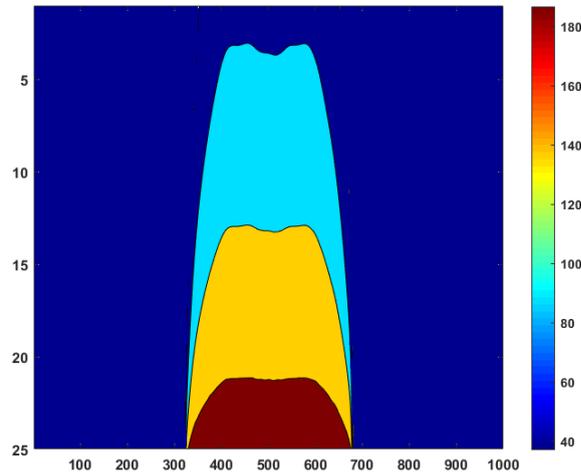


FIGURE 2.20: A typical frame with clustered temperature grid points: : Four colors corresponds to four clusters with high temperature 158.15 C - Brown, 116.38 C - Yellow, medium temperature 74.63 C - Cyan and low temperature 32.86 C - Blue with corresponding standard deviations of 56.86 C, 40.62 C, 24.39 C and 8.15 C.

following equation.

$$\Delta\sigma_V = \frac{1-2\nu}{1-\nu} \frac{e}{1+e} \Delta C$$

$$\Delta C = \alpha \Delta P : \text{Pore pressure change}$$

$$\Delta C = \frac{E\eta}{1-2\nu} \Delta T : \text{Pore temperature change}$$

Here  $\nu$  stands for Poisson's ratio,  $\eta$  is the thermal expansion coefficient and  $\alpha$  is the Biot coefficient. [37]. This illustrates the fact that there is a direct relationship between the pore pressure and temperature and the stress-strain profile in the caprock. An exhaustive analysis of the stress/strain profile based on detailed numerical calculations is outside the scope of our work in this manuscript, but will be pursued in our future work to develop a fully quantitative dynamic prediction for the factor of safety.

## 2.5 Comparison of proxy modelling methods

The Table 2.3 summarizes these two strategies and draws a comparison between them.

TABLE 2.3: POD verses Cluster Graph Comparison

Particular	POD	Graph	Comments
Linearity	Linear/Non-Linear	Linear	ARX models can be non-linear in nature but the cluster graph is inherently linear
Number of parameters	$n + 10(n - 1) + 10$ respectively for $A$ , $B_{1-10}$ and $k_{1-10}$ is $11n$	$2(n - 1)$	For a $n^{th}$ order model
System theoretic properties	Not obvious	Controllable and Observable	CG scores over POD
Nature	Generates the whole data field	Generates bincounts and lookup functions	POD scores over CG
Computational Complexity	Fairly Intense	Fairly simple	CG scores over POD
Parameter Estimation	Advanced optimization techniques	Ordinary Least Squares	CG scores over POD
Type of excitation used	PRBS	Step	ARX models inherently require PRBS excitation
Scalability	Fairly complex	Simple	CG scores over POD
Model Up-date	not straight forward	quite straight forward	CG scores over POD

## 2.6 Conclusions

The role of proxy models in reservoir management was discussed, and two computationally inexpensive proxy modelling approaches were discussed. A POD-based

---

model order reduction coupled with system identification was proposed to model and predict the pressure field. This approach demonstrated a promising result in modelling and predicting the entire pressure field and its response to variations in the reservoir inputs. Tensor analysis methods like Higher Order SVD and allied methods may provide farther mathematical insights into this dynamics, but the core idea behind the POD method was to generate a reduced order spatial basis and build a system identification framework to forecast the future frames with a low computational cost method. The clustering-based line graph model was proposed to model the pressure bin counts (regions of different ranges of pressures), which yielded a subspace model whose parameters were identified from the data obtained from detailed reservoir simulations. This model exhibited full controllability and observability properties and was successfully used in implementing an observer to track the bin counts by measuring the pressure in the outermost cluster by a sensor mounted on the ground surface. These methods can be implemented in forecasting the factor of safety dynamically based on predictions of the future injector and producer well bore pressures. The feasibility of the techniques discussed over the pore pressure were also used to model the pore temperature with similar results.

# Bibliography

- [1] G. van Essen, P. Van den Hof & J. D. Jansen (2013). A two-level strategy to realize life-cycle production optimization in an operational setting. *SPE Journal*, 18(06), 1-057.
- [2] A. I. Forrester, & A. J. Keane, (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3), 50-79.
- [3] A. Forrester, & A. Keane, (2008). *Engineering design via surrogate modeling: A practical guide*. John Wiley & Sons.
- [4] D. I. Zubarev, (2009, January). Pros and cons of applying proxy-models as a substitute for full reservoir simulations. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.
- [5] S. Wojtkiewicz, M. Eldred, Jr, R. Field, A. Urbina, & J. Red-Horse, (2001). Uncertainty quantification in large computational engineering models. In *19th AIAA Applied Aerodynamics Conference* (p. 1455).
- [6] J. W. Bandler, & K. Madsen, (2001). Surrogate modeling and space mapping for engineering optimization. *Optimization and Engineering*, 2(4), 367-368.
- [7] K. Lindhorst, M. C. Haupt, & P. Horst, (2014). Efficient surrogate modeling of nonlinear aerodynamics in aerostructural coupling schemes. *AIAA Journal*, 52(9), 1952-1966.
- [8] T. Heijn, R. Markovinovic and J. D. Jansen, Generation of low-order reservoir models using system-theoretical concepts. *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2003.
- [9] A. K. Tangirala, (2014). *Principles of system identification: Theory and practice*. CRC Press.
- [10] T. Soderstrom, & Stoica, P. (1989). *System identification*.

- 
- [11] A. C. Antoulas, Approximation of large-scale dynamical systems. Vol. 6. Siam, 2005.
- [12] S. Yao, J. J. Trivedi and V. Prasad, Proxy modeling of the production profiles of SAGD reservoirs based on system identification. *Industrial & Engineering Chemistry Research* 54.33 (2015): 8356-8367.
- [13] S. S. Vembadi, R. G. Patel, J. J. Trivedi, & V. Prasad, Real-time feedback control of SAGD wells using model predictive control to optimize steam chamber development under uncertainty. *The Canadian Journal of Chemical Engineering*, 96(6)(June 2018): 1290-1305.
- [14] S. Afra & E. Gildin, (2013). Permeability parametrization using higher order singular value decomposition (HOSVD). In *Machine Learning and Applications (ICMLA), 2013 12th International Conference on* (Vol. 2, pp. 188-193). IEEE.
- [15] G. R. Gavalas, P. C. Shah and J. H. Seinfeld, Reservoir history matching by Bayesian estimation. *Society of Petroleum Engineers Journal* 16.06 (1976): 337-350.
- [16] G. Chavent, M. Dupuy and P. Lemmonier, History matching by use of optimal theory. *Society of Petroleum Engineers Journal* 15.01 (1975): 74-86.
- [17] J. A. Jalali, S. D. Mohaghegh, & R. Gaskari, (2010). Coalbed methane reservoir simulation and uncertainty analysis with artificial neural networks. *Scientia Iranica. Transaction C, Chemistry, Chemical Engineering*, 17(1), 65.
- [18] P. A. Slotte, & E. Smorgrav, (2008, January). Response surface methodology approach for history matching and uncertainty assessment of reservoir simulation models. In *Europec/EAGE Conference and Exhibition*. Society of Petroleum Engineers.
- [19] G. Fishman, (2013). *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media.
- [20] A. Narasingam, P. Siddhamshetty, J. S. Kwon, Temporal clustering for order reduction of nonlinear parabolic PDE systems with time-dependent spatial domains: Application to a hydraulic fracturing process, *AIChE J.*, 2017, 63, 3818-3831.

- 
- [21] A. Narasingam, J. S. Kwon, Development of local dynamic mode decomposition with control: Application to model predictive control of hydraulic fracturing, *Comp. & Chem. Eng.*, 2017, 106, 501-511.
- [22] H. S. Sidhu, A. Narasingam, P. Siddhamshetty, J. S. Kwon, Model order reduction of nonlinear parabolic PDE systems with moving boundaries using sparse proper orthogonal decomposition: Application to hydraulic fracturing, *Comp. & Chem. Eng.*, 2018, 112, 92-100.
- [23] R. M. Butler, (1994). Steam-assisted gravity drainage: concept, development, performance and future. *Journal of Canadian Petroleum Technology*, 33(02), 44-50.
- [24] E. Gildin, & T. Lopez, (2011, January). Closed-loop reservoir management: Do we need complex models?. In *SPE Digital Energy Conference and Exhibition*. Society of Petroleum Engineers.
- [25] M. R. Carlson, (2012, January). An analysis of the caprock failure at Joslyn. In *SPE Heavy Oil Conference Canada*. Society of Petroleum Engineers.
- [26] J. Xiong, R. J. Chalaturnyk, Caprock Safety Factor Assessment of SAGD Projects. American Rock Mechanics Association (ARMA), 49th US Rock Mechanics / Geomechanics Symposium, San Francisco, CA, USA.
- [27] A. Azad, R. J. Chalaturnyk, A mathematical improvement to SAGD using geomechanical modeling. *Journal of Canadian Petroleum Technology* 49.10 (2010): 53-64.
- [28] A. Azad, R. Chalaturnyk & S. Movaghati, (2011). Reservoir Characterization: Application of Extended Kalman Filter and Analytical Physics-Based Proxy Models in Thermal Recovery. In *Proc., International Association for Mathematical Geosciences Conference IAMG*.
- [29] A. Azad, R. J. Chalaturnyk & S. Movaghati, Real-time reservoir model updating in thermal recovery: Application of analytical proxies and Kalman filtering. *Journal of Petroleum Science and Engineering* 127 (2015): 196-211.
- [30] J. D. Jansen, D. R. Brouwer, G. Naevdal & C. P. J. W. Van Kruijsdijk, (2005) Closed-loop reservoir management. *First Break*, 23(1), 43-48.
- [31] MATLAB version R2016a (9.0.0.341360) Natick, Massachusetts: The MathWorks Inc., 2016.

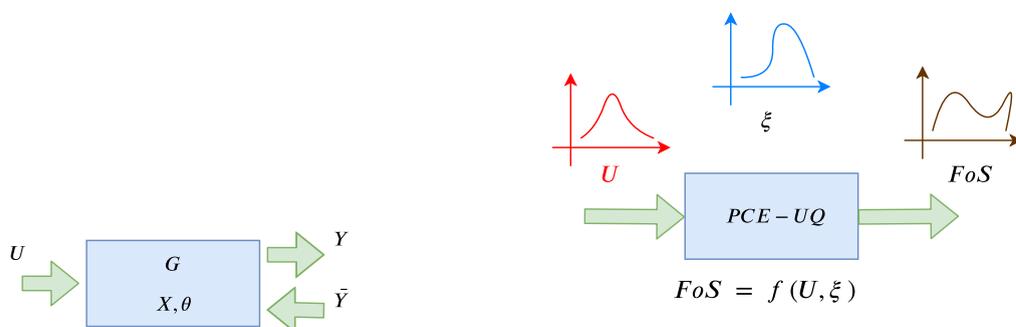
- 
- [32] A. Chatterjee, (2000). An introduction to the proper orthogonal decomposition. *Current Science*, 808-817.
- [33] Computer Modeling Group version 2015 (10.5715.22942) Calgary, Alberta, T2M 3Y7 Canada: Computer modeling Group Ltd., 2016.
- [34] M. H. Sahraei, M. A. Duchesne, P. G. Boisvert, R. W. Hughes & L. A. Ricardez-Sandoval, (2017). Reduced-Order Modeling of a Commercial-Scale Gasifier Using a Multielement Injector Feed System. *Industrial & Engineering Chemistry Research*, 56(25), 7285-7300.
- [35] M. H. Sahraei, M. A. Duchesne, R. W. Hughes & L. A. Ricardez-Sandoval, (2017). Dynamic reduced order modeling of an entrained-flow slagging gasifier using a new recirculation ratio correlation. *Fuel*, 196, 520-531.
- [36] M. Moharir, L. Kang, P. Daoutidis & A. Almansoori, (2017). Graph representation and decomposition of ODE/hyperbolic PDE systems. *Computers & Chemical Engineering*, 106, 532-543.
- [37] P. M. Collins (2005), Geomechanical Effects on the SAGD Process, SPE Reservoir Evaluations and Engineering Paper 97905: pp: 367-375.

## Chapter 3

Uncertainty quantification of the factor of safety in a steam-assisted gravity drainage process through polynomial chaos expansion

### 3.1 Introduction

Uncertainty and imprecision are unavoidable in all engineering problems, and two types of uncertainty quantification (UQ) can be carried out in model based optimal policy computation as shown in Figure 3.1(a): forward UQ and reverse UQ. Forward UQ deals with transferring input and parametric uncertainties to the outputs of the system and reverse UQ deals with propagating the input and output sensor uncertainties (noise) to the states [1]. This work deals with the former case for the factor of safety (FoS) in a steam assisted gravity drainage (SAGD) process as the output, with high parametric uncertainty in the reservoir permeability [2–5]. SAGD is a large scale nonlinear distributed parameter system spread across a vast geographical area and the limited availability of well logs to sample the heterogeneity makes it even more complex in UQ. Monte Carlo simulation [6] is a widespread method which is simple to implement but is computationally intensive, requiring a prohibitively large number of sampling points to apply effectively to SAGD [7]. Hence, we propose a computationally affordable technique, polynomial chaos expansion (PCE), that requires a handful of sampling points to achieve reasonable UQ of the FoS [8–10].



(a) **Forward UQ:** Propagating uncertainty in input  $U$  and system parameter  $\theta$  to output  $Y$ . **Reverse UQ:** Propagating uncertainty in input  $U$  and measured sensor output  $\bar{Y}$  to system states  $X$ .

(b) Forward uncertainty quantification of FoS with input  $U$  and petrophysical parameter  $\xi$  as the random variables.

FIGURE 3.1: Uncertainty quantification.

The SAGD process [11] shown in Figure 3.2 is the most widely used recovery technique for oil sands. Two parallel wells, the injector at the top and the producer at the bottom, run in parallel. The injector well injects steam into the reservoir, mobilizing the bitumen to flow down due to gravity to the producer. The development of the steam chamber applies a vertical stress field on the

caprock, which is an impermeable layer preventing the high pressure oil and gas from escaping to the surface. The FoS [12, 13] is a measure of the stability of the caprock and is computed as a function of the stress state and shear strength of the caprock layer. Numerically, it should be greater than 1 to assure caprock integrity. Caprock rupture incidents such as Joslyn [14] have proved the importance of caprock surveillance in SAGD operations and precisely quantifying the uncertainties in the FoS is essential to monitor it. The stress field on the caprock layer is mainly a function of reservoir pore parameters like pressure, temperature, well bottom hole pressure (well BHP) and the reservoir petrophysical parameters (mainly permeability). In this work, to have a parsimonious PCE representation, we only consider two of the factors affecting the stress field of the caprock, i.e. the well BHP and the reservoir permeability. Figure 3.1(b) shows

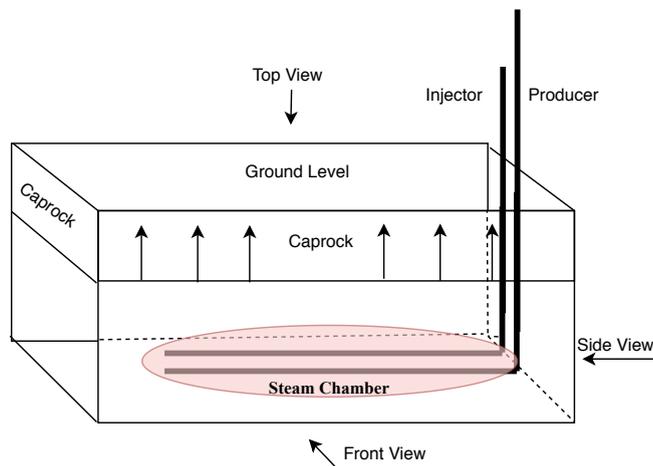


FIGURE 3.2: Schematic of a SAGD process showing two parallel wells (steam injector and the oil producer). The caprock layer shown at the top experiences stress due to the growth of the steam chamber.

the functional representation of the FoS as a function of the well bottom hole pressure as input ( $U$ ), and the petrophysical parameters ( $\xi$ ). An explicit form of this functional relationship ( $f$ ) is not directly available; hence, we are left with the option of analyzing this relationship through data obtained from a first-principles reservoir flow simulator CMG-STARs [15] sequentially coupled with FLAC3D [16], which is used to compute stress and strain (explained in detail in the next section). This work will present this simulator data-based technique to obtain a PCE model for the FoS. All of the computations are carried out in MATLAB [17].

PCE was introduced as a homogeneous chaos by Weiner and is derived

from the Cameron-Martin theorem to represent a random process in terms of orthogonal polynomials of standard random variables [18, 19]. As pointed out earlier, PCE is computationally efficient compared to Monte Carlo-based techniques [20]. PCE represents a random process as an infinite series of the products of orthogonal polynomials of standard random variables (RV) as basis functions and deterministic weighing coefficients:

$$Y = \sum_{i=1}^{\infty} y_i \psi_i(\xi) \quad (3.1)$$

$Y$  is the random process of interest,  $y_i$  are the deterministically computable weighting coefficients and  $\psi_i(\xi)$  are the orthogonal polynomials of RV  $\xi$  (basis functions) of order  $i$ . The recurrence relations of the orthogonal polynomials can be used to generate them and within a defined interval, they make an orthogonal basis for a defined inner product. The first  $N$  terms are considered depending upon the required accuracy. The most commonly used orthogonal polynomials include, but are not limited to, Hermite, Legendre and Chebyshev polynomials [21].

The proper orthogonal decomposition (POD) [22–24] refers to decomposing a given field (function) into orthogonal counterparts; this is practically equivalent to the singular value decomposition (SVD) [25, 26] for our case where the simulator FoS data is spatiotemporal data (also referred to as snapshots) [27]. The POD expansion allows us to estimate the spatial variation in the snapshots of FoS by decomposing the field into a set of basis functions and singular values (SVs). The SVs capture the variability present in the field hierarchically from the highest to the lowest (component-wise), and the variability in the spatiotemporal FoS data can be captured by the first few dominant SVs. Owing to this feature of the SVD, we consider the SVs of the FoS as the representatives of the uncertainty and model them with PCE while keeping the corresponding basis intact.

The organization of the chapter is as follows: Section 3.2 describes the SAGD system and the details of the sequentially coupled reservoir-geomechanics simulator setup and the obtained data format. Section 3.3 explains the analysis techniques, followed by Section 3.4 providing the corresponding results and discussion. Section 3.5 provides conclusions and future directions.

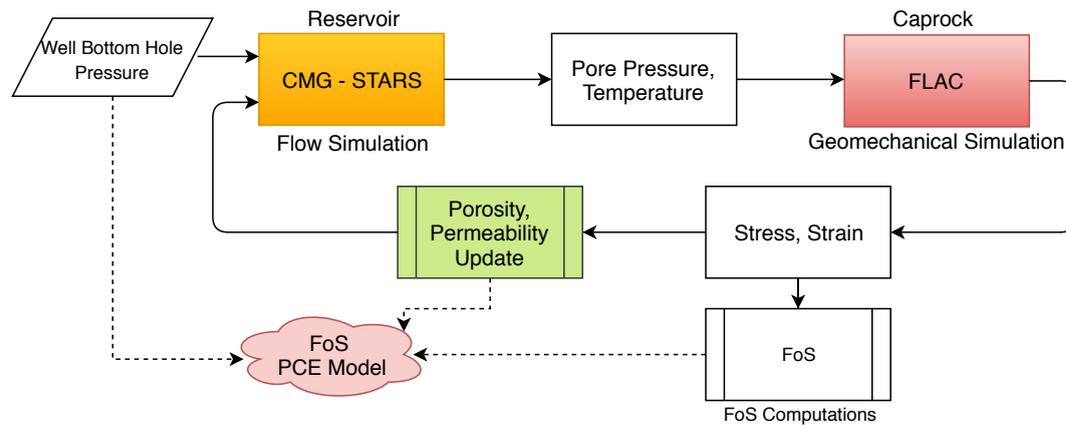


FIGURE 3.3: Sequentially coupled simulation between the CMG-STARS and FLAC3D; the pore parameters change affects the stress strain which in turn affects the petrophysical parameters, which changes the pore parameters.

## 3.2 Process modelling, data generation and visualization

This section describes the details of the simulation framework and the tools used in the data generation and visualization. To start with, we describe the details of the sequentially coupled simulation platform, which consists of CMG-STARS for reservoir flow simulation and FLAC3D for computing stress and strain. A sequential coupling platform facilitates the study of the effect of geomechanics and in turn the FoS in a SAGD process. This platform enables to visualize the interactions between flow and deformation response in subsurface modelling in the SAGD process [28]. Figure 3.3 shows the flow chart of the coupled reservoir-geomechanics simulation and the use of the PCE model for UQ. The pore pressure and temperature obtained from CMG-STARS are used to compute the stress and strain through FLAC3D. This information is transformed into the updates in the petrophysical parameters of the reservoir, and the pore pressure and temperature are re-computed with the updated parameters and the loop goes on (refer [29] for details).

A wellpad of MacKay River oil sand area is selected and applied in the coupled platform with geomechanical and reservoir models. The schematic of the reservoir model with the permeability distribution and the geomechanical model under consideration is shown in Figures 3.4 and 3.5, respectively. A total of 6 well pairs operate in the pad with well spacing of 100 m. The grid size and numbers vary in different formations to retain the inherent heterogeneity

due to the sedimentation process. The Wabiskaw and McMurray formations are included in both the reservoir and geomechanical model, and the Till, Clearwater, and Devonian formation are only included in the geomechanical model for computational efficiency, as the flow response in these three formations can be ignored due to the low permeability. The injectors open from day 1 for preheating of the cold bitumen near the wellbore and the producers open in 150 days for production; hence, for the purpose of modelling, data from day 160 and onwards is considered. The caprock region, which is the region of interest in this study, corresponds to the Clearwater formation, and the reservoir corresponds to the Wabiskaw and McMurray formations; the rest is the under and over burdens. A pseudorandom binary sequence (PRBS) input for the well BHP for injectors and producers is designed to provide excitation to the system and generate suitable data for the purposes of modelling and identification. The designed PRBS frequency range is between three to six months, sufficient enough to capture the variability in the FoS due to the well BHP. In this work, the variability in the FoS data is due to the uncertainty in the petrophysical and geomechanical parameters and the input PRBS of the bottom hole pressures. The PCE model developed serves as the UQ tool in both inputs and the parameters.

### 3.2.1 Data visualization

Here, we present the steam chamber evolution at different instants of time. Figure 3.6 shows the steam chamber growth with Figure 3.6(a), Figure 3.6(b) and Figure 3.6(c) showing the 3D steam chamber at  $t=1$ , 10 and 25 (time frames) respectively. Each time frame corresponds to 100 days of operation. The spread of the steam chamber across the reservoir and the heterogeneity in the permeability is clearly seen in the figure.

Working with PCE requires the generation of an ensemble of realizations of the random process of interest. In our case, it is the FoS realizations and the corresponding well BHP input. Since we consider working with a two-dimensional reservoir problem in this work, we conduct a single heterogeneous three-dimensional simulation and use each two-dimensional slice perpendicular to the well direction (XZ) as a separate realization. This approach is better as it considers the cross-flow terms capturing the true dynamic flow behaviour during the SAGD process, and is therefore our preferred approach for creating 2D realizations. To summarize, 200 XZ frames along the Y direction of the 3D

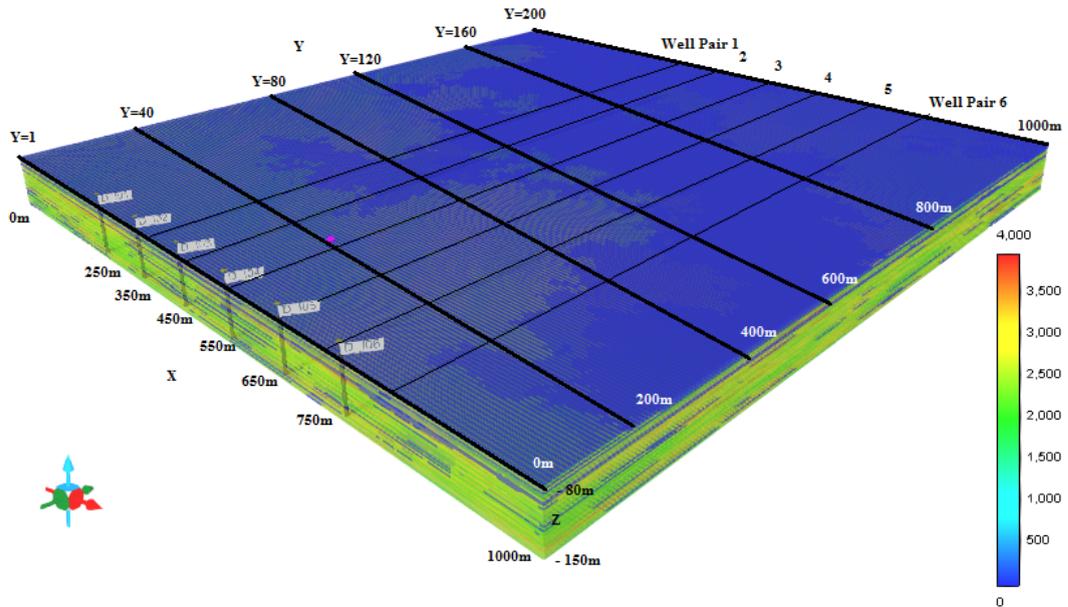


FIGURE 3.4: The working reservoir model in 3D with permeability I distribution in milli darcy. Well pairs 1-6 can also be seen.

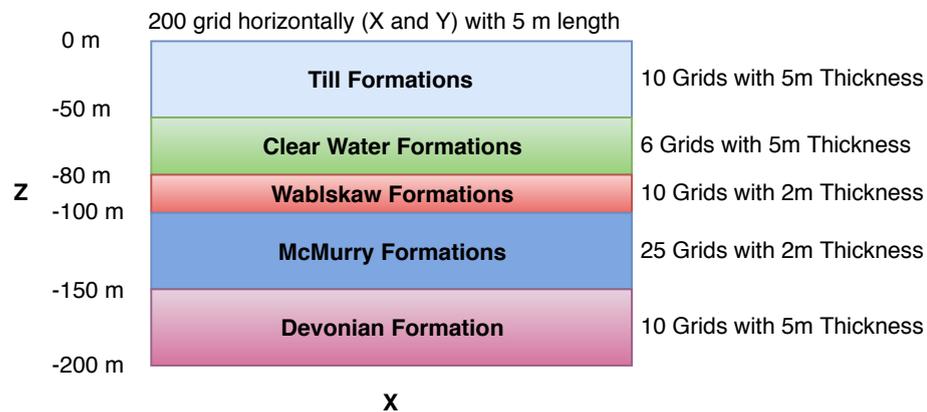


FIGURE 3.5: Reservoir model schematic (XZ plane description) with grid size and numbers. Caprock: Clear water Formations; Reservoir: Wablskaw and McMurry formations.

model are used as 200 2D realizations to construct the ensemble. Figure 3.7 demonstrates the formatting procedure to obtain the 2D realizations from a 3D simulation. Figure 3.8 shows the average change of the permeability across all the realizations at  $t=1$  (Figure 3.8(a)),  $t=10$  (Figure 3.8(b)) and  $t=25$  (Figure 3.8(c)) respectively. In sequential coupling, the permeability distribution is updated at

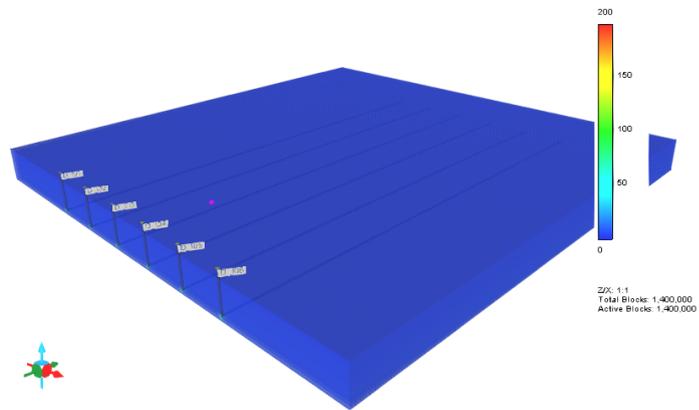
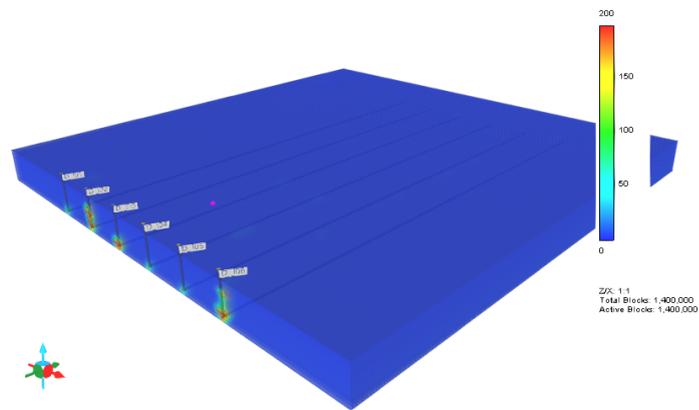
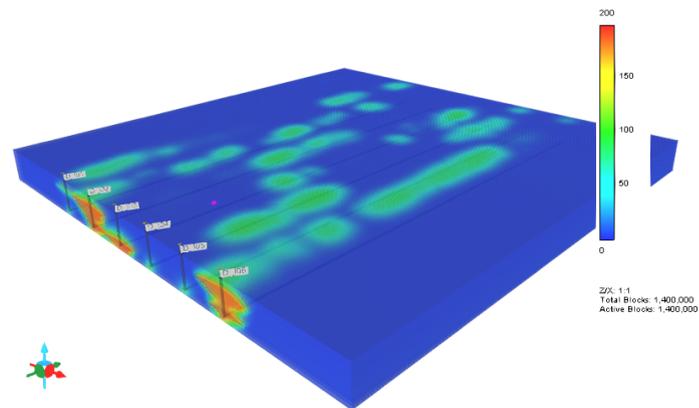
(a) Steam chamber growth across the reservoir at  $t=1$ .(b) Steam chamber growth across the reservoir at  $t=10$ .(c) Steam chamber growth across the reservoir at  $t=25$ .

FIGURE 3.6: Reservoir steam chamber visualization at different times.

each time step. The permeability distributions presented in Figure 3.8 are the ‘posterior’ permeability distributions obtained after the updates, which will be used in the PCE model development.

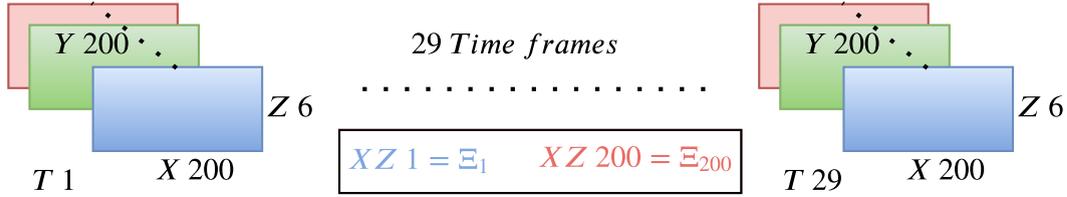


FIGURE 3.7: Working data format: 200 realizations ( $\Xi_1 - \Xi_{200}$ ) of spatiotemporal FoS data taken as the XZ frames. Data at every instant  $T$  has 200 grid points in  $X$ , 200 grid points in  $Y$  and 6 grid points in  $Z$  taken over 29 time frames. Every realization  $\{\Xi_k\}_1^{200}$  is of dimension  $200 \times 6$ . The PCE model is computed over the realizations for each time instant.

### 3.3 Working procedure

Figure 3.9 presents the modelling work flow. The first three columns to the left in the figure show the realizations of the reservoir permeability  $\{\Xi_k\}_1^{200}$ , the corresponding reservoir pressure and/or temperature fields computed from CMG-STARS and the FoS field realizations computed from the sequentially coupled FLAC3D data (refer to Figure 3.3 for more details on sequential coupling), respectively, at any instant of time  $t$ . The singular value decomposition of the FoS realizations yields basis  $\{U_k\}_1^{200}$  and  $\{V_k\}_1^{200}$  and the ensemble of singular values (SVs)  $\{S_{r,k}\}_{r=1,k=1}^{r=6,k=200}$ . The SV ensemble acts as the ‘representative’ of the FoS field, and is used as the output of the PCE model of the underlying FoS field.

#### 3.3.1 FoS computation

After obtaining the stress data from FLAC3D, the FoS can be computed for every grid point using

$$\tau_{strength} = (\sigma - P_p)\sin(\phi) + C\cos(\phi) \quad (3.2)$$

where  $\sigma_{max}$  and  $\sigma_{min}$  are the maximum and minimum stress component of the stress tensor, respectively.  $P_p$  is the pore pressure,  $C$  is the cohesion and  $\phi$  is the friction angle. All of these variables are directly available from the coupled simulations to compute the FoS.

$$\tau = \frac{\sigma_{max} - \sigma_{min}}{2} \quad (3.3)$$

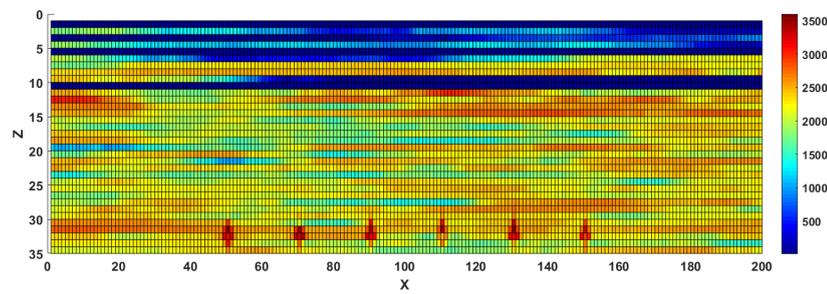
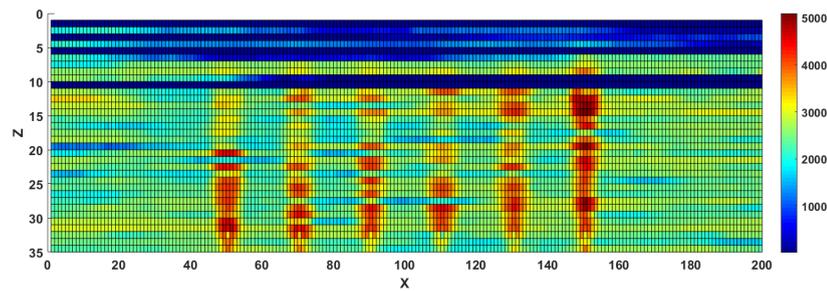
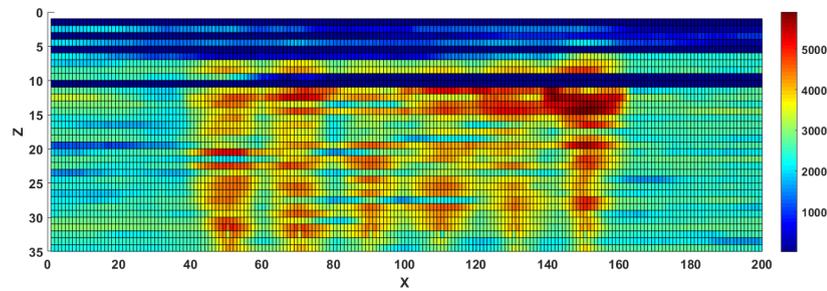
(a) Average reservoir XZ permeability over all realizations at time  $t=1$ .(b) Average reservoir XZ permeability over all realizations at time  $t=10$ .(c) Average reservoir XZ permeability over all realizations at time  $t=25$ .

FIGURE 3.8: Reservoir permeability field visualization at different times.

$$FoS = \frac{\tau_{strength}}{\tau} \quad (3.4)$$

### 3.3.2 Singular value decomposition: (SVD)

SVD decomposes a given matrix into a product of two basis vectors and the matrix of SV. SVD hierarchically captures the total variance in the data into the basis vectors scaled by the matrix of SVs (highest to lowest). In essence, the ratio of the magnitude of a particular SV to the sum of all SVs is proportional to the relative magnitude of the variance captured by it. Also, each SV along

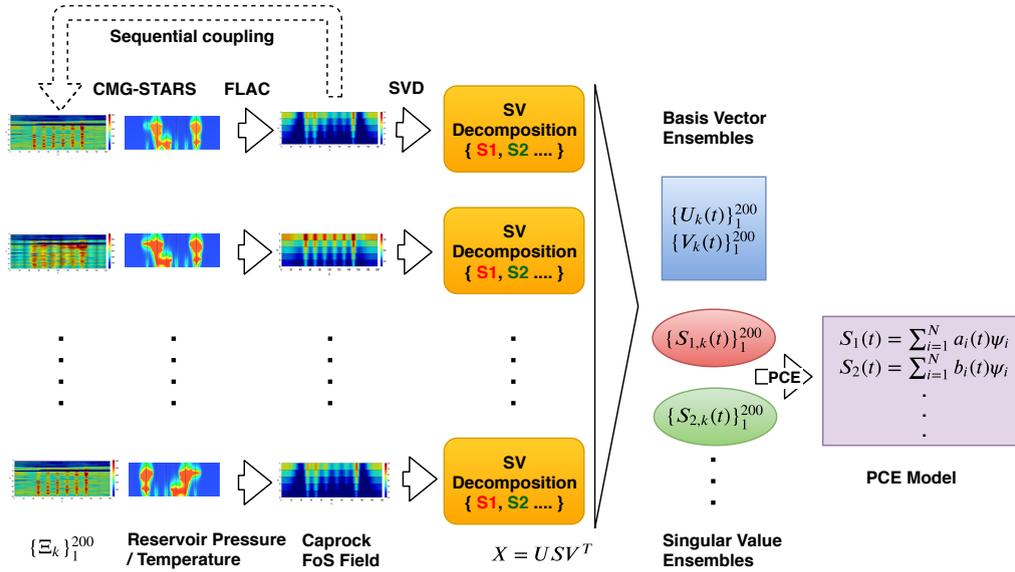


FIGURE 3.9: Overview of the work flow to develop a PCE model for the FoS field in a sequentially coupled reservoir simulation using SVD.

with its basis can be used to construct a low rank approximation of the original data and the highest SV gives the matrix 2-norm of the original data. In this context, we call the SVs as the representatives of the original data (in our case, the FoS field). SVD has been used extensively in data analysis both in deterministic, stochastic and spatiotemporal frameworks.[30–32]. We use the SVD for two purposes in our work. First, we use it to analyse the FoS field to determine how many SVs and corresponding basis vectors are required to reconstruct the field accurately. Secondly, we use the SVD on the reservoir permeability field realizations to quantify or to find a set of representatives of a particular realization to be used in building a PCE model.

The FoS data  $D$  is four dimensional (spatial dimension:  $200 \times 200 \times 6$  grid blocks, with 29 time frames) and is considered as 200 realizations of a frame of dimension  $200 \times 6$  with data available at 29 time instants (see Figure 3.7). The SVD for a particular realization  $\Xi_k$  of the frame at any time instant  $\{t\}_1^{29}$  is given by

$$D_{200 \times 6}(t) = U_{200 \times 6}(t) \Sigma_{6 \times 6}(t) V'_{6 \times 6}(t) \quad (3.5)$$

$$\Sigma(t) = \begin{bmatrix} \sigma_1(t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r(t) \end{bmatrix} \quad (3.6)$$

$U(t)$ ,  $V(t)$  are orthogonal matrices and  $\Sigma(t)$  is a diagonal matrix at any time frame  $t$ , with the SVs of  $D$  arranged in descending order. These basis vectors capture the variance in the data in the least squares sense and their corresponding SVs quantify the contribution of each basis vector. The ratio  $\frac{\sum_{i=1}^r \sigma_i}{\sum_{i=1}^6 \sigma_i}$  quantifies the percentage of variance captured by the first  $r$  SVs (out of 6 in our case). To keep the computational cost of building the PCE model low, we choose the first  $r$  SVs and fit a PCE model for each of them. Out of the  $r$  SVs, the first SV ( $S_1$ ) is the most important as it quantifies the 2-norm of the corresponding frame, which can be effectively taken as a representation of that frame.

$$S_r(t) = \{ F_r(\xi, \mu, t) \}_{r=1, t=1}^{r=6, t=20} \quad (3.7)$$

The reservoir permeability realizations  $\Xi_{1-200}$  are analyzed similarly; here, the data  $D$  is also four-dimensional (spatial dimensions  $200 \times 200 \times 35$  over 29 time frames) and is viewed as 200 realizations of frames of dimension  $200 \times 35$  at 29 time instants (see Figure 3.7). The SVD for a particular realization  $\Xi_k$  of the frame at any time instant  $\{t\}_1^{29}$  is given by:

$$D_{200 \times 35}(t) = U_{200 \times 35}(t) \Sigma_{35 \times 35}(t) V'_{35 \times 35}(t) \quad (3.8)$$

Here, we choose the first  $r$  dominant SVs to represent the  $k^{th}$  realization  $\Xi_k$ . Figure 3.10 explains the application of SVD in representing a realization  $\Xi_k$  by its dominant SVs ( $\xi_{k1}, \xi_{k2} \dots \xi_{kr}$ ). This procedure helps us to represent a 2D permeability field by a sequence of SVs, which is useful in computing the PCE model. It is worthwhile to mention that the  $\Xi_k$  represents the ‘posterior permeability’ of the sequential coupling process.



FIGURE 3.10: Representing each realization with a set of  $r$  representatives obtained from SV decomposition of the reservoir permeability field.

### 3.3.3 Polynomial chaos expansion

After computing the SV distribution of the FoS field over its realizations for every instant of time  $\{S_r(t)\}_1^r$ , we can develop a PCE model for each  $S_r(t)$  as a function

of the well BHP ( $\mu$ ) as the input and the permeability vector ( $\xi$ ) as the parametric uncertainty at every instant of time. The parametric uncertainty  $\xi$  incorporates the intrinsic heterogeneity in the permeability, the error in computing the permeability field due to the limited availability of well logs and the uncertainty arising in the sequential coupling. The input variable  $\mu$  is assumed to have a uniform distribution

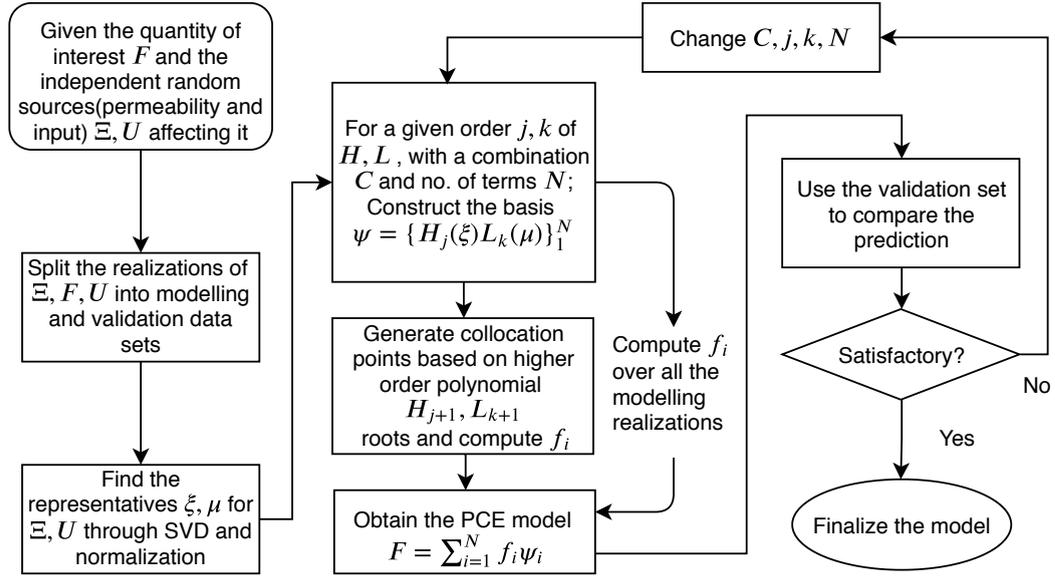


FIGURE 3.11: PCE workflow algorithm.

and is expanded with a sequence of Legendre polynomials that are orthogonal on a scaled interval  $[0, 1]$ . The uncertain parameters  $\xi_i$  (representatives) are normalized by scaling them as standard normal variables and expanded with a sequence of Hermite polynomials on the interval  $[-\infty, \infty]$ . The output variable of interest,  $S_r$  (SVs of the FoS), is expanded as a combination of the polynomials as follows:

$$S_r(t) = f_1(t)H_1(\xi) + \dots + f_{r+1}(t)L_1(\mu) + \dots + f_i(t)H_j(\xi)L_k(\mu) + \dots \quad (3.9)$$

Equation 3.9 represents the general PC expansion, where  $i$  represents the number of terms considered (based on truncation of equation 3.1), and  $j$  and  $k$  are the order of the polynomials  $H$  and  $L$ , respectively. The PCE coefficients  $f_i(t)$  are computed from the process data  $\{S_{r,k}\}_{r=1,k=1}^{r=6,k=200}$ , permeability data  $\{\Xi_k\}^{200}$  and the input  $\mu$  for every  $t$  using regression and collocation methods [20].

Figure 3.11 describes the details of the implementation of the PCE modelling. The realizations are split into modelling and validation data, with PCE coefficients being computed over all the modelling realizations and the developed

PCE model then being tested over the validation realizations to check the modelling adequacy. If the model's accuracy in validation is not satisfactory, the number of terms, the order of the polynomial basis, and the combination of the polynomials are changed; the new PCE model is rebuilt and checked again for adequacy over the validation realizations. This process is repeated until an adequate model is found. Regression analysis indices such as the error sum of squares (SSE), the coefficient of determination ( $R^2$ ) and the adjusted coefficient of determination ( $R_{adj}^2$ ) are used as the metrics for model adequacy.

A pair of polynomials,  $f(x)$  and  $g(x)$ , are said to be orthogonal in an interval  $[a, b]$  if

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx = 0 \quad (3.10)$$

where  $\langle f, g \rangle$  denotes the inner product of  $f(x)$  and  $g(x)$ . A sequence of such orthogonal polynomials can be generated for a given interval  $[a, b]$  using their recurrence relationship. Table 3.1 gives the Hermite and Legendre orthogonal polynomials of order 1-4 and their respective roots over the intervals  $[-\infty, \infty]$  and  $[0, 1]$ . The orthogonal polynomials can be generated using the Python library ORTHPOL [33].

TABLE 3.1: Hermite ( $H_i$ ) and Legendre ( $L_i$ ),  $i^{th}$  order polynomials defined respectively over the intervals  $[-\infty, \infty]$  and  $[0, 1]$  alongwith their roots.

Type	Polynomial	Roots
$H_1(\xi)$	$\xi$	0
$H_2(\xi)$	$\xi^2 - 1$	$\pm 1$
$H_3(\xi)$	$\xi^3 - 3\xi$	$0, \pm\sqrt{3}$
$H_4(\xi)$	$\xi^4 - 6\xi^2 + 3$	$\pm 2.3334, \pm 0.7420$
$L_1(\mu)$	$\mu - 0.5$	0.5
$L_2(\mu)$	$\mu^2 - \mu + \frac{1}{6}$	0.7887, 0.2113
$L_3(\mu)$	$\mu^3 - \frac{3}{2}\mu^2 + \frac{3}{5}\mu - \frac{1}{20}$	0.5, 0.8873, 0.1127
$L_4(\mu)$	$\mu^4 - 2\mu^3 + \frac{9}{7}\mu^2 - \frac{2}{7}\mu + \frac{1}{70}$	0.9306, 0.6700, 0.3300, 0.0694

### 3.4 Results and discussion

A set of randomly chosen 100 permeability realizations are used for training the PCE model and the other 100 realizations are used for validating it.

### 3.4.1 FoS computation

The factor of safety is computed using equations (3.2, 3.3 and 3.4) at every grid block for all realizations over time. Now we need a scalar measure of FoS for every frame to mathematically compare its magnitude with other frames and at other instants of time. For this purpose, we propose two schemes to obtain a scalar measure: in the first, we find the *minimum FoS* over all the grids points in a frame and call it the representative of that frame. In the second, the first SV ( $S_1$ ) of the FoS field (which represents the matrix 2-norm of the field) is normalized by dividing it by the 2-norm of a matrix of ones of the dimension of the frame ( $200 \times 6$ ). Figure 3.12 shows the evolution of the minimum FoS (first measure) of every frame over all realizations and time. Similarly, Figure 3.13 shows the evolution of the normalized highest SV (second measure) over realizations and time. These two visualizations present the two different scalar measures (representatives) of the FoS of a frame over all realizations and time. The trend in both of the representations is similar; the FoS measure decreases with time and the variation across the realizations increases over time. Figure 3.14 visualizes exact locations of occurrence of the minimum FoS in the X direction over all realizations (Y) (reported with X, Y and Z grid numbers). Figures 3.14(a), 3.14(b) and 3.14(c) shows the variation at  $t=1$ ,  $t=10$  and  $t=25$ , respectively. At each X location, the minimum FoS always occurs at the bottom layer of the caprock (at  $Z=6$ , the layer right above the reservoir). This is expected as the closeness of the steam chambers influences the FoS of the bottom of the caprock stronger than the other regions of the caprock. Also, it is interesting to note that the location of the minimum FoS is confined to the middle zone of X where the six well bores are placed.

### 3.4.2 SVD

The SV decomposition of the permeability field  $\Xi$  over all the realizations and time shows that only first three dominant SVs are good enough to capture most of the variance, as shown in Figure 3.15. The ratio  $\frac{\sum_{i=1}^3 \xi_i}{\sum_{i=1}^{35} \xi_i}$ , considering three of the total SVs averaged over all realizations and time, is 0.4912, which means the first three SVs capture 49.12% of the total variance in the permeability; this implies 3 random variables  $\xi_1, \xi_2$  and  $\xi_3$  can characterize the permeability field covering 49.12% of the total variance. Hence, every  $k^{th}$  realization of the permeability field  $\Xi_k$  at an instant of time  $t$  is characterized by the set  $\{\xi_{k1} \ \xi_{k2} \ \xi_{k3} \ \mu_t\}$ , where the random variable  $\mu$  represents the input (well BHP).

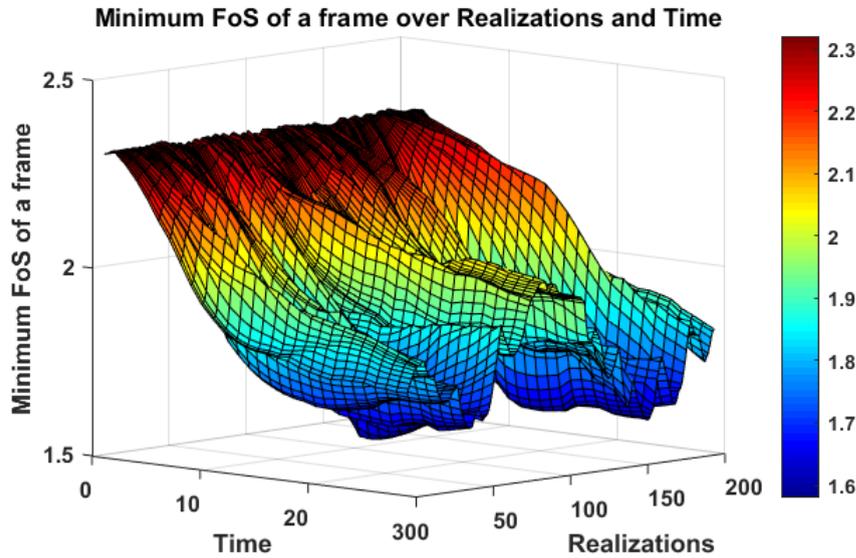


FIGURE 3.12: Evolution of actual minimum FoS over different realizations and time.

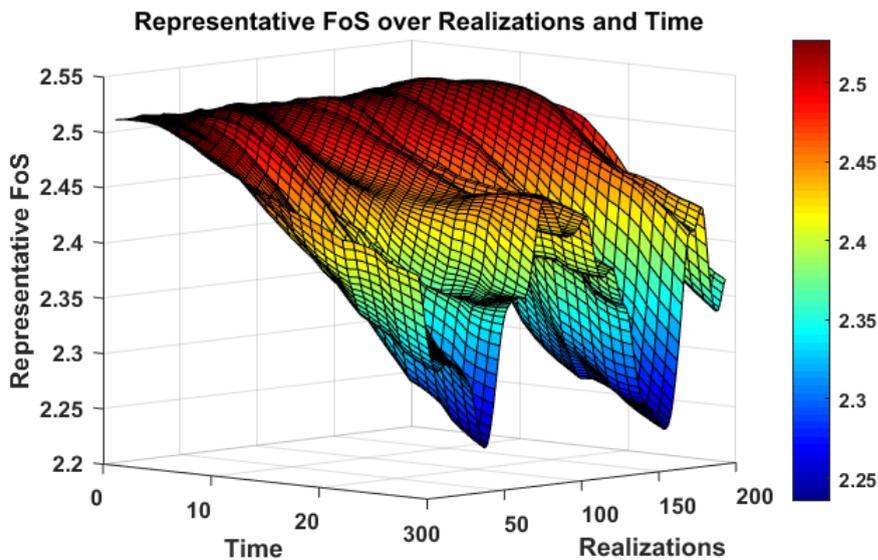


FIGURE 3.13: Evolution of representative FoS (normalized highest SV) over different realizations and time.

On the other hand, Figure 3.16 shows the SVs ( $S_1 \dots S_6$ ) of the FoS frame over all the realizations and time. The ratio  $\frac{\sum_{i=1}^2 S_i}{\sum_{i=1}^6 S_i}$  considering two of the total SVs averaged over all realizations and time, is 0.9995, and if only the first SV is considered, the ratio becomes  $\frac{S_1}{\sum_{i=1}^6 S_i} = 0.9985$ . That means the first SV captures over 99% of the total variance in the FoS data. This justifies the choice of three random variables to characterize the permeability: even though they capture only 49.12% of its variance, the corresponding variance of the FoS that is captured is

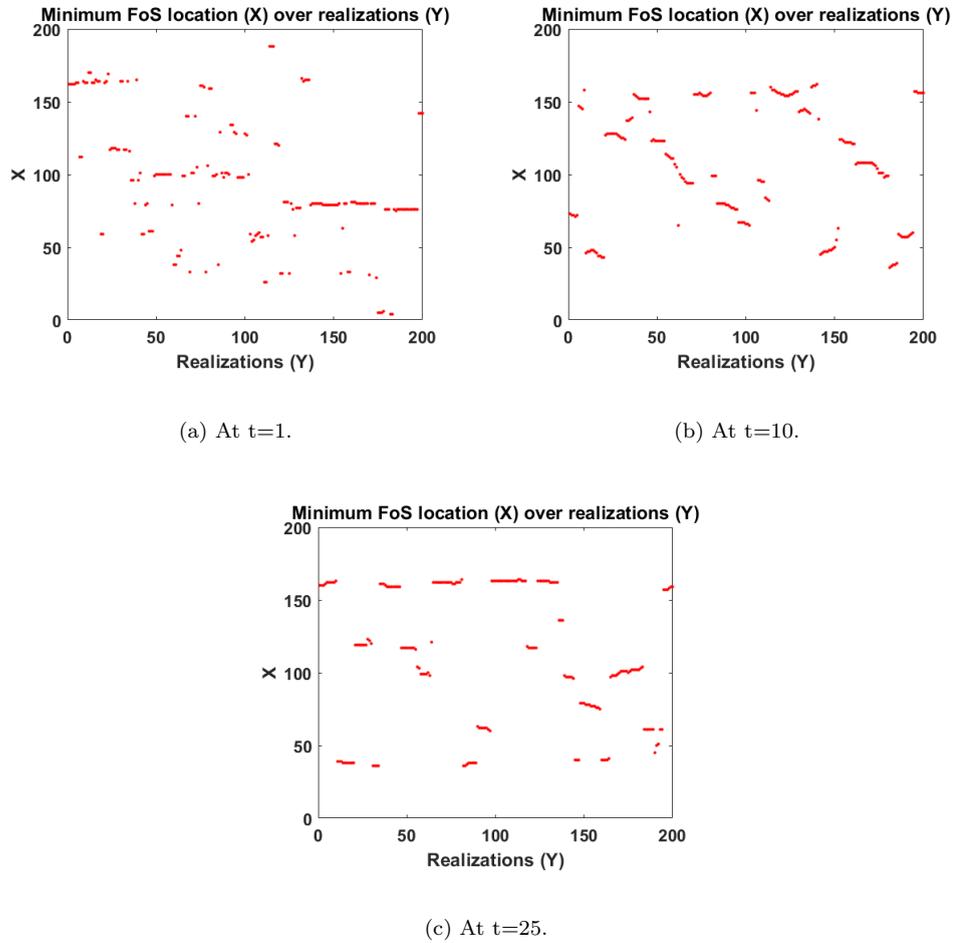


FIGURE 3.14: Minimum FoS locations in X over realizations (Y) at Z=6 (the bottom layer of the caprock, i.e. the layer adjacent to the reservoir). These are the locations in the XY plane where the FoS is the least at different time instants.

over 99%.

Once the number of SVs to be considered for the FoS field is decided, we can use the averaged basis over all the realizations  $U(t)$  and  $V(t)$  and reconstruct the FoS field from only two out of six SVs. Figures 3.17, 3.18 and 3.19 demonstrate this process at t=1, t=10 and t=25, respectively. Figures (3.17(a), 3.18(a), 3.19(a)), (3.17(b), 3.18(b), 3.19(b)) and (3.17(c), 3.18(c), 3.19(c)) show the actual FoS field, the reconstructed FoS field with the first two dominant SVs and the reconstruction error = [actual frame - reconstructed frame], respectively, at time instants 1, 10 and 25 instants. In all of the cases, the Frobenius norm of the error frame  $\|E\|_F$  is reported as a metric to evaluate the reconstruction error. The Frobenius norm is defined as  $\|E\|_F = (\sum_{i=1}^6 \sum_{j=1}^{100} |e_{ij}|^2)^{1/2}$ , where  $e_{ij}$  is the

error at each grid point.  $\|E\|_F$  is quite low at the start, increases later then and seems to saturate. The number of SVs and the effectiveness of the technique are deemed to be satisfactory, since  $\|E\|_F$  is relatively small at all times.

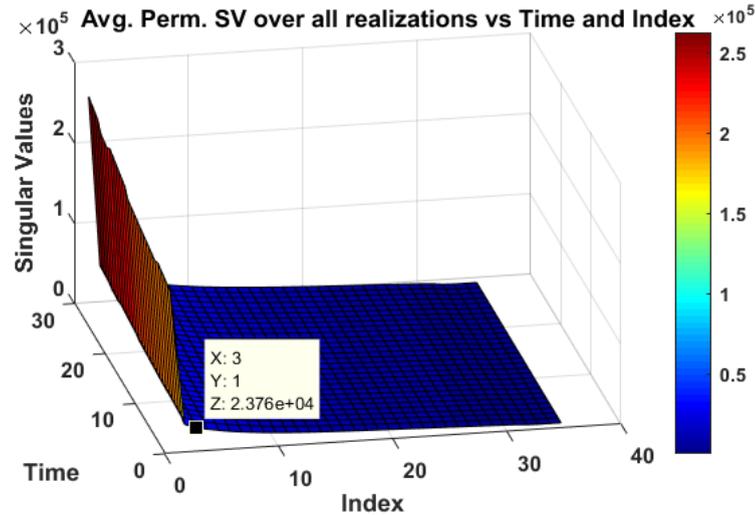


FIGURE 3.15: Average permeability SVs over realizations vs their index and time. A sharp decrease at the knee point is clearly seen at the third SV, where  $\frac{\sum_{i=1}^3 \xi_i}{\sum_{i=1}^{35} \xi_i} = 0.4912$ . Three SVs of each realization are used as its representative in PCE expansion.

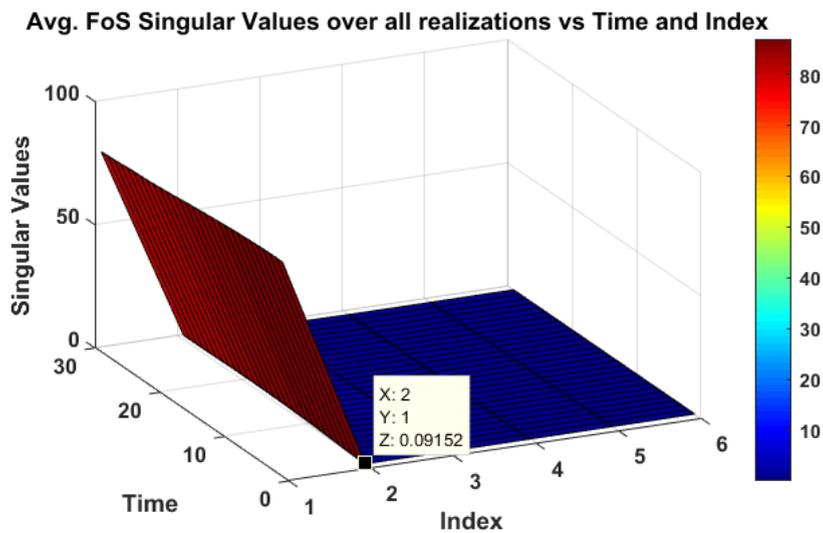
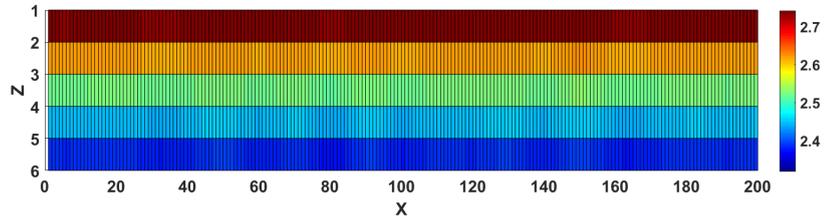
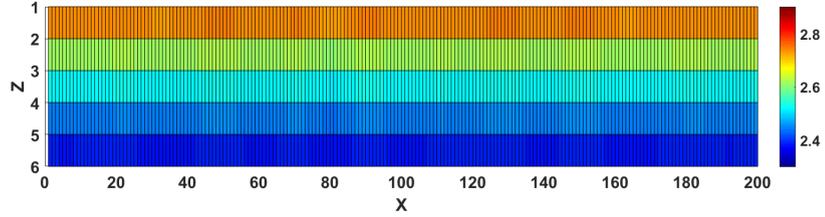


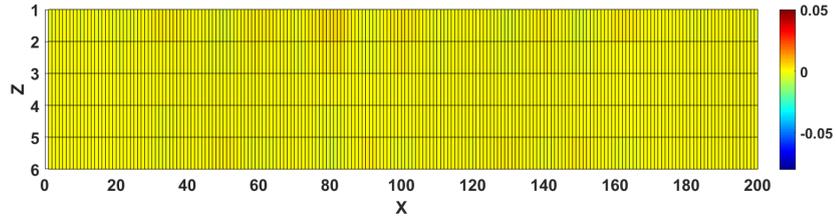
FIGURE 3.16: Average FoS SVs over all realizations versus their index and time. A sharp decrease at the knee point is clearly seen at the second SV, where  $\frac{\sum_{i=1}^2 S_i}{\sum_{i=1}^6 S_i} = 0.9995$ . Hence, at most two FoS SVs of each realization can capture 99.95% of the variance of the FoS field.



(a) Actual average FoS frame over all realizations.

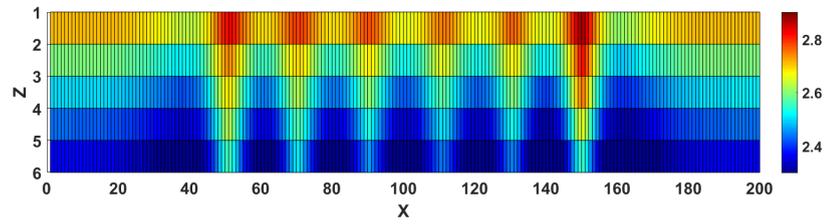


(b) SVD reconstructed average FoS frame over all realizations.

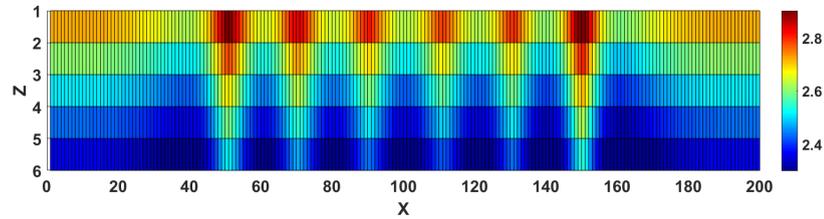
(c) Error frame = Actual - Reconstructed,  $\|E\|_F = 0.04$ .FIGURE 3.17: SVD-based FoS field reconstruction with two SVs; visualization at  $t=1$ .

### 3.4.3 Polynomial chaos expansion

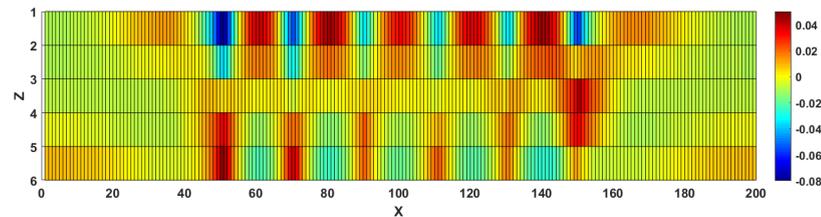
The SVs  $\{S_r\}_1^6$  of all realizations of permeability are computed and normalized as a standard Gaussian random variable  $\xi$  between  $[-\infty, \infty]$  for all time. The well BHP input is normalized as a standard uniform random variable  $\mu$  between  $[0, 1]$ . The realizations are characterized by  $\xi$  and  $\mu$  with the first three dominant SVs of the permeability field; this means that the set  $\{\xi_{r1}, \xi_{r2}, \xi_{r3}, \mu_t\}$  represents a particular realization  $\Xi_r$ . Equations 3.11 and 3.12 give the PCE expansions for  $\{S_r\}_1^6$  of first and second order, respectively, in terms of  $\mu$ ,  $\xi_1$ ,  $\xi_2$  and  $\xi_3$ . Each of the terms in the expansion and the corresponding estimated coefficients are tabulated in Tables 3.2 and 3.3. The model coefficients  $f_i$  are computed using regression or collocation methods. The regression coefficients are estimated by performing least



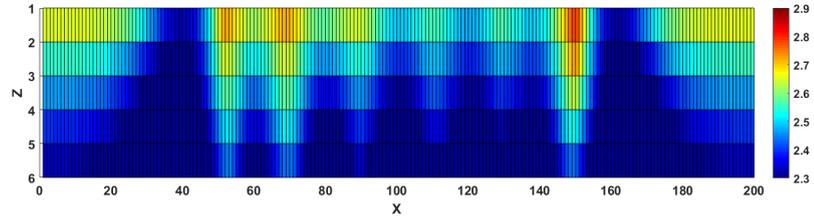
(a) Actual average FoS frame over all realizations.



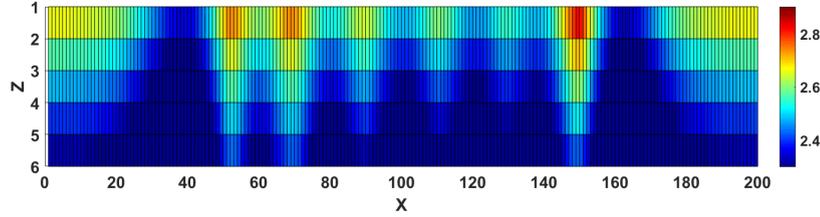
(b) SVD reconstructed average FoS frame over all realizations.

(c) Error frame = Actual - Reconstructed,  $\|E\|_F = 0.56$ .FIGURE 3.18: SVD-based FoS field reconstruction with two SVs; visualization at  $t=10$ .

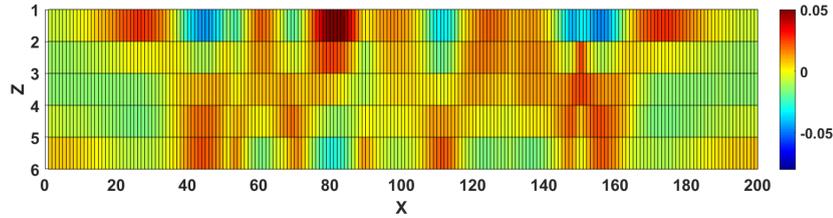
squares regression of the PCE model over all the modelling realizations and computing  $f_i$  over them. In the collocation method, specific realizations are chosen as the collocation points and  $f_i$  is computed over them. The Gaussian quadrature technique [34] is used to choose the collocation points. It suggests that the collocation points be chosen as the roots of the next higher order polynomials used in the PCE model. The collocation points (realizations) in terms of representative SVs for the first and second order PCE model are obtained through the roots of the second and third order polynomials, respectively (see Table 3.1). A combination of higher order polynomial roots (eg. roots of  $H_3$  in  $\xi_1, \xi_2, \xi_3$ ) are used as the collocation points and the corresponding realization and FoS are chosen as the



(a) Actual average FoS frame over all realizations.



(b) SVD reconstructed average FoS frame over all realizations.

(c) Error frame = Actual - Reconstructed,  $\|E\|_F = 0.52$ .FIGURE 3.19: SVD-based FoS field reconstruction with two SVs; visualization at  $t=25$ .

collocating realization to find  $f_i$ .

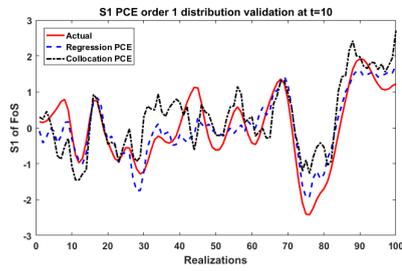
$$S_r(t) = f_1(t)H_1(\xi_1) + f_2(t)H_1(\xi_2) + f_3(t)H_1(\xi_3) + f_4(t)L_1(\mu) \quad (3.11)$$

$$S_r(t) = f_1(t)H_1(\xi_1) + \dots + f_4(t)H_2(\xi_1) + \dots \\ + f_8(t)H_1(\xi_1)H_1(\xi_2) + \dots + f_{11}(t)L_2(\mu) \quad (3.12)$$

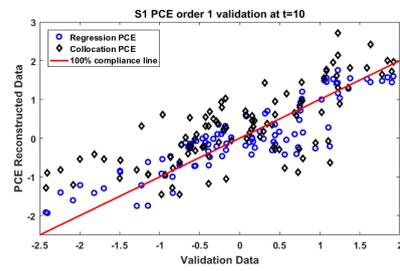
As pointed out earlier,  $S_1$  is the most important SV; hence, we show representative approximations for  $S_1$ . The coefficients  $f_i(t)$  for models of order 1, 2 and 3 computed using regression and collocation at times  $t=10$  and  $t=25$  are tabulated in Table 3.2 and Table 3.3, respectively. Figures 3.20, 3.21 and 3.22 show the

validation results of PCE for  $S_1$  for order 1, 2 and 3, respectively, for time  $t = 10$ . The validation data distribution and the scatter plots are shown side by side for easier visualization. Each point in the scatter plot corresponds to a realization  $\{\Xi_k\}_1^{100}$ . Similarly, Figures 3.23, 3.24 and 3.25 show the validation results of PCE for  $S_1$  for models of order 1, 2 and 3, respectively, for time  $t = 25$ . The  $R^2$ ,  $R_{adj}^2$  and the sum of square error  $SSE$  for each case is shown in Table 3.2 and Table 3.3 along with the coefficients. It is worth mentioning that each of the first, second and third order models have only one term (polynomial basis term in  $\mu$ ) of their respective order capturing the effect of the input. This is because the input being the same for all realizations restricts the model to have only one  $\mu$  basis term in order to consistently estimate its corresponding coefficient in the PCE model. Lower SSE and the higher  $R^2$  values are the indicators of the effectiveness of a model and it is evident that they increase with an increase in the model order. It is interesting to note that collocation-based models perform comparably in many cases to their regression counterparts, given the fact that the collocation coefficients are estimated only from a few chosen points. However, the regression-based PCE models provide superior performance overall. The adjusted coefficient of determination ( $R_{adj}^2$ ), which accounts for model accuracy but also penalizes extra degrees of freedom, indicates that the second order PCE models provide the best performance.

To completely validate the best performing PCE order 2 model, we will use it to replicate the validation realizations of the representative FoS over all times (Figure 3.12). As explained in Section 3.4.2 we can reconstruct the FoS field with 99.85% accuracy using only the first SV, the second order PCE model is used to reconstruct  $S_1$ , through which we can reconstruct the validation realizations of the representative FoS at all times. Figures 3.26(a), 3.26(b) and 3.26(c) show the minimum FoS surface for the actual validation realizations, reconstructed minimum FoS surface reconstructed using PCE order 2 model and the error between them respectively. The Frobenius norm of the reconstruction error is also reported. The error is highest when the steam chamber is dynamic and not fully established, i.e. around the middle part of the lifetime; the variation may also be due to the heterogeneity in the Y direction (over realizations). Overall, due to the low  $\|E\|_F$ , we can conclude that the PCE order 2 model has good performance.

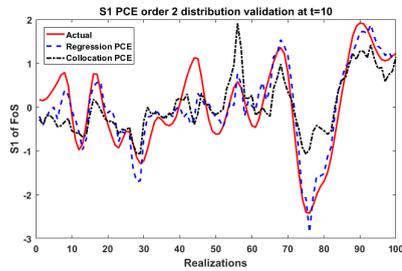


(a) Comparison of actual and PCE reconstructed validation data distribution for order 1 regression and collocation methods at  $t=10$ .

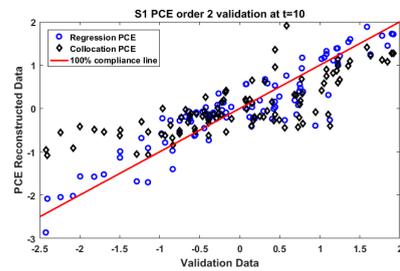


(b) Comparison of actual and PCE reconstructed validation data scatter plot for order 1 regression and collocation methods at  $t=10$ , each dot is a  $\{\epsilon_k\}_1^{100}$ .

FIGURE 3.20: Performance of PCE order 1 regression and collocation models for  $S_1$  at  $t=10$ .

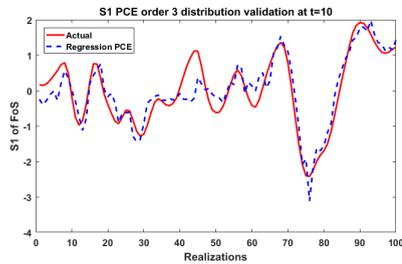


(a) Comparison of actual and PCE reconstructed validation data distribution for order 2 regression and collocation methods at  $t=10$ .

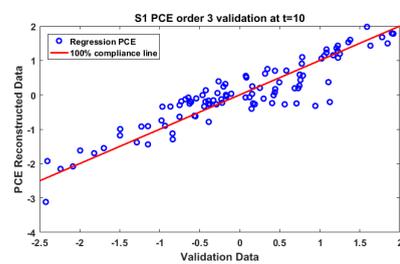


(b) Comparison of actual and PCE reconstructed validation data scatter plot for order 2 regression and collocation methods at  $t=10$ .

FIGURE 3.21: Performance of PCE order 2 regression and collocation models for  $S_1$  at  $t=10$ .

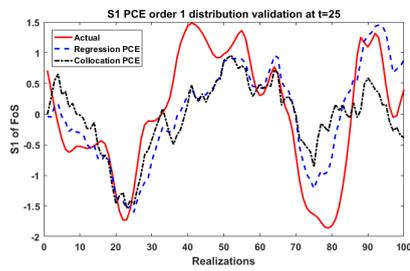


(a) Comparison of actual and PCE reconstructed validation data distribution for order 3 regression method at  $t=10$ .

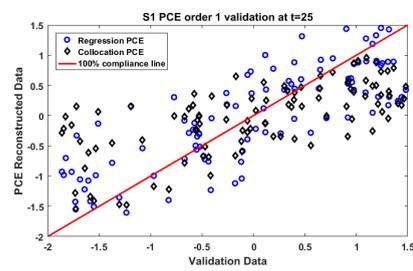


(b) Comparison of actual and PCE reconstructed validation data scatter plot for order 3 regression method at  $t=10$ .

FIGURE 3.22: Performance of PCE order 3 regression model for  $S_1$  at  $t=10$ .

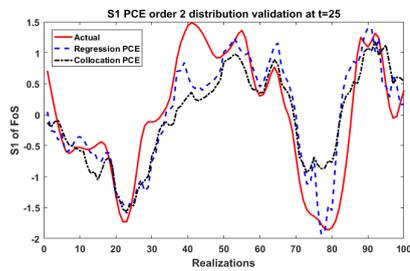


(a) Comparison of actual and PCE reconstructed validation data distribution for order 1 regression and collocation methods at  $t=25$ .

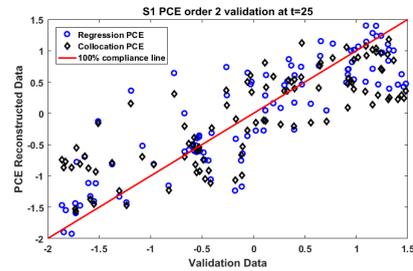


(b) Comparison of actual and PCE reconstructed validation data scatter plot for order 1 regression and collocation methods at  $t=25$ , each dot is a  $\{\Xi_k\}_1^{100}$ .

FIGURE 3.23: Performance of PCE order 1 regression and collocation models for  $S_1$  at  $t=25$ .

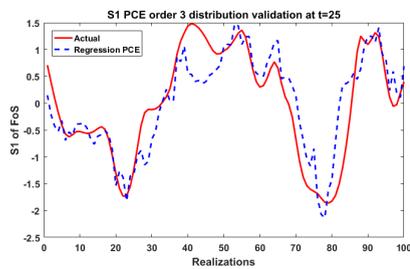


(a) Comparison of actual and PCE reconstructed validation data distribution for order 2 regression and collocation methods at  $t=25$ .

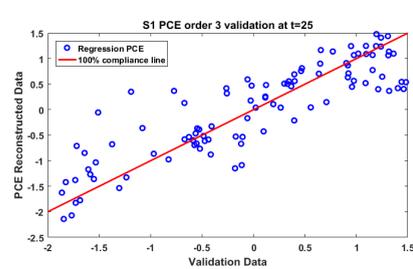


(b) Comparison of actual and PCE reconstructed validation data scatter plot for order 2 regression and collocation methods at  $t=25$ .

FIGURE 3.24: Performance of PCE order 2 regression and collocation models for  $S_1$  at  $t=25$ .



(a) Comparison of actual and PCE reconstructed validation data distribution for order 3 regression method at  $t=25$ .



(b) Comparison of actual and PCE reconstructed validation data scatter plot for order 3 regression method at  $t=25$ .

FIGURE 3.25: Performance of PCE order 3 regression model for  $S_1$  at  $t=25$ .

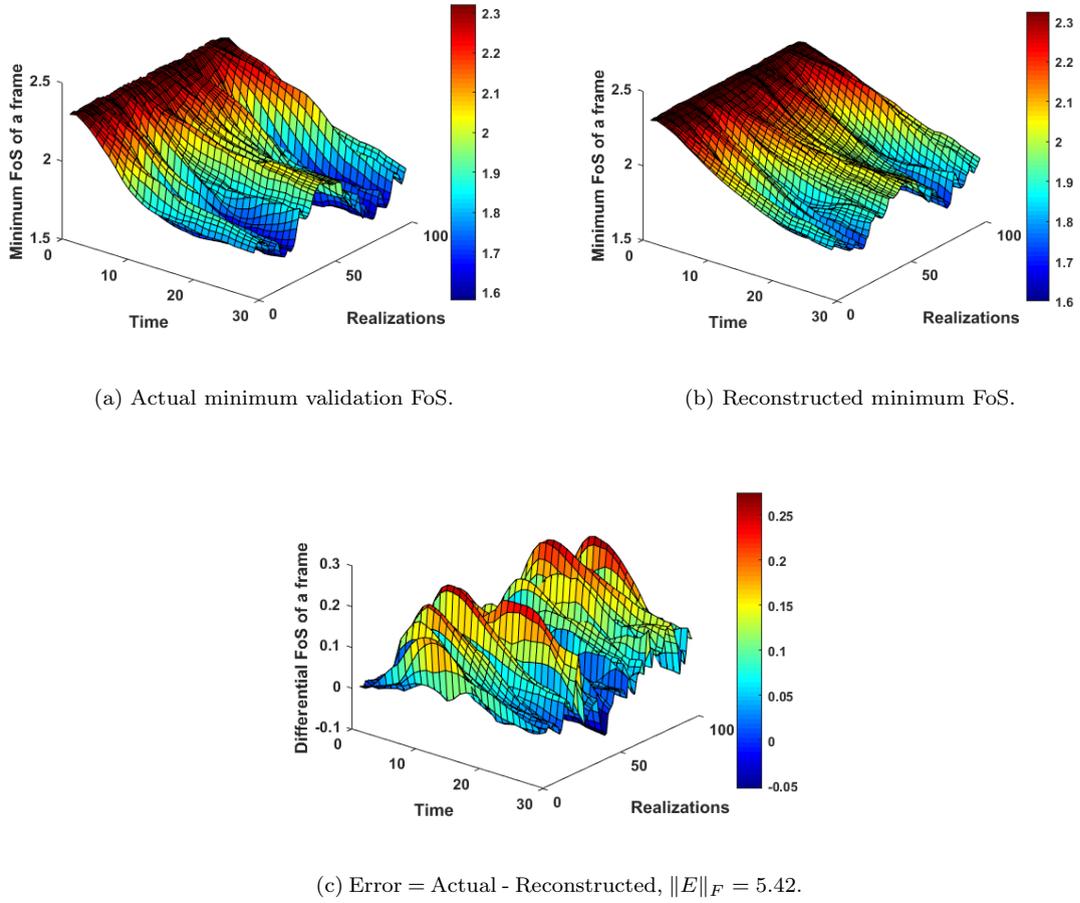


FIGURE 3.26: Performance of the PCE order 2 model in reconstructing the validation realizations of the representative FoS (minimum FoS) over all times. Only one SV is used in the reconstruction.

### 3.4.3.1 UQ through moments of the FoS

Another important feature of the PCE model is the ability to obtain explicit representation for the moments such as mean and variance of the underlying random process (FoS represented by  $S_r$ ). Consider the PCE model of equation 3.12, and take moments on both sides of the equation:

$$\begin{aligned}
 M[S_r(t)] &= M[f_1(t)H_1(\xi_1)] + \dots + M[f_4(t)H_2(\xi_1)] + \dots \\
 &\quad + M[f_8(t)H_1(\xi_1)H_1(\xi_2)] + \dots + M[f_{11}(t)L_2(\mu)] \quad (3.13)
 \end{aligned}$$

Here,  $M$  stands for any of the moments; the left hand side of the equation denotes the moment on the representative FoS ( $S_r$ ) expressed in terms of the moments of the individual terms in the right hand side. But we know that the individual terms (basis functions) in the PCE are polynomials of the standard random variables,

TABLE 3.2: Regression and collocation PCE coefficients for order 1, 2 and 3 models with the corresponding basis at time t=10.

$i$	Basis $\psi_i$	1 <sup>st</sup> order		2 <sup>nd</sup> order		3 <sup>rd</sup> order
		Reg.	Col.	Reg.	Col.	Reg.
1	$\xi_1$	-0.74	-0.54	-0.76	-0.47	-0.72
2	$\xi_2$	-0.35	-0.30	-0.36	0	-0.39
3	$\xi_3$	-0.04	-0.56	-0.11	-0.26	-0.14
4	$\mu - 0.5$	-0.01	-0.53	-	-	-
5	$\xi_1^2 - 1$	-	-	0.06	0	0.07
6	$\xi_2^2 - 1$	-	-	-0.12	0	-0.1
7	$\xi_3^2 - 1$	-	-	0.07	0	0.06
8	$\xi_1\xi_2$	-	-	-0.09	-0.08	-0.14
9	$\xi_1\xi_3$	-	-	-0.01	0.17	-0.01
10	$\xi_2\xi_3$	-	-	-0.18	-0.09	-0.19
11	$\mu^2 - \mu + \frac{1}{6}$	-	-	0.06	0	-
12	$\xi_1(\xi_2^2 - 1)$	-	-	-	-	-0.02
13	$\xi_2(\xi_3^2 - 1)$	-	-	-	-	-0.06
14	$\xi_3(\xi_1^2 - 1)$	-	-	-	-	-0.1
15	$\xi_1^3 - 3\xi_1$	-	-	-	-	0.04
16	$\xi_2^3 - 3\xi_2$	-	-	-	-	-0.02
17	$\xi_3^3 - 3\xi_3$	-	-	-	-	0.01
18	$\xi_1\xi_2\xi_3$	-	-	-	-	0.02
19	$\mu^3 - \frac{3}{2}\mu^2 + \frac{3}{5}\mu - \frac{1}{20}$	-	-	-	-	-0.68
SSE		4.63	7.54	4.12	6.54	3.99
$R^2$		0.53	0.25	0.59	0.35	0.6
$R_{adj}^2$		0.52	0.21	<b>0.54</b>	0.27	0.52

which can be easily computed using their moment generating functions. Also, the orthogonality of the basis functions forces the moments of the cross terms to zero. As an example, consider the usual operation of the reservoir, where the value of the well BHP input  $\mu$  is specified, but the randomness in the permeability is unknown and represented by standard zero-mean normal variables  $\xi_1$ ,  $\xi_2$  and  $\xi_3$ . Now the problem of quantifying the uncertainty in FoS ( $S_1$ ) is carried out by computing the moments of the PCE model. Below are the mean and variance computed for a first order PCE model at t=10 using the discussed procedure.

$$M[S_1] = M[-0.74\xi_1] + M[-0.35\xi_2] + M[-0.04\xi_3] + M[-0.01(\mu - 0.5)] \quad (3.14)$$

$$\begin{aligned} Mean[S_1] &= M[-0.74\xi_1] + M[-0.35\xi_2] + M[-0.04\xi_3] + M[-0.01(\mu - 0.5)] \\ &= -0.01(\mu - 0.5) \end{aligned} \quad (3.15)$$

TABLE 3.3: Regression and collocation PCE coefficients for order 1, 2 and 3 models with the corresponding basis at time  $t=25$ .

$i$	Basis $\psi_i$	1 <sup>st</sup> order		2 <sup>nd</sup> order		3 <sup>rd</sup> order
		Reg.	Col.	Reg.	Col.	Reg.
1	$\xi_1$	-0.84	-0.65	-0.85	-0.79	-1.09
2	$\xi_2$	0.05	0.32	-0.04	0	0.06
3	$\xi_3$	0.09	0.23	-0.13	-0.01	-0.23
4	$\mu - 0.5$	-0.01	0.08	-	-	-
5	$\xi_1^2 - 1$	-	-	-0.01	0	0.11
6	$\xi_2^2 - 1$	-	-	-0.18	0	-0.09
7	$\xi_3^2 - 1$	-	-	0.09	0	0.14
8	$\xi_1\xi_2$	-	-	-0.04	0.01	-0.12
9	$\xi_1\xi_3$	-	-	-0.11	-0.16	-0.19
10	$\xi_2\xi_3$	-	-	-0.1	-0.03	-0.29
11	$\mu^2 - \mu + \frac{1}{6}$	-	-	0.46	0	-
12	$\xi_1(\xi_2^2 - 1)$	-	-	-	-	-0.18
13	$\xi_2(\xi_3^2 - 1)$	-	-	-	-	-0.11
14	$\xi_3(\xi_1^2 - 1)$	-	-	-	-	0.02
15	$\xi_1^3 - 3\xi_1$	-	-	-	-	-0.07
16	$\xi_2^3 - 3\xi_2$	-	-	-	-	0.03
17	$\xi_3^3 - 3\xi_3$	-	-	-	-	-0.09
18	$\xi_1\xi_2\xi_3$	-	-	-	-	0.19
19	$\mu^3 - \frac{3}{2}\mu^2 + \frac{3}{5}\mu - \frac{1}{20}$	-	-	-	-	-5.18
SSE		6.05	7.51	5.10	5.95	4.93
$R^2$		0.39	0.25	0.49	0.40	0.51
$R_{adj}^2$		0.37	0.22	<b>0.43</b>	0.34	0.40

Since  $\xi_1$ ,  $\xi_2$  and  $\xi_3$  are zero mean, this means that the mean value of  $S_1$  is governed by the deterministic input, in other words, the mean uncertainty in FoS is due to the input. Similarly, we can compute the variance.

$$\begin{aligned}
\text{Variance}[S_1] &= V[-0.74\xi_1] + V[-0.35\xi_2] + V[-0.04\xi_3] + V[-0.01(\mu - 0.5)] \\
\text{Variance}[S_1] &= (-0.74)^2 + (-0.35)^2 + (-0.04)^2 + (-0.01(\mu - 0.5))^2
\end{aligned} \tag{3.16}$$

Equations 3.15 and 3.16 quantify the uncertainty in  $S_1$  due to the input  $\mu$ . These equations can be valuable from the perspective of designing a controller to control the FoS. As these equations quantify the uncertainty in FoS due to the input, these can be used in formulating a robust optimization problem often used in designing model predictive controllers [35–37].

## 3.5 Conclusions

In this work, we proposed two measures to quantify the caprock factor of safety based on the minimum FoS value and the normalized singular value of the FoS frame. The location of the weakest point (minimum FoS) in the caprock was also tracked and presented. We then demonstrated a framework to propagate the uncertainty in the petrophysical parameters and the well BHP of a SAGD process to the factor of safety of the caprock. We have proposed a computationally efficient technique to quantify uncertainty using singular value decomposition coupled with polynomial chaos expansion requiring a handful of realizations compared to Monte Carlo techniques requiring thousands of realizations. The uncertainty in petrophysical parameters was modelled by three representative Gaussian random variables and the input by one random variable with uniform distribution. PCE models of order 1, 2 and 3 were developed with regression and collocation approaches, the models were compared using various performance measures and a second order PCE model was chosen as the optimum. The model was used to reconstruct the representative FoS and the results were validated. Quantifying uncertainty in the FoS with respect to inputs and the permeability was demonstrated by determining the moments of the representative.

# Bibliography

- [1] H. N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Annual Review of Fluid Mechanics* 41 (2009) 35–52.
- [2] A. Ganesh, T. Jain, R. Chalaturnyk, V. Prasad, Robust optimization of the production profile of the steam assisted gravity drainage reservoir using a polynomial chaos expansion-based proxy model, in: *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, IEEE, 2018, pp. 1254–1259.
- [3] C. Yang, C. Card, L. Nghiem, Economic optimization and uncertainty assessment of commercial SAGD operations, in: *Canadian International Petroleum Conference*, Petroleum Society of Canada, 2007.
- [4] C. Yang, C. Card, L. X. Nghiem, E. Fedutenko, Robust optimization of SAGD operations under geological uncertainties, in: *SPE Reservoir Simulation Symposium*, Society of Petroleum Engineers, 2011.
- [5] F. A. Dilib, M. D. Jackson, Closed-loop feedback control for production optimization of intelligent wells under uncertainty, *SPE Production & Operations* 28 (04) (2013) 345–357.
- [6] G. Kimaev, L. A. Ricardez-Sandoval, Multilevel Monte Carlo applied to chemical engineering systems subject to uncertainty, *AIChE Journal* 64 (5) (2018) 1651–1661.
- [7] F. Xiong, B. Xue, Z. Yan, S. Yang, Polynomial chaos expansion based robust design optimization, in: *2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, IEEE, 2011, pp. 868–873.
- [8] P. A. Slotte, E. Smorgrav, Response surface methodology approach for history matching and uncertainty assessment of reservoir simulation models, in:

- Europec/EAGE Conference and Exhibition, Society of Petroleum Engineers, 2008.
- [9] G. Kimaev, L. A. Ricardez-Sandoval, A comparison of efficient uncertainty quantification techniques for stochastic multiscale systems, *AIChE Journal* 63 (8) (2017) 3361–3373.
- [10] R. G. Patel, T. Jain, J. J. Trivedi, Polynomial-chaos-expansion based integrated dynamic modelling workflow for computationally efficient reservoir characterization: a field case study, in: *SPE Europec featured at 79th EAGE Conference and Exhibition*, Society of Petroleum Engineers, 2017.
- [11] R. M. Butler, Steam-assisted gravity drainage: concept, development, performance and future, *Journal of Canadian Petroleum Technology* 33 (02) (1994) 44–50.
- [12] J. Xiong, R. Chalaturnyk, Caprock safety factor assessment of SAGD projects, in: *49th US Rock Mechanics/Geomechanics Symposium*, American Rock Mechanics Association, 2015.
- [13] J. Wu, G. Zambrano-Narvaez, R. Chalaturnyk, Newly developed centrifuge testing program of SAGD caprock integrity, in: *49th US Rock Mechanics/-Geomechanics Symposium*, American Rock Mechanics Association, 2015.
- [14] M. R. Carlson, An analysis of the caprock failure at Joslyn, in: *SPE Heavy Oil Conference Canada*, Society of Petroleum Engineers, 2012.
- [15] Computer Modeling Group, Calgary, Alberta, T2M 3Y7 Canada (2015 (10.5715.22942)).
- [16] FLAC3D-Fast Lagrangian Analysis of Continua in 3Dimensions, Minneapolis, USA (Version 7.0).
- [17] Mathworks MATLAB, Natick, MA, USA (9.0.0.341360 (R2016a)).
- [18] J. Mandur, H. Budman, A polynomial-chaos based algorithm for robust optimization in the presence of Bayesian uncertainty, *IFAC Proceedings Volumes* 45 (15) (2012) 549–554.
- [19] G. Senthamaraiikkannan, I. Gates, V. Prasad, Multiphase reactive-transport simulations for estimation and robust optimization of the field scale production of microbially enhanced coalbed methane, *Chemical Engineering Science* 149 (2016) 63–77.

- 
- [20] M. D. Webster, M. A. Tatang, G. J. McRae, Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model, MIT Joint program on the science and policy of global change, report Series No. 4.
- [21] K.-K. K. Kim, D. E. Shen, Z. K. Nagy, R. D. Braatz, Wiener's polynomial chaos for the analysis and control of nonlinear dynamical systems with probabilistic uncertainties [historical perspectives], *IEEE Control Systems Magazine* 33 (5) (2013) 58–67.
- [22] A. Chatterjee, An introduction to the proper orthogonal decomposition, *Current Science* (2000) 808–817.
- [23] G. Berkooz, P. Holmes, J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annual Review of Fluid Mechanics* 25 (1) (1993) 539–575.
- [24] A. Narasingam, P. Siddhamshetty, J. S.-I. Kwon, Identification of spatially varying geological properties in a heterogeneous reservoir using EnKF and POD based parameterization, in: 2018 Annual American Control Conference (ACC), IEEE, 2018, pp. 1144–1149.
- [25] V. Klema, A. Laub, The singular value decomposition: Its computation and some applications, *IEEE Transactions on Automatic Control* 25 (2) (1980) 164–176.
- [26] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, *SIAM Journal on Matrix Analysis and Applications* 21 (4) (2000) 1253–1278.
- [27] H. V. Ly, H. T. Tran, Modeling and control of physical processes using proper orthogonal decomposition, *Mathematical and Computer Modelling* 33 (1-3) (2001) 223–236.
- [28] B. Zhang, N. Deisman, M. Khajeh, R. Chalaturnyk, J. Boisvert, Numerical local upscaling of elastic geomechanical properties for heterogeneous continua, *Petroleum Geoscience* (2019) petgeo2018–159.
- [29] J. Kim, H. A. Tchelepi, R. Juanes, Stability, accuracy and efficiency of sequential methods for coupled flow and geomechanics, in: SPE reservoir simulation symposium, Society of Petroleum Engineers, 2009.

- 
- [30] M. E. Wall, A. Rechtsteiner, L. M. Rocha, Singular value decomposition and principal component analysis, in: *A practical approach to microarray data analysis*, Springer, 2003, pp. 91–109.
- [31] M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (3) (1999) 611–622.
- [32] A. Ganesh, J. Xiong, R. J. Chalaturnyk, V. Prasad, Proxy models for caprock pressure and temperature dynamics during steam-assisted gravity drainage process, *Computers & Chemical Engineering* 121 (2019) 594–607.
- [33] W. Gautschi, Algorithm 726: Orthpol—a package of routines for generating orthogonal polynomials and gauss-type quadrature rules, *ACM Transactions on Mathematical Software (TOMS)* 20 (1) (1994) 21–62.
- [34] C. W. Murray, N. C. Handy, G. J. Laming, Quadrature schemes for integrals of density functional theory, *Molecular Physics* 78 (4) (1993) 997–1014.
- [35] A. Mesbah, Stochastic model predictive control: An overview and perspectives for future research, *IEEE Control Systems Magazine* 36 (6) (2016) 30–44.
- [36] A. Mesbah, S. Streif, R. Findeisen, R. D. Braatz, Stochastic nonlinear model predictive control with probabilistic constraints, in: *2014 American Control Conference*, IEEE, 2014, pp. 2413–2419.
- [37] P. Siddhamshetty, J. S.-I. Kwon, Model-based feedback control of oil production in oil-rim reservoirs under gas coning conditions, *Computers & Chemical Engineering* 112 (2018) 112–120.

## Chapter 4

Factor of safety constrained  
model predictive controller design  
for closed loop reservoir  
management

## 4.1 Introduction

With the depletion of the oilsands available through surface mining, *in-situ* methods are becoming popular in Alberta where 80% of the oil sands have to be recovered using such enhanced oil recovery techniques. Steam Assisted Gravity Drainage (SAGD) [1] is one such thermal recovery technique used in the recovery process for heavy oil sands. A typical reservoir is of 500-1000 meters in length with three important layers viz. the *caprock layer*, which is the top most layer (300m from ground); the *impermeable* layer, which is the bottom most, and the middle layer, called the oil sand layer, which holds the bitumen. The SAGD process uses two parallel running wells; the upper well, called the *injector*, injects steam into the reservoir. As a result, the bitumen captured in the oil sands becomes less viscous, flows down due to gravity and gets collected by the lower well (called the producer). The reservoir can have multiple injector and producer wells. The more the injected steam, the greater is the oil recovery, but the injected steam creates a high pressure and temperature chamber which has to be monitored and kept well within the factor of safety (FoS) limits. FoS is the index of the reservoir operation stability defined as the ratio of allowed stress to the actual stress [2]; reservoir pressure, temperature and fluid flow profiles affect the caprock stress and strain profiles, which are the prime factors determining the FoS. If the injected steam exceeds the FoS limits, a catastrophic caprock damage may release harmful gases to the surroundings resulting in great environmental damage and financial losses [3]. A FoS of greater than unity is always desirable, but too high a FoS is also not economical from the perspective of oil production rate. Hence, FoS-constrained [4, 5] oil production rate (OP) maximized closed loop reservoir management [6–8] is valuable in the oil and gas industry. We will use the data obtained from commercial simulator CMG-STARS [9] and FLAC3D [10] to develop models for use in the closed loop reservoir management.

Model Predictive Controller (MPC) has been used widely in industrial applications to control large scale dynamical systems like reservoirs efficiently [11–13]. MPC uses a process model to forecast the output over the prediction horizon. An on-line optimizer computes the optimal control inputs over the control horizon by considering the given set points. An important feature of MPC is that it can handle input and constraints as a part of the control input design to yield a practically implementable control action. In our case, we use a deterministic model for computing the OP and a stochastic modelling

technique, polynomial chaos expansion, PCE, [14] for estimating the FoS. This approach balances the best OP with an acceptable FoS, providing an efficient solution to the SAGD management problem. We propose a framework and a solution technique to maximize the OP based on the injector steam flow as the decision variable taking into account the of the bounds on the FoS over uncertain petro-physical parameters [15]. The PCE model used for the FoS has inherently captured the information about the uncertainty in the petrophysical parameters and the input, hence there is a scope to obtain a static measure for the MOP based on the variance of the FoS over time. We present one such static measure for the maximum operating pressure (MOP) and compare it with the dynamic optimal MOP given by the MPC.

## 4.2 Model generation

Readers are referred to chapter 3 for all the details about the data, visualizations and procedure to obtain the PCE model of the FoS. The PCE model developed in that work will be used to design the MPC.

## 4.3 Proposed methodology

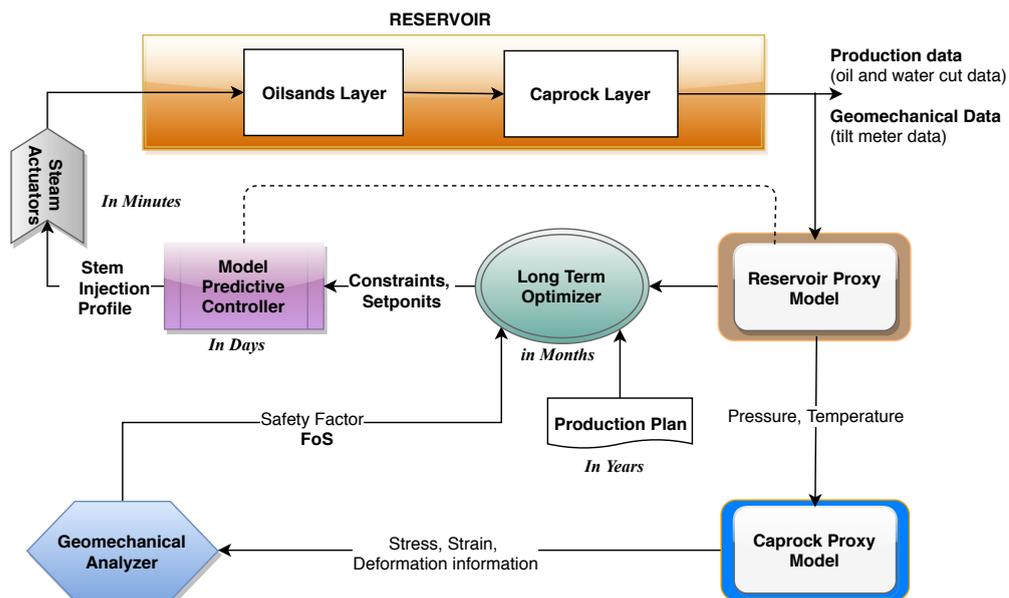


FIGURE 4.1: Process schematic for geomechanics based closed-loop reservoir control

Figure 4.1 explains the schematic of the geomechanics based closed loop reservoir management strategy, highlighting the role of every block. The proxy model of the reservoir and the caprock are updated dynamically from the field, production and geomechanical data to reflect the reservoir's current states. The geomechanical analyzer processes the stress, strain and other geomechanical responses to generate the required FoS-related information to update the optimizer (the update frequency is in months). The optimizer along with the production plan (planned in years) generates a new set of constraints and set points to the model predictive controller (that has a horizon of days). The controller then computes the optimal control inputs to operate the reservoir efficiently that are actuated through the steam actuators (with a time scale in hours). The main steps in the control loop are:

- Reservoir and caprock models are calibrated with production, geomechanical and other data whenever they become available to keep the models up to date
- A set of optimal well controls are determined to optimize the production process based on the updated models
- Model calibration and production optimization are carried out one after the other in the closed loop management process

Figure 4.2 shows the basic three components of the model predictive controller viz., the process model, control scheme and the optimization scheme. The deterministic process model relates the OP to the well bottom hole pressure (well BHP) and the PCE-based model describes the FoS. Figure 4.3 shows the schematic of MPC. The

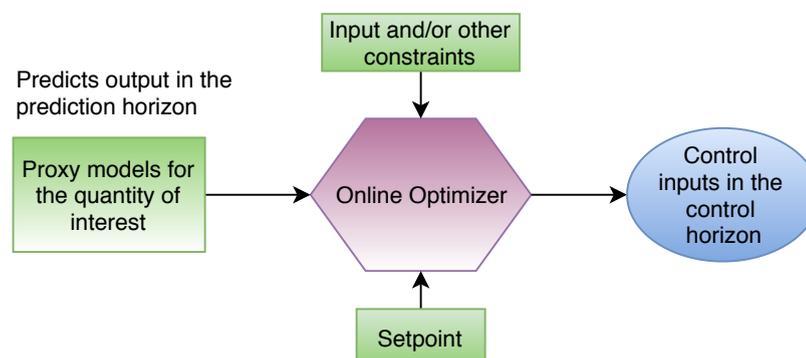


FIGURE 4.2: Schematic of a MPC

FoS and the OP are computed in the prediction horizon using the proxy models with the control inputs spanning the control horizon. Only the first input from

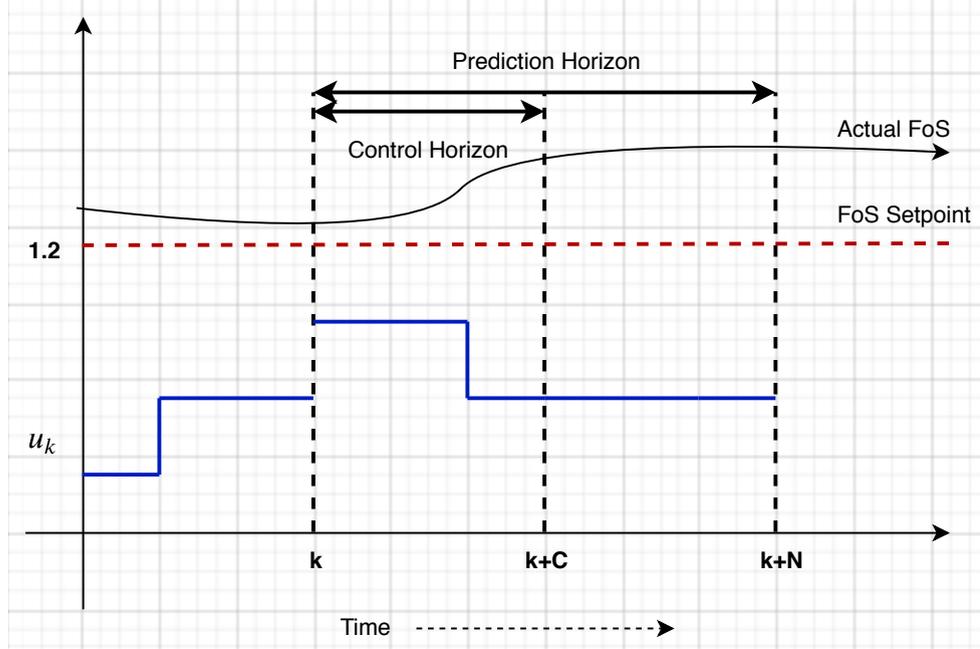


FIGURE 4.3: MPC control schematic

the optimal sequence is applied and the process is repeated at the next time step. In this work, the FoS set point is taken as  $1.2$ , which is just above the failure case ( $\text{FoS} = 1$ ).

### 4.3.1 Robust optimization based MPC formulation

A robust optimization (RO) framework incorporates the uncertainty  $\xi$  in constraints  $g(\xi, U) \leq b$  and the objective function  $f(U)$ . [15, 16].

$$\max_U \{ f(U) : g(\xi, U) \leq b \quad \forall \xi \in Z \} \quad (4.1)$$

Equation 4.1 formalizes the general framework of RO;  $Z$  is the primitive uncertainty set (ensemble) containing the different realizations of the uncertain parameter  $\xi$  over which the robustness is desired.  $f$  is the objective function over the optimization variable  $U$  with  $g(\xi, U) \leq b$  as the constraint having a deterministic upper bound  $b$ . The optimization of the moments of the  $f$  is carried out over the set of realizations ( $Z$ ) of  $\xi$ . A straightforward technique to solve this problem would be to use Monte Carlo simulations [17], but this requires a large number of computations, making it infeasible for our purpose. Hence, we propose a polynomial chaos expansion (PCE)-based technique to address this issue [18]. In our case the objective function  $f$  is a deterministic function, obtained by fitting an ARX model  $P(U, \xi)$  between the OP and the well bottom hole pressures input ( $U$ )

to the data obtained from a CMG-STARS FLAC3D simulator. A first order PCE proxy model for the FoS ( $F_{FoS}(U, \xi)$ ) over many permeability realizations ( $\xi$ ) is used for formulating the constraint  $g$  with a given bound  $b$  for the FoS for the robust optimization problem to maximize the OP  $f$  as described by equation 4.1. This formulation maximizes the OP while robustly maintaining the FoS within the prescribed limits.

### 4.3.2 Stochastic optimization based MPC formulation

The stochastic formulation [19, 20] proposes moments based formulation of the objective function instead of considering an ensemble.

$$MOP = \arg \max_U OP = \arg \max_U P(U, \xi) \quad (4.2)$$

$$a \leq \alpha E(F_{FoS}(U, \xi)) + \beta \sqrt{V(F_{FoS}(U, \xi))} \leq b \quad (4.3)$$

Here,  $P(U, \xi)$  is the OP function,  $F_{FoS}(U, \xi)$  is the PCE-based proxy model for the FoS.  $E$  and  $V$  are the expectation and variance of FoS as predicted by the proxy model,  $a$  and  $b$  are the prescribed FoS limits and  $\alpha$  and  $\beta$  are their weights in the constraint formulation. When  $\alpha = 1$  and  $\beta = 0$ , the constraint focuses on the nominal value of FoS. When  $\alpha = 0$  and  $\beta = 1$ , the constraint targets the standard deviation, i.e. the robustness. We may choose values of  $\alpha$  and  $\beta$  between 0 and 1 to achieve a practical solution for the MOP. The first two moments of  $F_{FoS}$  can be computed from its proxy model 4.4 as described by equations 4.5 and 4.6, and these expressions can be obtained explicitly based on the properties of the chosen polynomials  $H_j$  and  $L_k$ .

$$F = f_1 + f_2 H_1(\xi) + f_3 L_1(\mu) + \dots + f_i H_j(\xi) L_k(\mu) + \dots \quad (4.4)$$

$$E[F] = f_1 + f_2 E[H_1(\xi)] + f_3 E[L_1(\mu)] + \dots + f_i E[H_j(\xi) L_k(\mu)] + \dots \quad (4.5)$$

$$V[F] = f_1 + f_2 V[H_1(\xi)] + f_3 V[L_1(\mu)] + \dots + f_i V[H_j(\xi) L_k(\mu)] + \dots \quad (4.6)$$

Another stochastic formulation proposes an objective function incorporating the OP function and the FoS function together with the objective of maximizing the OP and keeping the FoS within the limit.

$$J(U) = Q(E(F_{FoS}(U, \xi)) - F_{FoS}^-) - R(OP(U)) \quad (4.7)$$

Equation 4.7 shows the objective function that needs to be minimized to find the MOP (U); here,  $F_{FoS}^-$  is the FoS set point and  $Q$  is penalty to the deviation in FoS and  $R$  is a scaling parameter for OP.

## 4.4 Results and discussion

This section presents static and dynamic measures for the maximum operating pressure (MOP) over time. The static measure is based on the FoS ensemble obtained from the PCE and the dynamic measure is based on stochastic optimization MPC (the second formulation presented in equation 4.7). A first order PCE proxy model (equation 4.8) for the FoS is taken from chapter 3 for the design purpose. Figure 3.20 and 3.23 shows the variation of the highest singular value of the FoS frame (representative FoS)  $S_1$  over all realizations at  $t=10$  and  $25$ .

$$S_1(t) = a_1(t)\xi_1 + a_2(t)\xi_2 + a_3(t)\xi_3 + a_4(t)(\mu - 0.5) \quad (4.8)$$

Here,  $\xi_1$ ,  $\xi_2$  and  $\xi_3$  are standard normal random variables and  $\mu$  is the standard uniform random variable over  $[0 \ 1]$ .

### 4.4.1 Static measure based on FoS ensemble

A static measure is obtained by projecting the MOP as inversely proportional to the variance or range of the current FoS ensemble as described by equation 4.9. Here,  $g$  is the FoS ensemble as in equation 4.1. This measure translates to the idea that the MOP should be decreased because the uncertainty in FoS is increasing over time.

$$\left\{ MOP(i) = \frac{k}{range(g_i)} \right\}_{i=1}^T \quad (4.9)$$

The FoS ensemble of a required size can be generated using the PCE model (equation 4.8) at every instant of time. We generated 100K points and computed the range of ensemble at every instant of time to compute the MOP. In the current data set (refer Figure 3.13), if we look at the trend in the FoS variation over different realizations and time, we see the FoS variance increasing with time and hence the static MOP decreases with time as shown in Figure 4.4.

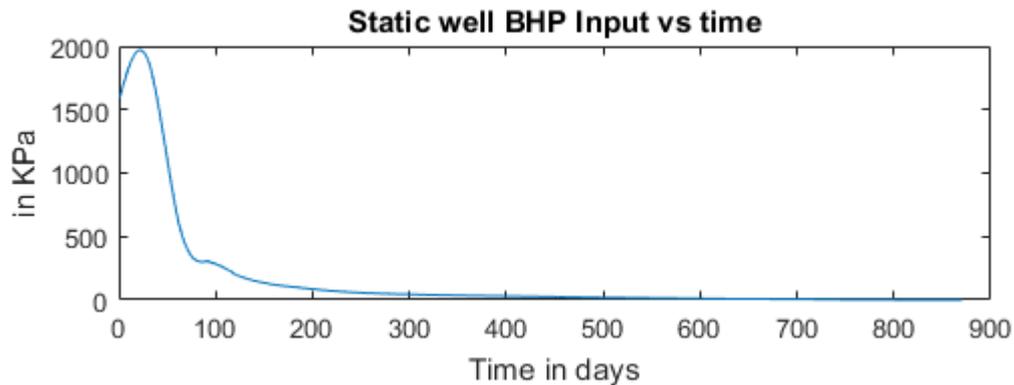


FIGURE 4.4: Static MOP measure based on FoS ensemble generated from PCE

## 4.4.2 Dynamic measure based on MPC design

This section presents the proxy models for the FoS and OP used in the design of the MPC.

### 4.4.2.1 Proxy models for FoS and OP

Figure 4.5 shows the mean of  $S_1$  (obtained from equation 4.8), well BHP and OP variation over time. Equations 4.10 and 4.11 shows two first order ARX styled proxy models for mean  $S_1$  and OP vs. well BHP. These models will be used for the purpose of dynamic prediction of FoS and OP in MPC design.  $S1_{mean}$  can be converted to the actual minimum FoS of the frame using the SVD equations (refer Chapter 3, section 3.4 for details).

$$OP(k+1) = a OP(k) + b MOP(k) \quad (4.10)$$

$$S1_{mean}(k+1) = c S1_{mean}(k) + d (2000 - MOP(k)) + \epsilon(k) \quad (4.11)$$

Here,  $\epsilon(k) = \alpha + \beta x$  is an additive noise term;  $x \in N(0, 1)$  is a standard normal random variable. The model coefficients are chosen as  $a=c=0.7$ ,  $b=0.09$ ,  $d=0.06$ , but can also be computed from data using the prediction error method.  $\alpha$  is offset arising from the plant to model mismatch, which is assumed to be different at each run. We chose a value of 2 for our simulation, but ideally we need to randomly choose a  $S1_{mean}$  realizations from the PCE model.  $\beta$  is the intrinsic noise variance present at every instant of time chosen to be 10 for the simulations. The maximum allowable MOP is chosen to be 2000KPa, hence the constant 2000 in equation 4.11.

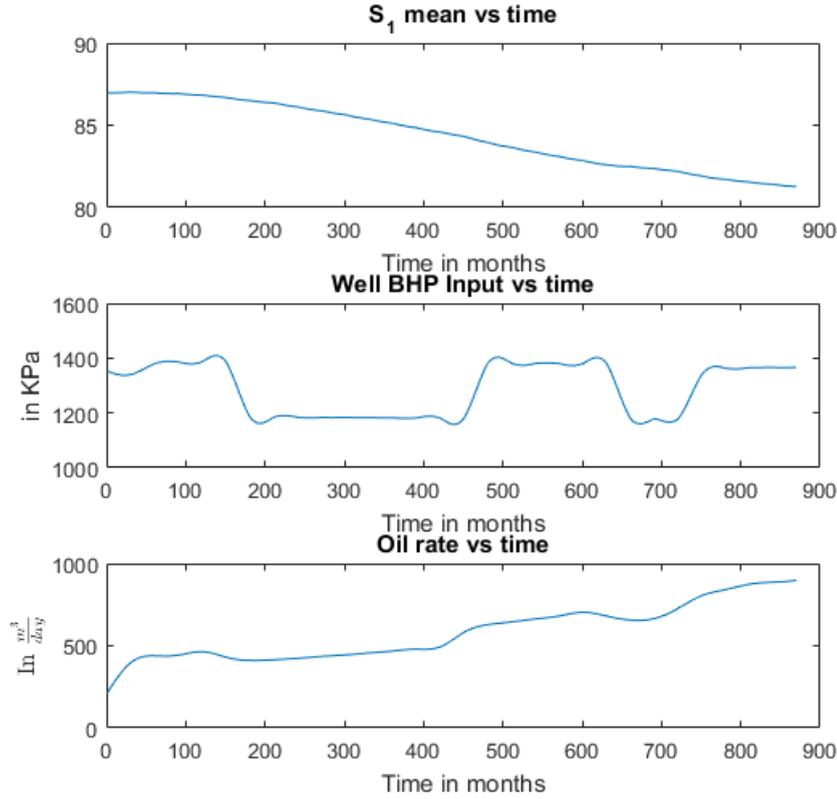


FIGURE 4.5: Input/output data set to build proxy models. Variation of representative FoS  $S_1$  (refer chapter 3), well BHP and OP over time.

#### 4.4.2.2 Stochastic MPC design

The objective function (equation 4.7) along with the prediction and control horizon can be represented as

$$J(\Delta u) = \sum_{j=1}^{N-k} (\tilde{X}_{k+j}^T q_{k+j} \tilde{X}_{k+j} - s_{k+j} OP(u_k)) \quad (4.12)$$

where  $\tilde{X}_k = [(FoS(S1_{mean}(u_k)) - F_{FoS}^-)]$ .  $FoS(S1_{mean}(u_k))$  is the actual minimum FoS of the frame computed from  $S1_{mean}(u_k)$  and  $F_{FoS}^-$  is the FoS set point.  $OP(u_k)$  is the output of the OP proxy model defined over the prediction horizon (N) and  $u_k$  is the MOP at instant  $k$ .  $q$  and  $r$  are the penalty factors for the output and input, respectively.  $s$  is the scaling factor for OP, in the prediction horizon. The objective of the optimizer is to minimize  $J$  and find the corresponding optimal input  $u_k$  over the control horizon ( $C \leq N$ ). We present the results for the following five cases to evaluate the effectiveness of MPC. Following cases are designed to

work in the extremes of the OP and FoS objectives, thereby serving as candidate cases to compare with the MPC strategy.

- Case 1: MPC based MOP; FoS always above the safety limit with optimal OP.
- Case 2: Static MOP (equation 4.9) with the corresponding FoS and OP.
- Case 3: MOP held constant at 1000 KPa ; High FoS but low OP.
- Case 4: MOP held constant at 1900 KPa ; Noisy unsafe FoS but high OP.
- Case 5: MOP held constant at 2000 KPa ; Low and unsafe FoS but highest OP.

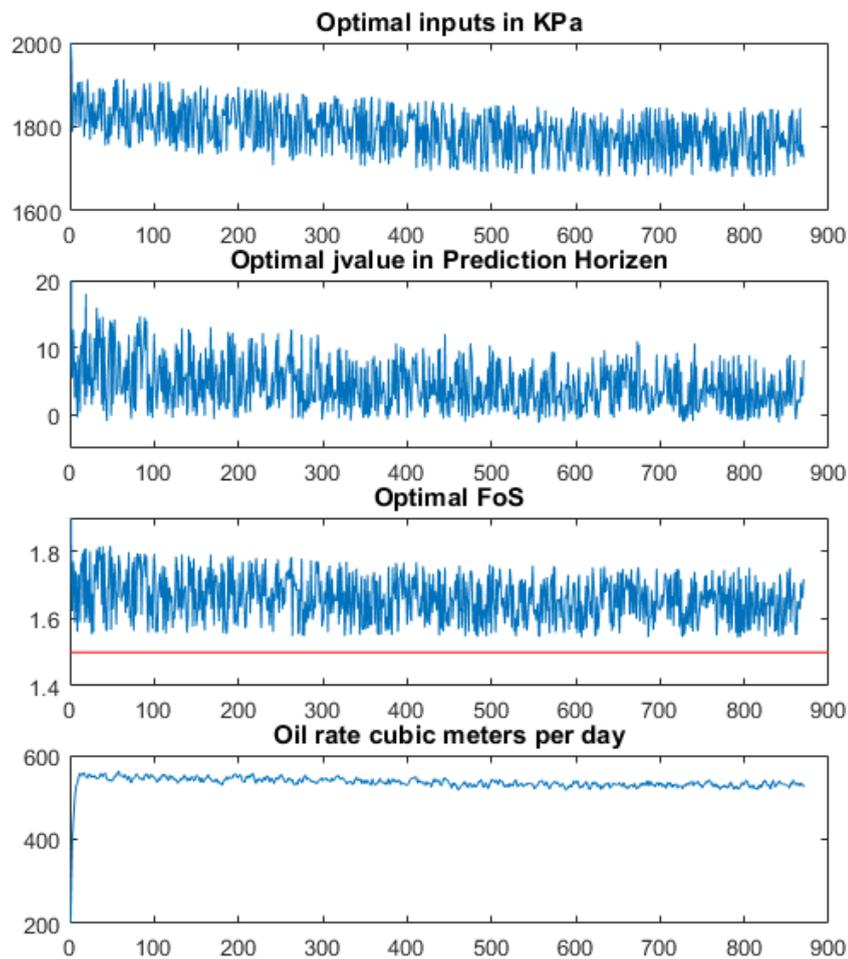


FIGURE 4.6: MPC based MOP with the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5.

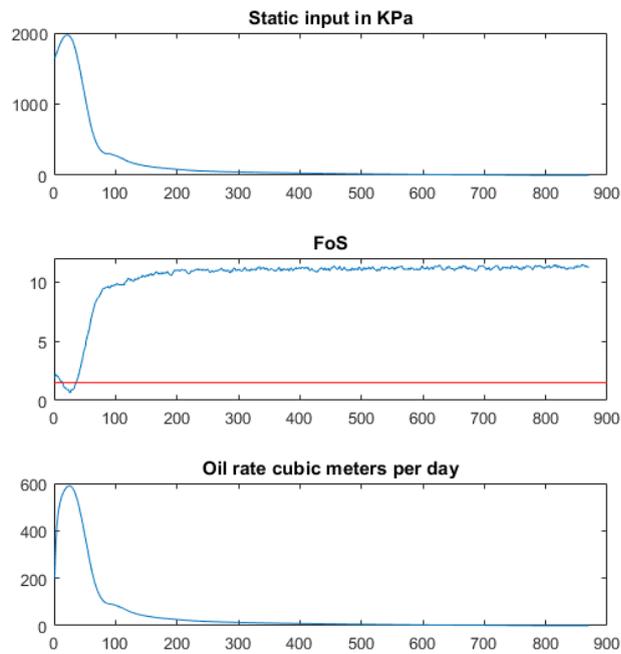


FIGURE 4.7: Static MOP with the corresponding FoS and OP. The red line indicates the FoS setpoint at 1.5.

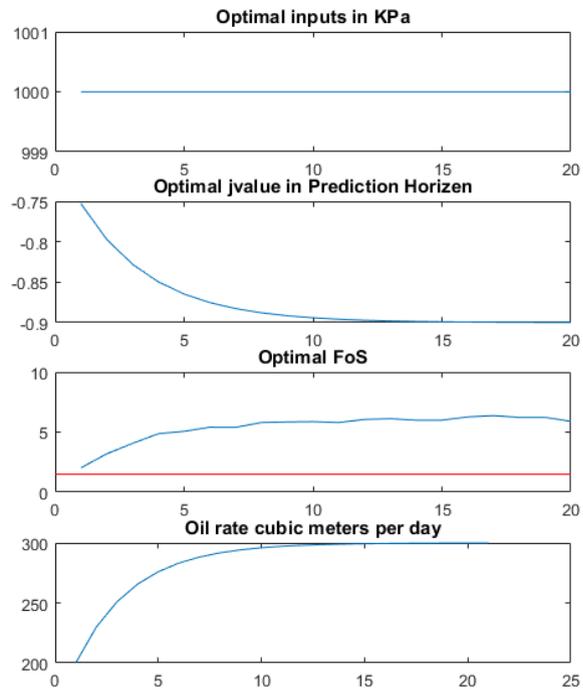


FIGURE 4.8: Fixed MOP at 1000KPa with the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5.

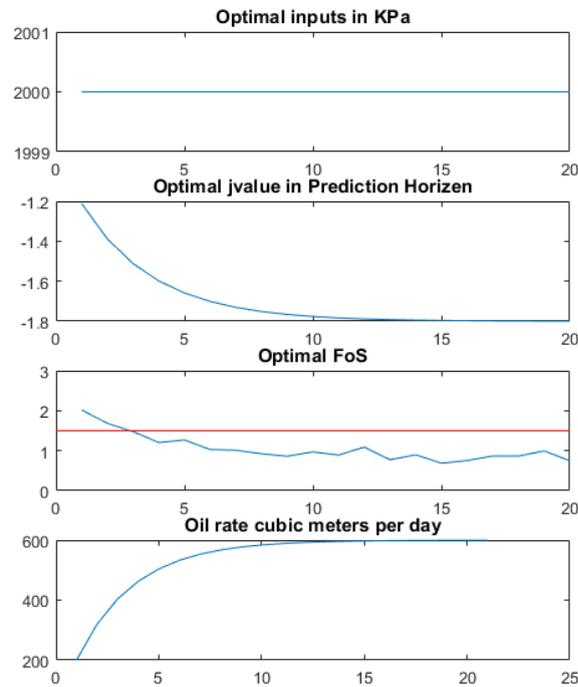


FIGURE 4.9: Fixed MOP at 2000KPa the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5.

TABLE 4.1: Comparison of different cases

Case	COP in $m^3$	Always FoS $\geq 1.5$ ?	Remarks
MPC	468317	Yes	Optimal COP with safe FoS limits
Static	42507	Almost	Poor COP and highly conservative FoS
Fixed at 1000KPa	260380	Yes	Lower COP and very conservative FoS
Fixed at 1900KPa	494114	Yes	Better COP but noisy unsafe FoS
Fixed at 2000KPa	520120	Yes	High COP but unsafe FoS

The Table 4.1 summarizes the performance related to the figures 4.6, 4.7, 4.8, 4.9 and 4.10. The MPC-based MOP formulation proves to be the best strategy for achieving high COP while maintaining the FoS above the safety limit.

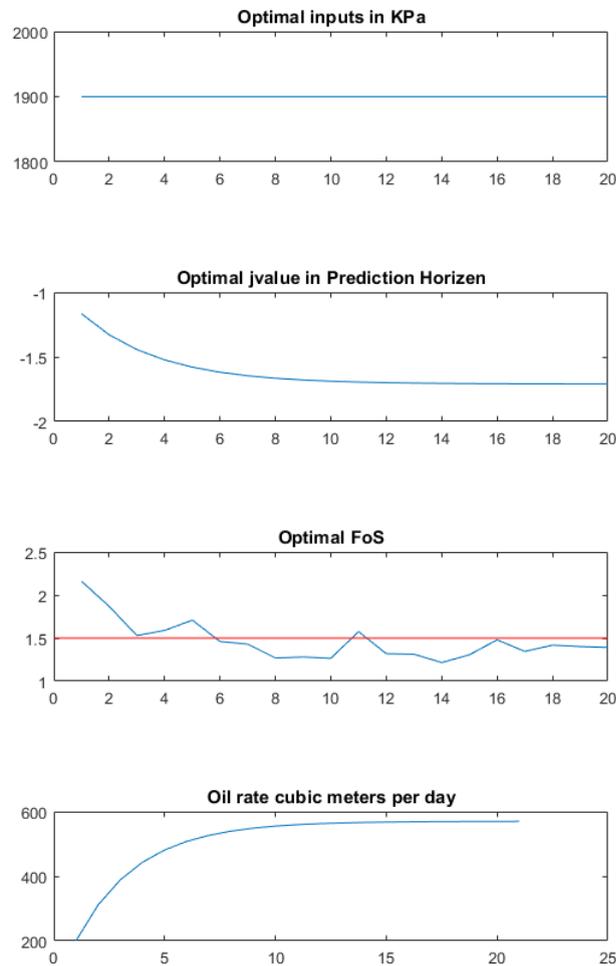


FIGURE 4.10: Fixed MOP at 1900KPa the corresponding FoS, OP and the Jvalue from the optimizer. The red line indicates the FoS setpoint at 1.5.

## 4.5 Conclusions

We presented a technique to develop a FoS-constrained and OP-maximized closed loop SAGD management framework. We presented a stochastic model predictive controller developed from a PCE-based ARX model of the factor of safety and an ARX model for the oil production. Maximum operating pressure measures based on the (static) FoS ensemble, fixed at certain values and based on OP maximized and FoS constrained stochastic MPC were presented and compared. The MPC strategy proves to be the best in balancing high COP while maintaining the FoS above the safety limits.

# Bibliography

- [1] R. M. Butler, Steam-assisted gravity drainage: concept, development, performance and future, *Journal of Canadian Petroleum Technology* 33 (02) (1994) 44–50.
- [2] J. Xiong, R. Chalaturnyk, Caprock safety factor assessment of SAGD projects, in: 49th US Rock Mechanics/Geomechanics Symposium, American Rock Mechanics Association, 2015.
- [3] M. R. Carlson, An analysis of the caprock failure at Joslyn, in: SPE Heavy Oil Conference Canada, Society of Petroleum Engineers, 2012.
- [4] A. Azad, R. Chalaturnyk, A mathematical solution to consider geomechanics in SAGD process, in: Canadian International Petroleum Conference, Petroleum Society of Canada, 2009.
- [5] A. Azad, R. Chalaturnyk, A mathematical improvement to SAGD using geomechanical modelling, *Journal of Canadian Petroleum Technology* 49 (10) (2010) 53–64.
- [6] J.-D. Jansen, D. Brouwer, G. Naevdal, C. Van Kruijsdijk, Closed-loop reservoir management, *First Break* 23 (1) (2005) 43–48.
- [7] F. A. Dilib, M. D. Jackson, Closed-loop feedback control for production optimization of intelligent wells under uncertainty, *SPE Production & Operations* 28 (04) (2013) 345–357.
- [8] E. Gildin, T. Lopez, Closed-loop reservoir management: Do we need complex models?, in: SPE Digital Energy Conference and Exhibition, Society of Petroleum Engineers, 2011.
- [9] Computer Modeling Group, Calgary, Alberta, T2M 3Y7 Canada (2015 (10.5715.22942)).

- 
- [10] FLAC3D-Fast Lagrangian Analysis of Continua in 3Dimensions, Minneapolis, USA (Version 7.0).
- [11] K. Patel, E. M. Aske, M. Fredriksen, et al., Use of model-predictive control for automating sagd well-pair operations: A simulation study, *SPE Production & Operations* 29 (02) (2014) 105–113.
- [12] S. J. Qin, T. A. Badgwell, A survey of industrial model predictive control technology, *Control engineering practice* 11 (7) (2003) 733–764.
- [13] S. S. Vembadi, Real-time feedback control of the sagd process using model predictive control to improve recovery: A simulation study.
- [14] H. N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Annual Review of Fluid Mechanics* 41 (2009) 35–52.
- [15] F. Xiong, B. Xue, Z. Yan, S. Yang, Polynomial chaos expansion based robust design optimization, in: *2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, IEEE, 2011, pp. 868–873.
- [16] B. L. Gorissen, İ. Yanıkoğlu, D. den Hertog, A practical guide to robust optimization, *Omega* 53 (2015) 124–137.
- [17] G. Kimaev, L. A. Ricardez-Sandoval, Multilevel Monte Carlo applied to chemical engineering systems subject to uncertainty, *AIChE Journal* 64 (5) (2018) 1651–1661.
- [18] C. Yang, C. Card, L. X. Nghiem, E. Fedutenko, Robust optimization of SAGD operations under geological uncertainties, in: *SPE Reservoir Simulation Symposium*, Society of Petroleum Engineers, 2011.
- [19] A. Mesbah, Stochastic model predictive control: An overview and perspectives for future research, *IEEE Control Systems Magazine* 36 (6) (2016) 30–44.
- [20] A. Mesbah, S. Streif, R. Findeisen, R. D. Braatz, Stochastic nonlinear model predictive control with probabilistic constraints, in: *2014 American Control Conference*, IEEE, 2014, pp. 2413–2419.
- [21] Mathworks MATLAB, Natick, MA, USA (9.0.0.341360 (R2016a)).

## Chapter 5

Graph-based determination of structural controllability and observability for pressure and temperature dynamics during steam-assisted gravity drainage operation

## 5.1 Introduction

Steam assisted gravity drainage (SAGD) is a thermal recovery technique commonly used for in-situ extraction and recovery of oil sands bitumen [1]. The assessment of observability and controllability, as with other distributed parameter systems, presents significant challenges. Historically, parametric sensitivity (the sensitivity of oil production to the permeability and/or porosity at different spatial locations in the reservoir) [2, 3], which actually quantifies identifiability, has been used as an alternative for observability analysis in these systems. The so-called covariance localization, [4, 5] which is used to determine which states and parameters to update in estimation often referred to as history matching [6, 7], is often accomplished by choosing regions of high sensitivity for updating. Our reading of the relevant literature in reservoir engineering has not unearthed a rigorous approach for controllability analysis in this field.

Determining the observability and controllability for these and other distributed parameter systems (DPS) [8] is not as straightforward as in the case of linear systems [9, 10], since these are infinite dimensional systems [11]. DPS pose significant challenges in modelling, monitoring and control. Prior work in quantifying the observability and controllability of DPS has focused on the analysis of first principles models. Pacharu, et al. [12] considered packed bed reactors and analyzed observability issues for the development of state estimates and reconstruction of spatial and temporal profiles. A number of earlier efforts [13, 14] have considered other quantitative criteria in conjunction with nonlinear first principles models to arrive at optimal sensor locations. Lainiotis and Ray [15] addressed controllability analysis in conjunction with spatiotemporal process models. While the above approaches do consider cause and effect in a rigorous manner and develop quantitative indices for observability and controllability, they are computationally intensive. Some of the approaches also tend to characterize these measures in a locally linearized sense [16].

Often, during the early stage of design for complex systems, it is perhaps adequate to consider qualitative measures of observability and controllability based on structural characteristics [17]. Graph theoretic approaches that represent qualitative relationships can prove to be a computationally cheap alternative to rigorous model-based approaches [18]. Early work in structural controllability approaches was represented in Lin [19] and Diop and Fliess [20]. More recently,

Liu, et al. [21, 22] have considered controllability and observability aspects of complex networks found in nature such as food webs, biochemical reaction networks, etc.

Figure 5.1 shows a system theoretic perspective of a SAGD reservoir [23] with injector and producer inputs, pore pressure/temperature as states and the measurement sensors to measure desired parameters. In this work, we present an approach to assess the structural controllability and observability of a SAGD process by using a graph theoretic approach. The method is data-driven in that it uses dynamic spatiotemporal data from detailed physical simulations of the process to identify the connectivity between different spatial locations/regions in the reservoir, thereby transforming the description of the process to a graph. A regionalization-based dynamically constrained agglomerative clustering algorithm [24] with partitioning is used to group spatially contiguous regions of the reservoir that display similar behaviour, and the connectivity structure is identified using Granger causality analysis. We then adopt some of the measures proposed in Liu, et al. [21, 22] to characterize observability and controllability in a structural sense through graph theoretic techniques [25–29]. Structural controllability and observability are analyzed by identifying the driver nodes and strongly connected components of the graph. The data required is generated in our work using the CMG-STARS simulator [30]. The original contributions in this work relate

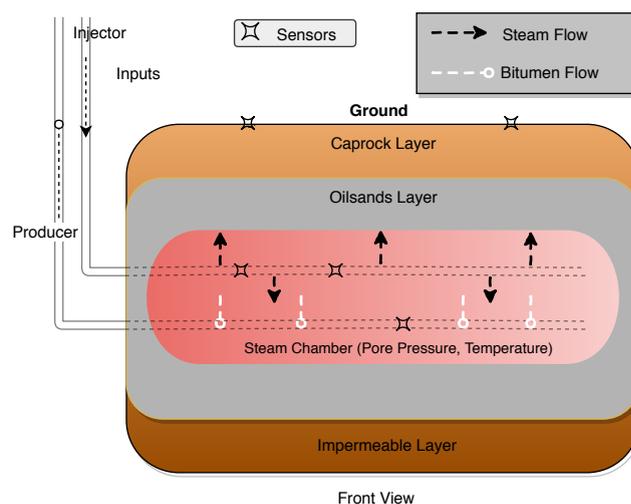


FIGURE 5.1: SAGD reservoir with schematic showing injector producer inputs, pore pressure/temperature and the measurement sensors on wells and at ground level.

to the development of a regionalization-based clustering algorithm to develop a

reduced-dimensional description of the spatiotemporal behaviour of the process based on pressure and temperature data, combining it with causality detection to obtain a graph-based description of the process, and the application of graph theoretic approaches to infer structural observability and controllability for this distributed parameter system. The results obtained for the SAGD process are original, and the data-driven method developed can be applied generally to a wide variety of distributed parameter systems. Note that this contribution extends our preliminary work on the topic [31], which used a standard density-based agglomerative clustering method that did not retain clusters as spatially contiguous regions and only analyzed the pressure field. In addition, the structural observability and controllability analysis is more detailed in this work, especially in terms of combining the pressure and temperature data and deriving an original algorithm for identifying the regions of influence (ranges) of actuators and sensors in a partially observable/controllable distributed parameter system.

The rest of the chapter is organized as follows: Section 5.2 describes the SAGD system, details of the reservoir simulation and details of the spatiotemporal pressure and temperature field data generated for the SAGD process. Section 5.3 explains the methods used for clustering, causality analysis and identification of the graph-based process description, and structural controllability and observability analysis. The results of the analysis are presented in section 5.4 and section 5.5 presents conclusions.

## 5.2 Reservoir modelling and data generation

The data used in the development of proxy models is gathered from reservoir simulations using CMG-STARs. The parameters and operating conditions used for the simulation are substantially similar to those used in our previous work [32], and we refer readers to that work for full details. For the sake of clarity, we present some essential details of the simulation conditions in this section.

A model of a heterogeneous SAGD reservoir having 5 injectors and 5 producers is provided with the designed inputs (described later) and the two dimensional cross section of the reservoir pressure and temperature fields at various time instants is gathered and analysed.

The spatiotemporal data consists of the cross section of the pressure/temperature field at equally spaced time intervals, referred to as frames. Each frame has  $1200 \times 81$  rectangular grid blocks (or points) in the X, Z direction corresponding to the width and the height of the reservoir, and 184 such frames (representing monthly data) are available. The pressure and temperature values are reported for each grid block at each time sample. The inputs variables are the well bottom-hole pressures of the injector and producer wells, so the dimensions of the caprock data (pressure/temperature) are  $1000 \times 25 \times 184$  and the dimensions of the reservoir data are  $1000 \times 45 \times 184$ . Details of the physical dimensions of the reservoir, caprock, underburden, overburden, well-pair spacing, shale distribution and porosity/permeability distributions are available in Ganesh, et al. [32].

The well bottom hole pressures in the injectors and producers form the inputs to the SAGD process. Persistence of excitation [33] is an essential consideration when designing inputs for the purpose of system identification. While our objective in this work is not system identification, the data is being used for dynamic clustering and causality analysis, which is why a pseudo random binary signal (PRBS) was applied at the inputs to the SAGD process.

The reservoir simulations provide dynamic data related to the pore pressure (P) and temperature (T) at all of the regions in the simulation domain. One of the main steps in the analysis is to identify a graph-based description from this data using Granger causality analysis (described in the next section on Methods). While vector autoregressive methods can be used for multivariate causality analysis [34, 35], since we only have two variables to consider, we have chosen to use principal component analysis (PCA) [36, 37] to identify uncorrelated directions in the two dimensional space. Since the first principal component explains 85% of the variance in the data, we only use the first principal component as a scalar input for Granger causality analysis. The individual pressure and temperature data visualizations are available in Ganesh, et al. [32].

### 5.3 Methods

The working data is the full reservoir data  $D$ , represented by a matrix of dimension  $1000 \times 70 \times 184$ . We regard this as 2D data (dimension:  $1000 \times 70$ ) having a time series of length 184 at every grid point in the simulation domain. This section

explains how this process data is modelled as a graph and the subsequent structural controllability and observability analysis. All the algorithms are developed in Python with the *statsmodels* module. [38].

### 5.3.1 Regionalization-based clustering

The simulation data is of high dimension, and it is possible that many grid points display similar behaviour. Clustering of the spatiotemporal data reduces the dimensionality and aids in the subsequent causality and structural controllability and observability analyses. Regionalization can be regarded as a specialized form of clustering for spatially distributed data. Along with placing similar samples into the same cluster while optimizing an objective function, it also makes sure that the samples in the same cluster are all in a spatially contiguous region. This can be achieved by introducing a specific constraint in agglomerative hierarchical clustering. The steps followed in agglomerative hierarchical clustering are as follows [39]:

1. Regard each data point (sample) as a separate cluster.
2. Merge the nearest two clusters into one (based on an appropriate similarity metric).
3. Repeat the merging process (step 2) until all samples are in the same cluster.

From the perspective of graph theory, a data set of  $n$  samples can be modelled as a weighted undirected graph  $G$  containing  $n$  nodes and a certain number of edges. The weight of an edge is the distance between the two nodes it connects. Agglomerative hierarchical clustering of a data set of this type is achieved by building a spanning tree from  $G$ . The spanning tree contains  $n - 1$  edges that connect all the nodes together.

We use the Euclidean distance between the two time series as the distance measure between the two data points. Of the three commonly used linkage methods used in agglomerative hierarchical clustering (single, average and complete linkage) [40], we use single linkage in this work. In single linkage, the distance between two clusters is defined as the distance between two closest pairs of data points from each cluster.

$$d = \min_{u \in L, v \in M} d_{uv} \quad (5.1)$$

where  $L$  and  $M$  are two clusters,  $u$  and  $v$  are data points within the clusters, and  $d$  is the distance between  $u$  and  $v$ . Because of the specific definition of distance between two clusters, we can simplify the process of building the hierarchy into building the minimum spanning tree of the graph  $G$ . The algorithm to build the hierarchy is:

1. Create an empty set  $T = \phi$
2. Sort all edges in graph  $G$  in ascending order.
3. For each edge  $e$  in the sorted list: if  $e$  connects two different clusters, add it to set  $T$  until all nodes in  $G$  are connected by edges in  $T$ .

To make the minimum spanning tree spatially contiguous, we add a constraint to the edges in graph  $G$  to *only use the edges connecting two neighbors*[24]. This is illustrated in Figure 5.2, where sample  $A$  has two neighbours, and the two edges connecting  $A$  and its neighbours are considered. Sample  $B$  has four neighbours, and the four edges connecting  $B$  and its neighbours are considered. By adding this

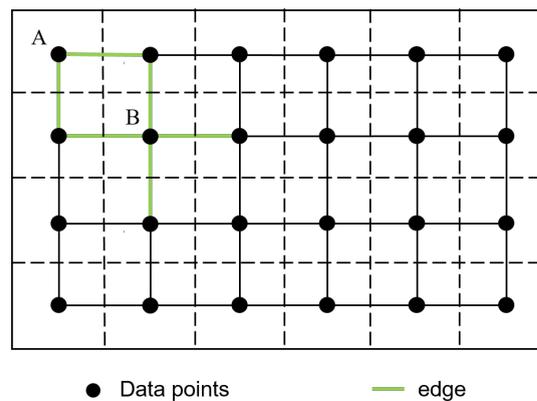


FIGURE 5.2: The edges considered for regionalization

constraint to the edges in graph  $G$ , we can ensure that at each level of the hierarchy, the sub-trees we get from the minimum spanning tree are spatially contiguous [24].

After obtaining the minimum spanning tree, we carry out partitioning to obtain a desired number ( $K$ ) of clusters. At each step, one sub-tree is split while maximizing the homogeneity gain  $h_g$ . All data points in the same sub-tree belong to the same cluster.  $h_g^*$  is the maximum of  $h_g$ ,  $H(\star)$  is the heterogeneity of a

cluster, and  $k$  is the number of data points in one cluster.

$$h_g^* = \max(H(R) - H(R_a) - H(R_b)) \quad (5.2)$$

$$H(*) = \sum_{j=1}^k (x_j - \bar{x})^T (x_j - \bar{x}) \quad (5.3)$$

$$\bar{x} = \sum_{j=1}^k x_j \quad (5.4)$$

The partitioning algorithm for obtaining  $K$  clusters can be described as follows:

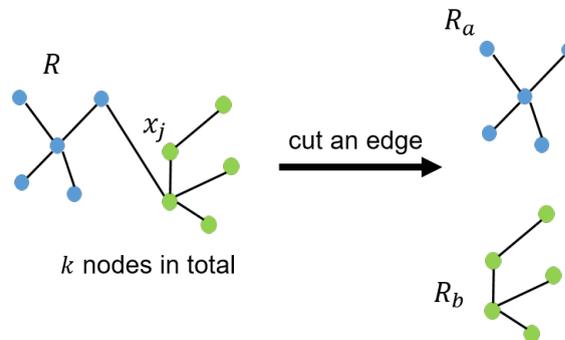
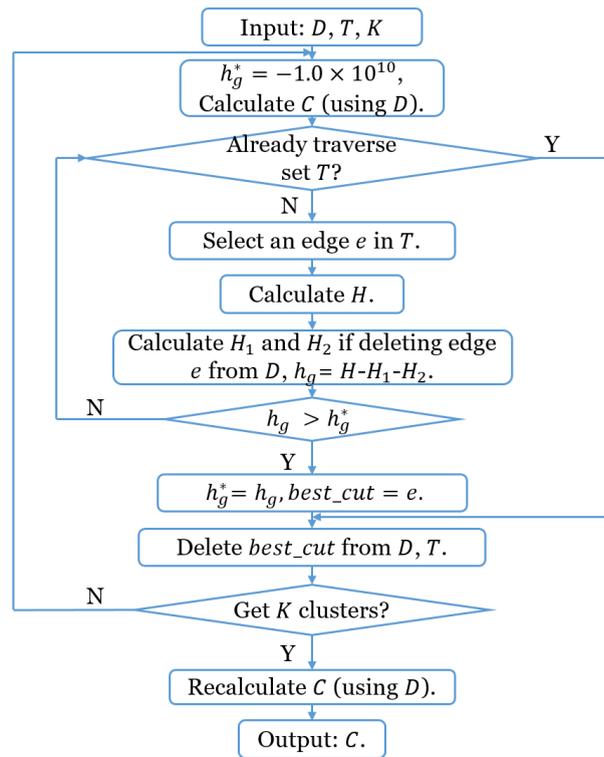


FIGURE 5.3: Illustration of partitioning.

1. Input: The adjacency matrix  $D$ ; initially, it stores the reachability in the minimum spanning tree. The set  $T$  stores all edges in the minimum spanning tree.
2. Set  $h_g^* = -1.0 \times 10^{10}$  (a large value). Compute the undirected strongly connected components (USCCs) (maximal subset of connected nodes in an undirected graph, Breadth First Search/ Depth First Search is used to compute them [28], Python library has inbuilt functions for it. Section 5.3.3 explores more on SCCs for directed graphs) of  $D$ . Store the USCCs to which each node belongs in  $C$ .
3. For each edge  $e$  in set  $T$ :
  - Obtain the USCC to which its start node and end node belong and calculate the heterogeneity  $H$  of this combined USCC.
  - Delete  $e$  from  $D$  temporarily, calculate the USCCs of  $D$ , and get the USCCs to which its start node and end node belong separately. Calculate the heterogeneity  $H_1$  and  $H_2$  of the two USCCs.
  - Calculate the homogeneity gain after cutting this edge:  $h_g = H - H_1 - H_2$

- If  $h_g > h_g^*$ , then  $h_g^* = h_g$ ,  $best\_cut = e$
  - Re-add edge  $e$  to  $D$ .
4. Delete  $best\_cut$  from  $D$  and  $T$ .
  5. If the number of USCCs in  $D$  is  $K$ , recalculate the USCCs of  $D$ , return  $C$ .  
Else, jump to step 2.

Figure 5.3 illustrates the principle of partitioning and Figure 5.4 presents the flowchart of the algorithm.



$D$ : Adjacency matrix of the minimum spanning tree.  
 $T$ : A set storing all edges in the minimum spanning tree.  
 $K$ : The number of clusters.  
 $C$ : A matrix storing which USCC each node belongs to.  
 $H$ : Heterogeneity of a USCC that contains edge  $e$ .  
 $H_1, H_2$ : Heterogeneity of the two USCCs that contain the start node and end node of edge  $e$ .

FIGURE 5.4: Partition algorithm flowchart

### 5.3.2 Granger causality-based graph construction

Granger causality uses a statistical hypothesis test to determine whether one time series is useful in forecasting another [41]. Assume that  $x$  and  $y$  are two stationary time series. If the autoregressive estimate of  $y$  is inferior to the estimate including lagged values of  $x$  along with the autoregressive component, this implies that

$x$  has significant explanatory power with respect to the future behaviour of  $y$ , which can be established by a F-test (null hypothesis: lagged values of  $x$  add no explanatory power). If the test establishes that  $x$  has significant explanatory power for the future behaviour of  $y$ , we say that  $x$  Granger-causes  $y$  (with the additional difficult-to-test condition that this specific explanatory power is unique to  $x$ ). After applying the regionalization-based clustering and partitioning algorithms to the SAGD simulation data, we have a certain number of clusters  $K$ . Each cluster can be regarded as a node in a graph or network. The value of this node is the average of all time series within this cluster and the geographic location is the average of co-ordinates of all data points within this cluster. Granger causality is used to create directed edges between the nodes. Note that the first principal component of the (pressure+temperature) dynamic data for the spatial locations is used as the time series that is averaged within each cluster. If node  $x$  in the network is the Granger-cause of node  $y$ , we draw an edge from  $x$  to  $y$ . In this work, we have set the threshold to infer Granger causality to be  $1.0 \times 10^{-7}$ , and we calculate the F-statistic for up to 15 lags. If the maximum of the 15 values is less than the threshold, the null hypothesis is rejected. To obtain all directed edges, we apply the Granger causality test on all node pairs in the network. We also calculate the in-degrees and out-degrees of each node in the network and identify the top and bottom 20% nodes with highest/lowest in-degree and out-degree. This helps us visualize the highly connected and sparsely connected spatial regions of the reservoir and caprock.

### 5.3.3 Structural controllability and observability

The classical definition of controllability relates to the ability to navigate the system from an arbitrary initial state to any desired final state in a finite amount of time with the given inputs [10], and the Kalman controllability rank criterion [9] can be used to determine complete controllability for linear time-invariant systems. Similarly, observability relates to the ability to reconstruct the initial state from the current output measurements, and a similar rank criterion is used in its assessment for linear time-invariant systems. However, the determination of controllability and observability for nonlinear systems, especially those which are distributed parameter systems, presents significant challenges [42]. To alleviate this difficulty, we propose to employ structural controllability and observability analyses for these systems; these are assessed on the basis of the system's structure alone, and the assessment is satisfied for all the possible

choices of the parameters of the system's dynamic matrices (except for zero and some pathological cases). The graph-based description of the system allows us to use graph theoretic algorithms to assess structural controllability and observability.

Note that there are two aspects of interest related to controllability and observability for us: the actuator/sensor locations for which the system is shown to be fully controllable/observable, and the regions of influence for a specific choice of actuator/sensor locations (referred to as the control range and the observe range in subsequent discussion). The first question relates to optimal design or placement of sensor/actuators for a system, while the second addresses the question of which subset of states/parameters to update in an estimation framework for an already established process.

### 5.3.3.1 Driver nodes

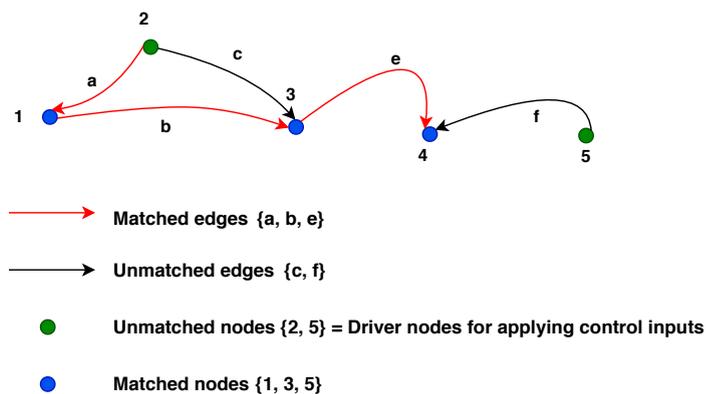


FIGURE 5.5: Illustration of matched edges, matched nodes and driver nodes in an example directed graph. A set of edges (maximum) are said to be matched if they do not have a source (arrow going out) or sink (arrow coming in) node in more than one edge. Matched nodes are the ones pointed by the matched edges, remaining nodes are the driver nodes where the input is applied to achieve full controllability.

Driver nodes are the set of nodes that, if driven by different signals, can offer full control over the network. According to Liu, et al. [21], the algorithm to find driver nodes can be described as finding the ‘maximum matching’ of the network. If all the matched nodes are excluded from the set containing all nodes, the remaining nodes are the driver nodes. The ‘maximum matching’ in the network is the maximum set of edges that do not share start or end nodes. A node is said to be matched if an edge in the maximum matching points at it; otherwise it is unmatched. Figure 5.5 illustrates the notion of a driver node.

### 5.3.3.2 Root SCCs

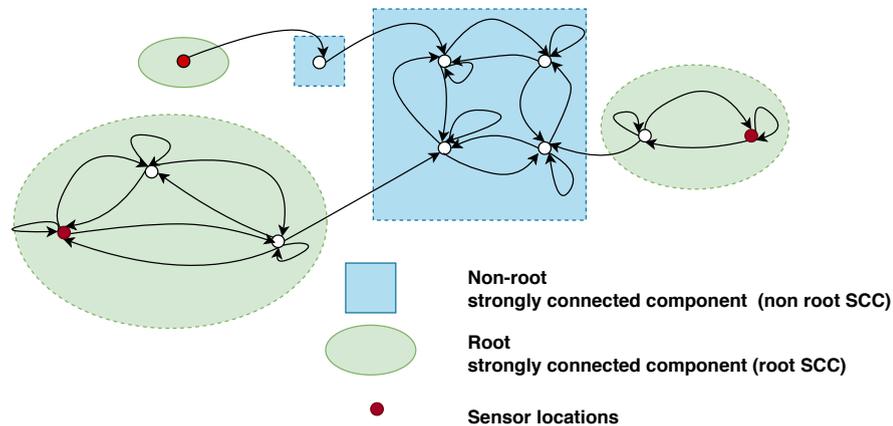


FIGURE 5.6: Illustration of an inference diagram with connected components, strongly connected components (SCCs) and root SCCs in an example directed graph. A graph can be split into largest connected components (sub-graphs) which are individually connected and having no edges amongst the connected components. In SCCs, all nodes are bidirectionally connected to each other including a self loop. SCC having no incoming edge is a root SCC. Sensors are placed in any one node of each of the root SCC to achieve full observability.

As proposed in Liu, et al. [22], the minimum number of sensors required for full observability in a graph-based model is equal to the number of root strongly connected components (SCCs). In a directed graph, SCCs are the largest subgraphs in which each node has a directed path to every other node. This means that if any one node in an SCC is observable, all the nodes in that SCC are observable. Figure 5.6 illustrates the concepts of connected components, SCCs and root SCCs. The procedure for identifying the sensor locations which provide complete observability of the system is [22]:

1. Obtain an inference diagram by flipping the direction of all edges in the constructed graph.
2. Decompose the inference diagram into a unique set of strongly connected components (SCCs).
3. Find the SCCs that have no incoming edges, which are called root SCCs. Place sensors on at least one node from each root SCC to ensure observability of the whole system.

As mentioned earlier, we can control the complete system by placing actuators on all driver nodes. Figures 5.5 and 5.6 provides an example that illustrates driver nodes and SCCs for a graph and the placement of sensors/actuators for complete

observability and controllability. The system is made completely controllable by placing actuators at the driver nodes, and it is made completely observable by placing sensors on at least one node from each root SCC.

### 5.3.4 Evaluation of actuator and sensor reachability range

For the SAGD process, sensor/actuator locations are often in the vicinity of the well-pairs, as shown in Figure 5.7. For this situation, it is of interest to find the actuator and sensor reachability range from the obtained graph model.

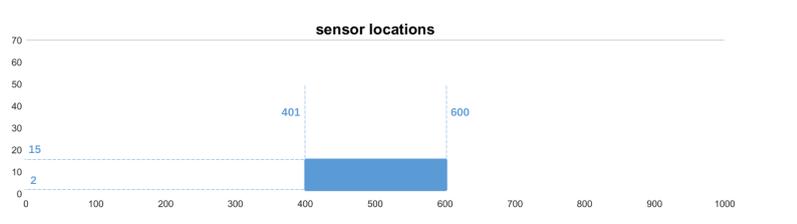
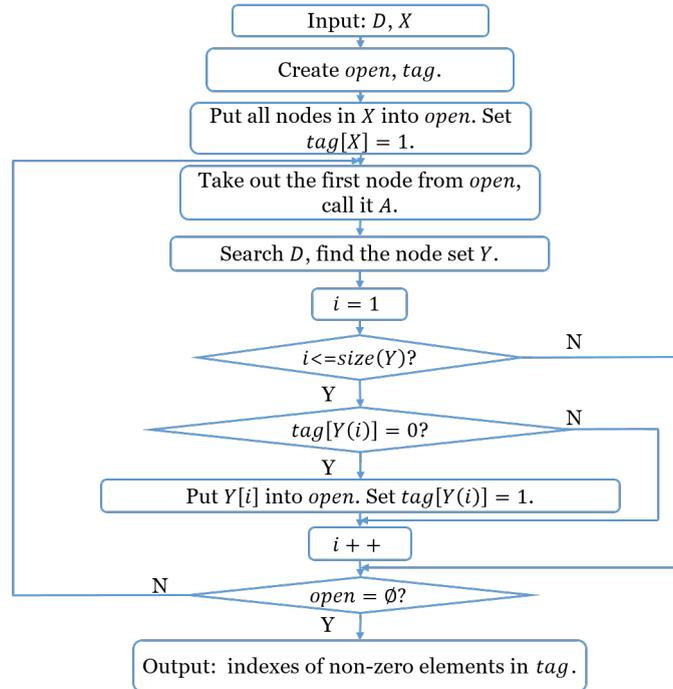


FIGURE 5.7: Sensor/actuator locations in the neighbourhood of well pairs

#### 5.3.4.1 Actuator reachability range

This algorithm (presented in Figure 5.8) identifies, for a specific set of actuator locations such as the ones shown in Figure 5.7, the actuation reachability regions (range) of the system. The algorithm can be described as follows:

1. Input: a  $N \times N$  adjacency matrix  $D$ , where  $N$  is the number of nodes in the network. The adjacency matrix reflects the reachability between nodes. Node set  $X$  stores the nodes at which we place actuators (which correspond to the clusters identified using the regionalization-based approach).
2. Create an empty set  $open$  and a  $1 \times N$  zero matrix  $tag$ .  $open$  records nodes to be processed.  $tag$  records whether a node has been touched. Initialize  $open$  by placing the node set  $X$  into it. Set  $tag[X] = 1$ .
3. Take out the first node from  $open$  and call it node  $A$ .
4. Search  $D$  to find the node set  $Y$  (there are directed edges that start from node  $A$  and point to nodes in  $Y$ ).
5. For the  $i^{th}$  node in  $Y$  (with  $i$  taking values from 1 to the number of nodes in  $Y$ ), if  $tag[Y(i)] = 0$ , place this node into  $open$  and set  $tag[Y(i)] = 1$ .



$D$ : A  $N \times N$  adjacency matrix of the network.  $N$  is the number of nodes in the network.  
 $X$ : A node set, it stores the nodes where we put actuators.  
 $open$ : Initially, it is  $\emptyset$ .  
 $tag$ : A  $1 \times N$  zero matrix.  
 $Y$ : A node set (there are directed edges start from node  $A$  and point to nodes in  $Y$ ).

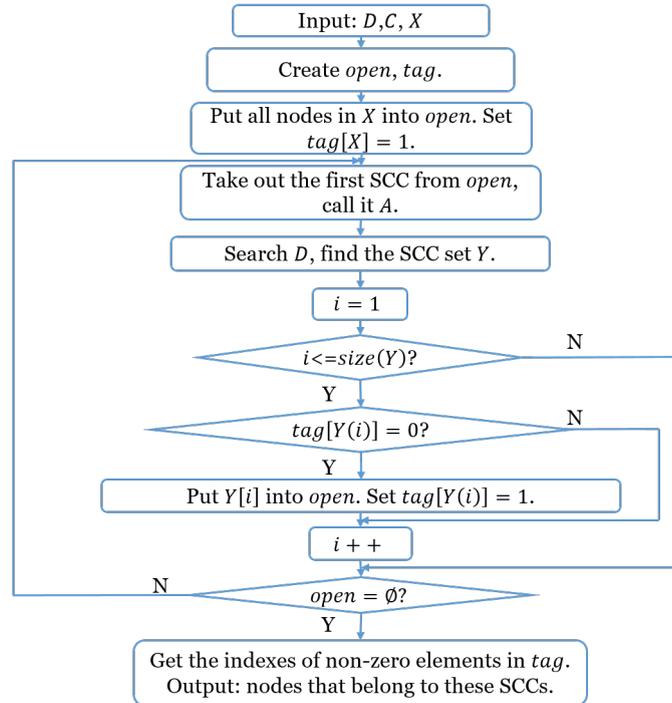
FIGURE 5.8: Algorithm to find the control range.

6. If  $open = \emptyset$ , the algorithm finishes; return the indices of non-zero elements in  $tag$ . Else, jump to step 3.

### 5.3.4.2 Sensor reachability range

This algorithm identifies, for a specific set of sensor locations such as the ones shown in Figure 5.7, the sensor reachability regions (range) of the system. The algorithm can be described as follows and is also presented in Figure 5.9:

1. Input: a  $S \times S$  adjacency matrix, where  $S$  is the number of SCCs in the network. The adjacency matrix reflects the reachability between SCCs.  $C$  is a  $1 \times N$  matrix, where  $N$  is the number of nodes in the network, that reflects the SCC to which each node belongs.  $X$  is the set of SCCs, and we place sensors on nodes within these SCCs set.
2. Create an empty matrix  $open$  and a  $1 \times S$  zero matrix  $tag$ .  $open$  records SCCs to be processed.  $tag$  records whether a SCC has been touched. Initialize  $open$  by placing the SCCs in  $X$  into it. Set  $tag[X] = 1$ .
3. Take out the first SCC from  $open$  and call it SCC  $A$ .



$D$ : A  $S \times S$  adjacency matrix reflecting reachability between SCCs.  $S$  is the number of SCCs in the network.

$C$ : It reflects which SCC each node belongs to.

$X$ : A SCC set (we put sensors on nodes within these SCCs).

$open$ : Initially, it is  $\emptyset$ .

$tag$ : A  $1 \times S$  zero matrix.

$Y$ : A SCC set (there are directed edges start from SCC  $A$  and point to SCCs in  $Y$ ).

FIGURE 5.9: Algorithm to find the observe range.

4. Search  $D$  to find the SCC set  $Y$  (there are directed edges that start from SCC  $A$  and point to SCCs in  $Y$ ).
5. For the  $i^{th}$  SCC in  $Y$  (with  $i$  taking values from 1 to the number of SCCs in  $Y$ ), if  $tag[Y(i)] = 0$ , put this node into  $open$  and set  $tag[Y(i)] = 1$ .
6. If  $open$  is empty, the algorithm finishes; obtain the indices of non-zero elements in matrix  $tag$  (the indices reflect which SCCs can be reached), and return the nodes that belong to these SCCs. Else, jump to step 3.

## 5.4 Results and discussion

This section discusses the results obtained using the techniques described above for the SAGD process simulated as described in section 5.3. Note again that the data corresponds to the PCA-based fusion of the pressure and temperature data.

### 5.4.1 Numbers of driver nodes and root SCCs and regionalization-based clustering

The clustering technique used is hierarchical, and this leaves the choice of the number of clusters to the user. The number of clusters controls the dimension reduction achieved and the size of the regions that are placed into single clusters; thus, it affects the subsequent analysis of observable and controllable regions of the reservoir. Table 5.1 shows the dependence of the number of driver nodes and root SCCs on the number of clusters chosen (in the range from 50 to 300 clusters). There is a roughly linear dependence of the number of driver nodes and root SCCs on the level of clustering; however, it should be noted that the size of the spatial regions corresponding to each cluster changes with the number of clusters. When a smaller number of clusters is chosen, the average size of the clusters is much larger than when a larger number of clusters is chosen. The issue of determining the optimal level of clustering is not trivial and is not addressed in this work. Figure

TABLE 5.1: Number of driver nodes and root SCCs

number of clusters	number of driver nodes	number of root SCCs
50	33	28
100	54	49
150	85	72
200	107	96
250	123	112
300	143	130

5.10 presents the results of regionalization-based clustering for the different choices of the number of clusters, with the  $X - Z$  plane of the reservoir and caprock being visualized. The changes in the average size of clusters as a function of the number of clusters can be seen. Each of the identified cluster is treated as a potential node in the graph-based description of the system identified in subsequent analysis. The density of clusters is highest in the central part of the visualized region, which corresponds to the location of the well-pairs and is where the steam chamber is developed. This is where one would expect the highest variation and diversity in the pressure/temperature dynamics with respect to spatial location. Given that this is region with the largest dissimilarity between samples (grid points), it is understandable that a larger number of clusters is situated in this relatively small region.

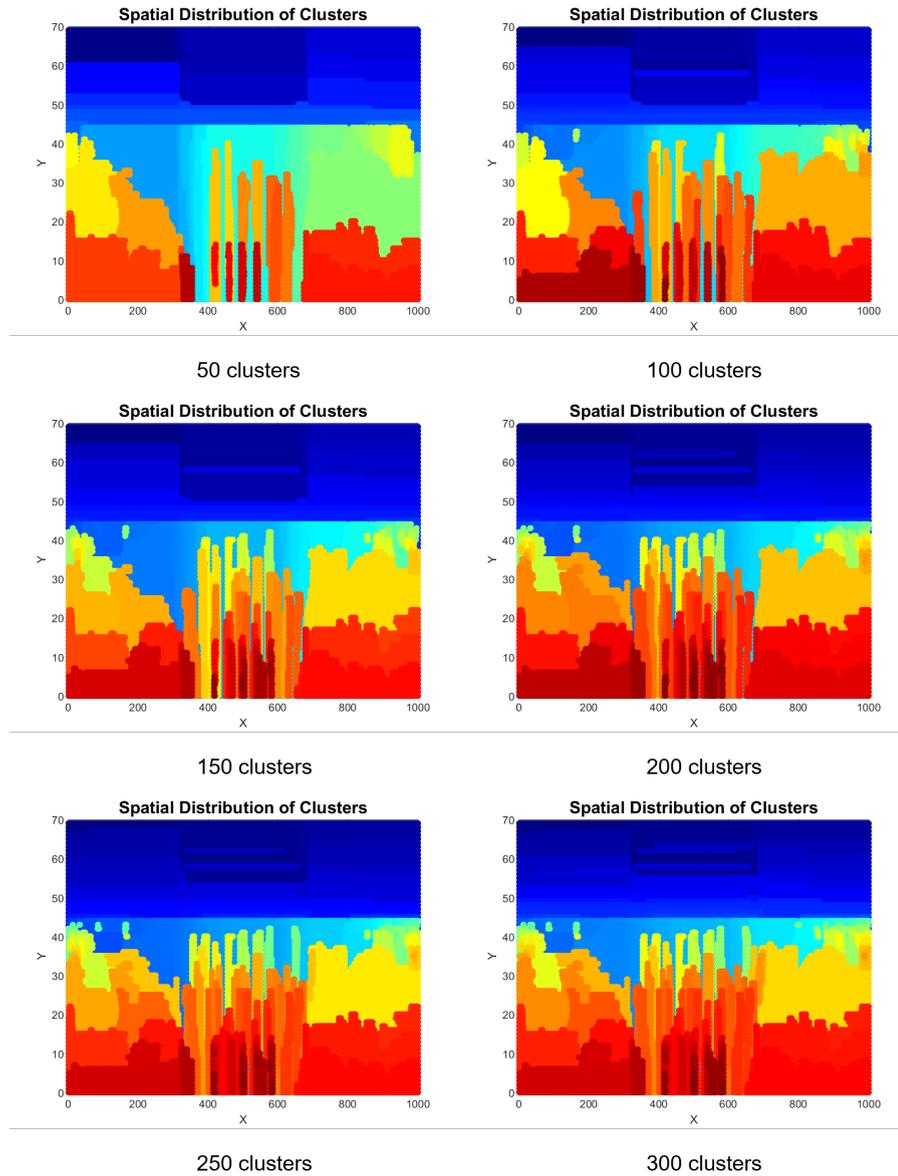


FIGURE 5.10: The variation of the spatial distribution of clusters with the number of clusters. Each color corresponds to a separate (spatially contiguous) cluster and the trend shows that the density of clusters increases in the central region, which is the region that exhibits large variations in pressure and temperature in the original data.

### 5.4.2 Graph-theoretic analysis

Figure 5.11 presents the graph model for the SAGD system, with nodes and the directed edges connecting them. As mentioned earlier, the directed edges were identified using Granger causality analysis with the threshold and lags specified in Section 5.3.2. As expected, the number of edges increases significantly with the number of clusters (nodes). Note that each node location represents the arithmetic mean of the coordinates of all data points within the corresponding cluster, and

the node connectivity graph is visualized in the spatial representation in the same  $X - Z$  frame as for the regionalization-based clustering.

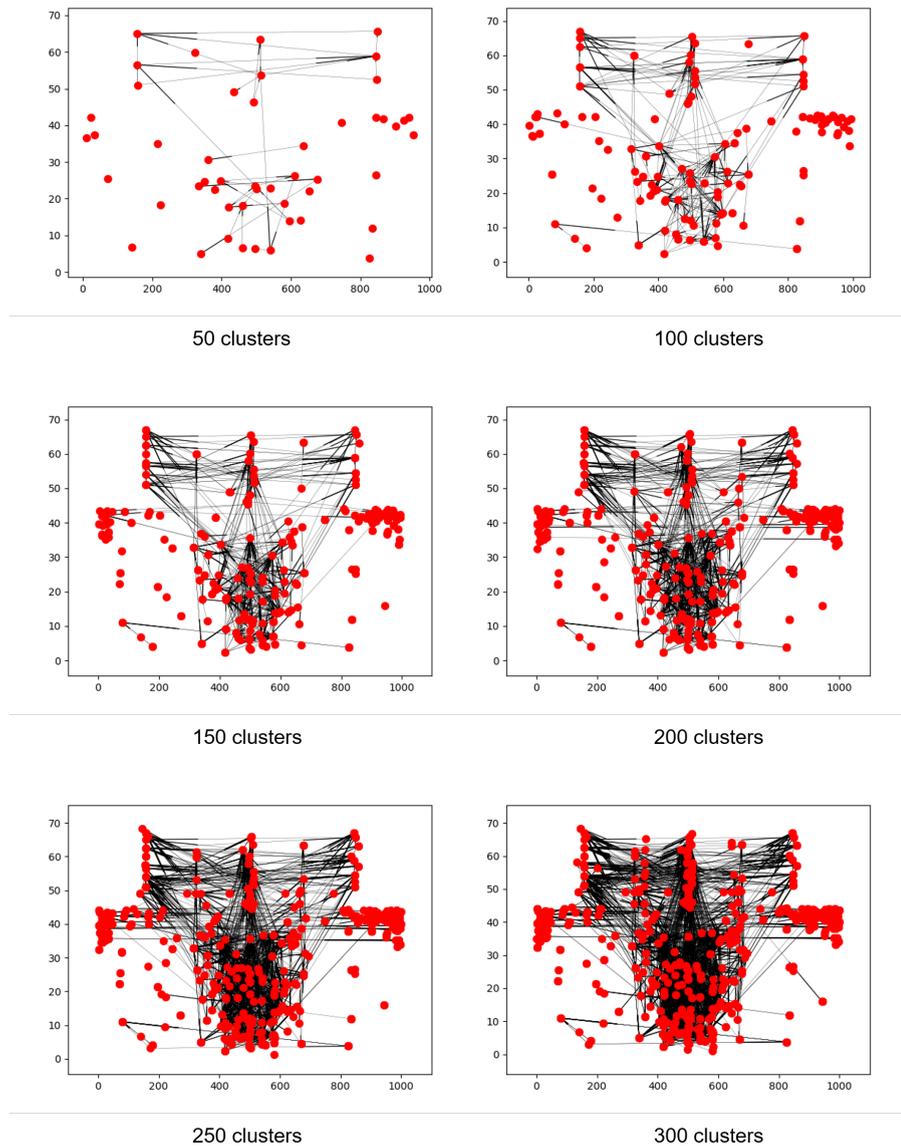


FIGURE 5.11: The graph model for the system (reservoir+caprock) based on the spatiotemporal pressure and temperature data, with nodes and edges identified.

To aid visualization and analysis, we present a measure of the degree distribution of the nodes for different numbers of clusters in Figures 5.12 and 5.13. In the figures, we identify the nodes with the most, least and medium connectivity with respect to in-degree and out-degree. The in-degree refers to the number of edges pointing towards the node, while the out-degree refers to the number of edges pointing away from the node. This helps us visualize the densely and sparsely connected regions of the graph and their spatial location in the reservoir. Figure

5.12 shows the in- and out- degree distribution for 50-150 clusters and Figure 5.13 shows the same distributions for 200-300 clusters. As expected, the in- and out-degrees both are highest, in general, for nodes in the central region of the figure, which corresponds to the well-pair locations and the development of the steam chamber.

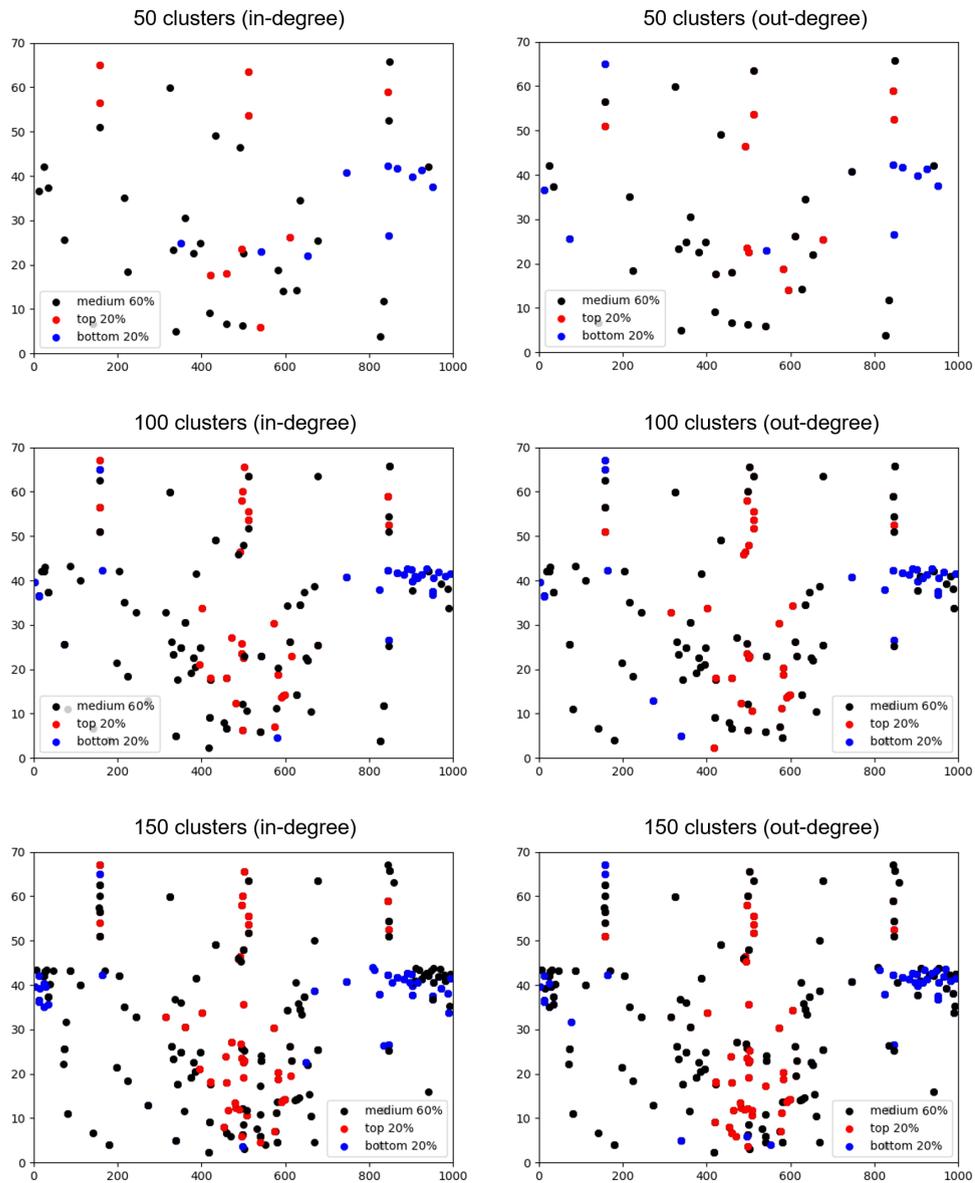


FIGURE 5.12: (In- and out-) Degree distribution of the nodes of the graph as a function of the number of clusters (nodes) in the range 50-150. Red markers correspond to the top 20% nodes in terms of in- or out- degree, blue markers to the bottom 20%, and black markers to the middle 60% in terms of connectivity.

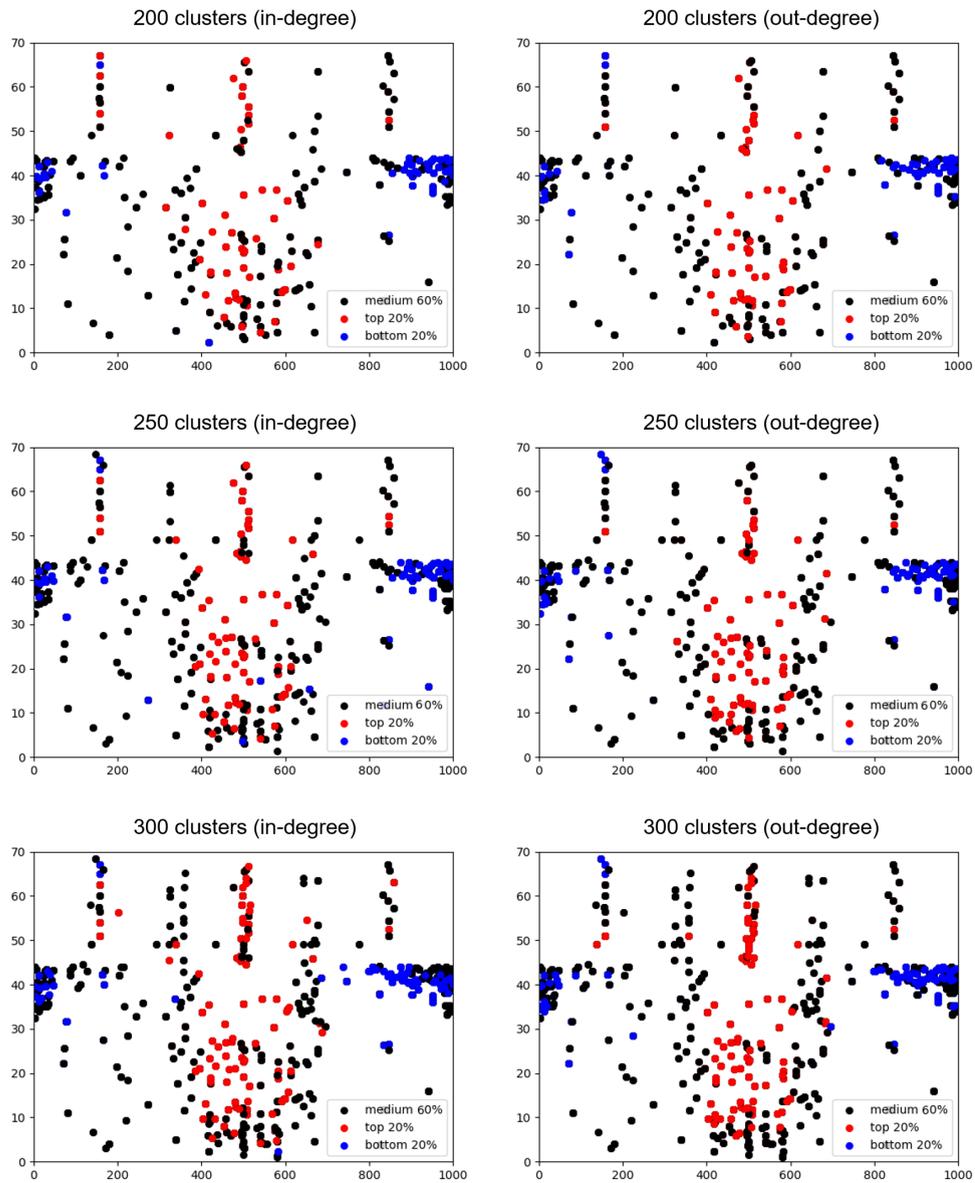


FIGURE 5.13: (In- and out-) Degree distribution of the nodes of the graph as a function of the number of clusters (nodes) in the range 200-300. Red markers correspond to the top 20% nodes in terms of in- or out- degree, blue markers to the bottom 20%, and black markers to the middle 60% in terms of connectivity.

### 5.4.3 Complete structural controllability and observability

As described earlier, the driver nodes identify the set of actuator locations for which complete structural controllability of the graph is obtained. Figure 5.14 presents the clusters that represent driver nodes that provide complete controllability for the SAGD reservoir studied in this work. Placing actuators in any location within each of the identified clusters will enable complete controllability.

of the system. The highest density of driver nodes is within the region where the well-pairs are located, as expected.

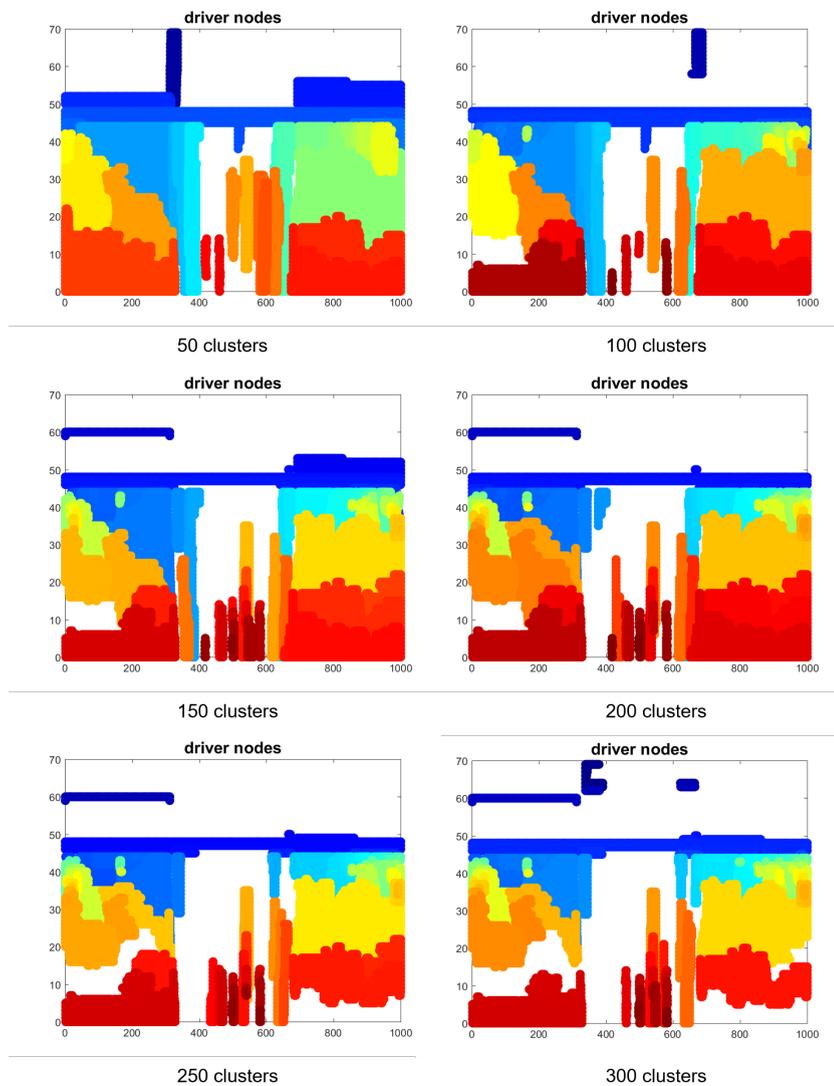


FIGURE 5.14: Location of driver nodes for complete controllability as a function of the number of clusters. Each colour corresponds to a separate cluster identified as a driver node. Blank regions do not contain driver nodes.

Similarly, Figure 5.15 shows the root SCCs, which identifies the clusters where sensors need to be placed in order to obtain complete observability of the system. More root SCCs are concentrated in the central region of the reservoir, which is where the steam chamber is developed and the larger spatiotemporal variations in pressure and temperature occur, which makes sense.

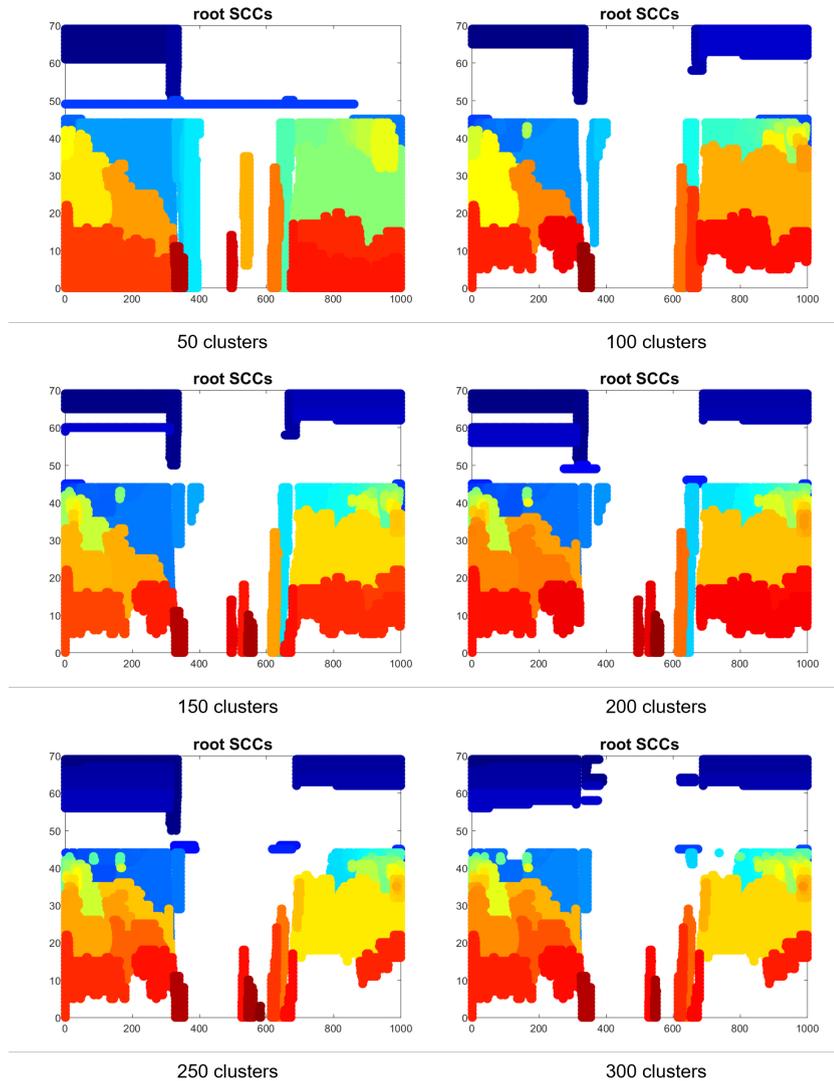


FIGURE 5.15: Location of the root strongly connected components as a function of the number of clusters. Each colour corresponds to a separate cluster identified as a root SCC.

#### 5.4.4 Actuator and sensor range

As mentioned earlier, the actuator and sensor ranges identify the regions of actuation and sensor reachability in a partially observable system with spatiotemporal variation. In this section, we present the results for specific actuator and sensor locations. If actuators are placed in the shaded area shown in Figure 5.7, Figure 5.16 shows the actuated and non-actuated regions as a function of the number of clusters. Similarly, if sensors are placed in the same region as the actuators, Figure 5.17 shows the regions in-sensor-range and out-of-sensor-range as a function of the number of clusters. It is noticeable that these ranges are larger for graphs with larger numbers of nodes (clusters), which indicates that the smallest amount of

clustering/lumping of spatial regions possible is preferable from the viewpoint of maximizing the ranges. However, not performing clustering presents significant computational demands because of the large dimension of the set of variables to be considered for constructing the graph based on Granger causality. It is gratifying to note that, as expected, both of the ranges are in the central part of the reservoir, where the well-pairs are located and the steam chamber develops.

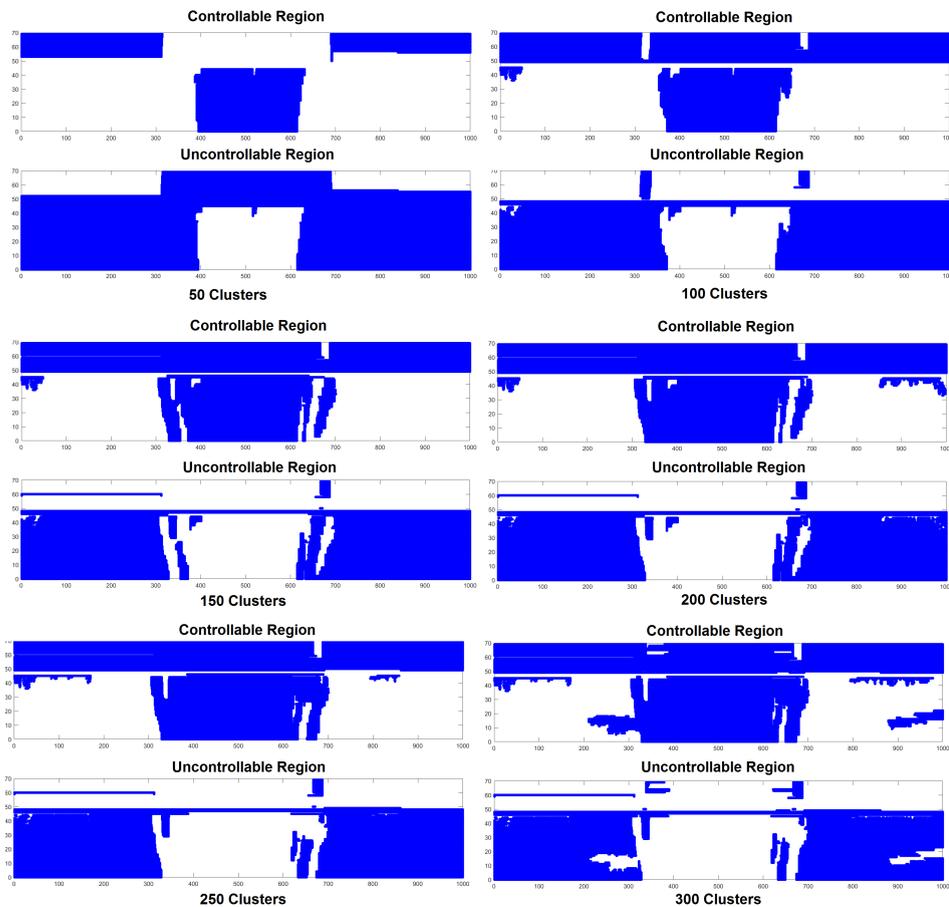


FIGURE 5.16: Actuation range for the actuator locations as shown in Figure 5.7. The graphs with higher numbers of nodes yield larger controllable regions.

## 5.5 Conclusions

We have demonstrated a method to investigate the structural controllability and observability of distributed parameter systems for which a fundamental model or simulations are available based on deriving a directed graph using regionalization-based clustering, and connectivity determination in a graph model based on Granger causality. Algorithms for maximum matching and the identification of strongly connected components were used to identify locations at which sensors

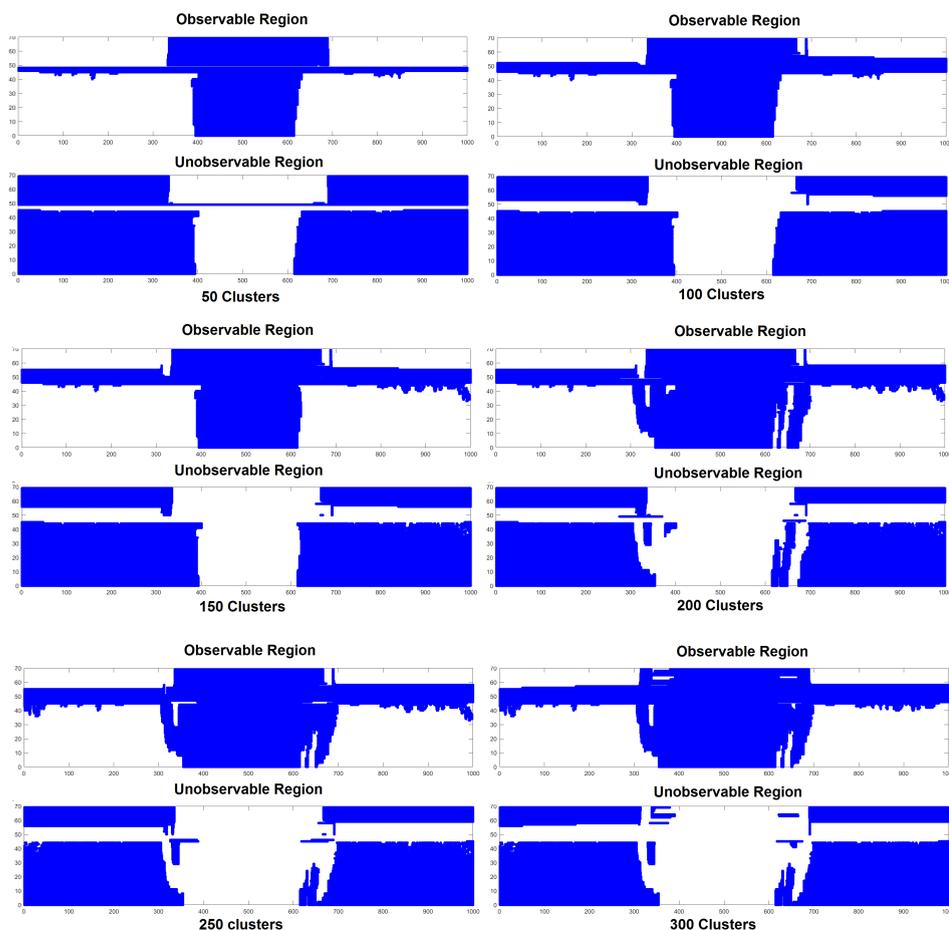


FIGURE 5.17: In-sensor-range for the sensor locations as shown in Figure 5.7. The graphs with higher numbers of nodes yields larger observable regions.

or actuators needed to be placed for full observability or controllability, respectively. In addition, methods were developed to identify the reachable regions on the graph (indicating the partially actuated/in-sensor-range regions) based on a specified set of actuator/sensor locations. The method was demonstrated for the analysis of the pressure and temperature fields of a heterogeneous steam assisted gravity drainage reservoir process, but is applicable to a wide variety of distributed parameter systems. Our future work in this area will focus on more rigorous analysis of the optimal level of clustering, applying the observability results in covariance localization for history matching in SAGD operation, and on applying the controllability results in model predictive control of the SAGD process.

# Bibliography

- [1] R. M. Butler, Steam-assisted gravity drainage: concept, development, performance and future, *Journal of Canadian Petroleum Technology* 33 (02) (1994) 44–50.
- [2] Y. Souraki, M. Ashrafi, H. Karimaie, O. Torsaeter, Experimental analyses of Athabasca bitumen properties and field scale numerical simulation study of effective parameters on SAGD performance, *Energy and Environment Research* 2 (1) (2012) 140.
- [3] A. Nabilou, Study of the parameters of steam assisted gravity drainage (SAGD) method for enhanced oil recovery in a heavy oil fractured carbonate reservoir, *American Journal of Engineering and Applied Sciences* 9 (2016) 647–658.
- [4] A. Raghu, X. Yang, S. Khare, J. Prakash, B. Huang, V. Prasad, Reservoir history matching using constrained ensemble Kalman filtering, *The Canadian Journal of Chemical Engineering* 96 (1) (2018) 145–159.
- [5] V. Gervais, M. Le Ravalec, Identifying influence areas with connectivity analysis—application to the local perturbation of heterogeneity distribution for history matching, *Computational Geosciences* 22 (1) (2018) 3–28.
- [6] K. Sasaki, S. Akibayashi, N. Yazawa, Q. Doan, S. Ali, Numerical and experimental modelling of the steam-assisted gravity drainage (SAGD) process, in: *Annual Technical Meeting, Petroleum Society of Canada*, 1999.
- [7] O. R. Ayodele, T. N. Nasr, J. J. Ivory, G. Beaulieu, G. Heck, Testing and history matching of ES-SAGD (using hexane), in: *SPE Western Regional Meeting, Society of Petroleum Engineers*, 2010.
- [8] E. Aulisa, D. Gilliam, *A practical guide to geometric regulation for distributed parameter systems*, Chapman and Hall/CRC, 2015.

- 
- [9] R. E. Kalman, Mathematical description of linear dynamical systems, *Journal of the Society for Industrial and Applied Mathematics, Series A: Control* 1 (2) (1963) 152–192.
- [10] D. G. Luenberger, *Introduction to dynamic systems: Theory, models and applications* (1979).
- [11] R. F. Curtain, H. Zwart, *An introduction to infinite-dimensional linear systems theory*, Vol. 21, Springer Science & Business Media, 2012.
- [12] S. R. Pacharu, R. Gudi, S. Patwardhan, Advanced state estimation techniques for packed bed reactors, *IFAC Proceedings Volumes* 45 (15) (2012) 519–524.
- [13] C. Antoniadis, P. D. Christofides, Integrated optimal actuator/sensor placement and robust control of uncertain transport-reaction processes, *Computers & Chemical Engineering* 26 (2) (2002) 187–203.
- [14] D. Ucinski, M. A. Demetriou, An approach to the optimal scanning measurement problem using optimum experimental design, in: *Proceedings of the 2004 American Control Conference*, Vol. 2, IEEE, 2004, pp. 1616–1621.
- [15] D. G. Lainiotis, W. H. Ray, *Distributed Parameter Systems: Identification, Estimation, and Control*, Marcel Dekker, 1978.
- [16] U. Boscain, J.-P. Gauthier, F. Rossi, M. Sigalotti, Approximate controllability, exact controllability, and conical eigenvalue intersections for quantum mechanical systems, *Communications in Mathematical Physics* 333 (3) (2015) 1225–1239.
- [17] W. B. Rouse, Engineering complex systems: Implications for research in systems engineering, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 33 (2) (2003) 154–156.
- [18] B. Cai, H. Liu, M. Xie, A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks, *Mechanical Systems and Signal Processing* 80 (2016) 31–44.
- [19] C.-T. Lin, Structural controllability, *IEEE Transactions on Automatic Control* 19 (3) (1974) 201–208.

- 
- [20] S. Diop, M. Fliess, Nonlinear observability, identifiability, and persistent trajectories, in: [1991] Proceedings of the 30th IEEE Conference on Decision and Control, IEEE, 1991, pp. 714–719.
- [21] Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks, *Nature* 473 (7346) (2011) 167.
- [22] Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Observability of complex systems, *Proceedings of the National Academy of Sciences* 110 (7) (2013) 2460–2465.
- [23] L. Chow, R. Butler, Numerical simulation of the steam-assisted gravity drainage process (SAGD), *Journal of Canadian Petroleum Technology* 35 (06).
- [24] D. Guo, Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP), *International Journal of Geographical Information Science* 22 (7) (2008) 801–823.
- [25] Z. Galil, Efficient algorithms for finding maximum matching in graphs, *ACM Computing Surveys (CSUR)* 18 (1) (1986) 23–38.
- [26] J. E. Hopcroft, R. M. Karp, An  $n^{5/2}$  algorithm for maximum matchings in bipartite graphs, *SIAM Journal on Computing* 2 (4) (1973) 225–231.
- [27] J. Edmonds, R. M. Karp, Theoretical improvements in algorithmic efficiency for network flow problems, *Journal of the ACM (JACM)* 19 (2) (1972) 248–264.
- [28] R. Tarjan, Depth-first search and linear graph algorithms, *SIAM Journal on Computing* 1 (2) (1972) 146–160.
- [29] A. Gibbons, *Algorithmic graph theory*, Cambridge university press, 1985.
- [30] Computer Modeling Group, Calgary, Alberta, T2M 3Y7 Canada (2015 (10.5715.22942)).
- [31] A. Ganesh, K. Lyu, R. J. Chalaturnyk, R. Gudi, V. Prasad, Graph-based structural controllability and observability of steam assisted gravity drainage pressure chamber: A data driven approach, *IFAC-PapersOnLine* 52 (1) (2019) 766–771.

- 
- [32] A. Ganesh, J. Xiong, R. J. Chalaturnyk, V. Prasad, Proxy models for caprock pressure and temperature dynamics during steam-assisted gravity drainage process, *Computers & Chemical Engineering* 121 (2019) 594–607.
- [33] T. Söderström, P. Stoica, *System identification*, Prentice-Hall, Inc., 1988.
- [34] G. Deshpande, S. LaConte, G. A. James, S. Peltier, X. Hu, Multivariate Granger causality analysis of fMRI data, *Human Brain Mapping* 30 (4) (2009) 1361–1373.
- [35] J. P. Hamilton, G. Chen, M. E. Thomason, M. E. Schwartz, I. H. Gotlib, Investigating neural primacy in major depressive disorder: multivariate Granger causality analysis of resting-state fMRI time-series data, *Molecular Psychiatry* 16 (7) (2011) 763.
- [36] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems* 2 (1-3) (1987) 37–52.
- [37] B. Moore, Principal component analysis in linear systems: Controllability, observability, and model reduction, *IEEE Transactions on Automatic Control* 26 (1) (1981) 17–32.
- [38] S. Seabold, J. Perktold, Statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference, 2010.
- [39] W. H. Day, H. Edelsbrunner, Efficient algorithms for agglomerative hierarchical clustering methods, *Journal of Classification* 1 (1) (1984) 7–24.
- [40] M. Steinbach, G. Karypis, V. Kumar, A comparison of document clustering techniques, in: *KDD workshop on text mining*, Vol. 400, Boston, 2000, pp. 525–526.
- [41] C. W. Granger, Testing for causality: a personal viewpoint, *Journal of Economic Dynamics and Control* 2 (1980) 329–352.
- [42] J.-J. E. Slotine, W. Li, *Applied nonlinear control*, Vol. 199, Prentice hall Englewood Cliffs, NJ, 1991.

# Chapter 6

## Conclusions

## 6.1 Conclusion

The aim of this thesis is to demonstrate a framework to optimally manage the SAGD reservoir using proxy models. For this purpose, two computationally inexpensive proxy modelling approaches were presented. A POD-based model order reduction coupled with system identification was proposed to model and predict the pressure field. This approach demonstrated a promising result in modelling and predicting the entire pressure field and its response to variations in the reservoir inputs. A clustering-based line graph model was proposed to model the pressure bin counts (regions of different ranges of pressures), which yielded a subspace model whose parameters were identified from the data obtained from detailed reservoir simulations. These methods can be implemented in forecasting the pressure/temperature fields dynamically based on predictions of the future injector and producer well bore pressures. A polynomial chaos expansion based proxy model was then proposed to quantify the caprock factor of safety based on the minimum FoS value and the normalized singular value of the FoS frame. The location of the weakest point (minimum FoS) in the caprock was also tracked and presented. We then demonstrated a framework to propagate the uncertainty in the petrophysical parameters and the well BHP of a SAGD process to the factor of safety of the caprock. We proposed a computationally efficient technique to quantify uncertainty using singular value decomposition coupled with PCE, requiring a handful of realizations compared to Monte Carlo techniques requiring thousands of realizations. The PCE model was validated by reconstructing the representative FoS. Quantifying uncertainty in the FoS with respect to inputs and the permeability was demonstrated by determining the moments of the representative. A model predictive controller was proposed based on the previous proxy models to track the FoS and to maximize the production, thus demonstrating an optimal closed loop reservoir management scheme based on proxy models. Apart from this, we also investigated the structural controllability and observability of distributed parameter systems to identify locations at which sensors or actuators needed to be placed for full observability or controllability, respectively. We used regionalization-based clustering, and connectivity determination in a graph model based on Granger causality. Algorithms for maximum matching and the identification of strongly connected components were used to determine the structural controllability and observability. In addition, methods were developed to identify the reachable regions on the graph (indicating the partially actuated/in-sensor-range regions) based on a specified set of actuator/sensor locations.

## 6.2 Future work

Some possibilities for future work are highlighted below:

- Incorporating the PCE approach to model the reservoir states, input and outputs and coupling with an ensemble Kalman filter to realize a hybrid estimator for the reservoir geomechanical optimization studies
- Updating the proxy model parameters with the real time geomechanical data like tiltmeter data and/or SAGD well log sensors data
- Rigorous analysis of the optimal level of clustering, applying the observability results in covariance localization for history matching in SAGD operation, and on applying the controllability results in model predictive control of the SAGD process
- Exploring HO-SVD based techniques to proxy model the spatiotemporal data.
- Improvising the MPC with more detailed proxy models to realize stochastic and robust formulations.

# Bibliography

- [1] A. I. Forrester & A. J. Keane, (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3), 50-79.
- [2] A. Forrester & A. Keane, (2008). *Engineering design via surrogate modeling: A practical guide*. John Wiley & Sons.
- [3] A. C. Antoulas, *Approximation of large-scale dynamical systems*. Vol. 6. SIAM, 2005.
- [4] E. Gildin & T. Lopez, (2011, January). Closed-loop reservoir management: Do we need complex models?. In *SPE Digital Energy Conference and Exhibition*. Society of Petroleum Engineers.
- [5] R. M. Butler, (1994). Steam-assisted gravity drainage: concept, development, performance and future. *Journal of Canadian Petroleum Technology*, 33(02), 44-50.
- [6] M. R. Carlson, (2012, January). An analysis of the caprock failure at Joslyn. In *SPE Heavy Oil Conference Canada*. Society of Petroleum Engineers.
- [7] Computer Modeling Group, Calgary, Alberta, T2M 3Y7 Canada (2015 (10.5715.22942)).
- [8] *FLAC3D-Fast Lagrangian Analysis of Continua in 3Dimensions*, Minneapolis, USA (Version 7.0).
- [9] A. Azad, and R. J. Chalaturnyk, A mathematical improvement to SAGD using geomechanical modeling. *Journal of Canadian Petroleum Technology* 49.10 (2010): 53-64.
- [10] G. van Essen, P. Van den Hof & J. D. Jansen (2013). A two-level strategy to realize life-cycle production optimization in an operational setting. *SPE Journal*, 18(06), 1-057.

- 
- [11] D. I. Zubarev, (2009, January). Pros and cons of applying proxy-models as a substitute for full reservoir simulations. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.
- [12] S. Wojtkiewicz, M. Eldred, Jr, R. Field, A. Urbina, & J. Red-Horse, (2001). Uncertainty quantification in large computational engineering models. In 19th AIAA Applied Aerodynamics Conference (p. 1455).
- [13] J. W. Bandler, & K. Madsen, (2001). Surrogate modeling and space mapping for engineering optimization. *Optimization and Engineering*, 2(4), 367-368.
- [14] K. Lindhorst, M. C. Haupt, & P. Horst, (2014). Efficient surrogate modeling of nonlinear aerodynamics in aerostructural coupling schemes. *AIAA Journal*, 52(9), 1952-1966.
- [15] T. Heijn, R. Markovinovic and J. D. Jansen, Generation of low-order reservoir models using system-theoretical concepts. SPE Reservoir Simulation Symposium. Society of Petroleum Engineers, 2003.
- [16] A. K. Tangirala, (2014). Principles of system identification: Theory and practice. CRC Press.
- [17] T. Soderstrom, & Stoica, P. (1989). System identification.
- [18] S. Yao, J. J. Trivedi and V. Prasad, Proxy modeling of the production profiles of SAGD reservoirs based on system identification. *Industrial & Engineering Chemistry Research* 54.33 (2015): 8356-8367.
- [19] S. S. Vembadi, R. G. Patel, J. J. Trivedi, & V. Prasad, Real-time feedback control of SAGD wells using model predictive control to optimize steam chamber development under uncertainty. *The Canadian Journal of Chemical Engineering*, 96(6)(June 2018): 1290-1305.
- [20] S. Afra & E. Gildin, (2013). Permeability parametrization using higher order singular value decomposition (HOSVD). In *Machine Learning and Applications (ICMLA)*, 2013 12th International Conference on (Vol. 2, pp. 188-193). IEEE.
- [21] G. R. Gavalas, P. C. Shah and J. H. Seinfeld, Reservoir history matching by Bayesian estimation. *Society of Petroleum Engineers Journal* 16.06 (1976): 337-350.

- [22] G. Chavent, M. Dupuy and P. Lemmonier, History matching by use of optimal theory. *Society of Petroleum Engineers Journal* 15.01 (1975): 74-86.
- [23] J. A. Jalali, S. D. Mohaghegh, & R. Gaskari, (2010). Coalbed methane reservoir simulation and uncertainty analysis with artificial neural networks. *Scientia Iranica. Transaction C, Chemistry, Chemical Engineering*, 17(1), 65.
- [24] P. A. Slotte, & E. Smorgrav, (2008, January). Response surface methodology approach for history matching and uncertainty assessment of reservoir simulation models. In *Europec/EAGE Conference and Exhibition*. Society of Petroleum Engineers.
- [25] G. Fishman, (2013). *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media.
- [26] A. Narasingam, P. Siddhamshetty, J. S. Kwon, Temporal clustering for order reduction of nonlinear parabolic PDE systems with time-dependent spatial domains: Application to a hydraulic fracturing process, *AIChE J.*, 2017, 63, 3818-3831.
- [27] A. Narasingam, J. S. Kwon, Development of local dynamic mode decomposition with control: Application to model predictive control of hydraulic fracturing, *Comp. & Chem. Eng.*, 2017, 106, 501-511.
- [28] H. S. Sidhu, A. Narasingam, P. Siddhamshetty, J. S. Kwon, Model order reduction of nonlinear parabolic PDE systems with moving boundaries using sparse proper orthogonal decomposition: Application to hydraulic fracturing, *Comp. & Chem. Eng.*, 2018, 112, 92-100.
- [29] J. Xiong, R. J. Chalaturnyk, Caprock Safety Factor Assessment of SAGD Projects. *American Rock Mechanics Association (ARMA), 49th US Rock Mechanics / Geomechanics Symposium*, San Francisco, CA, USA.
- [30] A. Azad, R. Chalaturnyk & S. Movaghati, (2011). Reservoir Characterization: Application of Extended Kalman Filter and Analytical Physics-Based Proxy Models in Thermal Recovery. In *Proc., International Association for Mathematical Geosciences Conference IAMG*.
- [31] A. Azad, R. J. Chalaturnyk & S. Movaghati, Real-time reservoir model updating in thermal recovery: Application of analytical proxies and Kalman filtering. *Journal of Petroleum Science and Engineering* 127 (2015): 196-211.

- 
- [32] J. D. Jansen, D. R. Brouwer, G. Naevdal & C. P. J. W. Van Kruijsdijk, (2005) Closed-loop reservoir management. *First Break*, 23(1), 43-48.
- [33] MATLAB version R2016a (9.0.0.341360) Natick, Massachusetts: The MathWorks Inc., 2016.
- [34] A. Chatterjee, (2000). An introduction to the proper orthogonal decomposition. *Current Science*, 808-817.
- [35] M. H. Sahraei, M. A. Duchesne, P. G. Boisvert, R. W. Hughes & L. A. Ricardez-Sandoval, (2017). Reduced-Order Modeling of a Commercial-Scale Gasifier Using a Multielement Injector Feed System. *Industrial & Engineering Chemistry Research*, 56(25), 7285-7300.
- [36] M. H. Sahraei, M. A. Duchesne, R. W. Hughes & L. A. Ricardez-Sandoval, (2017). Dynamic reduced order modeling of an entrained-flow slagging gasifier using a new recirculation ratio correlation. *Fuel*, 196, 520-531.
- [37] M. Moharir, L. Kang, P. Daoutidis & A. Almansoori, (2017). Graph representation and decomposition of ODE/hyperbolic PDE systems. *Computers & Chemical Engineering*, 106, 532-543.
- [38] P. M. Collins (2005), Geomechanical Effects on the SAGD Process, SPE Reservoir Evaluations and Engineering Paper 97905: pp: 367-375.
- [39] H. N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Annual Review of Fluid Mechanics* 41 (2009) 35–52.
- [40] A. Ganesh, T. Jain, R. Chalaturnyk, V. Prasad, Robust optimization of the production profile of the steam assisted gravity drainage reservoir using a polynomial chaos expansion-based proxy model, in: 2018 18th International Conference on Control, Automation and Systems (ICCAS), IEEE, 2018, pp. 1254–1259.
- [41] C. Yang, C. Card, L. Nghiem, Economic optimization and uncertainty assessment of commercial SAGD operations, in: Canadian International Petroleum Conference, Petroleum Society of Canada, 2007.
- [42] C. Yang, C. Card, L. X. Nghiem, E. Fedutenko, Robust optimization of SAGD operations under geological uncertainties, in: SPE Reservoir Simulation Symposium, Society of Petroleum Engineers, 2011.

- 
- [43] F. A. Dilib, M. D. Jackson, Closed-loop feedback control for production optimization of intelligent wells under uncertainty, *SPE Production & Operations* 28 (04) (2013) 345–357.
- [44] F. Xiong, B. Xue, Z. Yan, S. Yang, Polynomial chaos expansion based robust design optimization, in: 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, IEEE, 2011, pp. 868–873.
- [45] P. A. Slotte, E. Smorgrav, Response surface methodology approach for history matching and uncertainty assessment of reservoir simulation models, in: *Europec/EAGE Conference and Exhibition*, Society of Petroleum Engineers, 2008.
- [46] G. Kimaev, L. A. Ricardez-Sandoval, A comparison of efficient uncertainty quantification techniques for stochastic multiscale systems, *AIChE Journal* 63 (8) (2017) 3361–3373.
- [47] R. G. Patel, T. Jain, J. J. Trivedi, Polynomial-chaos-expansion based integrated dynamic modelling workflow for computationally efficient reservoir characterization: a field case study, in: *SPE Europec featured at 79th EAGE Conference and Exhibition*, Society of Petroleum Engineers, 2017.
- [48] J. Wu, G. Zambrano-Narvaez, R. Chalaturnyk, Newly developed centrifuge testing program of SAGD caprock integrity, in: 49th US Rock Mechanics/-Geomechanics Symposium, American Rock Mechanics Association, 2015.
- [49] J. Mandur, H. Budman, A polynomial-chaos based algorithm for robust optimization in the presence of Bayesian uncertainty, *IFAC Proceedings Volumes* 45 (15) (2012) 549–554.
- [50] G. Senthamaraiikkannan, I. Gates, V. Prasad, Multiphase reactive-transport simulations for estimation and robust optimization of the field scale production of microbially enhanced coalbed methane, *Chemical Engineering Science* 149 (2016) 63–77.
- [51] M. D. Webster, M. A. Tatang, G. J. McRae, Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model, MIT Joint program on the science and policy of global change, report Series No. 4.

- [52] K.-K. Kim, D. E. Shen, Z. K. Nagy, R. D. Braatz, Wiener's polynomial chaos for the analysis and control of nonlinear dynamical systems with probabilistic uncertainties [historical perspectives], *IEEE Control Systems Magazine* 33 (5) (2013) 58–67.
- [53] G. Berkooz, P. Holmes, J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annual Review of Fluid Mechanics* 25 (1) (1993) 539–575.
- [54] A. Narasingam, P. Siddhamshetty, J. S.-I. Kwon, Identification of spatially varying geological properties in a heterogeneous reservoir using EnKF and POD based parameterization, in: 2018 Annual American Control Conference (ACC), IEEE, 2018, pp. 1144–1149.
- [55] V. Klema, A. Laub, The singular value decomposition: Its computation and some applications, *IEEE Transactions on Automatic Control* 25 (2) (1980) 164–176.
- [56] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, *SIAM Journal on Matrix Analysis and Applications* 21 (4) (2000) 1253–1278.
- [57] H. V. Ly, H. T. Tran, Modeling and control of physical processes using proper orthogonal decomposition, *Mathematical and Computer Modelling* 33 (1-3) (2001) 223–236.
- [58] B. Zhang, N. Deisman, M. Khajeh, R. Chalaturnyk, J. Boisvert, Numerical local upscaling of elastic geomechanical properties for heterogeneous continua, *Petroleum Geoscience* (2019) petgeo2018–159.
- [59] J. Kim, H. A. Tchelepi, R. Juanes, Stability, accuracy and efficiency of sequential methods for coupled flow and geomechanics, in: SPE reservoir simulation symposium, Society of Petroleum Engineers, 2009.
- [60] M. E. Wall, A. Rechtsteiner, L. M. Rocha, Singular value decomposition and principal component analysis, in: *A practical approach to microarray data analysis*, Springer, 2003, pp. 91–109.
- [61] M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (3) (1999) 611–622.

- [62] A. Ganesh, J. Xiong, R. J. Chalaturnyk, V. Prasad, Proxy models for caprock pressure and temperature dynamics during steam-assisted gravity drainage process, *Computers & Chemical Engineering* 121 (2019) 594–607.
- [63] W. Gautschi, Algorithm 726: Orthpol—a package of routines for generating orthogonal polynomials and gauss-type quadrature rules, *ACM Transactions on Mathematical Software (TOMS)* 20 (1) (1994) 21–62.
- [64] C. W. Murray, N. C. Handy, G. J. Laming, Quadrature schemes for integrals of density functional theory, *Molecular Physics* 78 (4) (1993) 997–1014.
- [65] A. Mesbah, S. Streif, R. Findeisen, R. D. Braatz, Stochastic nonlinear model predictive control with probabilistic constraints, in: *2014 American Control Conference, IEEE, 2014*, pp. 2413–2419.
- [66] P. Siddhamshetty, J. S.-I. Kwon, Model-based feedback control of oil production in oil-rim reservoirs under gas coning conditions, *Computers & Chemical Engineering* 112 (2018) 112–120.
- [67] K. Patel, E. M. Aske, M. Fredriksen, et al., Use of model-predictive control for automating sagd well-pair operations: A simulation study, *SPE Production & Operations* 29 (02) (2014) 105–113.
- [68] S. J. Qin, T. A. Badgwell, A survey of industrial model predictive control technology, *Control engineering practice* 11 (7) (2003) 733–764.
- [69] S. S. Vembadi, Real-time feedback control of the sagd process using model predictive control to improve recovery: A simulation study.
- [70] B. L. Gorissen, İ. Yanıkoğlu, D. den Hertog, A practical guide to robust optimization, *Omega* 53 (2015) 124–137.
- [71] G. Kimaev, L. A. Ricardez-Sandoval, Multilevel Monte Carlo applied to chemical engineering systems subject to uncertainty, *AIChE Journal* 64 (5) (2018) 1651–1661.
- [72] A. Mesbah, Stochastic model predictive control: An overview and perspectives for future research, *IEEE Control Systems Magazine* 36 (6) (2016) 30–44.
- [73] Y. Souraki, M. Ashrafi, H. Karimaie, O. Torsaeter, Experimental analyses of Athabasca bitumen properties and field scale numerical simulation study

- of effective parameters on SAGD performance, *Energy and Environment Research* 2 (1) (2012) 140.
- [74] A. Nabilou, Study of the parameters of steam assisted gravity drainage (SAGD) method for enhanced oil recovery in a heavy oil fractured carbonate reservoir, *American Journal of Engineering and Applied Sciences* 9 (2016) 647–658.
- [75] A. Raghu, X. Yang, S. Khare, J. Prakash, B. Huang, V. Prasad, Reservoir history matching using constrained ensemble Kalman filtering, *The Canadian Journal of Chemical Engineering* 96 (1) (2018) 145–159.
- [76] V. Gervais, M. Le Ravalec, Identifying influence areas with connectivity analysis—application to the local perturbation of heterogeneity distribution for history matching, *Computational Geosciences* 22 (1) (2018) 3–28.
- [77] K. Sasaki, S. Akibayashi, N. Yazawa, Q. Doan, S. Ali, Numerical and experimental modelling of the steam-assisted gravity drainage (SAGD) process, in: *Annual Technical Meeting*, Petroleum Society of Canada, 1999.
- [78] O. R. Ayodele, T. N. Nasr, J. J. Ivory, G. Beaulieu, G. Heck, Testing and history matching of ES-SAGD (using hexane), in: *SPE Western Regional Meeting*, Society of Petroleum Engineers, 2010.
- [79] E. Aulisa, D. Gilliam, *A practical guide to geometric regulation for distributed parameter systems*, Chapman and Hall/CRC, 2015.
- [80] R. E. Kalman, Mathematical description of linear dynamical systems, *Journal of the Society for Industrial and Applied Mathematics, Series A: Control* 1 (2) (1963) 152–192.
- [81] D. G. Luenberger, *Introduction to dynamic systems: Theory, models and applications* (1979).
- [82] R. F. Curtain, H. Zwart, *An introduction to infinite-dimensional linear systems theory*, Vol. 21, Springer Science & Business Media, 2012.
- [83] S. R. Pacharu, R. Gudi, S. Patwardhan, Advanced state estimation techniques for packed bed reactors, *IFAC Proceedings Volumes* 45 (15) (2012) 519–524.

- 
- [84] C. Antoniadou, P. D. Christofides, Integrated optimal actuator/sensor placement and robust control of uncertain transport-reaction processes, *Computers & Chemical Engineering* 26 (2) (2002) 187–203.
- [85] D. Ucinski, M. A. Demetriou, An approach to the optimal scanning measurement problem using optimum experimental design, in: *Proceedings of the 2004 American Control Conference*, Vol. 2, IEEE, 2004, pp. 1616–1621.
- [86] D. G. Lainiotis, W. H. Ray, *Distributed Parameter Systems: Identification, Estimation, and Control*, Marcel Dekker, 1978.
- [87] U. Boscain, J.-P. Gauthier, F. Rossi, M. Sigalotti, Approximate controllability, exact controllability, and conical eigenvalue intersections for quantum mechanical systems, *Communications in Mathematical Physics* 333 (3) (2015) 1225–1239.
- [88] W. B. Rouse, Engineering complex systems: Implications for research in systems engineering, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 33 (2) (2003) 154–156.
- [89] B. Cai, H. Liu, M. Xie, A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks, *Mechanical Systems and Signal Processing* 80 (2016) 31–44.
- [90] C.-T. Lin, Structural controllability, *IEEE Transactions on Automatic Control* 19 (3) (1974) 201–208.
- [91] S. Diop, M. Fliess, Nonlinear observability, identifiability, and persistent trajectories, in: [1991] *Proceedings of the 30th IEEE Conference on Decision and Control*, IEEE, 1991, pp. 714–719.
- [92] Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks, *Nature* 473 (7346) (2011) 167.
- [93] Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Observability of complex systems, *Proceedings of the National Academy of Sciences* 110 (7) (2013) 2460–2465.
- [94] L. Chow, R. Butler, Numerical simulation of the steam-assisted gravity drainage process (SAGD), *Journal of Canadian Petroleum Technology* 35 (06).

- 
- [95] D. Guo, Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP), *International Journal of Geographical Information Science* 22 (7) (2008) 801–823.
- [96] Z. Galil, Efficient algorithms for finding maximum matching in graphs, *ACM Computing Surveys (CSUR)* 18 (1) (1986) 23–38.
- [97] J. E. Hopcroft, R. M. Karp, An  $n^{5/2}$  algorithm for maximum matchings in bipartite graphs, *SIAM Journal on Computing* 2 (4) (1973) 225–231.
- [98] J. Edmonds, R. M. Karp, Theoretical improvements in algorithmic efficiency for network flow problems, *Journal of the ACM (JACM)* 19 (2) (1972) 248–264.
- [99] R. Tarjan, Depth-first search and linear graph algorithms, *SIAM Journal on Computing* 1 (2) (1972) 146–160.
- [100] A. Gibbons, *Algorithmic graph theory*, Cambridge university press, 1985.
- [101] A. Ganesh, K. Lyu, R. J. Chalaturnyk, R. Gudi, V. Prasad, Graph-based structural controllability and observability of steam assisted gravity drainage pressure chamber: A data driven approach, *IFAC-PapersOnLine* 52 (1) (2019) 766–771.
- [102] T. Söderström, P. Stoica, *System identification*, Prentice-Hall, Inc., 1988.
- [103] G. Deshpande, S. LaConte, G. A. James, S. Peltier, X. Hu, Multivariate Granger causality analysis of fMRI data, *Human Brain Mapping* 30 (4) (2009) 1361–1373.
- [104] J. P. Hamilton, G. Chen, M. E. Thomason, M. E. Schwartz, I. H. Gotlib, Investigating neural primacy in major depressive disorder: multivariate Granger causality analysis of resting-state fMRI time-series data, *Molecular Psychiatry* 16 (7) (2011) 763.
- [105] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems* 2 (1-3) (1987) 37–52.
- [106] B. Moore, Principal component analysis in linear systems: Controllability, observability, and model reduction, *IEEE Transactions on Automatic Control* 26 (1) (1981) 17–32.

- 
- [107] S. Seabold, J. Perktold, Statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference, 2010.
- [108] W. H. Day, H. Edelsbrunner, Efficient algorithms for agglomerative hierarchical clustering methods, *Journal of Classification* 1 (1) (1984) 7–24.
- [109] M. Steinbach, G. Karypis, V. Kumar, A comparison of document clustering techniques, in: *KDD workshop on text mining*, Vol. 400, Boston, 2000, pp. 525–526.
- [110] C. W. Granger, Testing for causality: a personal viewpoint, *Journal of Economic Dynamics and Control* 2 (1980) 329–352.
- [111] J.-J. E. Slotine, W. Li, *Applied nonlinear control*, Vol. 199, Prentice hall Englewood Cliffs, NJ, 1991.

## .1 Appendix: Temperature data analysis

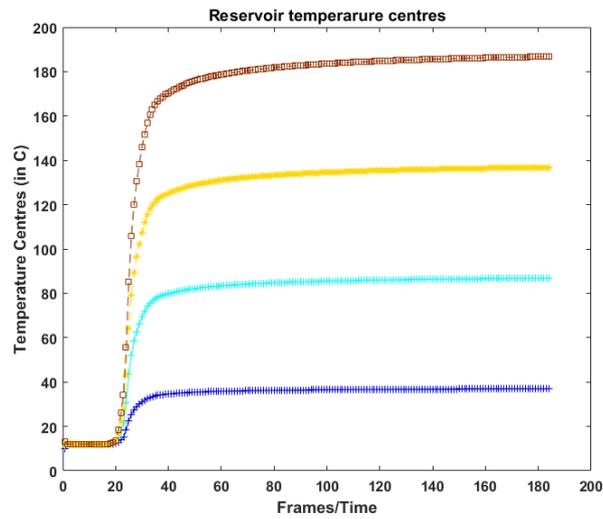


FIGURE 1: Variation of temperature centres of every cluster with frames. Color map same as the temperature cluster

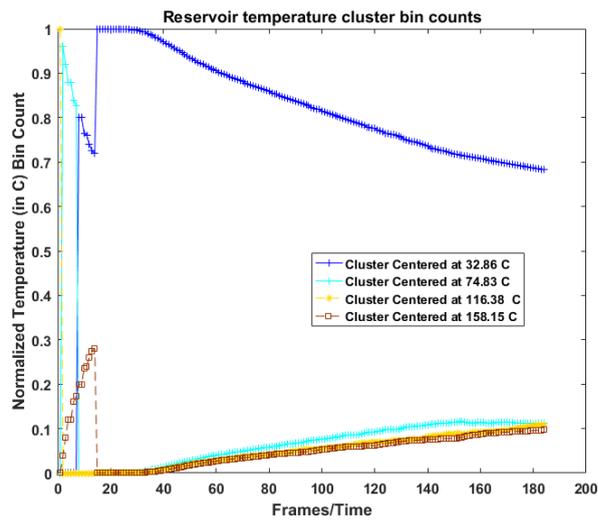


FIGURE 2: The variation of normalized bin count of each cluster vs frames. Trend shows that, with time clusters with higher pressure value overtakes the lower ones

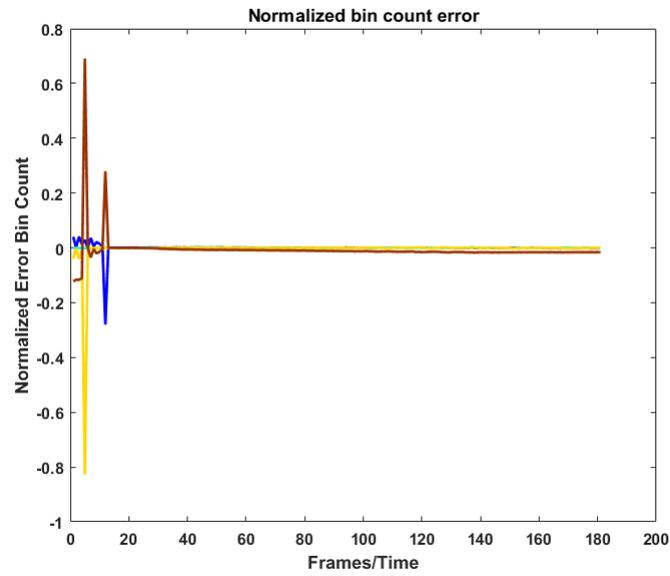
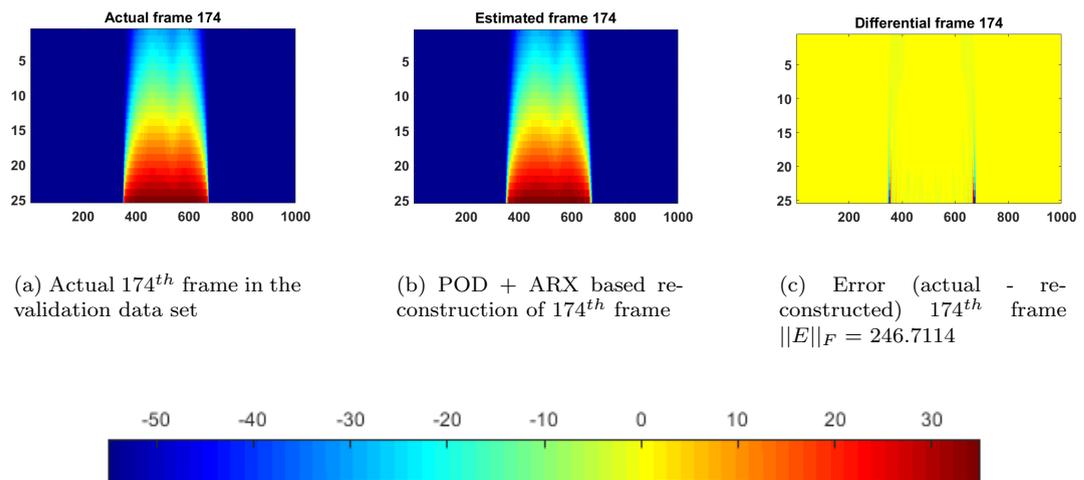


FIGURE 3: Normalized bin count error (actual - estimated) for all frames



(a) Actual  $174^{th}$  frame in the validation data set

(b) POD + ARX based reconstruction of  $174^{th}$  frame

(c) Error (actual - reconstructed)  $174^{th}$  frame  
 $\|E\|_F = 246.7114$

(d) Color map for the differential frame

FIGURE 4: POD+ARX based reconstruction visualization

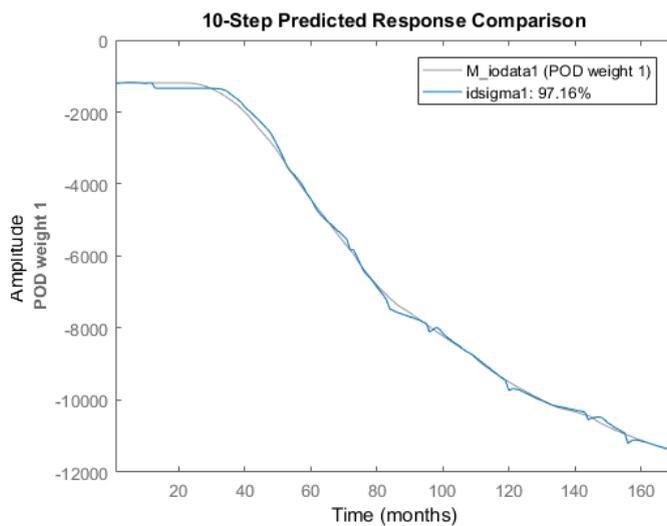


FIGURE 5: System identification on training dataset result: Actual and modeled weight  $w_1(l)$ , dotted line indicates identified model and solid line indicates the actual data.

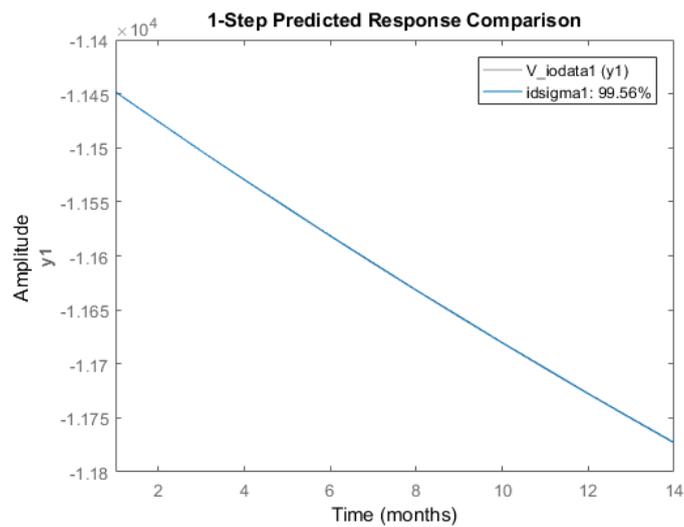


FIGURE 6: System identification on validation dataset result: Actual and modeled weight  $w_1(l)$ , dotted line indicates identified model and solid line indicates actual.