



HAL
open science

Reduced Order Modelling and Uncertainty Propagation Applied to Water Distribution Networks

Mathias Braun

► **To cite this version:**

Mathias Braun. Reduced Order Modelling and Uncertainty Propagation Applied to Water Distribution Networks. Modeling and Simulation. Université de Bordeaux, 2019. English. NNT: 2019BORD0050 . tel-02278297

HAL Id: tel-02278297

<https://theses.hal.science/tel-02278297>

Submitted on 4 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET
D'INFORMATIQUE

par **Mathias BRAUN**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES ET CALCUL
SCIENTIFIQUE

**Modélisation réduite et propagation d'incertitudes
pour les réseaux d'alimentation en eau potable.**

Date de soutenance : 04 Aril 2019

Devant la commission d'examen composée de :

Angelo IOLLO	Prof., Université de Bordeaux	Directeur
Iraj MORTAZAVI	Prof., CNAM Paris	Co-Directeur
Olivier PILLER	Chargé de recherche, IRSTEA	Co-encadrant
Irène VIGNON-CLEMENTEL	Directrice de recherche, INRIA	Rapporteur
Zoran KAPELAN	Prof., TU Delft	Rapporteur
Olivier LE MAÎTRE	Directeur de recherche, CNRS	Président
Pierre-Olivier MALATERRE	Ingénieur X-IPEF, IRSTEA	Examineur
Jochen DEUERLEIN	Chercheur principal, 3S Consult GmbH	Examineur

Titre Modélisation réduite et propagation d’incertitudes pour les réseaux d’alimentation en eau potable.

Résumé Les réseaux de distribution d’eau consistent en de grandes infrastructures réparties dans l’espace qui assurent la distribution d’eau potable en quantité et en qualité suffisantes. Les modèles mathématiques de ces systèmes sont caractérisés par un grand nombre de variables d’état et de paramètres dont la plupart sont incertains. Les temps de calcul peuvent s’avérer conséquents pour les réseaux de taille importante et la propagation d’incertitude par des méthodes de Monte Carlo. Par conséquent, les deux principaux objectifs de cette thèse sont l’étude des techniques de modélisation à ordre réduit par projection ainsi que la propagation spectrale des incertitudes des paramètres.

La thèse donne tout d’abord un aperçu des méthodes mathématiques utilisées. Ensuite, les équations permanentes des réseaux hydrauliques sont présentées et une nouvelle méthode de calcul des sensibilités est dérivée sur la base de la méthode adjointe. Les objectifs spécifiques du développement de modèles d’ordre réduit sont l’application de méthodes basées sur la projection, le développement de stratégies d’échantillonnage adaptatives plus efficaces et l’utilisation de méthodes d’hyper-réduction pour l’évaluation rapide des termes résiduels non linéaires. Pour la propagation des incertitudes, des méthodes spectrales sont introduites dans le modèle hydraulique et un modèle hydraulique intrusif est formulé. Dans le but d’une analyse plus efficace des incertitudes des paramètres, la propagation spectrale est ensuite évaluée sur la base du modèle réduit.

Les résultats montrent que les modèles d’ordre réduit basés sur des projections offrent un avantage considérable par rapport à l’effort de calcul. Bien que l’utilisation de l’échantillonnage adaptatif permette une utilisation plus efficace des états système pré-calculés, l’utilisation de méthodes d’hyper-réduction n’a pas permis d’améliorer la charge de calcul. La propagation des incertitudes des paramètres sur la base des méthodes spectrales est comparable aux simulations de Monte Carlo en termes de précision, tout en réduisant considérablement l’effort de calcul.

Mots-clés Modèles d’ingénierie, Réseaux de distribution d’eau, Modèle d’ordre réduit, Méthodes basées sur la projection, Quantification d’incertitude, Développement de chaos polynomial.

Title Reduced Order Modelling and Uncertainty Propagation Applied to Water Distribution Networks

Abstract Water distribution systems are large, spatially distributed infrastructures that ensure the distribution of potable water of sufficient quantity and quality. Mathematical models of these systems are characterized by a large number of state variables and parameter. Two major challenges are given by the time constraints for the solution and the uncertain character of the model parameters. The main objectives of this thesis are thus the investigation of projection based reduced order modelling techniques for the time efficient solution of the hydraulic system as well as the spectral propagation of parameter uncertainties for the improved quantification of uncertainties.

The thesis gives an overview of the mathematical methods that are being used. This is followed by the definition and discussion of the hydraulic network model, for which a new method for the derivation of the sensitivities is presented based on the adjoint method. The specific objectives for the development of reduced order models are the application of projection based methods, the development of more efficient adaptive sampling strategies and the use of hyper-reduction methods for the fast evaluation of non-linear residual terms. For the propagation of uncertainties spectral methods are introduced to the hydraulic model and an intrusive hydraulic model is formulated. With the objective of a more efficient analysis of the parameter uncertainties, the spectral propagation is then evaluated on the basis of the reduced model.

The results show that projection based reduced order models give a considerable benefit with respect to the computational effort. While the use of adaptive sampling resulted in a more efficient use of pre-calculated system states, the use of hyper-reduction methods could not improve the computational burden and has to be explored further. The propagation of the parameter uncertainties on the basis of the spectral methods is shown to be comparable to Monte Carlo simulations in accuracy, while significantly reducing the computational effort.

Keywords Engineering models, Water distribution networks, Reduced order model, Projection based methods, Uncertainty quantification, Polynomial chaos expansion.

Laboratoire d'accueil IRSTEA IRSTEA Centre de Bordeaux, Equipe GPIE, 50 Avenue de Verdun, 33610 Cestas-CEDEX

Laboratoire d'accueil INRIA INRIA Bordeaux - Sud-Ouest, Equipe MEMPHIS, 200 Avenue de la Vieille Tour, 33405 Talence-CEDEX

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisors Prof. Angelo Iollo from the University of Bordeaux and Prof. Iraj Mortazavi from the CNAM Paris for the continuous support of my PhD study and related research, for their patience and motivation. Their guidance helped me in all the time of research and writing of this thesis.

Special thanks goes to my supervisor at IRSTEA Dr. Olivier Piller, who played a major role in bringing me to Bordeaux by offering me the PhD position in the ResiWater project.

I would further like to thank the rest of my thesis committee: Dr. Irène Vignon-Clementel, Prof. Zoran Kapelan, Dr. Olivier Le Maître, Dr. Pierre-Olivier Malaterre and Dr. Jochen Deuerlein, for their insightful comments and encouragement, but also for the hard question which incited me to widen my research from various perspectives.

I thank my fellow office mates David Ayala-Cabrera and Camille Chambon for the stimulating discussions and for the fun times during and outside of work.

In general, I would like to thank all the colleagues of the research unit ETBX, especially the members of the group GPIE, at IRSTEA and the INRIA group MEMPHIS for the warm welcome in France and their support.

Last but not the least, I would like to thank my flat mate, my family and friends for supporting me spiritually throughout creation of this thesis.

The thesis was created in the context of the French-German collaborative research project ResiWater that was funded by the French National Research Agency (ANR; project: ANR-14-PICS-0003) and the German Federal Ministry of Education and Research (BMBF; project: BMBF-13N13690).

"In mathematics you don't understand things. You just get used to them."

– Johann von Neumann

Acronyms

CDF	Cumulative Distribution Function
CFD	Computational Fluid Dynamics
CLT	Central Limit Theorem
CVT	Centroidal Voronoi Tessellation
DDM	Demand Driven Modelling
DoE	Design of Experiment
EIM	Empirical Interpolation Method
FAVAD	Fixed and Variable Area Discharges
FCV	Flow Control Valve
FOSM	First Order Second Moment
GIS	Geographic Information System
KDE	Kernel Density Estimation
KLT	Karhunen–Loève Theorem
LOO	Leave One Out
LLN	Law of Large Numbers
LHS	Latin Hypercube Sampling
LAR	Least Angle Regression
MAE	Mean Average Error
MCS	Monte Carlo Simulation
NISP	Non-Intrusive Spectral Projection

PCA Principle Component Analysis
PCE Polynomial Chaos Expansion
PDF Probability Density Function
PDM Pressure Driven Modelling
POD Proper Orthogonal Decomposition
POR Pressure Outflow Relationship
QoI Quantity of Interest
RBF Radial Basis Function
RMSE Root Mean Square Error
ROM Reduced Order Model
SCEM Shuffled Complex Evolution Metropolis
SVD Singular Value Decomposition
VEDIF Veolia Eau d'Île-de-France
WDN Water Distribution Network

Contents

Résumé	1
Introduction	7
1 Mathematical Basis	17
1.1 Graph Theory	17
1.2 Adjoint Method	20
1.3 Singular Value Decomposition	21
1.4 Delaunay Triangulation	23
1.5 Statistics	25
1.5.1 Random variables	25
1.5.2 Evaluation of uncertainties	26
1.6 Orthogonal Polynomials	28
1.6.1 Polynomial Projection	29
1.6.2 Quadrature Formulas	30
1.7 Conclusion	30
2 Model synthesis	31
2.1 Hydraulic Network Modelling	32
2.1.1 Head-Loss	32
2.1.2 Hydraulic Equations	35
2.1.3 Sensitivity	37
2.1.4 Alternative Formulations	39
2.2 Pressure Driven Modelling	40
2.2.1 Pressure Outflow Relationship	40
2.2.2 Pressure Driven Problem Description	41
2.3 Model Deficiencies	42
2.4 Network Models	43
2.4.1 Tree Network	44
2.4.2 Highly Aggregated Looped Network	44
2.4.3 Realistic Network Model	45
2.5 Conclusion	46

3	Reduced Order Modelling	47
3.1	Interpolation-Based Models	49
3.1.1	Response Surface Models	49
3.1.2	Radial Basis Interpolation	50
3.2	Projection Based Methods	51
3.2.1	Proper Orthogonal Decomposition	51
3.2.2	Galerkin Projection	52
3.2.3	Global and Local Reduced Basis	54
3.3	Empirical Interpolation Method	55
3.4	Model Validation	57
3.4.1	Error Definition	58
3.4.2	Validation Procedure	60
3.5	Adaptive Parameter Sampling	62
3.6	Evaluation on the medium size network	64
3.6.1	Evaluating Interpolation	65
3.6.2	Evaluating POD	66
3.6.3	Evaluating EIM	74
3.6.4	Evaluating Adaptive Sampling Strategy	78
3.7	Conclusion	85
4	Uncertainty Quantification	89
4.1	Classical Uncertainty Quantification	90
4.1.1	Perturbation Methods	90
4.1.2	Monte Carlo	91
4.2	Polynomial Chaos Expansion	93
4.2.1	Intrusive stochastic hydraulic model	96
4.2.2	Non-Intrusive Spectral Projection	97
4.2.3	Pseudo-Spectral Head-Loss Function	98
4.3	Applications	100
4.3.1	Illustrative Model	101
4.3.2	Strong Non-Linearities	106
4.3.3	High Dimensional Parameter-Space	109
4.4	Conclusion	115
5	Efficient Uncertainty Propagation	119
5.1	Realistic Network Model	119
5.1.1	Scenario	120
5.1.2	Results	120
5.1.3	Discussion	122
5.1.4	Conclusion	123
	Conclusion	125

List of Figures

1	Sources for parameter uncertainties in hydraulic water distribution network models.	3
1.1	Example digraph D	18
1.2	Forest contained in example graph D . Forest is displayed by full elements. The root nodes connect the trees to the opaque grid elements.	19
1.3	Illustration of the grid contained in example digraph D . The grid is displayed by full elements. Parts that are not included in the grid are opaque.	20
1.4	Validity of a Delaunay triangulation based on the empty circle property.	24
1.5	Duality of Delaunay triangulation and Voronoi diagram for a given point set.	24
2.1	Moody diagram. [Beck et Collins, 2008]	34
2.2	Tree shaped water distribution network.	44
2.3	Highly aggregated realistic water distribution network.	45
2.4	Realistic water distribution network from the Paris region, consists of a total number of 2,175 pipes, 1,822 nodes and one reservoir (indicated by the blue square).	46
3.1	Approximation error convergence for 9 reduced order models as a function of the validation set size. The surrogates are based on three realizations for the Cartesian sampling of the snapshot matrix $n_{samples} = [121, 441, 1681]$ and with three different degrees of freedom $n_{modes} = [10, 40, 120]$	61
3.2	Average RMSE in the validation set as a function of the snapshot database size using linear, bi-linear, bi-cubic and RBF interpolation with a cubic element.	65
3.3	Normalized singular values for Cartesian samplings of the parameter-space \mathcal{P} with 121, 441, 1681 and 6561 high-fidelity simulations.	67

3.4	Average root mean square error for the approximation error and the projection error in the validation points as a function of the degrees of freedom in the global reduced order model and for different Cartesian sampled snapshot databases.	68
3.5	Average relative number of iterations for the solution of the global reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.	69
3.6	Average relative number of iterations for the solution of the global reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.	69
3.7	Average relative number of iterations for the solution of the global reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.	70
3.8	Average root mean square error for the reduced order model and the projection of the high-fidelity solution onto the reduced basis in the validation points as a function of the degrees of freedom in the local reduced order model and for different Cartesian sampled snapshot databases.	71
3.9	Average relative evaluation time for the solution of the local reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.	71
3.10	Average relative number of function evaluations for the solution of the local reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.	72
3.11	Average relative number of iterations for the solution of the local reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.	72
3.12	Average root mean square error for the reduced order model and the projection of the high-fidelity solution onto the reduced basis in the validation points as a function of the degrees of freedom in the local reduced order model and for different Cartesian sampled snapshot databases.	73
3.13	Normalized singular values for the reduced basis built on the 441, 1681 and 6561 snapshot database in the head-loss domain.	75
3.14	Projection and empirical interpolation of the head-loss function in the converged solution of the hydraulic system.	76

LIST OF FIGURES

3.15	Projection of the head-loss function onto the reduced basis in the solver steps of the hydraulic system equations.	76
3.16	Ordered normalized singular values for the sampling of the extended snapshot database that includes the solver trajectory. . .	77
3.17	Projection of the head-loss function in the solver steps, using an extended snapshot matrix.	78
3.18	Delaunay starting triangulations for Cartesian grid discretizations of the parameter-space \mathcal{P}	79
3.19	Average leave-one-out error development as a function of the snapshot database sample size for the adaptive sampling strategy with a global 80 mode and a global 120 mode reduced basis.	80
3.20	Triangulation of the global adaptive sampling strategy starting from a 121 element Cartesian grid.	80
3.21	Average RMSE approximation and projection error on the validation set as a function of the size of the reduced model for Cartesian sampling and the 80 mode and 120 mode adaptive sampling strategy.	81
3.22	Average leave-one-out error development as a function of the snapshot database sample size for the adaptive sampling strategy with a local 10 mode reduced basis.	82
3.23	Triangulation of adaptive sampling strategy with a local 10 mode reduced basis starting from a 121 element Cartesian grid.	82
3.24	Average root mean square error for the reduced order model in the validation points as a function of the degrees of freedom, calculated on a Cartesian and adaptive snapshot database using a 80 mode reduced order model.	83
3.25	Average root mean square error for the reduced order model in the validation points as a function of the degrees of freedom, calculated on a Cartesian and adaptive snapshot database using a 120 mode reduced order model.	83
3.26	Average root mean square error for a global and a local reduced order model in the validation points as a function of the degrees of freedom, calculated on a local adaptive snapshot database using a 80 mode reduced order model and compared to a Cartesian sampling.	84
4.1	First five Legendre polynomials as a function of the uniform germ distribution $Z \sim \mathcal{U}(-1, 1)$	94
4.2	First five Hermite polynomials as a function of the normal germ distribution $X \sim \mathcal{N}(0, 1)$	95
4.3	Pseudo-spectral approximation of the absolute value function of a flow variable distributed as $q \sim \mathcal{N}(0, 1)$ for different orders P of Hermite polynomials.	99

4.4	Pseudo-spectral approximation of the quadratic term in the head-loss function of a flow variable distributed as $q \sim \mathcal{N}(0, 1)$ for different orders P of Hermite polynomials.	99
4.5	Pseudo-spectral approximation of the absolute value function of a flow variable distributed as $q \sim \mathcal{N}(1, 0.5)$ for different orders P of Hermite polynomials.	100
4.6	Pseudo-spectral approximation of the quadratic term in the head-loss function of a flow variable distributed as $q \sim \mathcal{N}(1, 0.5)$ for different orders P of Hermite polynomials.	100
4.7	Parameter uncertainties for demand $Z \sim \mathcal{N}(1, 0.3)$	101
4.8	Estimated probability density function for the flow rate through link 5 using FOSM, MCS and a 1 st -order PCE.	102
4.9	Convergence of mean estimated flow rate through link 5 using MCS and a 1 st -order PCE.	103
4.10	Convergence of estimated standard deviation for the flow rate through link 5 using MCS and a 1 st -order PCE.	103
4.11	Estimated probability density function for the head at node 5 using FOSM, MCS and a 8 th -order PCE.	104
4.12	Convergence of estimated mean value for the head at node 5 using MCS and a 8 th -order PCE	104
4.13	Convergence of estimated standard deviation for the head at node 5 using MCS and a 8 th -order PCE.	105
4.14	Estimated probability density for the flow rate through the pipe 2 in the small network loop, using Monte Carlo simulation and a 12 th order PCE on the DDM and PDM approach.	107
4.15	Estimated probability density for the head at node 7 in the network loop of the small model, using Monte Carlo simulation and a 12 th order PCE on the DDM and PDM approach.	108
4.16	Clusters for parameter uncertainties in the medium size network.	110
4.17	Exemplary parameter uncertainties for demand multiplier $Z_{1,\dots,6} \sim \mathcal{N}(2, 0.6)$ and roughness multiplier $Z_{7,\dots,12} \sim \mathcal{N}(2, 0.3)$	111
4.18	Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6 th -order PCE.	112
4.19	Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6 th -order PCE. Comparison of convergence in mean and variance as a function of sample size.	112
4.20	Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6 th -order PCE. Comparison of convergence in mean and variance as a function of sample size.	113

LIST OF FIGURES

4.21 Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 114

4.22 Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 114

4.23 Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 115

4.24 Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 115

4.25 Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 116

4.26 Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 116

5.1 MC and PCE estimates of the probability density function for the loop flow, using the high fidelity model. 121

5.2 MC and PCE estimates of the probability density function for the loop flow, using the reduced 10 mode model and the 1681 element adaptive sampling. 122

5.3 MC and PCE estimates of the probability density function for the loop flow, using the reduced 40 mode model and the 1681 element adaptive sampling. 123

5.4 Relative computational effort for the reduced order model MC sampling and the 12th order PCE. 124

List of Tables

2.1	Dependency of the friction coefficient on the flow regime.	33
4.1	Correspondence between the type of generalized polynomial chaos and their underlying random variables [Xiu, 2010].	94
4.2	Calculated coefficients for the polynomial chaos expansion using the intrusive Galerkin projection and the non-intrusive matrix inversion approaches for the flow rate through pipe 5 q_5 and the head at node 5 h_5	105

Résumé

Les réseaux de distribution d'eau sont des infrastructures essentielles, réparties dans l'espace, qui fournissent de l'eau potable en quantité, qualité et pression appropriées à la population des villes et des villages du monde entier. Ces systèmes sont complets en ce sens qu'ils traitent de la collecte, du traitement, du stockage, de la distribution et de l'évacuation de l'eau. La première étape consiste à recueillir l'eau brute à partir d'eaux souterraines ou de sources d'eau de surface comme les rivières et les réservoirs d'eau naturels. Au cours du traitement, les composants indésirables sont éliminés de l'eau, dans le but d'assurer une utilisation économique en tant qu'eau potable ou dans le cadre d'utilisations industrielles et médicales. Cette étape est consommatrice d'énergie et comprend de nombreuses étapes qui dépendent des sources d'eau brute disponibles et de la qualité de l'eau souhaitée. Le stockage et la distribution de l'eau traitée sont assurés par le réseau de distribution d'eau. Ses éléments clés sont le réseau de distribution, les stations de pompage et les réservoirs d'eau. Le réseau de distribution est un réseau de canalisations distribué dans l'espace et généralement en boucle haute qui relie les consommateurs à la station d'épuration et fournit les moyens de distribution de l'eau potable. Les stations de pompage et les réservoirs d'eau fournissent la hauteur manométrique nécessaire pour assurer un service fiable sur l'ensemble du réseau. Pour un fonctionnement efficace et économique du réseau, un certain nombre d'appareils supplémentaires de régulation de débit et de pression sont installés à des endroits stratégiques. L'étape finale ferme le cycle de l'eau, en collectant les eaux usées et en les acheminant vers une station de traitement des eaux usées.

L'exploitation pratique des réseaux de distribution d'eau est régie par de nombreuses conditions limites et exigences. Il s'agit notamment de la satisfaction des consommateurs, de la gestion de ressources naturelles limitées, de l'entretien technique et des coûts d'exploitation. L'utilisation de modèles mathématiques pour évaluer l'état hydraulique du réseau est un élément clé de la gestion d'un réseau de distribution d'eau et fournit des informations essentielles aux décideurs. Ces modèles sont formulés au moyen de systèmes d'équations non linéaires caractérisés par un grand nombre de paramètres, dont certains sont donnés par la consommation d'eau nodale, le diamètre et la rugosité des conduites.

Motivation

Deux des principaux problèmes que pose aujourd’hui le modèle de réseau d’un réseau de distribution d’eau sont l’effort de calcul nécessaire pour résoudre les équations du réseau et l’incertitude inhérente aux paramètres du modèle.

La modélisation des réseaux de distribution de villes comme Strasbourg, Berlin ou Paris conduit à la formulation de grands systèmes d’équations non linéaires. L’évaluation de ces modèles peut prendre un temps considérable allant de quelques minutes à plusieurs heures. Bien que cela ne soit pas problématique en soi, cela peut devenir prohibitif dans le contexte d’opérations plus complexes. De nombreuses applications pratiques comme la conception optimale d’un réseau, l’estimation de paramètres ou la quantification de l’incertitude nécessitent l’évaluation répétée des équations hydrauliques qui peuvent devenir prohibitives sur le plan informatique. D’autres applications comme la simulation en ligne et l’identification en ligne des sources nécessitent une vitesse d’évaluation en temps réel ou presque, ce qui peut ne pas être réalisable avec un modèle haute fidélité. La modélisation à ordre réduit est une méthode conçue pour alléger la charge de calcul des modèles de simulation numérique. L’idée fondamentale est de construire un modèle de substitution avec un nombre drastiquement réduit de degrés de liberté, qui se rapproche du vecteur d’état complet. Ce modèle de commande réduite doit répondre à trois exigences majeures: Premièrement, l’erreur de l’approximation doit être raisonnable par rapport au modèle à ordre complet. Deuxièmement, il doit conserver les caractéristiques et les propriétés du modèle d’ordre complet. Troisièmement, le temps de calcul doit être nettement inférieur à celui du modèle d’ordre complet.

La quantification de la variabilité de l’état d’un modèle hydraulique, introduite par l’incertitude de ses paramètres, est une tâche importante qui peut grandement bénéficier d’une modélisation à ordre réduit. Jusqu’à présent, l’état hydraulique d’un réseau de distribution d’eau était généralement calculé à l’aide d’un modèle déterministe. Cependant, les équations contiennent de nombreuses sources d’incertitudes, dont certaines sont données par les incertitudes des paramètres, la variabilité des paramètres ou l’incertitude structurelle qui découlent des simplifications des modèles et des imprécisions algorithmiques. Le modèle mathématique d’un système de distribution d’eau contient un grand nombre de paramètres, dont la plupart ne sont pas connus exactement. La figure 1 illustre les trois principales sources d’incertitude des paramètres : Les exigences du consommateur, la rugosité et le diamètre effectif du tube. La variabilité de ces paramètres peut être définie plus précisément par le processus stochastique sur lequel ils sont basés. Bien que la demande des consommateurs présente une incertitude aléatoire inhérente et fluctue avec une fréquence élevée, d’autres paramètres comme le diamètre ou la rugosité de la conduite subissent des changements en raison du processus de vieillissement.

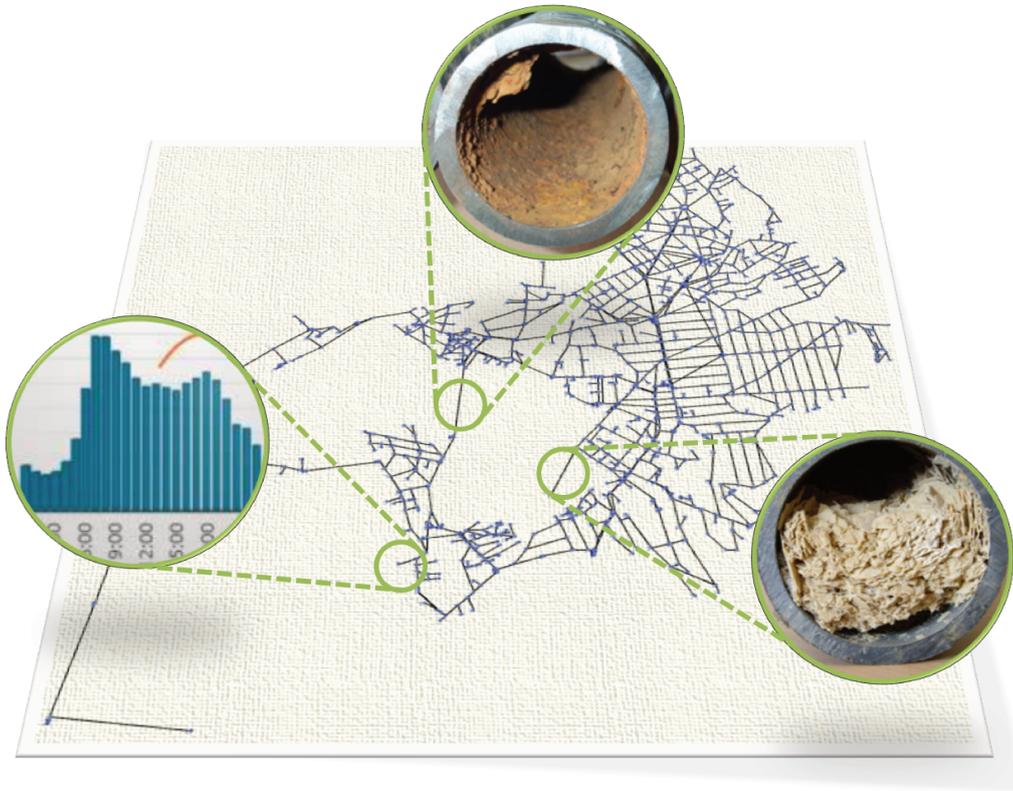


Figure 1: Sources for parameter uncertainties in hydraulic water distribution network models.

L'objectif de la quantification de l'incertitude est d'évaluer l'effet des incertitudes des paramètres sur l'état hydraulique, afin de fournir une mesure de la fiabilité. Cette évaluation est réalisée par la propagation des incertitudes des paramètres à l'aide du modèle mathématique. Les méta-modèles stochastiques sont l'un des outils les plus importants dans ce processus.

L'utilisation d'un modèle de substitution déterministe en combinaison avec un méta-modèle stochastique a le potentiel d'être un avantage énorme dans de telles applications et sera donc l'objet du travail présenté.

Objectif et grandes lignes de la thèse

Avec le grand nombre de variables d'état et par la présence d'un grand nombre de paramètres incertains dans le modèle hydraulique d'un réseau de distribution d'eau se pose deux grandes questions. Premièrement, comment les incertitudes des paramètres influencent-elles l'état du système et, deuxièmement, comment les déterminer d'une manière efficace sur le plan du calcul.

Pour répondre à ces questions, l'objectif principal de cette thèse est le développement et l'application d'outils pour la propagation efficace et la quan-

tification des incertitudes dans le modèle hydraulique des réseaux de distribution d'eau. Ce faisant, l'utilisation de méta-modèles déterministes basés sur la décomposition orthogonale appropriée et de méta-modèles stochastiques basés sur l'expansion du chaos polynomial sont étudiés et évalués pour leur capacité à réduire le coût de calcul. Les objectifs spécifiques du méta-modèle déterministe sont:

- L'application et la comparaison des techniques d'interpolation et de projection basées sur la modélisation à ordre réduit,
- Élaboration d'une stratégie d'échantillonnage adaptative pour la construction à base réduite,
- Évaluation des méthodes d'hypperréduction pour les termes résiduels non linéaires,

et pour la propagation de l'incertitude à l'aide du méta-modèle stochastique, les objectifs spécifiques sont :

- L'application et la comparaison de FOSM, des simulations de Monte Carlo et des méthodes spectrales au modèle hydraulique,
- Élément Analyse des sensibilités hydrauliques basée sur la méthode adjointe,
- Le développement d'un modèle stochastique intrusif.

La structure de la thèse est la suivante :

Chapitre 1 : Base mathématique

Le chapitre présente les outils mathématiques de base qui sont utilisés tout au long de la thèse dans l'ordre thématique. Tout d'abord, une brève introduction est donnée à la théorie des graphes. Ceci est pertinent dans le processus de modélisation hydraulique des réseaux de distribution d'eau, la réduction hiérarchique de ces modèles et la triangulation Delaunay. La méthode adjointe est présentée comme un outil pratique pour la détermination des dérivées du système au chapitre 2. L'UDS est présentée comme l'une des méthodes permettant d'obtenir une base orthogonale appropriée pour les méthodes de réduction du modèle de projection utilisées au chapitre 3 et la triangulation de Delaunay constitue la pierre angulaire de la stratégie d'échantillonnage adaptatif. Enfin, pour l'utilisation dans le chapitre 4, quelques définitions de base pour les statistiques sont introduites avec le cadre des polynômes orthogonaux pour leur utilisation dans l'expansion du chaos polynomial.

Chapitre 2 : Synthèse du modèle

Ce chapitre présente le modèle général du réseau hydraulique. La section 2.1 présente le paradigme de la modélisation axée sur la demande. Il contient la présentation de la perte de charge induite par frottement, la formulation des équations hydrauliques statiques et la formulation de la matrice de sensibilité. La section 2.2 présente ensuite le paradigme de la modélisation axée sur la pression, avec l'introduction de la relation pression-débit et la formulation des équations hydrauliques modifiées. Les capacités des paradigmes de modélisation axés sur la demande et sur la pression pour la modélisation des réseaux déficients sont examinées à la section 2.3. Le chapitre se termine par la présentation des graphiques des réseaux d'exemples utilisés tout au long de la thèse.

Chapitre 3 : Modélisation à ordre réduit

Une catégorisation des techniques courantes de modélisation d'ordre réduit est présentée avec des méthodes basées sur l'interpolation, des méthodes basées sur la projection et des modèles hiérarchiques. Suit la définition d'une procédure de validation commune sur la base de l'erreur d'approximation et de l'erreur de projection dans la section 3.4. Pour l'application aux modèles de réseaux hydrauliques, les méthodes basées sur l'interpolation et la projection sont présentées plus en détail. La section 3.1 présente les modèles classiques d'interpolation de surface de réponse et les modèles d'interpolation basés sur les fonctions de base radiale. Pour les méthodes basées sur la projection, la section 3.2 se concentre sur l'introduction de la décomposition orthogonale correcte basée sur la SVD. De plus, les équations du modèle hydraulique réduit sont dérivées pour les équations de débit de tête et de débit en boucle, ainsi que la forme projetée pour l'évaluation directe de la matrice jacobienne. Enfin, il introduit l'utilisation d'une base réduite globale et locale. La section 3.3 introduit l'application de la méthode d'interpolation empirique pour l'application dans les équations hydrauliques, suivie par l'introduction d'une stratégie d'échantillonnage adaptative basée sur les erreurs de projection dans la section 3.5.

Chapitre 4 : Quantification de l'incertitude

Le chapitre traite de la propagation des incertitudes des paramètres à l'aide de modèles mathématiques. Il donne d'abord une caractérisation des approches de propagation de base avec les méthodes de perturbation, d'échantillonnage ou de collocation stochastique et les méthodes spectrales stochastiques. Dans la section 4.1 la méthode de perturbation et la simulation de Monte Carlo sont présentées plus en détail, suivies de l'introduction de l'expansion du chaos

polynomial dans la section 4.2. La section se poursuit sur la formulation des équations spectrales stochastiques pour le modèle hydraulique.

Chapitre 5 : Quantification efficace de l'incertitude

Dans ce chapitre, l'application combinée des techniques de modélisation d'ordre réduit et de propagation de l'incertitude spectrale est évaluée sur l'exemple d'un réseau de distribution d'eau réaliste avec un espace paramètre bidimensionnel.

Introduction

Water distribution systems are critical, spatially distributed infrastructures that supply potable water of proper quantity, quality and pressure to the population of cities and towns all over the world. These systems are comprehensive in a sense that they deal with the collection, treatment, storage, distribution and disposal of water. The first step raw water is collected from groundwater or surface water sources like rivers and natural water reservoirs. During treatment unwanted constituents are removed from the water, with the objective to ensure the safe use as drinking water or in the context of industrial and medical uses. This step is energy consuming and includes numerous steps that depend on the available raw water sources and the desired water quality. Storage and distribution of the treated water is provided through the water distribution system. Its key elements are the distribution network, pumping stations and water tanks. The distribution network is a spatially distributed and usually highly looped system of pipes that connects the consumers to the treatment plant and provides the means for the distribution of potable water. Pumping stations and water tanks provide the necessary pressure head to ensure the reliable service throughout the network. For an efficient and economic operation of the network a number of additional appliances for flow and pressure control are installed at strategic positions. The final step closes the water cycle, by collecting waste water and transporting it to a waste water treatment plant.

The practical operation of water distribution networks is driven by numerous boundary conditions and requirements. Some of those are the consumer satisfaction, the management of limited natural resources, technical maintenance and operational cost. The use of mathematical models to evaluate the hydraulic state of the network is a key element in managing a water distribution system and provides crucial information to decision makers. These models are formulated through systems of non-linear equations that are characterized by a big number of parameters, some of which are given by nodal water consumption, pipe diameters and pipe roughnesses.

Motivation

Two of the major problems in dealing with the network model of a water distribution system today are coping with the computational effort for the solution of the network equations and dealing with the inherent uncertainty in the model parameters.

Modelling the distribution networks of cities like Strasbourg, Berlin or Paris leads to the formulation of large systems of non-linear equations. The evaluation of such models can take a considerable amount of time ranging from several minutes to hours. While this by itself is not problematic, it can become prohibitively expensive in the context of more complex operations. Many practical applications like optimal network design, parameter estimation or uncertainty quantification require the repeated evaluation of the hydraulic equations which may become computationally prohibitive. Other applications like on-line simulation and on-line source identification require real-time or close to real-time evaluation speed which may not be achievable with a high fidelity model. Reduced order modelling is a method designed to alleviate the computational burden of numerical simulation models. The fundamental idea is to build a surrogate model with a drastically reduced number of degrees of freedom, that approximates the full state vector. Such a reduced order model has to fulfil three major requirements: First, the error of the approximation has to be reasonable, in comparison to the full order model. Second, it has to conserve the characteristics and properties of the full order model. Third, the calculation time has to be significantly lower than for the full order model.

Quantifying the variability in the state of a hydraulic model, introduced through uncertainties in its parameters is an important task that can greatly benefit from reduced order modelling. To date, the hydraulic state of a water distribution network is commonly calculated using a deterministic model. However, these contain numerous sources for uncertainties, some of which are given by parameter uncertainties, parameter variability or structural uncertainty that stem from model simplifications and algorithmic inaccuracies. The mathematical model of a water distribution system contains a huge number of parameters, most of which are not known exactly. Figure 1 illustrates the three most important sources for parameter uncertainties: The consumer demands, pipe roughness and the effective pipe diameter. The variability in these parameters can be defined more precisely through the stochastic process they are based on. While the consumer based demand has an inherent aleatory uncertainty and fluctuates with a high frequency, other parameters like the pipe diameter or the roughness undergo changes due to the ageing process. The objective of the uncertainty quantification is to evaluate the effect of parameter uncertainties on the hydraulic state, in order to provide a measure of reliability. This evaluation is achieved through the propagation of the parameter uncertainties using the mathematical model. Stochastic meta-models are

one of the most important tools in this process.

Employing the use of a deterministic surrogate model in combination with a stochastic meta model has the potential to be a huge benefit in such applications and will thus be the focus of the presented work.

State of the Art

The classical demand driven hydraulic model describes a looped water distribution network by one dimensional flow based on the first and second Kirchhoff laws. In this demand-driven model (DDM), nodal demands and the head potential at a selection of resource nodes are given as boundary conditions. One way to solve the resulting system of equations is given by the Hardy-Cross algorithm [Cross, 1936]. It is defined as a relaxed Newton method that solves the equations successively for each loop. With the advent of personal computers more complex solution algorithms have been applied with trust-region Newton algorithms or the global gradient algorithm introduced by Todini et Pilati [1988]. However, with the introduction of complex pressure and flow regulation devices the convergence of the hydraulic solver is not guaranteed. The use of variational methods by authors like Birkhoff [1963], Collins [1978] and Carpentier et Cohen [1993] lead to the formulation of the primal and dual models. This model is more robust in the presence of hydraulic appliances and allows for the solution through the use of sophisticated optimization methods.

Originally being developed for planning and sizing of water distribution networks the applications have since been extended to areas like sensor placement, leakage reduction, water security and on-line system management. This introduces more complex requirements to the hydraulic model. Issues, like modelling networks that lose their connectivity due to operational boundary conditions or the ill conditioned non-linear flow/tension problem under zero-flow conditions, demonstrate the need for more robust models. In dealing with zero-flow several changes have been proposed to improve demand driven modelling. Elhay et Simpson [2011] present a regularization approach that introduces small flow rates for a more robust solution of the hydraulic equations. Another possibility, presented by Elhay et al. [2014] is the application of loop method solution algorithms that are not as vulnerable to zero-flows as classical solution methods. The introduction of pressure driven modelling (PDM) provides an efficient tool for calculating the hydraulic state in disconnected or pressure deficient networks. Based on the fact first voiced by Wagner et al. [1988] that outflow at demand nodes is not a fixed, but rather a pressure dependent boundary condition, a number of approaches have been developed. Several authors introduced functions called pressure outflow relationships (PORs) to determine the actual flow based on the available pressure. Early approaches like Bhave [1981] use these PORs in an iterative approach to solve a series of DDM problems while adjusting the demands to be compatible

with the pressure. [Piller et Van Zyl \[2007\]](#) presented a mathematical formulation of the pressure-driven model that does not rely on the definition of any head flow relationships. Instead, the authors use modified mass-balance constraints at consumption nodes to allow reduced demands in case the pressure is insufficient [Elhay et al. \[2015\]](#).

Reduced order modelling in water distribution networks can take a number of different forms. Depending on the level of information that is required for a particular application water distribution networks or parts thereof can be modelled by Computational Fluid Dynamics (CFD) using the Navier-Stokes equations. This is often the case in water quality problems, especially in questions of mixing behavior at pipe junctions [[Gilbert et al., 2017](#); [van Bloemen Waanders et al., 2005](#)]. However, the evaluation of a complete network in this level of detail is infeasible. For this reason in most of the applications a simplified physical model is used that employs a one-dimensional flow model along discrete pipe elements and can be described using the mass and energy balance of the Kirchhoff laws [Cross \[1936\]](#); [Walski et al. \[2001\]](#).

The need for more efficient evaluation of the hydraulic equations has inspired numerous publications with the objective to decrease the computational effort or increase the computational performance. These methods can generally be described as conservative or non-conservative, depending on the fact if information is lost in the process. Some of the most promising and used approaches are the network skeletonization, the development of non-linear algebraic surrogate models or graph decomposition. Network skeletonization is the process of selecting only parts of a network for inclusion in the reduced model that have significant effect on the behavior of the system [[Walski et al., 2003](#)]. This process was introduced by [Eggener et Polkowski \[1976\]](#) who tested the sensitivity of model results to the systematic removal of pipes and found that, under normal conditions, removing large numbers of pipes did not significantly effect the pressure. The investigation of the rules for model skeletonization has been published by [Hamberg et Shamir \[1988a,b\]](#). In general this method does not exclude the data, but agglomerates the data from deleted elements in the reduced model. The construction of non-linear algebraic approximation models linearises the hydraulic model in a defined working state. For this state, the known non-linear relations are then rebuild through the use of algebraic equations [[Ulanicki et al., 1996](#)]. Graph decomposition methods use the intrinsic properties of the network graph in order to divide linear and non-linear components in the system equations for a more efficient solution [[Deuerlein, 2008](#)]. This can be achieved through the decomposition of the network into the forest graph governed by a linear set of equations and the non-linear grid. Further improvements can be made through the formulation of super nodes, that reduce the graph complexity [[Deuerlein et al., 2012b](#)]. While the skeletonization and the use of non-linear algebraic models allow for the loss of a certain degree of information, the graph decomposition permits the complete

reconstruction of the full order state vector.

The use of reduced order models is a well studied topic in the context of CFD [Benner *et al.*, 2013]. Projection based modelling techniques that are based on the singular value decomposition (SVD) have found application in many problems, with one of the most popular representatives being the Proper Orthogonal Decomposition (POD) [Antoulas, 2005]. Building a reduced order approximation of a parametric system, uses a database of pre-sampled snapshots that gives an appropriate representation of the system behaviour. This snapshot database is then used to formulate a set of orthogonal basis vectors. Projecting the high-fidelity model on this basis allows to formulate a reduced order model. Common methods are the Galerkin and the Petrov-Galerkin projection. Especially the preservation of physical model properties due to the projection of the full order system gives the POD an advantage over purely data-driven methods [Carlberg *et al.*, 2015]. As projection based methods are strongly influenced by the quality of the snapshot database, the use of efficient sampling techniques can greatly influence the accuracy of the resulting model. Popular approaches use a greedy algorithm for the selection of efficient parameter samples [Haasdonk et Ohlberger, 2008] or residual based centroidal voronoi tessellation (CVT) [Lombardi *et al.*, 2009]. The further use of hyper reduction techniques, like the empirical interpolation method (EIM) introduced by Chaturantabut et Sorensen [2010], greatly improve the computational efficiency of projection based reduced order models in a non-linear context.

Analysing uncertainties is generally divided into the quantification and the propagation of parameter uncertainties. The inherent uncertainties in water distribution systems are well documented and numerous approaches have been undertaken in order to find an appropriate description. The underlying aleatory variability of the nodal demand on the scale of minutes, hours and days or even on monthly and annual time-scales is discussed by Herrera *et al.* [2010] and Buchberger et Wells [1996]. Further, a number of modelling assumptions introduce epistemic errors to the demand. This includes the nodal agglomeration of demands that occur distributed along a pipe [Walski *et al.*, 2003] or due to simplification of the network graph [Perelman *et al.*, 2008] and unrealistic demands due to model deficiencies [Braun *et al.*, 2017]. Pipe diameter and roughness are influenced by corrosive processes and will change over time Boulos et Karney [2004]. For fast changing parameters like the nodal demand there exist models that try to predict the value for the next time step. This may be achieved either by data driven models Herrera *et al.* [2010]; Braun *et al.* [2014] or physical models that simulate the stochastic nature of consumers Blokker et Van der Schee [2006]. An efficient way to reduce the parameter uncertainties is through the application of calibration methods. The goal is, to try to identify a set of parameters that give the best fit between measured and simulated data. The majority of calibration algorithms are based on optimization and least-squares approaches Savic *et al.* [2009]; Piller *et al.*

[2010].

The most popular approach for deriving information on the probability distribution of the parameters uses the Gauss-linear First Order Second Moment (FOSM) method, which is limited to linear systems with Gaussian error or in the case of non-linear systems to small errors. An alternative approach that takes full account of the model non-linearities is presented by [Kapelan *et al.* \[2007\]](#). They use a shuffled complex evolution metropolis (SCEM-UA) algorithm in order to solve the inverse problem by the means of a Bayesian approach which results in a more realistic estimate for the probability distribution. Given an accurate description of the input uncertainties, the influence on the QoIs is determined through their propagation using the hydraulic model. Popular approaches are perturbation methods like the FOSM that use the sensitivity of the network model [Razavi et Gupta \[2015\]](#); [Lu et Vesselinov \[2015\]](#) and Monte Carlo simulation [Lansey \[1997\]](#). Several applications have shown that taking parameter uncertainties into account may have significant influence on the results. For example [Pasha et Lansey \[2010\]](#) investigate the effect of parameter uncertainties on water quality using the Monte Carlo method and [Perelman *et al.* \[2013\]](#) investigate the influence of uncertainties in the context of least-cost design for water distribution networks. While perturbation methods are computationally very efficient, the FOSM propagates errors only linearly. This may be sufficient for models with small non-linear effects or uncertainties with a small variances. For more general cases the Monte Carlo simulation is commonly used. The interest of spectral propagation methods lays in the improved convergence behaviour compared to the Monte Carlo approach, while preserving the non-linear system behaviour.

Objective and Outline of the Thesis

With the large number of state variable and through the presence of the big number of uncertain parameters in the hydraulic model of a water distribution network poses two major questions. First, how do the parameter uncertainties influence the system state and second, how to determine it in an computationally efficient way.

To answer these questions, the main objective of this thesis is the development and application of tools for the efficient propagation and quantification of uncertainties in the hydraulic model of water distribution networks. In doing so, the use of deterministic meta-modelling approaches based on the Proper Orthogonal Decomposition and stochastic meta-modelling approaches based on the Polynomial Chaos Expansion are investigated and evaluated for their capacity to reduce the computational cost. The specific objectives for the deterministic meta-model they are:

- The application and comparison of interpolation and projection based reduced order modelling techniques,

- Development of an adaptive sampling strategy for reduced basis construction,
- Evaluation of hyper-reduction methods for non-linear residual terms,

and for the uncertainty propagation using the stochastic meta-model the specific aims are:

- The application and comparison of FOSM, Monte Carlo simulations and spectral methods to the hydraulic model,
- Analysis of hydraulic sensitivities based on the adjoint method,
- The development of an intrusive stochastic model.

The structure of the thesis is as follows:

Chapter 1: Mathematical Basis

The chapter introduces the basic mathematical tools that are used throughout the thesis in topical order. First, a short introduction is given for graph theory. This is relevant in the hydraulic modelling process for water distribution networks, the hierarchical reduction of these models and the Delaunay triangulation. The adjoint method is presented as a convenient tool for the determination of the system derivatives in Chapter 2. The SVD is introduced as one possible method to obtain a suitable orthogonal basis for the projection based model reduction methods used in Chapter 3 and the Delaunay triangulation builds the corner stone of the adaptive sampling strategy. Finally, for the use in Chapter 4 some basic definitions for statistics are introduced together with the framework of orthogonal polynomials for their use in the polynomial chaos expansion.

Chapter 2: Model Synthesis

The chapter introduces the general hydraulic network model. Section 2.1 introduces the demand-driven modelling paradigm. This contains the presentation of the friction induced head-loss, the formulation of the static hydraulic equations and the formulation of the sensitivity matrix. This is followed by the presentation of the pressure-driven modelling paradigm in Section 2.2, with the introduction of the pressure-outflow relationship and the formulation of the altered hydraulic equations. The capabilities of both the demand-driven and pressure-driven modelling paradigms for modelling deficient networks are discussed in Section 2.3. The chapter closes with the presentation of the graphs for the example networks that are used throughout the thesis.

Chapter 3: Reduced Order Modelling

A categorization of common reduced order modelling techniques is presented with interpolation based methods, projection based methods and hierarchical models. This is followed by the definition of a common validation procedure on the basis of the approximation error and the projection error in Section 3.4. For the application to the hydraulic network models, the interpolation based and projection based methods are introduced in further detail. Section 3.1 introduces classical response surface interpolation models and interpolation models based on the radial basis functions. For the projection based methods Section 3.2 concentrates on the introduction of the SVD based proper orthogonal decomposition. Further, the reduced hydraulic model equations are derived for the head-flow and the loop-flow equations, together with the projected form for the direct evaluation of the Jacobian matrix. Finally it introduces the use of global and local reduced basis. Section 3.3 introduces the application of the empirical interpolation method for the application in the hydraulic equations, followed by the introduction of an adaptive projection error based sampling strategy in Section 3.5.

Chapter 4: Uncertainty Quantification

The chapter deals with the propagation of parameter uncertainties using mathematical models. It first gives a characterization of the basic propagation approaches with the perturbation methods, sampling or stochastic collocation methods and the stochastic spectral methods. In Section 4.1 the perturbation method and the Monte Carlo simulation are presented in more detail followed by the introduction of the polynomial chaos expansion in Section 4.2. The section continues to discuss the formulation of the stochastic spectral equations for the hydraulic model.

Chapter 5: Efficient Uncertainty Quantification

In this chapter the combined application of reduced order modelling techniques and spectral uncertainty propagation is evaluated on the example of a realistic water distribution network with a two dimensional parameter-space.

Related Publications

Much of the work and results presented in this thesis has been published in scientific peer-reviewed journals. A full list of the publications is given in the following:

- Braun, M., Piller, O., Deuerlein, J. and Mortazavi, I., "Limitations of Demand- and Pressure-Driven Modelling for Large Deficient Networks," *Drinking Water Engineering and Science*, No.10(2), pp.93-98, 2017.

- Braun, M., Piller, O., Deuerlein, J., Mortazavi, I. and Iollo, A., "A Spectral Approach to Uncertainty Quantification in Water Distribution Networks," *Journal of Water Resources Planning and Management*, accepted.
- Braun, M., Piller, O., Deuerlein, J., Mortazavi, I. and Iollo, A., "Uncertainty Quantification of Water Age in Water Supply Systems by use of Spectral Propagation," *Journal of Hydroinformatics*, accepted.
- Braun, M., Piller, O., Iollo, A. and Mortazavi, I., "Use of Projection Based Reduced Order Models for Hydraulic State Estimation," *Journal of Hydroinformatics*, in preparation.

Chapter 1

Mathematical Basis

Motivation: *The presentation of the main mathematical tools that are used in the thesis.*

Principal Elements:

- *Graph theory*
- *Adjoint method*
- *Singular value decomposition*
- *Delaunay triangulation*
- *Statistics*
- *Orthogonal polynomials*

This chapter introduces the necessary basis for the mathematical methods that are used in this thesis in topical order. First, a short introduction is given for graph theory for the application in the hydraulic modelling process of water distribution networks. Graph theory is also relevant in the context of the hierarchical model reduction and the Delaunay triangulation. The adjoint method is presented as a convenient tool for the determination of a systems sensitivity to parameters. The SVD is introduced as one possible method to obtain a suitable orthogonal basis for the projection based model reduction methods and the Delaunay triangulation is used for the adaptive sampling strategy. The chapter closes with some basic definitions for statistics and the framework of orthogonal polynomials for their use in the polynomial chaos expansion.

1.1 Graph Theory

Graph theory is a field of discrete mathematics that was first introduced by Leonard Euler in 1736 during his work on the Seven Bridges of Königsberg problem. Since then it has found use in numerous scientific applications, en-

compassing computer science, physics, mathematics and engineering among others.

A *graph* G is made up of a finite set of *vertices* or *nodes* $V = \{v_i\}$, $i \in 1, \dots, n_j$ and a set of *edges* or *links* $E = \{e_j\}$, $j \in 1, \dots, n_p$. Each edge is a pair of unordered vertices $e_j = (v_i, v_k)$ that is determined by a mapping $H : E \mapsto V \times V$. The graph is then defined by the triple $G = (V, E, H)$. In contrast to an undirected graph G a directed graph or *digraph* D as the one shown in Figure 1.1 defines an edge as an ordered pair with a starting and an ending vertex.

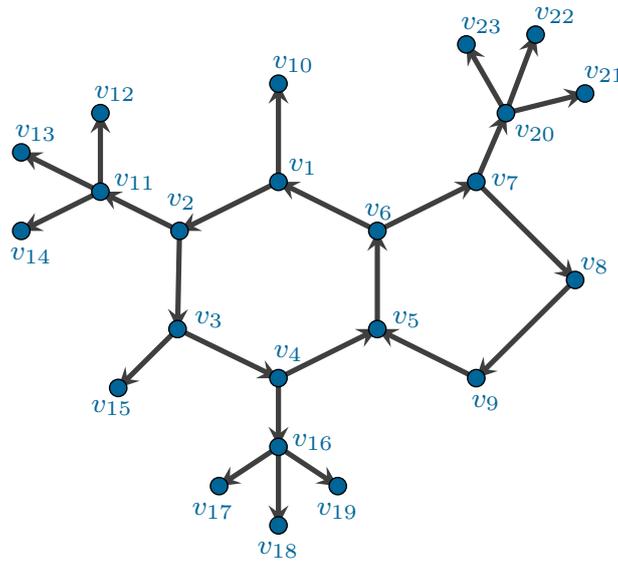


Figure 1.1: Example digraph D .

For a given digraph $D = (V, E, H)$ one way to describe its structure is defined by the *incidence matrix*. The incidence matrix $\mathbf{A} \in \mathcal{M}^{n_j \times n_p}$ defines the relation between vertices and edges, where its elements are defined as

$$A_{i,j} = \begin{cases} -1 & , \text{ if edge } j \text{ ends at vertex } i \\ 0 & , \text{ if edge } j \text{ is not connected to vertex } i \\ +1 & , \text{ if edge } j \text{ starts vertex } i. \end{cases}$$

for $i = 1, 2, \dots, n_j$ and $j = 1, 2, \dots, n_p$. Due to the fact that a graph usually contains more edges than vertices the $(n_j \times n_p)$ matrix generally is not square.

Given a graph D a sub-graph D' is defined by the sub-sets $V(D')$ and $E(D')$ such that $V(D) \subseteq V(D')$ and $E(D) \subseteq E(D')$. Some links or sub-graphs contained in the graph D have a special location or structure that is linked with certain properties.

A bridge is defined as a link that connects two parts of a graph. If the link is deleted the graph is divided into two independent parts. Based on this

definition it is possible to define the substructure of a *tree*. Given be the set of edges contained in a tree $E_i^t \subseteq E$ and the set of edges not included in the tree $\bar{E} = E \setminus E_i^t$. The root node of the tree is defined as the intersection $R_i = V_i^t \cap \bar{V}$ of the sets $V_i^t \subseteq V$ and $\bar{V} \subseteq V$. If further all of the edges $e \in E_i^t$ are bridges, then the sub-graph $T_i = (V_i^t, E_i^t, H_i^t)$ is called a tree and $r_i \in V_i^t \wedge r_i \in \bar{V}$ is the root of T_i . The *spanning tree* of the digraph D is defined as the tree T_s for which the set of vertices is equal to the set of vertices in the digraph $V_i^t = V$.

The digraph D can be decomposed into two important sub-graphs: Its *forest* and the *grid*. The forest $F \subset D$ of the digraph D is defined as the union $F = (V^f, E^f, H^f) = (\cup V_i^t, \cup E_i^t, \cup H_i^t)$ of all trees $T_i, i = 1, \dots, n$. The forest of the example graph is illustrated in Figure 1.2.

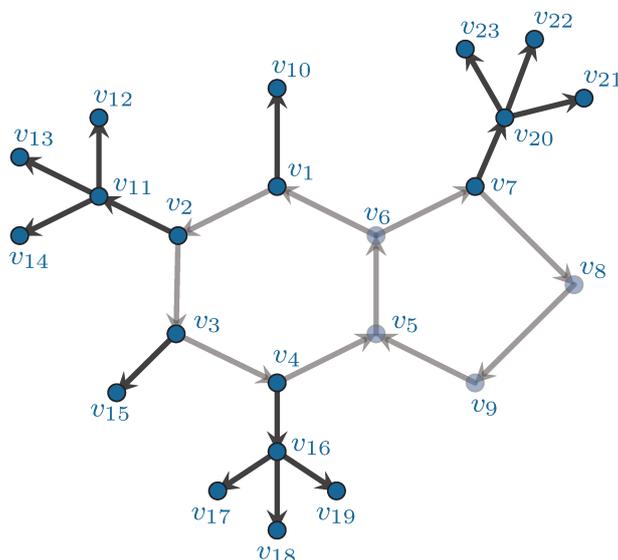


Figure 1.2: Forest contained in example graph D . Forest is displayed by full elements. The root nodes connect the trees to the opaque grid elements.

Removing the edges of the forest from the digraph D leads to the definition of the grid $G = (V^g, E^g, H^g)$ with $E^g = (E \setminus E^f)$, illustrated in Figure 1.3. The grid of the graph D is composed of a number of *cycles*. Given be the graph of the grid G and a sequence of edges $S = (e_1^g, \dots, e_n^g) \in E^g$ with $H(e_j^g) = (a_{j-1}, a_j)$ for $j = 1, \dots, n$ and $a_j \in V^g, i = 1, \dots, n$. If the starting vertex a_0 and the end node a_n are identical, the sequence S is called a circle or loop C .

The definition of the cycles introduces an alternative description of the graph D through the loop matrix $\mathbf{C} \in \mathcal{M}^{n_l \times n_p}$, where $n_l = n_p - n_j + 1$ for a connected graph. The components of \mathbf{C} are defined according to the

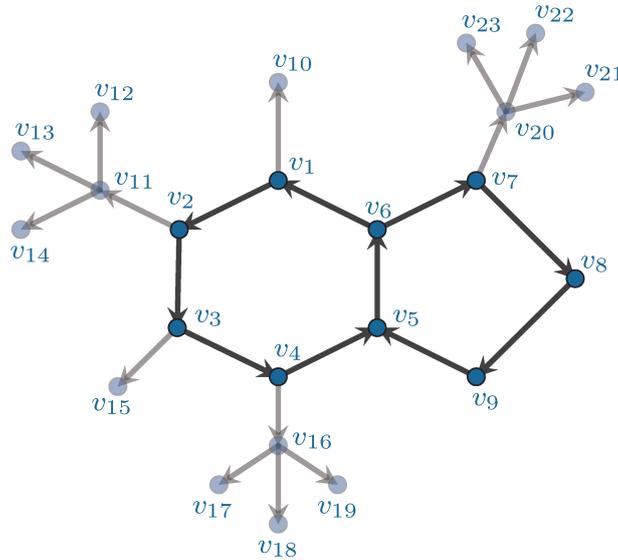


Figure 1.3: Illustration of the grid contained in example digraph D . The grid is displayed by full elements. Parts that are not included in the grid are opaque.

orientation of an edge with respect to a cycle:

$$[C_{i,j}] = \begin{cases} -1 & , \text{ if edge } j \text{ lies against the direction of cycle } i \\ 0 & , \text{ if edge } j \text{ is not part of cycle } i \\ +1 & , \text{ if edge } j \text{ lies in the direction of cycle } i. \end{cases}$$

Reordering the links of the network, allows for the alternative representation of the loop matrix as a concatenation of the spanning tree \mathbf{T}_S and the chords given as the identity \mathbf{I} .

$$\mathbf{C} = [\mathbf{T}_S | \mathbf{I}] \tag{1.1}$$

A complete and more basic definition of the theorems in graph theory can be found in [Godsil et Royle \[2013\]](#).

1.2 Adjoint Method

The adjoint method is a powerful approach for the use with parametric systems. It is frequently used for the efficient evaluation of the sensitivities or in the solution of inverse problems. Given a parametric system that is defined in a state vector $\mathbf{x} \in \mathbb{R}^{n_x}$, a parameter vector $\mathbf{p} \in \mathbb{R}^{n_p}$ and an objective vector $\mathbf{o} \in \mathbb{R}^{n_o}$. Suppose a response function is defined as $\mathbf{f}(\mathbf{x}) : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_o}$ and a residual function is given as $\mathbf{g}(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ by definition, where $\mathbf{g} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \mapsto \mathbb{R}^{n_x}$. In general the state vector \mathbf{x} is a function of the parameter vector \mathbf{p}

and the size of the parameter vector n_v is bigger than the size of the objective vector n_o . A usual objective in numerical methods is the evaluation of the derivative $d_p \mathbf{f}$.

One possible approach to the derivation of the adjoint equation is through the formulation of the *Lagrangian*

$$\mathcal{L}(\mathbf{x}, \mathbf{p}, \boldsymbol{\lambda}) \equiv \mathbf{f}(\mathbf{x}) - \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}, \mathbf{p}), \quad (1.2)$$

with a vector of Lagrangian multipliers $\boldsymbol{\lambda}$. The Lagrangian is equivalent to the response function \mathbf{f} for arbitrary values of $\boldsymbol{\lambda}$, as the residual function \mathbf{g} is equal to zero on the whole domain.

Considering the chain rule, the derivative of \mathbf{f} with respect to the parameter vector \mathbf{p} can be expressed as

$$\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{p}} = \frac{d\mathcal{L}}{d\mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\mathbf{p}} + \frac{d\boldsymbol{\lambda}^T}{d\mathbf{p}} \mathbf{g} + \boldsymbol{\lambda}^T \left(\frac{\partial \mathbf{g}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\mathbf{p}} + \frac{\partial \mathbf{g}}{\partial \mathbf{p}} \right). \quad (1.3)$$

As $\mathbf{g} = 0$ the second term on the right-hand side can be omitted, resulting in

$$\frac{d\mathcal{L}}{d\mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\mathbf{p}} + \boldsymbol{\lambda}^T \left(\frac{\partial \mathbf{g}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\mathbf{p}} + \frac{\partial \mathbf{g}}{\partial \mathbf{p}} \right). \quad (1.4)$$

By reordering (1.4) the derivative of \mathbf{f} with respect to the parameters can be isolated

$$\frac{d\mathcal{L}}{d\mathbf{p}} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \boldsymbol{\lambda}^T \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right) \frac{d\mathbf{x}}{d\mathbf{p}} + \boldsymbol{\lambda}^T \frac{\partial \mathbf{g}}{\partial \mathbf{p}}. \quad (1.5)$$

Choosing the Lagrangian multipliers λ in a way that

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \boldsymbol{\lambda}^T \frac{\partial \mathbf{g}}{\partial \mathbf{x}} = 0, \quad (1.6)$$

allows to express the derivative of \mathbf{f} with respect to \mathbf{p} as

$$\frac{d\mathbf{f}}{d\mathbf{p}} = \boldsymbol{\lambda}^T \frac{\partial \mathbf{g}}{\partial \mathbf{p}}, \quad (1.7)$$

which is computationally more efficient in cases where the number of parameters is lower than the degrees of freedom in the forward system [Marchuk, 2013].

1.3 Singular Value Decomposition

The Singular Value Decomposition is often cited as one of the most important matrix decompositions in the field of applied mathematics [Stewart, 2000]. In linear algebra, it is used to determine a factorization of real and complex matrices. Its uses range from applications in signal processing to the calculation of

the pseudo-inverse of a matrix and it is closely tied to the principal component analysis (PCA) and the Karhunen–Loève transform (KLT).

For a given matrix $\mathbf{A} \in \mathbb{C}^{n \times m}$, $n \leq m$ the square roots of the eigenvalues of $\mathbf{A}\mathbf{A}^T$ are given by the ordered non-negative numbers $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$. There also exist two unitary matrices $\mathbf{U} \in \mathbb{C}^{n \times n}$ and $\mathbf{V} \in \mathbb{C}^{m \times m}$ with $\mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ and $\mathbf{V}\mathbf{V}^T = \mathbf{I}_m$, where \mathbf{I}_k defines a $k \times k$ identity matrix. From this it is possible to express matrix \mathbf{A} as

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (1.8)$$

where the diagonal of the $n \times m$ matrix $\mathbf{\Sigma}$ contains the ordered singular values $\Sigma_{ii} = \sigma_i, i = 1, \dots, n$ and all other entries are equal to zero. The columns of matrix \mathbf{U} are also termed the left singular vectors, whereas the columns of \mathbf{V} are called the right singular vectors.

Properties of the SVD

The matrices \mathbf{U} , \mathbf{V} and $\mathbf{\Sigma}$ given by the SVD as stated in (1.8) can be represented in a block structure, under the condition that $\sigma_r > 0$ and $\sigma_{r+1} = 0$:

$$\mathbf{U} = [\mathbf{U}_1 \mathbf{U}_2], \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = [\mathbf{V}_1 \mathbf{V}_2], \quad (1.9)$$

where

$$\mathbf{\Sigma}_1 = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} > \mathbf{0} \quad \text{and} \quad \mathbf{\Sigma}_2 = \mathbf{0}.$$

Other than the orthogonality of the matrices \mathbf{U} and \mathbf{V} , three of the most important derived properties as mentioned by Antoulas [2005] are stated in the following. First, the matrix has the rank $(\mathbf{A}) = r$. Second, it can be expressed in the form of the dyadic decomposition:

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T, \quad (1.10)$$

using the left and right eigenvectors. And third, the Frobenius norm of the matrix \mathbf{A} is defined as:

$$\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}. \quad (1.11)$$

Based on the matrix rank and the dyadic product it is possible to formulate a short form of the matrix \mathbf{A} , given as

$$\mathbf{A} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T, \quad (1.12)$$

without the loss of information.

SVD based approximation

One of the most frequent uses for the SVD is the construction of low-rank matrix approximations. Given a matrix $\mathbf{A} \in \mathbb{C}^{n \times m}$, $\text{rank } \mathbf{A} = r \leq n \leq m$ it is possible to find a low rank approximation $\mathbf{X} \in \mathbb{C}^{n \times m}$, $\text{rank } \mathbf{X} = k < r$. One such approximation \mathbf{X}_* that is based on the SVD can be obtained through the truncation of (1.10)

$$\mathbf{X}_* = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \quad (1.13)$$

which contains the first k terms of the dyadic product. It is possible to show that for a general approximation \mathbf{X} the error \mathbf{X}_* is minimal with respect to the Frobenius norm.

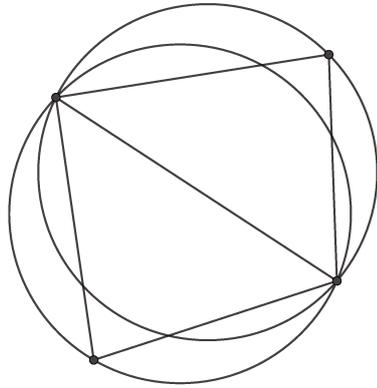
$$\|\mathbf{A} - \mathbf{X}\|_2 \leq \|\mathbf{A} - \mathbf{X}_*\|_2 \quad (1.14)$$

The sum of the normalized singular values that are not included in the construction of \mathbf{X}_* can be used as a conservative estimate for the approximation error.

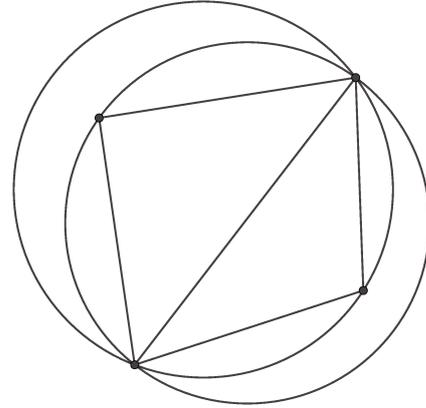
1.4 Delaunay Triangulation

Given a set of points a triangulation constructs a grid of triangles for the convex hull of the set. Let V be a set of $n \geq 3$ vertices and assume that none of these vertices are co-linear. There exists a set E of $\binom{n}{2}$ possible edges between the vertices. Two of the edges $e_i, e_j \in E$ for $e_i \neq e_j$ intersect properly, if they intersect in a point other than one of the vertices. A triangulation of the point set V is then defined as a graph $T(V, E')$, where E' is a subset of E such that none of the edges in E' intersect properly. For the point set V there exist a number of different triangulations and algorithms to construct them. Simple methods like the scan triangulation do not demand any special properties of the final graph and are random to a certain degree. In contrast to this approach, there exist triangulations that are design to optimize certain properties.

The Delaunay triangulation is defined as the maximum angle or angle optimal triangulation [Delaunay \[1934\]](#). Given the triangulation T the angle vector $\mathbf{\Gamma}(T) = (\gamma_1, \gamma_2, \dots, \gamma_{3m})$ contains the angles of all the triangles, ordered from smallest to largest. The angle vector $\mathbf{\Gamma}(T)$ is said to be bigger than $\mathbf{\Gamma}(T')$ if there exists an index i for which $\gamma_j = \gamma'_j$ for $j = 1, \dots, i - 1$ and $\gamma_j > \gamma'_j$ for $j = i, \dots, 3m$. Following this definition, the angle optimal triangulation has the biggest angle vector. This angle optimality is closely related to the *empty circle* property, which gives a geometric interpretation for the construction of a valid Delaunay triangulation. Given the four point set V in [Figure 1.4](#), the circumcircles of the triangles for the valid Delaunay triangulation in [Figure 1.4a](#) do not contain any other points, as opposed to the invalid Delaunay triangulation in [Figure 1.4b](#).



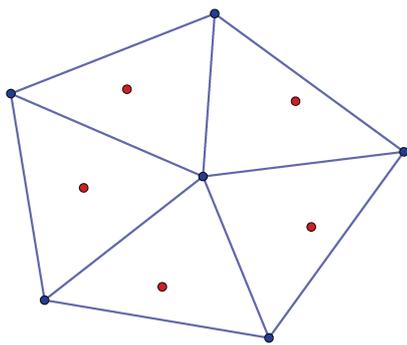
(a) Delaunay triangulation: Circle through 3 points does not contain any additional points.



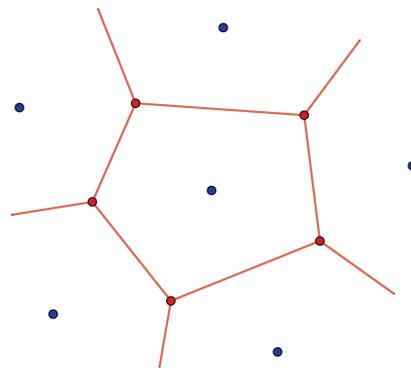
(b) Non-Delaunay triangulation: Circle through 3 contains an additional point.

Figure 1.4: Validity of a Delaunay triangulation based on the empty circle property.

The Delaunay triangulation is closely related to the centroidal voronoi tessellation through the concept of duality. This relation is visualized in Figure 1.5. Given the blue convex point set Figure 1.5a shows the Delaunay graph, where the red points indicate the points of intersection for the perpendicular bisectors of each triangle. Figure 1.5b shows the dual Voronoi diagram, where the centroids are defined by the same point set as for the Delaunay triangulation



(a) Delaunay triangulation.



(b) Voronoi diagram.

Figure 1.5: Duality of Delaunay triangulation and Voronoi diagram for a given point set.

For the calculation of a valid Delaunay triangulation there exist a number

of algorithms. One popular method is called the *Lawson flip algorithm*. Given a point set P it is performed in two basic steps.

1. Compute a general triangulation of V (for example, the scan triangulation).
2. While there exists an invalid sub-triangulation of four convex points with respect to the empty circle property demonstrated in Figure 1.4, replace the sub-triangulation by the valid one.

Given be a set $V \subseteq \mathbb{R}^2$ of n points and a general triangulation T . The iterative Lawson flip algorithm needs $\mathcal{O}(n^2)$ flips for the worst case scenario [Lawson, 1972]. Shamos et Hoey [1975] introduced an $\mathcal{O}(n \log n)$ algorithm for the creation of the Voronoi diagram on the same point set, which can be converted to the Delaunay triangulation in an additional $\mathcal{O}(n)$ steps. With the presentation of an algorithm that uses a divide-and-conquer approach Lee et Schachter [1980] introduced a direct approach for the creation of the Delaunay triangulation that runs in $\mathcal{O}(n \log n)$ time, which is asymptotically optimal.

1.5 Statistics

this section gives a short introduction to the most important concepts from statistics and probability theory that are used in the scope of this thesis.

1.5.1 Random variables

Measurement theory defines a random variable by the triple (Ω, \mathcal{F}, P) containing the sample space Ω , the σ -field or σ -algebra \mathcal{F} and the probability measure P . The sample space Ω of an experiment is defined as the set of all possible outcomes $\Omega = \{\omega\}$, where ω is one specific outcome. The σ -algebra \mathcal{F} is a subset of the sample space that contains all relevant events. In this context an event may be defined as a set of outcomes, including the empty set \emptyset and all combinations of other events in the σ -field. Probability is a concept to measure the likelihood of occurrence for a certain event $P : \mathcal{F} \rightarrow [0, 1]$. It has to satisfy the definitions $P(\emptyset) = 0$, $P(\Omega) = 1$ and if A_i and $A_j \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$, $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. On this basis a random variable $X = X(\omega)$ assigns a number to each outcome ω of a random experiment with a quantifiable probability. Based on the nature of the experiment the sample space may be defined by a discrete set of abstract outcomes like in a coin toss or as in the parameters of a water distribution network by a continuous range of values which can be used directly as the random variable.

There are a number of different ways for characterizing random variables. A common tool is the distribution as a function of the random variable. Also

known as the cumulative distribution function (CDF) is defined as $F_X \rightarrow [0, 1]$ by

$$F_X(x) = P \{ \omega \in \Omega \mid X(\omega) \leq x \} \quad (1.15)$$

and describes the probability that a realization of the random variable has a value lower than x . An illustrative derivation of the CDF is the probability density function (PDF) defined by $f_X(x)$ which describes directly the probability of a certain realization. The PDF and the CDF are linked by the integral

$$F_X(x) = \int_{-\infty}^x f_X(t) dt. \quad (1.16)$$

1.5.2 Evaluation of uncertainties

The uncertainty in a systems quantities of interest (QoI) can be expressed in a number of different ways. This includes the stochastic moments that contain important information on the probability distribution, probability intervals and the probability density function. This section introduces the definition of the stochastic moments and common approaches for the calculation and estimation of the PDF.

Stochastic moments: In many applications the probability distribution of random variable is characterized by a number of derived parameters called stochastic moments. The 1-st moment of a distribution is defined by

$$\mu_1 = \mathbb{E} [|X|] = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (1.17)$$

It is also known as the mean and gives the balance point of the distribution. With the use of the first moment it is possible to define the k -th central moment as:

$$\sigma_k = \mathbb{E} [|X - \mu_1|^k] = \int_{-\infty}^{\infty} (x - \mu_1)^k f_X(x) dx. \quad (1.18)$$

The central moments give a characterization for the shape of a distribution. For simple distributions a good characterization may be given by the mean and the second to fourth central moments also known as the variance, skewness and kurtosis.

Marginal distribution: One is often interested in the marginal density distribution of a QoI. The marginal distribution can be interpreted as a projection of the multivariate distribution on one of the output variables. This allows for a more comprehensible evaluation, however additional information like the covariance is lost in this representation. The marginal density is defined as

$$f_{X_j}(x_j) = \int_{D_{x_{\sim j}}} f_X(y_{\sim j}) dy_{\sim j} \quad (1.19)$$

with the simplified notation $y_{\sim j} = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_M)^T$. Their spectral representation of the integral may be formulated using the PCE and the marginal of a QoI in a M dimensional parameter space is given by

$$x_N(Z_j) = \int_{D_{x_{\sim j}}} \sum_{k=0}^N x_k \Psi_k(Z) dZ_{\sim j}. \quad (1.20)$$

In general this integral is evaluated using the Monte Carlo algorithm by sampling the multivariate basic random variable $Z_{\sim j}$.

Estimation of the probability density and marginal probability density function: The uncertainty propagation gives a characterization for the result random variables that allows for further evaluation and the estimation of the confidence intervals. One of the most common ways to visualize sampling data, which is generated by Monte Carlo type algorithms or from the Polynomial Chaos Expansion, is a histogram. In a histogram, the parameter domain x is divided into n equidistant sections and the density for each section is approximated by

$$\tilde{f}(x) = \frac{1}{n} \frac{\text{Number of } x_j \text{ in same section as } x}{\text{Width of section}} \quad (1.21)$$

Kernel Density Estimation: A more general approach is the kernel density estimation (KDE). It achieves a smooth and continuous approximation for the probability density function based on a chosen kernel function K

$$\tilde{f}(x) = \frac{1}{Mh} \sum_{i=1}^M K\left(\frac{x - x_i}{h}\right). \quad (1.22)$$

Here M is the sample size and $h > 0$ is a smoothing parameter. The choice of the kernel function greatly influences the final result. One of the most common examples is the KDE with a Gaussian kernel function.

Pearson distributions: A theoretic way for the reconstruction of a probability density function that does not depend on sampling is given by the Pearson distributions. The Pearson distributions are a set of five functions that, based on the first four moments of a random variable, give a direct expression for the probability density function. Although this approach works quite well for random variables that have been modelled on a one dimensional parameter space, the application to more complex distribution of a random variable modelled on a two dimensional parameter space fails to give an accurate description of the real probability distribution.

Interval Estimation: The objective of an interval estimate is to determine the values f_L and f_R that bound the location of the true value $f_L \leq f \leq f_R$. The estimate is based on a set of realizations $f = [f_1, \dots, f_M]$ of the random variable and the interval $[f_L, f_R]$ is called an interval estimator. A confidence interval is the combination of an interval estimator and a confidence coefficient

α . The confidence coefficient can be interpreted as the probability that the interval estimator contains the true value f . The $(1 - \alpha) \times 100\%$ confidence interval for $[f_L, f_R]$ is defined such that for all $f \in \mathbb{F}$,

$$P[f_L(X) \leq f] = 1 - \frac{\alpha}{2} \text{ and } P[f_R(X) \leq f] = \frac{\alpha}{2}. \quad (1.23)$$

1.6 Orthogonal Polynomials

A general polynomial of the order n is defined by $Q_n(x)$ as

$$Q_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0. \quad (1.24)$$

Systems of polynomials are defined by sets of polynomials $\{Q_n(x), n \in \mathbb{N}_0\}$, \mathbb{N}_0 being the set of non-negative integers including 0. The set is called orthogonal if the polynomials are defined in a way, such that

$$\int_I Q_n(x) Q_m(x) \rho(x) dx = \gamma_n \delta_{nm} \quad (1.25)$$

applies, using a real positive measure ρ on the support interval I . Here, δ_{nm} is defined as the Kronecker Delta

$$\delta_{nm} = \begin{cases} 0 & \text{if } n \neq m, \\ 1 & \text{if } n = m. \end{cases} \quad (1.26)$$

The normalization constant γ_n of the polynomial Q_n is defined by the integral

$$\gamma_n = \int_I Q_n(x) Q_n(x) \rho(x) dx, \quad (1.27)$$

using the measure ρ . These constants can be used to define an orthonormal set of polynomials with $\tilde{Q}_n(x) = Q_n(x) / \sqrt{\gamma_n}$. Through the definition of a weighted inner product with the continuous for

$$\langle u, v \rangle_\rho = \int_I u(x) v(x) \rho(x) dx \quad (1.28)$$

and the discrete form

$$\langle u, v \rangle_\rho = \sum_i u(x_i) v(x_i) \rho_i \quad (1.29)$$

the aforementioned orthogonality condition may be expressed as

$$\langle Q_n, Q_m \rangle_\rho = \gamma_n \delta_{nm} \quad (1.30)$$

and the normalization constant as

$$\gamma_n = \langle Q_n, Q_n \rangle_\rho = \|Q_n\|^2. \quad (1.31)$$

The components $\{Q_n(x)\}$ of a set of orthogonal polynomials can be created from the three-term recurrence relation defined in Favard's theorem [Chihara, 2011]. Favard's theorem states that, given a sequence of real numbers a_n , b_n and c_n , the subsequent orthogonal polynomial Q_{n+1} in the set is defined as

$$Q_{n+1}(x) = (a_n x + b_n) Q_n(x) - c_n Q_{n-1}(x), \quad (1.32)$$

starting with $Q_0(x) = 1$ and $Q_{-1}(x) = 0$. An attempt to order the orthogonal polynomials into an hierarchical structure was published by Askey et Wilson [1985] in the form of the Askey scheme.

1.6.1 Polynomial Projection

For a linear space of polynomials with a maximum degree of n

$$\mathbb{P}_n = \text{span} \{x^k : k = 0, 1, \dots, n\}, \quad (1.33)$$

classical approximation theory, based on the Weierstrass theorem states that, for any function $f \in C^0(I)$ defined on an interval I , there exists a polynomial $P_n(x)$ with sufficiently high degree n such that

$$|f(x) - P_n(x)| < \epsilon \quad \text{for } a \leq x \leq b, \quad (1.34)$$

for any error bound $\epsilon > 0$. More detail and the formal proof of the theorem are given by Cheney [1966]. From this theorem, the best approximation problem for a function f in the n^{th} -degree polynomial $\Phi_n(f)$ can be formulated as:

$$\lim_{n \rightarrow \infty} \|f - \Phi_n(f)\|_{L^2_\rho} = 0. \quad (1.35)$$

Polynomial projection is an efficient way to address the best approximation problem. Formulating it in a way that is not limited to bounded intervals, a positive weight function $\rho(x)$, $x \in I$ is introduced, on the weighted L^2 space

$$L^2_\rho(I) \triangleq \left\{ v : I \rightarrow \mathbb{R} \left| \int_I v^2(x) \rho(x) dx < \infty \right. \right\}. \quad (1.36)$$

On this L^2 space a weighted inner product is defined as

$$\langle u, v \rangle_{L^2_\rho(I)} = \int_I u(x) v(x) \rho(x) dx, \quad (1.37)$$

and the norm

$$\|u\|_{L^2_\rho(I)} = \left(\int_I u^2(x) \rho(x) dx \right)^{1/2}. \quad (1.38)$$

The use of orthogonal polynomials allows for an efficient evaluation.

1.6.2 Quadrature Formulas

The numerical evaluation of integrals through numerical quadrature rules is a widely used approach that is extremely efficient when dealing with orthogonal polynomials.

Given be the integral

$$I[f] \triangleq \int_I f(x) w(x) dx, \quad (1.39)$$

for the product of a function $f(x)$ and a weight function $w(x)$. A discrete integration formula with $k \geq 1$ can be defined

$$U^q \triangleq \sum_{i=1}^k f(x^{(i)}) w^{(i)}, \quad (1.40)$$

where $x^{(i)}$ are a set of points on the interval $I = [a, b]$ and $w^{(i)}$ are the according values of the weighting function for $i = 1, \dots, k$. It is possible to determine a set of parameters $\{x^{(i)}, w^{(i)}\}$ such that $U^q \approx I[f]$.

$$\int_I f(x) w(x) dx \approx \sum_{i=1}^k f(z_i^{(N)}) w_i^{(n)} \quad (1.41)$$

The approximation of the integral in (1.41) becomes an equality if the function $f(x)$ is a polynomial of the degree $2N - 1$ or higher, which is demonstrated in the formal proof by [Xiu \[2010\]](#).

1.7 Conclusion

The chapter presented the basic mathematical tools that are used throughout the thesis. They are introduced in topical order. Section 1.1 gives a short introduction on graph theory, which is relevant in the hydraulic modelling process for water distribution networks, the hierarchical reduction of these models, as well as the Delaunay triangulation. The adjoint method is introduced in Section 1.2 as a convenient tool for the determination of the system derivatives with respect to model parameters in Chapter 2. For the development of the projection based model reduction methods used in Chapter 3 the SVD is introduced as one possible method to obtain a suitable orthogonal basis. As a corner stone of the adaptive sampling strategy the Delaunay triangulation is introduced in Section 1.4. Finally, for the use in Chapter 4 some basic definitions for statistics are introduced in Section 1.5 together with the framework of orthogonal polynomials for their use in the polynomial chaos expansion in 1.6.1.

Chapter 2

Model synthesis

Motivation: *Presentation of the mathematical models that are the subject of the thesis.*

Principal Elements:

- *Demand-driven hydraulic model*
- *Derivation of sensitivities with the adjoint method*
- *Pressure-driven hydraulic model*
- *Discussion of model error sources*

Calculating the flow in hydraulic networks has a long history starting with the work presented by Cross [1936]. Today more than ever it is an important component in managing the distribution of potable water. Originally being developed for planning and sizing of water distribution networks (WDNs) the applications have since been extended to areas like sensor placement, leakage reduction, water security and on-line system management.

In Section 2.1 the classic hydraulic equations are derived for the demand driven modelling approach. This approach models the outflows at demand nodes as fixed boundary conditions. The section further demonstrates the derivation of the parameter sensitivities by use of the adjoint method and presents two important alternatives to the classical Kirchhoff formulation of the head-flow model. Section 2.2 introduces the pressure-driven modelling paradigm. Modelling deficient hydraulic networks in the presence of extreme events, like considerable technical accidents or natural disasters requires more robust modelling approach, with relaxed boundary conditions for pressure dependent demands. A discussion of the deficiencies in these models under certain circumstances follows in Section 2.3 and the chapter closes with the introduction of the three example networks in Section 2.4, that are used for the development and validation of the methods introduced in Chapters 3 and 4.

2.1 Hydraulic Network Modelling

A water distribution system is made up of a network of conduits, reservoirs and hydraulic appliances like pumps, valves and other flow and pressure regulating devices. The topology of a network is usually known to a certain degree and is described by Geographic Information Systems (GIS) in the form of a graph. The graph contains a number of n_p links or pipes, n_j free junction nodes or junctions and n_r resource nodes that encode the topological information like pipe lengths and diameters or nodal elevation. The degree of freedom for the system is given by the number of links n_p and the number of free junction nodes n_j . It follows, that the state of the hydraulic model is described in its entirety by the state vector $\mathbf{x} = [\mathbf{q}^T; \mathbf{h}^T]^T$ containing the vector of flow rates along the links \mathbf{q} and the vector of piezometric heads at the free junction nodes \mathbf{h} .

In hydraulics, the use of the piezometric head \mathbf{h} is a functional alternative of describing the nodal potential in the system. It is defined as an equivalent height of the fluid column in $h[m_{H_2O}]$

$$h = \frac{p}{\rho g} + z, \quad (2.1)$$

with the pressure $p[Pa]$, the gravitational constant $g[m/s^2]$, the density of the fluid $\rho[kg/m^3]$ and the junction elevation $z[m]$.

2.1.1 Head-Loss

As fluids flow through a pipe they experience a loss in hydraulic head due to friction. These friction losses are caused by the shear stress between the stationary pipe wall and the fluid. They depend on the conditions of the flow as well as the properties of the system.

The condition of the flow along a pipe is generally characterized by the Reynolds number Re , which is defined as

$$Re = \frac{\rho v D}{\mu} = \frac{\rho q D}{\mu A}. \quad (2.2)$$

Here, the hydraulic diameter of the pipe $D[m]$, the volumetric flow rate $q[m^3/s]$, the pipe cross-sectional area $A[m^2]$, mean velocity of the fluid $v[m/s]$ and the dynamic viscosity of the fluid $\mu[Pa \cdot s]$. The flow is considered to be laminar for $Re < 2300$, transitional for $2300 < Re < 4000$ and turbulent if $Re > 4000$. A reliable quantity for the characterization of the system properties is given by the relative roughness ϵ/D , with the average roughness parameter $\epsilon[mm]$ that depends on the material and the age of the pipe wall.

Popular methods for the evaluation of the head-loss are introduced in the following with the Darcy-Weisbach equation and the Hazen-Williams equation.

Darcy-Weisbach

The Darcy-Weisbach equation is valid for steady state flow of an incompressible fluid. The head-loss is calculated as

$$\Delta h_{DW} = \lambda \frac{L}{D} \frac{v^2}{2g}, \quad (2.3)$$

with the friction factor λ and the pipe length $L[m]$.

In general the pipe friction factor λ depends of the Reynolds number Re and the relative pipe roughness ϵ/D . Based on the flow regime its dependency on these factors changes, as shown in Table 2.1. This shows that, while laminar

Table 2.1: Dependency of the friction coefficient on the flow regime.

Flow regime	Dependency
laminar	$\lambda = f(Re)$
laminar-turbulent transition	$\lambda = f(Re, \epsilon/D)$
turbulent	$\lambda = f(\epsilon/D)$

flow is dominated by the viscous forces, for turbulent flows the wall shear stress is the main source of friction losses.

Common ways to determine the friction factor are the use of the Moody diagram or its approximation through the Colebrook equation.

Moody Diagram

The Moody diagram, introduced by [Moody \[1944\]](#), is an empirical chart that provides a method for finding the Darcy-Weisbach friction factor. As [Figure 2.1](#) shows, the friction factor can in general be expressed as a function of the Reynolds number Re and the relative pipe roughness ϵ/D . The three afore mentioned regions for laminar, transitional and complete turbulent flow can be identified in the Moody diagram.

For laminar flow rates the friction coefficient can be evaluated according to the Hagen-Poiseuille equation

$$\lambda = \frac{64}{Re}. \quad (2.4)$$

According to [Trüeb \[1961\]](#) the transition from laminar to turbulent flow occurs above Reynolds number of $Re = 2320$, completely developed turbulent flows are usually reached at Reynolds number between $Re = 6000$ and $Re = 10000$ or even more for hydraulically smooth pipes.

Colebrook

In an attempt to find a closed formulation for the friction coefficient, [Colebrook et al. \[1939\]](#) combined the previously known definitions of the friction coefficient

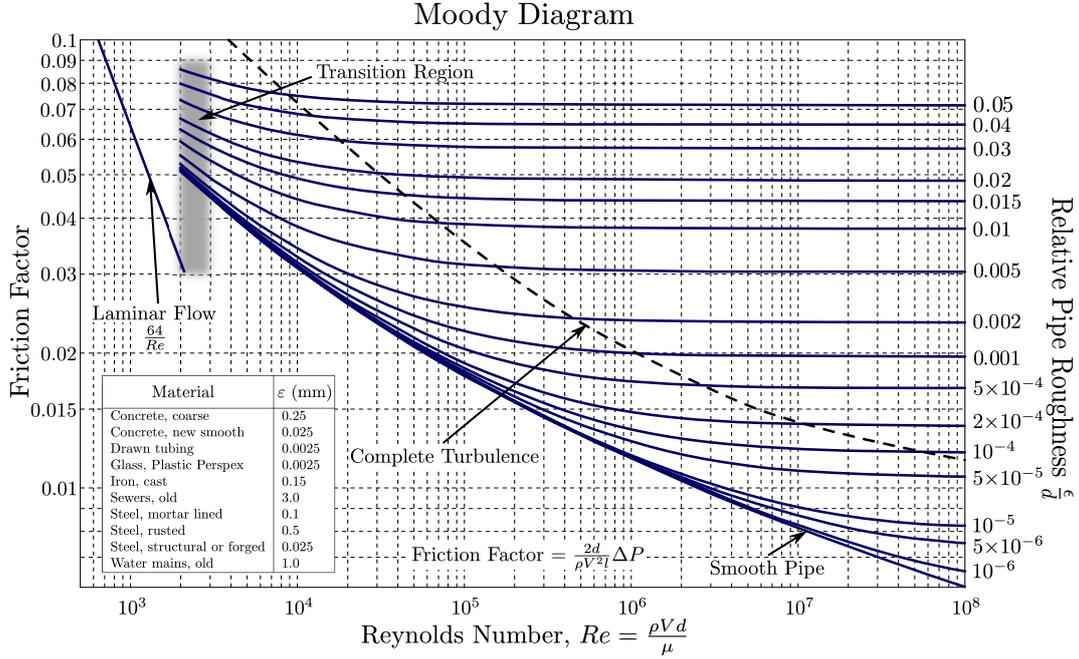


Figure 2.1: Moody diagram. [Beck et Collins, 2008]

for hydraulically smooth pipes

$$\frac{1}{\sqrt{\lambda}} = 2 \log_{10} Re \sqrt{\lambda} - 0.8 \quad (2.5)$$

and hydraulically rough pipes

$$\frac{1}{\sqrt{\lambda}} = -2 \log_{10} \left(\frac{\epsilon}{D} \right) + 1.14 \quad (2.6)$$

by Prandtl [1961] and Nikuradse [1933]. This leads to the formulation of Equation 2.7 for the friction factor, that closes the gap between smooth and rough pipes.

$$\frac{1}{\sqrt{\lambda}} = -2 \log_{10} \left(\frac{2.51}{Re \sqrt{\lambda}} + \frac{\epsilon}{3.71D} \right) \quad (2.7)$$

The friction factor is defined implicitly which means that the solution of Equation 2.7 is usually performed iteratively. A popular approach in numerical simulation environments like Porteau and EPANet is the use of explicit approximations to the Colebrook equation.

Hazen-Williams

While the Darcy-Weisbach equation in conjunction with the Moody diagram and the Prandtl-Colebrook formula is considered to be the most accurate description for the head-loss due to pipe friction, the implicit definition of the friction coefficient is inconvenient for its numerical evaluation.

The Hazen-Williams equation is an empirical formula for the calculation of the head-loss that is frequently used in the hydraulic modelling of water distribution networks. It is defined as:

$$\Delta h_{HW} = 10.69 \frac{Lq^{1.852}}{D^{4.87}C^{1.852}}, \quad (2.8)$$

where $L[m]$ is length of the pipe, $D[m]$ the pipe diameter, $q[m^3/s]$ the flow rate, and C Hazen-Williams coefficient. The Hazen-Williams coefficient characterizes the roughness of a pipe and can as such be derived from the roughness ϵ according to existing tables. For older pipes this method is inaccurate due to degradation of the material, which means that the coefficient has to be estimated or determined through calibration.

Quadratic Approximation

In some mathematical applications the aforementioned calculation methods for the head-loss are not feasible. The Darcy-Weisbach equation, for example is not continuously differentiable for the transition between laminar and turbulent flow regime and the iterative determination of the friction coefficient is inefficient. On the other hand, the derivative of the Hazen-Williams equation is equal to zero under zero flow conditions. Using these formulations in the context of optimization problems or for the derivation of the stochastic hydraulic equations is infeasible or even impossible.

In order to address this problem [Pecci *et al.* \[2017\]](#) evaluated the use of quadratic head-loss formulation

$$\Delta h = aq + \text{sign}(q) bq^2, \quad (2.9)$$

which attempts to approximate the Hazen-Williams head-loss formula over a predefined region in the flow-rate q by minimizing the absolute or the relative errors. The authors show, that for the approximation of the head-loss function in a reasonable range of flow rates the relative error is less than 1%.

2.1.2 Hydraulic Equations

The formulation of the hydraulic equations follows suit to the construction of lumped electrical circuit models with the first and second Kirchhoff laws.

Continuity

The principle of conservation of mass states, that the change in mass over time t enclosed in a control volume Ω has to be equal to the mass flow through the surface of the control volume $\partial\Omega$:

$$\int_{\Omega} \frac{\partial \rho}{\partial t} dV + \int_{\partial\Omega} \rho v dA = 0. \quad (2.10)$$

Here ρ stands for the density and v for the local velocity. In the case of stationary flow for an incompressible fluid the first term of Equation 2.10 is equal to zero, as $\rho = \text{const}$. For the flow along a pipe with the cross-sections A_1 at its inflow and A_2 at its outflow the equation simplifies to

$$v_1 A_1 - v_2 A_2 = q_1 - q_2 = 0. \quad (2.11)$$

For the modelling of pipe junctions in distribution network models the continuity at a node i can be expressed as

$$\sum_j^{deg(i)} a_{i,j} q_{i,j} - d_i = 0, \quad i = 1, \dots, n_j. \quad (2.12)$$

Here, $deg(i)$ is the number of pipes entering node i , $a_{i,j}$ is the coefficient from the incidence matrix specifying the defined direction of the pipe, $q_{i,j}$ is the flow through pipe j exiting node i and d_i is the outflow at node i that can be seen as a source term. In the hydraulic models there exist two types of nodes. At junction or consumption nodes the inflow and outflow d is modelled as a fixed boundary condition and the hydraulic head is one of the degrees of freedom in the system. For resource nodes it is the inverse, as the pressure head is a boundary condition and the inflow and outflow d_r is determined during the solution of the equations. According to this, the transpose of the incidence matrix can be partitioned as $[\mathbf{A}|\mathbf{A}_r]^T$, with \mathbf{A}_r encoding the fixed potential nodes. Following Kirchhoffs current law and using the incidence matrix, the continuity for the hydraulic model can be expressed as

$$\mathbf{A}\mathbf{q} + \mathbf{d} = \mathbf{0}. \quad (2.13)$$

In the general case of looped networks, the number of pipes n_l is higher than the number of nodes n_n . In this case Equation 2.13 is underdetermined.

Compatibility

In order to find a unique solution to Equation 2.13, further conditions are needed. Following Kirchhoffs second law, this is done through the introduction of the energy equation or compatibility equation

$$\Delta\mathbf{h}(\mathbf{r}, \mathbf{q}) - \mathbf{A}^T \mathbf{h} - \mathbf{A}_r^T \mathbf{h}_r = \mathbf{0}. \quad (2.14)$$

It includes the nodal heads at the fixed potential nodes $\mathbf{h}_r \in \mathbb{R}^{nr}$ and junction nodes $\mathbf{h} \in \mathbb{R}^{nj}$, setting them into relation through the head-loss $\Delta\mathbf{h}$, which is a function of the flow rate vector $\mathbf{q} \in \mathbb{R}^{np}$ and the resistance vector $\mathbf{r} \in \mathbb{R}^{np}$. In other words the compatibility states that the potential difference in the starting and the end node has to be equal to the losses experienced along the pipe.

Steady State Hydraulic Solution

The resulting formulation of the steady state hydraulic equations is given as follows:

$$\begin{aligned} \mathbf{A}\mathbf{q} + \mathbf{d} &= \mathbf{0} \\ \Delta\mathbf{h}(\mathbf{r}, \mathbf{q}) - \mathbf{A}^T\mathbf{h} - \mathbf{A}_r^T\mathbf{h}_r &= \mathbf{0} \\ \Delta\mathbf{h}: \mathbb{R}^{np} \times \mathbb{R}^{np} &\rightarrow \mathbb{R}^{np} \\ (\mathbf{r}, \mathbf{q}) &\mapsto \Delta\mathbf{h}(\mathbf{r}, \mathbf{q}). \end{aligned} \quad (2.15)$$

The system contains a linear part for the determination of the flow rates \mathbf{q} and a non-linear part through the inclusion of the head-loss function $\Delta\mathbf{h}$.

In the discussion of the graph theory in Section 1.1 it has been shown that the network graph can be decomposed into grid and forest. Calculation of the forest can be performed very efficiently through the solution of a linear system

$$\mathbf{A}_f\mathbf{q}_f + \mathbf{d}_f = \mathbf{0}, \quad (2.16)$$

which possesses a unique solution in that case. Knowledge of the flow rates allows for the direct evaluation of the head-loss function $\Delta\mathbf{h}$, which allows to calculate of the nodal heads \mathbf{h}_f according to the linear system

$$\mathbf{h}_f = \mathbf{A}_f^{-T} (\Delta\mathbf{h} - \mathbf{A}_f^T\mathbf{h}_f). \quad (2.17)$$

The concept of graph decomposition, for a more efficient solution of the hydraulic system, has been extended even further by [Deuerlein \[2008\]](#).

2.1.3 Sensitivity

In many applications knowledge of the sensitivities for the hydraulic equations comes with a benefit. One way to determine the sensitivity of the hydraulic state vector \mathbf{x} with respect to the model parameters \mathbf{p} is through the adjoint method. To this end, the response function is defined as the state vector

$$\mathbf{f}(\mathbf{x}, \mathbf{p}) = \begin{pmatrix} \mathbf{q}(\mathbf{p}) \\ \mathbf{h}(\mathbf{p}) \end{pmatrix}. \quad (2.18)$$

The linearised form of the residual function is given as:

$$\mathbf{g}(\mathbf{x}, \mathbf{p}) = \begin{pmatrix} \mathbf{D} & -\mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{q}(\mathbf{p}) \\ \mathbf{h}(\mathbf{p}) \end{pmatrix} + \begin{pmatrix} -\mathbf{A}_r^T\mathbf{h}_r \\ \mathbf{d} \end{pmatrix} = \mathbf{0}. \quad (2.19)$$

Here, $\mathbf{D}(\mathbf{q}, \alpha)$ is the diagonal Jacobian matrix containing the derivatives of the head-loss $\Delta\mathbf{h}(\mathbf{r}, \mathbf{q})$ with respect to the flow rate \mathbf{q} and α is the exponent of the flow for the chosen head-loss formula. In the following the constant $-\mathbf{A}_r^T\mathbf{h}_r$ is expressed as \mathbf{e} .

Based on these definition the Lagrangian is be formulated as

$$\mathcal{L} = \begin{pmatrix} \mathbf{q} \\ \mathbf{h} \end{pmatrix} + \left\langle \boldsymbol{\lambda}, \begin{pmatrix} \mathbf{D} & -\mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \mathbf{h} \end{pmatrix} + \begin{pmatrix} \mathbf{e} \\ \mathbf{d} \end{pmatrix} \right\rangle, \quad (2.20)$$

which can be reordered as

$$\mathcal{L} = \boldsymbol{\lambda}^T \begin{pmatrix} \mathbf{e} \\ \mathbf{d} \end{pmatrix} + \left(\mathbf{I} + \boldsymbol{\lambda}^T \begin{pmatrix} \mathbf{D} & -\mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \right) \begin{pmatrix} \mathbf{q} \\ \mathbf{h} \end{pmatrix}. \quad (2.21)$$

In order to avoid explicit dependency of the derivatives on the state vector in Equation 2.21, suppose that

$$\mathbf{I} + \boldsymbol{\lambda}^T \begin{pmatrix} \mathbf{D} & -\mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} = \mathbf{0}. \quad (2.22)$$

From this the Lagrangian multipliers can be calculated as:

$$\boldsymbol{\lambda} = \begin{pmatrix} -\mathbf{D} & -\mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix}^{-1}. \quad (2.23)$$

Since it is a diagonal matrix, the transposed of $\mathbf{D} = \mathbf{D}^T$.

With the use of the Schur complement the inverse matrix in Equation 2.23 can be expressed as

$$\begin{pmatrix} -\mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{D}^{-1} & -\mathbf{D}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1} \\ (\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{D}^{-1} & -(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1} \end{pmatrix}. \quad (2.24)$$

From Equation 1.7 a direct formulation of the sensitivity can then be given as

$$\mathbf{S}_d = \frac{\partial \mathcal{L}}{\partial \mathbf{d}} = \begin{pmatrix} \frac{\partial \mathbf{q}}{\partial \mathbf{d}} \\ \frac{\partial \mathbf{h}}{\partial \mathbf{d}} \end{pmatrix} = \begin{pmatrix} -\mathbf{D}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1} \\ -(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1} \end{pmatrix}, \quad (2.25)$$

for the derivative with respect to the demand. A similar formulation for the sensitivities with respect to the pipe resistance is given as

$$\mathbf{S}_r = \frac{\partial \mathcal{L}}{\partial \mathbf{r}} = \begin{pmatrix} \frac{\partial \mathbf{q}}{\partial \mathbf{r}} \\ \frac{\partial \mathbf{h}}{\partial \mathbf{r}} \end{pmatrix} = \begin{pmatrix} (\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B} \\ \mathbf{D}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B} - \mathbf{D}^{-1}\mathbf{B} \end{pmatrix}, \quad (2.26)$$

Where the matrix \mathbf{B} is the diagonal Jacobian matrix for the derivatives of the head-loss $\Delta \mathbf{h}(\mathbf{r}, \mathbf{q})$ with respect to the roughness \mathbf{r} . These results are in agreement with the derivation of the sensitivities using the forward procedure given by Piller *et al.* [2016].

2.1.4 Alternative Formulations

Apart from the full problem formulation in flow rates \mathbf{q} and heads \mathbf{h} , a number of alternative representations for the hydraulic state can be conceived that allow for the reconstruction of the full state vector. One possible form of the state vector is given by the flow rate vector $\mathbf{x}_f = \mathbf{q}$, which is closely tied to the description in the head-loss vector $\mathbf{x}_{hl} = \mathbf{\Delta h}$ and the reduced loop-flow vector $\mathbf{x}_{lf} = \mathbf{q}_c$. A second possibility is the description of the hydraulic state through the nodal heads $\mathbf{x}_h = \mathbf{h}$, which is more compact than \mathbf{x}_f .

Some of these unique state vectors are linked to a specific formulation of the hydraulic equations, the most important of which are introduced in the following. A more complete discussion of the alternative can be found in [Filler \[1995\]](#).

Flow Description

The loop-flow description uses the orthogonality of the loop matrix $\mathbf{C}^T \mathbf{A} = \mathbf{0}$ introduced in Section 1.1. Going from the second Kirchoff law in Equation 2.14, it is possible to eliminate the head from the system

$$\mathbf{C} \mathbf{\Delta h}(\mathbf{r}, \mathbf{q}) - \mathbf{C} \mathbf{A}_r^T \mathbf{h}_r = \mathbf{0}. \quad (2.27)$$

The flow based problem description is thus given as

$$\begin{aligned} \mathbf{A} \mathbf{q} + \mathbf{d} &= \mathbf{0} \\ \mathbf{C} \mathbf{\Delta h}(\mathbf{r}, \mathbf{q}) - \mathbf{C} \mathbf{A}_r^T \mathbf{h}_r &= \mathbf{0}. \end{aligned} \quad (2.28)$$

The head can be reconstructed from the linear system (2.14) as

$$\mathbf{h} = (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{\Delta h}(\mathbf{r}, \mathbf{q}) - \mathbf{A}_r^T \mathbf{h}_r). \quad (2.29)$$

Loop-Flow Description

From the flow description it is possible to reach a reduced form with a minimal number of components for the state vector. The flow rate in all of the pipes for a network is actually determined by a reduced number of $n_p - n_j$ flow rates. These flow rates are also termed the loop-flow rates \mathbf{q}_c . A physical interpretation of this representation is that the total flow distribution in a network can be seen as the superposition of flows around each loop. The complete flow state \mathbf{q} can be calculated from the flow in the network tree structure \mathbf{q}_d and the loop-flow rates \mathbf{q}_c .

$$\mathbf{q} = \mathbf{C}^T \mathbf{q}_c + \mathbf{q}_d. \quad (2.30)$$

Here, the flows \mathbf{q}_d are linked to the nodal demands and can be evaluated as

$$\mathbf{q}_d = \begin{bmatrix} \mathbf{A}_S^{-1} \mathbf{d} \\ \mathbf{0} \end{bmatrix}, \quad (2.31)$$

where \mathbf{A}_S is the part of the incidence matrix describing the networks spanning tree.

With \mathbf{q}_c the full flow rate vector \mathbf{q} can be determined from (2.30). From this the head-loss can be calculated for the evaluation of the (2.27). The head can once again be reconstructed from Equation 2.29. Due to its description in a minimal number of degrees of freedom and the sparse structure of matrix \mathbf{C} , the solution of this system can be performed very efficiently.

2.2 Pressure Driven Modelling

In the case of pressure dependent demand, experience has shown that under certain conditions the Demand Driven Model can lead to non-physical solutions with negative pressures. This is the case in pressure deficient networks where, under realistic conditions, the requested outflow at a demand node cannot be met. From hydrostatics it is known that the maximum flow volume depends on the difference between the nodal and the atmospheric pressure. To take this into account the Pressure Driven Modelling approach loosens up the demand boundary conditions and the fixed consumption is replaced by the set of inequality conditions $0 \leq c \leq d$. They state that the actual discharge at the node lays in between zero and the desired service demand based on the state of the network. By far the most popular approach to close the pressure dependent formulation is the introduction of a pressure outflow relation $c(h)$ that quantifies discharges based on the present head.

2.2.1 Pressure Outflow Relationship

The selection of the pressure outflow relation is a design choice made for the model. One of the first publications on the topic by [Bhave \[1981\]](#) uses the Heaviside function as pressure outflow relation. This means that for a head lower than the minimum head there is no supply and above the full demand is met. [Elhay et al. \[2015\]](#) give a overview of different modelling choices. In general the flow-rate q is assumed to be proportional to a power n of the nodal pressure head. [Van Zyl et Clayton \[2007\]](#) and [Cheung et al. \[2005\]](#) estimate the value of the exponent to lie in an interval of $n \in [0.5, 2.79]$. Alternatives to these models are given by the cubic consumption function defined by [Fujiwara et Ganesharajah \[1993\]](#), the exponential consumption function presented by [Tanyimboh et Templeman \[2004\]](#) and the Regularized Wagner function given in [Piller \[1995\]](#).

For the work presented in this thesis, the pressure outflow relation introduced by [Wagner et al. \[1988\]](#) is chosen as it is derived directly from the energy conservation law for the free flow boundary condition in hydrostatics.

This leads to the function:

$$c(h) = \begin{cases} 0 & , \text{ if } h \leq h_m \\ \left(\frac{h-h_m}{h_s-h_m}\right)^{\frac{1}{2}} d & , \text{ if } h_m < h < h_s \\ d & , \text{ if } h_s \leq h. \end{cases} \quad (2.32)$$

This equation complies with the inequality conditions and is based on a physical model for the outflow. In (2.32) h is the calculated head. The minimum head necessary is given by h_m . In general the minimum head is defined by the nodal elevation and depends on the specific characteristics of the connected consumers. The required head for servicing the full requested demand is defined by h_s .

In the case of pressure dependent leakage for pipe ruptures Schwaller et Van Zyl [2014] describes a concept called Fixed and Varied Area Discharge (FAVAD) which defines an orifice function as follows.

$$c(h) = C_d \sqrt{2g} (A_0 h^{0.5} + m h^{1.5}) \quad (2.33)$$

Here C_d is a discharge coefficient, g the gravitational constant, A_0 is the area of the opening if no head is present and m is a linear value describing the growth in surface area. In contrast to the demand emitter function, this orifice function not only defines the discharge based on the pressure but also the change in surface for the rupture due to the elasticity of materials and the pressure.

2.2.2 Pressure Driven Problem Description

Using a general pressure outflow relationship a modified set of Kirchhoff's equations has been published by Piller *et al.* [2003].

$$\begin{aligned} \mathbf{A}\mathbf{q} + \mathbf{c}(\mathbf{h}) &= \mathbf{0} \\ \Delta\mathbf{h}(\mathbf{r}, \mathbf{q}) - \mathbf{A}^T\mathbf{h} - \mathbf{A}_r^T\mathbf{h}_r &= \mathbf{0}. \end{aligned} \quad (2.34)$$

While the pressure driven problem formulation gives a more realistic result, especially in modelling water distribution networks with low pressure zones, it is no longer possible to solve with a loop flow method. However, Piller *et al.* [2016] managed to provide a direct formulation for the systems sensitivities, with respect to the pressure outflow relation. The inclusion of the POR according to Equation 2.34 usually leads to an iterative solution process, that is computationally more demanding than the simple solution of a non-linear system. In order to improve the solver efficiency, Piller *et al.* [2003] also introduce the variational pressure driven equations of the hydraulic system through the dual Content and Co-Content formulations. This allows for the direct solution of the problem through an optimization approach.

2.3 Model Deficiencies

From literature the notion of deficient networks can take a number of different definitions. These definitions may be divided into model, mathematical and physical deficiencies. Model deficiencies are errors in the creation, conversion or transfer of the network graph. A mathematical deficiency can be defined as a maximal connected network where, due to some boundary condition the set of feasible solutions is reduced to the empty set or the solution is not unique. In contrast to mathematical deficiencies, in the case of a hydraulic deficiency a unique solution exists, but it is physically incorrect. With respect to the demand-driven and the pressure-driven modelling paradigms different phenomena have to be classified as deficient. In the following a number of reoccurring deficiency phenomena are presented and evaluated with respect to demand and pressure driven modelling.

Conflicting boundary conditions

The first scenario is given if the boundary conditions are in conflict for certain parts of the network. This occurs if flow regulating devices are incorporated into the model and introduce additional boundary conditions to the mathematical model. In unfortunate cases these additional boundary conditions may conflict with the demand request of the consumption nodes. Simply put, the controlled flow entering a region of the network is not satisfying the required demand. For the optimization problem formulated on the variational content and co-content models for the demand-driven model defined in [Carpentier *et al.* \[1985\]](#), this reduces the set of feasible solutions to zero, as demonstrated by [Deuerlein *et al.* \[2012a\]](#). The publication also suggests an algorithm to determine if a feasible solution exists for the particular scenario. Looking at the pressure dependent calculation of the same system, it can be shown that by loosening the demand boundary conditions the system becomes solvable again, but the consumers will be supplied with a reduced flow.

Ambiguous boundary conditions

Another example for a mathematical deficiency is given if the boundary condition allow for an infinite number of solutions. [Gorev *et al.* \[2016\]](#) describe a scenario where two flow control valves (FCV) are installed in series. Here the two FCVs create a combined head-loss, but due to the ambiguous nature of this problem it is not possible to determine which of the two FCVs contributes how much. This phenomenon is neither addressed by DDM nor by PDM approaches.

Pipe rupture

In respect to resilience, phenomena like pipe-rupture or -bursts are of special interest. In this cases the massive water loss dominates the flow in the network. Recent research has shown that the fixed and varied area discharge (FAVAD) model as described in (2.33) provides a good description of leakage behavior for elastic materials Van Zyl *et al.* [2011]. Due to the pressure dependent nature of the phenomenon, in demand driven modelling it is not possible to adequately handle the problem. In contrast, there have been a number of successful applications to in the PDM framework.

High lying nodes / vapour pressure

The fourth scenario is correlated with the occurrence of low pressure zones in the network. This may for instance be triggered by a pipe burst and the subsequent pressure loss. If the pressure drops below a certain value water starts to change the phase. In general this can be described in three stages. First water starts to boil and form vapour bubbles that are distributed in the water column. In the second stage both phases start to separate and a stratified flow forms in the pipe. For extreme cases in the third stage water has completely changed phase and the pipe is filled by vapour. For modelling purpose this complex behavior is simplified by defining that with reaching the vapour pressure water instantly changes phase and the hydraulic connection between two nodes is severed. Looking at current demand and pressure driven models this behavior is not taken into account. In the case of zero or negative pressure, software packages like Porteau and EPANet will give a warning notifying the user that pressure dropped below zero, but the hydraulic connection is still intact and disconnected network parts will still be supplied. A conceptually simple way to solve this problem in the PDM framework may be implemented an iterative approach that analyses the pressure and deletes all links connected to the deficient node. A different approach has been proposed by Piller *et Van Zyl* [2009]. They introduce an additional constraint to the optimization formulation that reduces the flow on deficient pipes to zero.

2.4 Network Models

The network graph gives a natural interpretation of the network structure, where pipes are represented by the edges and the pipe junctions are represented by the vertices. The mathematical description of this graph is given by the incidence matrix $\mathbf{A} \in \mathcal{M}^{n_j \times n_p}$. For the application of the methods developed in this thesis, in the following three basic water distribution networks are introduced. Each of the networks is designed or chosen due to a special attribute, which is further elaborated in the subsequent sections.

2.4.1 Tree Network

As mentioned in Section 2.1.2 the calculation of the forest is performed through the solution of a linear system of equations in the flow rates \mathbf{q} . This is a convenient attribute especially for the first implementation and validation of the complex methods that are developed in this thesis. For this reason the tree shaped exemplary network model is defined according to the topology shown in Figure 2.2. The graph has resource node at the trees root node (0). From there 10 branches and 10 nodes are connected without forming a closed loop. The network contains a total number of 20 degrees of freedom. As the continuity is described by a system of 10 equations defined in 10 variables, the flow rates are evaluated by a linear system. Methods developed on this

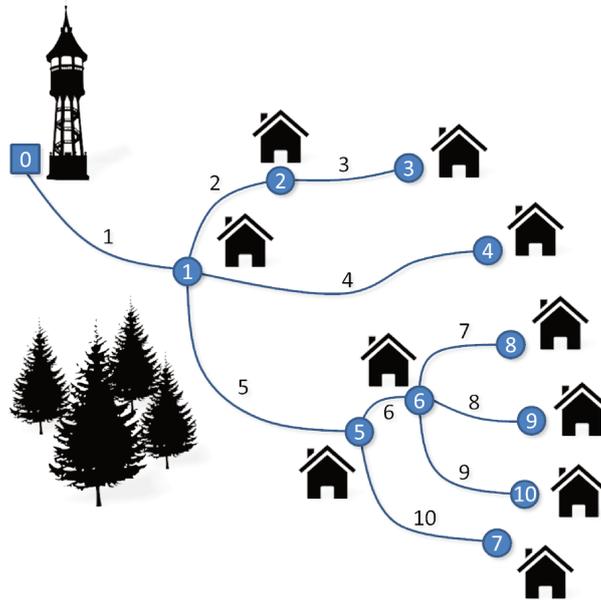


Figure 2.2: Tree shaped water distribution network.

model can be extended to more complex network structures without any loss of generality.

2.4.2 Highly Aggregated Looped Network

In order to advance the problem complexity from the tree structure network, a simple looped network is introduced next. Its topology is shown in Figure 2.3. It is derived from a large regional water transportation network in the Bordeaux region. Through the aggregation of a large number of nodes it is simplified to a simple network with one major loop. The graph contains 9 free potential nodes, 12 edges and 3 fixed potential nodes resulting in a total number of 21 degrees of freedom. Looking at the loop-flow formulation the

complexity of the model can be greatly reduced to a total number of 3 degrees of freedom. Through the addition of the loop, the flow cannot be solved solely

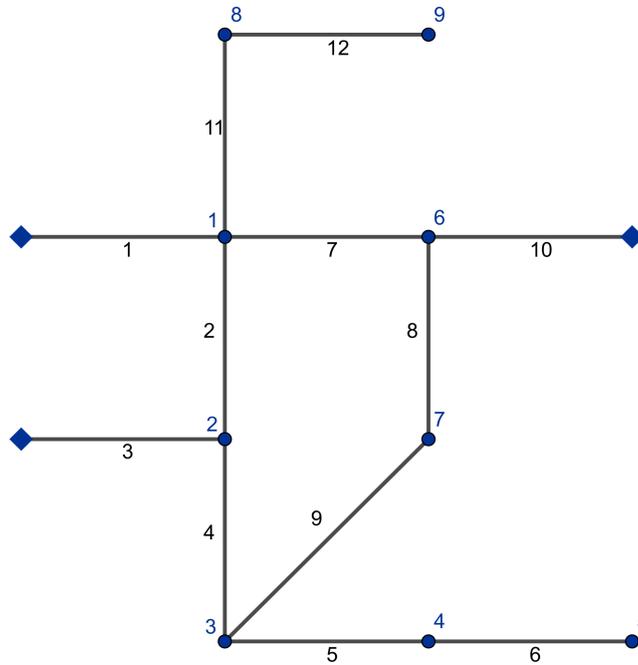


Figure 2.3: Highly aggregated realistic water distribution network.

through the solution of the continuity equation. This means that the hydraulic problem also becomes non-linear in the flow vector \mathbf{q} .

2.4.3 Realistic Network Model

In order to validate the methods developed in this thesis, a full size model of a realistic water distribution network is needed. This network model was provided by Veolia Eau d'Île-de-France (VEDIF) through the ResiWater project [ResiWater, 2019] and is depicted in Figure 2.4. It contains part of a network from the Paris region and consists of a total number of 2,175 pipes and 1,822 nodes, resulting in nearly 4,000 degrees of freedom for the head-flow formulation. Inside the highly looped region lays one reservoir. Once again, applying the loop-flow formulation massively reduces the networks degrees of freedom to a total number of 348. The model defined two consumer types with an industrial and a domestic demand patterns which are evaluated during the peak demand phase.



Figure 2.4: Realistic water distribution network from the Paris region, consists of a total number of 2,175 pipes, 1,822 nodes and one reservoir (indicated by the blue square).

2.5 Conclusion

This chapter discussed the major aspects of the hydraulic network model, which is the subject of the studies presented in this thesis. In Section 2.1 the demand-driven modelling paradigm is introduced. This contains the presentation of the friction induced head-loss, the formulation of the static hydraulic equations and the formulation of the sensitivity matrix based on the adjoint method. This is followed by the presentation of the pressure-driven modelling paradigm in Section 2.2, with the introduction of the pressure-outflow relationship and the formulation of the altered hydraulic equations. The capabilities of both, the demand-driven and pressure-driven modelling paradigms for modelling deficient networks are discussed in Section 2.3. The chapter closes with the presentation of the graphs for the example networks that are used throughout the thesis.

Chapter 3

Reduced Order Modelling

Motivation: *Dimensional reduction for the more efficient evaluation of the hydraulic model.*

Principal Elements:

- *Interpolation based reduced order modelling techniques*
- *Projection based reduced order modelling techniques*
- *Empirical interpolation hyper reduction*
- *Adaptive parameter sampling*

Conclusions:

- *Interpolation based methods are inefficient in the use of pre-calculated system states.*
- *Projection based methods are efficient in the use of pre-calculated system states, but inefficient in the evaluation of the non-linear residuals.*
- *The empirical interpolation method for the efficient evaluation of the non-linear residuals tends to be unstable without further measures.*
- *Adaptively sampled models use fewer sampling points to reach the accuracy of a Cartesian sampled model.*

The process of modelling parametric, spatially distributed systems leads, in its most general case, to the formulation of a non-linear system of equations with a considerable number of unknowns. In the case of partial differential equations this is achieved through the application of spacial discretization methods like the finite element method, finite difference method or the finite volume method [Chung, 2010]. For the discretization of a water distribution network model a system of non-linear equations is formulated based on a lumped component model that models pipes as discrete elements as described in Chapter 2. The solution of these systems of equations can take somewhere from seconds up to hours. As many practical applications impose certain limi-

tations with respect to the evaluation time, the use of such models may become computationally prohibitive. This is especially the case for methods that either require a big number of evaluations, like optimization, Bayesian parameter estimation or uncertainty quantification, or cases where evaluation times have to be short, like real-time capable for applications. In order to mediate the computational effort for such applications reduced order models (ROM) can be used to formulate a surrogate system that, compared to the high-fidelity model, uses a greatly reduced number of variables and thus accelerating the solution process of the non-linear solver [Antoulas, 2005].

Smith [2013] states that these surrogate models can be roughly categorized in three classes given by interpolation or regression-based, projection-based and hierarchical models.

- Interpolation based models: These treat the high-fidelity model as a black box that is used to create data samples. The surrogate is then created using data driven techniques like for example radial basis functions, Gaussian process modelling or Artificial Neural Networks. This purely data-driven approach implies that the method is non-intrusive, however it is strongly dependent on the sampling of the parameter space [Baur *et al.*, 2011].
- Projection based models: This approach is what is usually referred to as reduced order models. They are constructed by projecting the system states onto a low-order subspace. Examples for this are given by eigenfunction expansion as it is used in modal analysis of dynamic systems or the proper orthogonal decomposition [Benner *et al.*, 2013; Antoulas, 2005]. In contrast to purely data-driven methods these projection-based approaches retain valuable information from the equations of the problem formulation through the projection.
- Hierarchical models: This third group of surrogate models usually use efficient grid techniques in spatially distributed systems, a simplified physical model or linearisation based approaches [Walski *et al.*, 2003; Deuerlein, 2008; Ulanicki *et al.*, 1996; Paluszczyszyn *et al.*, 2013].

The use of these methods is commonplace in numerical engineering problems and have proven their benefit in countless applications. Objective of the work presented in this chapter is to apply reduced order methods to the hydraulic water distribution network model and investigate if similar improvements on computational efficiency can be achieved with acceptable accuracy. In Section 3.4 the error metrics and the general validation procedure are described for the head-flow and the loop-flow formulation of the hydraulic system. Further, it describes the different styles of error plots that are used in the subsequent evaluation of the reduced order models. Section 3.2 then introduces the concept of the proper orthogonal decomposition, defines the reduced model

in head-flow and loop-flow description. Special treatment for a more efficient evaluation of the non-linearities introduced by the head-loss function is discussed in Section 3.3 by introducing the empirical interpolation method. Section 3.5 discusses the influence of the sampling strategy on the reduced order model and introduces an adaptive sampling strategy, that selects parameter combinations for the snapshot simulations based on the leave-one-out error. Finally, in Section 3.6, the methods are evaluated through the application to the medium sized test network that was introduced in Section 2.4.

3.1 Interpolation-Based Models

For the regression and interpolation approach to the construction of a surrogate model the high-fidelity model is assumed to be a black box system $y = f(x)$. This system is evaluated at M realizations of the parameter vector x :

$$y_m = f(x^m) \quad , \quad m = 1, \dots, M. \quad (3.1)$$

The resulting sample data (x^m, y_m) is then used to create an emulator $\tilde{f}(x)$ that approximates the full model $f(x)$ with a certain accuracy, making this a purely data-driven approach.

Algebraic interpolation methods can use a big assortment of algebraic functions for the construction of the meta-model. In its easiest form this describes the linear interpolation. However, depending on the problem orthogonal polynomials [Xiu, 2010] or Lagrangian interpolating polynomials are used for piecewise interpolations. Kernel based interpolation methods like the radial basis interpolation or Gaussian process modelling, also known as Kriging, are a popular alternative to these algebraic interpolation methods, that are well suited for irregular grid applications.

In general, the quality of such a meta-model depends in large parts on the appropriate sampling of the parameter space. Popular strategies include the Monte Carlo and Latin Hypercube sampling as well as sparse grid techniques. A particular method that is used in this thesis is the adaptive error driven sampling strategy detailed in section 3.5.

In the following, a closer look is taken at the construction of response surface models and the radial basis interpolation.

3.1.1 Response Surface Models

The construction of polynomial response surface models can be formulated as a linear regression problem. For a bi-linear estimator the polynomial form is given as

$$f(\mathbf{x}, \mathbf{a}) \approx a_0 + \sum_{i=1}^k a_i x_i + \sum_{i=1}^k \sum_{j>i}^k a_{ij} x_i x_j, \quad (3.2)$$

with the state vector $\mathbf{x} = [x_1, \dots, x_k]$ and the unknown parameter vector \mathbf{a} .

For the example of a two-dimensional rectilinear grid, the coefficient vector \mathbf{a} is determined from the four neighbouring points $[P_{11} \ P_{12} \ P_{21} \ P_{22}]^T$, posing the linear system of equations:

$$\underbrace{\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_1^{(1)}x_2^{(1)} \\ 1 & x_1^{(1)} & x_2^{(2)} & x_1^{(1)}x_2^{(2)} \\ 1 & x_1^{(2)} & x_2^{(1)} & x_1^{(2)}x_2^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_1^{(2)}x_2^{(2)} \end{bmatrix}}_{\mathbf{X}} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \underbrace{\begin{bmatrix} f(P_{11}) \\ f(P_{12}) \\ f(P_{21}) \\ f(P_{22}) \end{bmatrix}}_{\mathbf{y}_m}, \quad (3.3)$$

with the design matrix \mathbf{X} and the sample vector \mathbf{y}_m .

Given that the size of the sample vector is at least equal to the size of the parameter vector and the matrix $[\mathbf{X}^T \mathbf{X}]$ is invertible, linear regression theory allows to determine the parameter vector as

$$\mathbf{a} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y}_m. \quad (3.4)$$

Bi-linear interpolation models are popular surrogates as they include low order non-linearities, making them more efficient than pure linear interpolation. However, the approach can easily be extended using higher order polynomial terms.

3.1.2 Radial Basis Interpolation

The use of Radial Basis Functions (RBF) for the construction of an interpolation meta-model is an alternative approach to the polynomials basis expansion. It can be expressed as

$$\tilde{f}(x) = \sum_{m=1}^M f_m \Psi_m(x) + P(x), \quad (3.5)$$

where $\Psi_m(x) = \psi(\|x^m - x\|)$ give the radial basis function, f_m are the weights and $P(x)$ is a bias function.

The radial basis functions are real valued kernel functions whose value depends on the Euclidean distance of x from its origin x^m . Common choices for Ψ are:

$$\Psi(r_m)_i = \begin{cases} e^{r_m/2\epsilon_i^2} & , \text{ Gaussian} \\ \sqrt{1 + (\epsilon_i r)^2} & , \text{ Multiquadric} \\ \frac{1}{\sqrt{1 + (\epsilon_i r)^2}} & , \text{ Inverse multiquadric} \\ r_m^k & , \text{ Polyharmonic spline.} \\ r_m^k \ln(r_m) & \end{cases} \quad (3.6)$$

where ϵ are the shape parameters and $r_m = \|x^m - x\|$ is the radial distance. The value of the shape parameter influences the area of effect for each kernel function and has to be chosen carefully as it influences the fitting quality of the RBF to the sample data.

The coefficients f_m are determined on the basis of the interpolation condition

$$\tilde{f}(x^m, a_0) = y_m, \quad m = 1, \dots, M,$$

sufficing the constraint

$$\sum_{m=1}^M f_m = 0,$$

due to the inclusion of the bias function $P(x)$.

The radial basis interpolation function can thus be expressed as

$$\tilde{f}(x, a_0) = a_0 + \Psi^T(x) \Phi^{-1} [\mathbf{y}_s - a_0 \mathbf{1}], \quad (3.7)$$

where $a_0 = [\mathbf{1}^T \Phi^{-1} \mathbf{1}]^{-1} \mathbf{1}^T \Phi^{-1} \mathbf{y}_s$ and $\Phi_{mk} = \Psi_k(x^m) = \psi(\|x^m - x\|)$.

The interpolation with Radial Basis Functions is especially well suited for functions that depend on a multivariate parameter space and are defined by a large number of data points that are scattered in the domain. While [Smith \[2013\]](#) states that the basic formulation of the RBF interpolation and the Gaussian process regression is essentially the same, he points out that the RBF approach does not include a definition for the uncertainty bounds in the model.

3.2 Projection Based Methods

The proper orthogonal decomposition is a popular method for the approximation of general stationary and dynamic systems. Given be an ensemble of states \mathbf{x} in \mathbb{R}^n , where n is the number of state variables, of a system parametrized in the parameter-space \mathcal{P} . The POD seeks to approximate the full state vector with one in a lower-dimensional space \mathbb{R}^k , with $k < n$. In order to do so, the original system is reduced by projecting it on the lower-dimensional space via the *Galerkin* or *Petrov-Galerkin* method. The subsequent problem is then to determine how well the reduced state approximates the original one for states that are not included in the ensemble.

3.2.1 Proper Orthogonal Decomposition

Given the state vector $\mathbf{x}(\boldsymbol{\mu}) \in \mathbb{R}^n$ of the problem as a function of the parameter vector $\boldsymbol{\mu}$, a finite set of samples $\mathbf{x}(\boldsymbol{\mu}_i)$ for $\boldsymbol{\mu} = \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_N$ can be evaluated. Collecting the resulting data in matrix form leads to the *snapshot database*

$$\mathbf{X} = [\mathbf{x}(\boldsymbol{\mu}_1) \quad \dots \quad \mathbf{x}(\boldsymbol{\mu}_N)] \in \mathbb{R}^{n \times N}. \quad (3.8)$$

Given the snapshot matrix \mathbf{X} a set of orthonormal basis vectors $\mathbf{u}_j \in \mathbb{R}^n$ can be constructed in a way that its entries may be expressed as a linear combination

$$\mathbf{x}_i = \sum_{j=1}^n \gamma_{ji} \mathbf{u}_j. \quad (3.9)$$

The snapshot matrix can thus be expressed as

$$\underbrace{[\mathbf{x}_1 \ \cdots \ \mathbf{x}_N]}_{\mathbf{X}} = \underbrace{[\mathbf{u}_1 \ \cdots \ \mathbf{u}_n]}_{\mathbf{U}} \overbrace{\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nN} \end{bmatrix}}^{\mathbf{\Gamma}}, \quad \mathbf{U}^T \mathbf{U} = \mathbf{I}_n \quad (3.10)$$

using a coefficient matrix $\mathbf{\Gamma}$. The vectors \mathbf{u}_i are usually referred to as the *empirical eigenfunctions* and may be interpreted as the principal directions of the data set \mathbf{X} . Further, it is required that the truncated reconstruction of the snapshots using a reduced number k of empirical eigenfunctions

$$\underbrace{[\tilde{\mathbf{x}}_1 \ \cdots \ \tilde{\mathbf{x}}_N]}_{\tilde{\mathbf{X}}} = \underbrace{[\tilde{\mathbf{u}}_1 \ \cdots \ \tilde{\mathbf{u}}_n]}_{\tilde{\mathbf{U}}} \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{k1} & \cdots & \gamma_{kN} \end{bmatrix}, \quad k < n \quad (3.11)$$

approximates the set \mathbf{X} optimally. Optimality is usually defined through the minimization of the Frobenius norm on the approximation error $\|\mathbf{X} - \tilde{\mathbf{X}}\|_2$.

A solution to this problem is given by the singular value decomposition of $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. The columns of \mathbf{U} define the empirical eigenfunctions and the coefficient matrix is given as $\mathbf{\Gamma} = \mathbf{\Sigma}\mathbf{V}^T$ [Antoulas, 2005].

As the POD is working with the collected data in the snapshot matrix, it is important to have a sufficiently rich basis in the sense that it is able to represent the system behaviour with sufficient accuracy. A closely related topic that is still subject to research are sampling strategies that assure the construction of a good data set. Current approaches are often inspired by Design of Experiments (DoE) as well as Monte Carlo sampling, sparse grid sampling. Section 3.5 further discusses the topic and introduces a greedy algorithm that chooses parameter combination based on a maximal error estimate on the leave-one-out projection error.

3.2.2 Galerkin Projection

One of the most popular methods to create a reduced order model is the Galerkin projection. The Galerkin approach projects the system on the same reduced basis that is used in the low order approximation in order to formulate a reduced system. The benefit over reduction methods that do not include a projection of the original system is that it preserves the problem structure [Lall *et al.*, 2003; Carlberg *et al.*, 2015]

Reduced Head-Flow Model

Based on the truncated basis Φ it is possible to formulate a reduced order system. One of the most popular model reduction techniques is the Galerkin projection.

In Chapter 2 the hydraulic model has been defined in the form of the head-flow equations (2.15), for which the state vector is given as

$$\mathbf{x}_{hf} = \begin{bmatrix} \mathbf{q} \\ \mathbf{h} \end{bmatrix}.$$

Using the reduced basis Φ_{hf} the approximation $\tilde{\mathbf{x}}_{hf}$ of the high-fidelity state vector \mathbf{x}_{hf} is expressed as

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{q}} \\ \tilde{\mathbf{h}} \end{bmatrix}}_{\tilde{\mathbf{x}}_{hf}} = \underbrace{\begin{bmatrix} \Phi_q \\ \Phi_h \end{bmatrix}}_{\Phi_{hf}} \hat{\mathbf{x}} \quad (3.12)$$

with the reduced state vector $\hat{\mathbf{x}} \in \mathbb{R}^k$. Starting from (2.15)

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}^T \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \mathbf{h} \end{bmatrix} + \begin{bmatrix} \mathbf{K}\boldsymbol{\mu} \\ \Delta\mathbf{h}(\mathbf{r}, \mathbf{q}) - \mathbf{A}_f^T \mathbf{h}_f \end{bmatrix} = \mathbf{0},$$

the state vector \mathbf{x}_{hf} is replaced with the reduced order approximation $\hat{\mathbf{x}}$

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}^T \end{bmatrix} \Phi_{hf} \hat{\mathbf{x}} + \begin{bmatrix} \mathbf{K}\boldsymbol{\mu} \\ \Delta\mathbf{h}(\mathbf{r}, \Phi_q \hat{\mathbf{x}}) - \mathbf{A}_f^T \mathbf{h}_f \end{bmatrix} = \mathbf{0}. \quad (3.13)$$

Projecting the system on the reduced basis Φ_{hf} gives the residual function of the reduced order model in the head-flow formulation

$$\hat{\mathbf{r}}_{hf}(\hat{\mathbf{x}}, \boldsymbol{\mu}) = \Phi_{hf}^T \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}^T \end{bmatrix} \Phi_{hf} \hat{\mathbf{x}} + \Phi_{hf}^T \begin{bmatrix} \mathbf{K}\boldsymbol{\mu} \\ \Delta\mathbf{h}(\mathbf{r}, \Phi_q \hat{\mathbf{x}}) - \mathbf{A}_f^T \mathbf{h}_f \end{bmatrix}. \quad (3.14)$$

The Jacobian matrix is derived as

$$\hat{\mathbf{J}}_{hf} = \Phi_{hf}^T \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}^T \end{bmatrix} \Phi_{hf} + \Phi_{hf}^T \begin{bmatrix} \mathbf{0} \\ \mathbf{D}\Phi_q \end{bmatrix}, \quad (3.15)$$

where \mathbf{D} is a diagonal matrix containing the derivatives of the head-loss function with respect to the full state vector \mathbf{x} . The equation can be summarized to

$$\hat{\mathbf{J}}_{hf} = \Phi_{hf}^T \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{D} & -\mathbf{A}^T \end{bmatrix} \Phi_{hf}. \quad (3.16)$$

Reduced Loop-Flow Model

In contrast to the head-flow model the state of the loop-flow equations (2.27) is defined in the loop-flow vector

$$\mathbf{x}_{lf} = \mathbf{q}_c,$$

that may be approximated using a reduced basis Φ_{lf} derived from a snapshot database in the loop-flow state

$$\tilde{\mathbf{x}}_{lf} = \Phi_{lf} \hat{\mathbf{x}}. \quad (3.17)$$

Replacing this in (2.27)

$$\mathbf{C} \Delta \mathbf{h}(\mathbf{r}, \mathbf{C}^T \mathbf{q}_c + \mathbf{q}_d(\boldsymbol{\mu})) - \mathbf{C} \mathbf{A}_f^T \mathbf{h}_f = \mathbf{0}$$

gives the approximation

$$\mathbf{C} \Delta \mathbf{h}(\mathbf{r}, \mathbf{C}^T \Phi_{lf} \hat{\mathbf{x}} + \mathbf{q}_d(\boldsymbol{\mu})) - \mathbf{C} \mathbf{A}_f^T \mathbf{h}_f = \mathbf{0} \quad (3.18)$$

The residuals of the reduced Galerkin system is then derived through the projection

$$\hat{\mathbf{r}}_{hf}(\hat{\mathbf{x}}, \boldsymbol{\mu}) = \Phi_{lf}^T \mathbf{C} \Delta \mathbf{h}(\mathbf{r}, \mathbf{C}^T \Phi_{lf} \hat{\mathbf{x}} + \mathbf{q}_d(\boldsymbol{\mu})) - \Phi_{lf}^T \mathbf{C} \mathbf{A}_f^T \mathbf{h}_f. \quad (3.19)$$

Using the chain rule

$$\frac{d\hat{\mathbf{r}}_{hf}}{d\hat{\mathbf{x}}} = \frac{d\hat{\mathbf{r}}_{hf}}{d\mathbf{q}} \frac{d\mathbf{q}}{d\mathbf{q}_c} \frac{d\mathbf{q}_c}{d\hat{\mathbf{x}}}, \quad (3.20)$$

with

$$\begin{aligned} \frac{d\hat{\mathbf{r}}_{hf}}{d\mathbf{q}} &= \Phi_{lf}^T \mathbf{C} \mathbf{D}, \\ \frac{d\mathbf{q}}{d\mathbf{q}_c} &= \mathbf{C}^T, \\ \frac{d\mathbf{q}_c}{d\hat{\mathbf{x}}} &= \Phi_{lf}, \end{aligned}$$

the Jacobian matrix is given as

$$\hat{\mathbf{J}}_{hf} = \Phi_{lf}^T \mathbf{C} \mathbf{D} \mathbf{C}^T \Phi_{lf}. \quad (3.21)$$

3.2.3 Global and Local Reduced Basis

For parametrized reduced order models, a reduced basis Φ for the parameter $\boldsymbol{\mu}$ has to be created. This parameter is defined on a single domain $\Omega \subset \mathbb{R}^d$. For a global reduced basis a single basis matrix is computed over the range of parameters. The concatenation approach creates the global basis from the ensemble

of simulations in the parameter-space via the SVD or Krylov subspace methods [Liesen et Strakos, 2013]. Alternative methods use a bi-linearization approach to approximate the reduced system matrix as a affine function of the parameter [Benner et Breiten, 2011]. However, recently a number of approaches have been suggested that subdivide the domain Ω and create a reduced basis for each of the sub-domains [Amsallem et Haasdonk, 2016; Haasdonk et al., 2010; Maday et Stamm, 2013]. In contrast to the construction of one fixed global basis, a number of local basis is constructed. Numerous publications discuss the interpolation of these local basis on the parameter-space for the further improvement of the reduced basis [Amsallem et Farhat, 2008; Degroote et al., 2010; Lohmann et Eid, 2007].

3.3 Empirical Interpolation Method

Section 3.2.2 has shown how to effectively reduce the system in the state variables and its application to the hydraulic network model in Section 3.6.2 shows to give accurate results for a sensible choice of parameters and a significant reduction in the computational effort. One of the main factors leading to the reduced evaluation time was the reduced number in function evaluation. This reduction is for one due to a lower number of solver iterations and the faster convergence of the reduced system, but mainly due to the reduced number of function evaluations during the calculation of the Jacobian matrix. However, the evaluation of the non-linear residual function still has to be performed in the dimension of the high-fidelity model.

Hyper-reduction methods like the empirical interpolation method are designed to create a surrogate model for the non-linear function, significantly reducing the computational effort in each function evaluation. The application of the EIM to the POD model of a system of non-linear ordinary differential equations by Chaturantabut et Sorensen [2010] showed to greatly reduce the problem dimension with negligible error over long term integration. Similar reductions could be shown by Fritzen et al. [2018], who applied the EIM to the FEM model of heat transfer problem and compare it to other hyper-reduction approaches and Peherstorfer et al. [2014] present an extension of the EIM, that introduces a localized EIM approach for further improvement of the performance in a reduced reacting flow model.

The EIM is a sampling based interpolation approach. Given a snapshot database, that samples the solution space of a non-linear function in an appropriate manner, it creates a collateral reduced basis $\mathbf{V} = [\mathbf{\Xi}_1, \dots, \mathbf{\Xi}_M]$ with the orthogonal basis vectors $\mathbf{\Xi}_i$. Using this basis and the evaluation of the non-linear function at a number of support points the full non-linear result vector is approximated through interpolation. The method can be used to approximate either the non-linear part of residual function equation as shown

in [Chaturantabut et Sorensen \[2009\]](#) or the residual function as a whole.

Given be the non-linear function \mathbf{f} of a set of arguments $\boldsymbol{\tau} = \{\mathbf{x}, \boldsymbol{\mu}\}$ containing parameters $\boldsymbol{\mu}$ and the state vector \mathbf{x} .

$$\mathbf{f}(\boldsymbol{\tau}) = \mathbf{f}(\mathbf{x}, \boldsymbol{\mu}) \quad (3.22)$$

Based on a selection of realisations for the parameter vector $\boldsymbol{\tau}$ a snapshot database \mathcal{Y}_f is created from which an orthogonal basis \mathbf{V} is determined, for example by use of the singular value decomposition. The orthogonal basis allows for the construction of a truncated basis \mathbf{V}_m by selecting the first m eigenvectors from \mathbf{V} . The selection of m influences directly the efficiency of the reduction and the accuracy of the interpolation and is used as a design variable for the rest of this section. Using the reduced basis \mathbf{V}_m the full non-linear function can be approximated through

$$\mathbf{f}(\boldsymbol{\tau}) \approx \mathbf{V}_m \mathbf{c}(\boldsymbol{\tau}) \quad (3.23)$$

with the coefficient vector $\mathbf{c} \in \mathbb{R}^{m \times 1}$. As this system is overdetermined, the full solution can be determined from the evaluation of a reduced number of m equations. These equations are selected by the selection matrix $\mathbf{P} \in \mathbb{R}^{n \times m}$ with the dimension of the full function n and the number of evaluation points m . \mathbf{P} contains m unit vectors $\mathbf{e}_i = [0 \ \dots \ 1 \ \dots \ 0]^T \in \mathbb{R}^{n \times 1}$.

$$\mathbf{P} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_m] \quad (3.24)$$

With the construction of matrix \mathbf{P} a number of m equations are selected from (3.23)

$$\mathbf{P}^T \mathbf{f}(\boldsymbol{\tau}) = \mathbf{P}^T \mathbf{V}_m \mathbf{c}(\boldsymbol{\tau}) \quad (3.25)$$

and if $\mathbf{P}^T \mathbf{V}$ is invertible the coefficient vector $\mathbf{c}(\boldsymbol{\tau})$ can be evaluated as

$$\mathbf{c}(\boldsymbol{\tau}) = (\mathbf{P}^T \mathbf{V}_m)^{-1} \mathbf{P}^T \mathbf{f}(\boldsymbol{\tau}). \quad (3.26)$$

The full non-linear vector $\hat{\mathbf{f}}(\boldsymbol{\tau})$ is now determined by replacing $\mathbf{c}(\boldsymbol{\tau})$ in (3.23) with (3.26)

$$\hat{\mathbf{f}}(\boldsymbol{\tau}) = \mathbf{V}_m \mathbf{c}(\boldsymbol{\tau}) = \mathbf{V}_m (\mathbf{P}^T \mathbf{V}_m)^{-1} \mathbf{P}^T \mathbf{f}(\boldsymbol{\tau}) = \mathbf{W}_m \mathbf{f}(\boldsymbol{\tau}). \quad (3.27)$$

The fact that \mathbf{V}_m and \mathbf{P} are constant matrices that are determined during the off-line step of the EIM allows for the definition of the interpolation matrix \mathbf{W}_m . The interpolation matrix \mathbf{W} has rank m . In the calculation of the interpolation this allows to use a reduced $\mathbb{R}^{n \times m}$ matrix that eliminates the zero columns and that uses only m values that have to be evaluated by the original function \mathbf{f} .

The EIM algorithm is divided into an off-line and an on-line phase. During the off-line phase, described in Algorithm 1, the selection matrix \mathbf{P} is determined and the interpolation points \mathbf{O} are chosen. The interpolation points are

chosen by a greedy algorithm that minimizes the interpolation error on the snapshot database.

Data: reduced basis vectors \mathbf{v}_k from SVD of the snapshot database \mathcal{Y}_f
Result: selection matrix \mathbf{P} , index set $\mathbf{O} := \{i_1, \dots, i_m\}$
 set $\mathbf{P}_0 = \mathbf{I}$, $\mathbf{V}_0 = \mathbf{I}$, $\mathbf{O}_0 := \emptyset$; // initialization
for $j \leftarrow 1$ **to** m **do**
 $\mathbf{V}_0 \leftarrow [\mathbf{v}_1, \dots, \mathbf{v}_j]$; // truncated basis
 $\boldsymbol{\epsilon}_j \leftarrow (\mathbf{I} - \mathbf{W}_{j-1}) \mathbf{v}_j$; // interpolate function
 $i_j \leftarrow \arg \max_{i \in \{1, \dots, n\}} |(\boldsymbol{\epsilon}_j)_i|$; // maximum of error
 $\mathbf{P}_j \leftarrow [\mathbf{P}_{j-1}, \boldsymbol{\epsilon}_{i_j}]$; // add point to selection
end
 set $\mathbf{P} := \mathbf{P}_m$, $\mathbf{O} := \mathbf{O}_m$.

Algorithm 1: Empirical interpolation method off-line phase

In the on-line step the non-linear input function is evaluated at the m interpolation points defined by the index set \mathbf{O} and the estimate for the full vector is calculated based on (3.27).

Hydraulic Model Implementation

For the application of the EIM to the hydraulic model, the non-linear function $\mathbf{f}(\boldsymbol{\tau})$ can be either defined as the residual function r or the head-loss function Δh . While a successful implementation of the EIM to the residual function is the more efficient approach from a computational point of view, the application to the head-loss function offers a certain adaptability to the different formulations introduced in section 2.1.1. In both cases, the set of arguments $\boldsymbol{\tau}$ is defined by the hydraulic state vector \mathbf{x} and the parameter vector $\boldsymbol{\mu}$, which can contain variables like the nodal demands and pipe roughnesses.

Including the state vector as dimensions for the parameter space makes complete sampling problem infeasible by any stretch of imagination, as even small reduced order models need at a bare minimum around 20 degrees of freedom. A solution to this problem proposed by Fritzen *et al.* [2018] is to reduce the samples of the state space to those points that are used during the iterative solution process of the reduced Galerkin system during the sampling of the parameter space for the POD. Given that the reduced order model based on the EIM uses a solver trajectory close to the ones captured in the sampling space, this approach gives an accurate basis for the approximation.

3.4 Model Validation

In the development of any model the validation process plays a key role. Smith [2013] defines validation as the process of determining the accuracy with which

mathematical models quantify the physical processes of interest. In the context of this chapter the physical process is assumed to be modelled with sufficient accuracy by the hydraulic equations defined in Chapter 2 and the objective of the validation is to determine if the reduced order model approximates the high-fidelity solution with sufficient accuracy to satisfy the analysis objectives. In order to ensure this, the following section defines a repeatable process for the estimation of the approximation error in the hydraulic state vector on the parameter-space.

3.4.1 Error Definition

Depending on the choice for the meta-modelling approach, it is possible to define types of errors that allow to evaluate different aspects of the surrogate model. While the approximation error is used to assess the accuracy of the reduced order model, the projection error evaluates the quality of the reduced basis for projection based methods.

Approximation Error

In order to evaluate the approximation quality of a reduced order model an error measure has to be defined on the state vector. The simplest definition of the approximation error is given as

$$\epsilon_a = \|\mathbf{x}_{exact} - \mathbf{x}_{approx}\| \quad (3.28)$$

However, the definition of this error measure is directly dependent on the choice of the system equations and the physical units that are used in the evaluation. As described in Chapter 2 the hydraulic state of a water distribution network can be described in head-flow (2.15) or the loop-flow (2.27) domain. While the head-flow formulation gives a complete description of the system state in the flow rates \mathbf{q} [m^3/s] and the piezometric heads \mathbf{h} [m_{H_2O}], the difference in their units requires to use a relative error measure.

$$\epsilon_n = \frac{\|\mathbf{x}_{exact} - \mathbf{x}_{approx}\|}{\|\mathbf{x}_{exact}\|} \quad (3.29)$$

Further complexity is introduced through the possibility that the state of several elements in the network, depending on the combination of the input parameters, may have zero or close to zero values. In this case, normalizing the error using (3.29) may exaggerate the error of certain elements. An effective measure to counteract this, is the introduction of a constant \mathbf{c} that is added to the denominator of the relative error definition which leads to the formulation of robust normalized error:

$$\epsilon_{rn} = \frac{\|\mathbf{x}_{exact} - \mathbf{x}_{approx}\|}{\|\mathbf{x}_{exact}\| + \mathbf{c}} \quad (3.30)$$

The state for the loop-flow formulation is defined by the loop-flow vector $\mathbf{q}_c [m^3/s]$. In contrast to the head-flow formulation this description uses a reduced number of degrees of freedom and a single unit of measurement. Given (2.29) and (2.30) the flow and head vectors are a set of derived variables. The evaluation of errors in the loop-flow formulation of the hydraulic problem is straight forward in comparison. As the state vector in (2.27) is only defined in loop-flow rates there is no added complexity for the evaluation of the state error in two different units. Further, the residuals of (2.27) are the sum of the head-loss over the network loops. This means that the unit of the residuals is directly given as $[m_{H_2O}]$ for which an error tolerance of $\epsilon_h = \mathcal{O}(10^{-2})$ is acceptable. For the tolerance in the flow rates this translates to an error tolerance of $\epsilon_q = \mathcal{O}(10^{-3})$ in l/s or $\epsilon_q = \mathcal{O}(10^{-6})$ in m^3/s . Such a strong definition of the tolerance implies that an evaluation of relative errors like for the head-flow formulation is not necessary.

Projection Error

For projection based reduced order modelling techniques it is possible to evaluate the quality of the snapshot database and the reduced basis through the projection error ϵ_p . The projection error is a way to evaluate the fidelity of the sample data and the reduced basis that has been constructed from it. As described in Section 3.2 the basis for a reduced model is constructed by applying the SVD to the snapshot database \mathbf{X} and selecting a truncated basis Φ from the left singular vector matrix \mathbf{U} and the solution in the reduced system (3.9) is expressed by a linear combination of the reduced basis vectors. Projecting the solution for the state vector \mathbf{x}_{hf} from the high-fidelity model onto the reduced basis Φ results in the coefficient vector γ of the best possible approximation in that particular truncated basis.

$$\gamma = \Phi^T \mathbf{x}_{hf} \quad (3.31)$$

Using these coefficients, the best approximation on Φ can be written as

$$\mathbf{x}_p = \Phi \gamma = \Phi \Phi^T \mathbf{x}_{hf}. \quad (3.32)$$

On the basis of this projected solution the projection error can be calculated as

$$\epsilon_p = (\mathbf{I} - \Phi \Phi^T) \mathbf{x}_{hf}. \quad (3.33)$$

The projection error is strongly influenced by the number of eigenvectors that has been chosen for the reduced basis. A higher number of singular vectors will lead to a reduced projection error, as the columns of Φ are orthonormal vectors. This means that the product $\Phi \Phi^T$ will converge to the identity matrix and the projection error is converging to zero. However, from the point of reduced order modelling this is not desirable as the higher eigenvectors contain

a low amount of information and can basically be considered as noise. The selection of the size of the reduced basis is part of the design process. A good indicator for the dimension of truncated basis are the normalized singular values, that show the relative amount of signal energy that is contained in each eigenvector.

Error Metrics

As the solution of the hydraulic equations is expressed in a vector valued variable, it is preferable to evaluate the error on the basis of an error metric. Two of the most common error metrics for continuous or vector valued quantities are the mean absolute error (MAE) and the root mean squares error (RMSE). The MAE is defined as a linear score function, that is closely related to the L_1 -norm. It expresses the average magnitude of the errors with equal weights.

$$\text{MAE} = \frac{\sum_{i=1}^n |\epsilon_i|}{n} \quad (3.34)$$

The RMSE is a quadratic scoring rule for measuring the average error magnitude. In contrast to the MAE it is closely related to the L_2 -norm or the sum of squares.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\epsilon_i)^2}{n}}. \quad (3.35)$$

While both metrics express the average error in the units of the input variable they offer different interpretations of the error. While the MAE weighs all errors equally, the RMSE puts higher weight on bigger errors and is closely related to the signal energy.

Due to the specific distribution of errors in the case of the reduced order hydraulic model, the evaluation on the basis of a L_2 -norm has been chosen over the L_1 -norm. The error between the high-fidelity state vector and its reconstruction from the reduced state vector usually contains a large number of states that are approximated with reasonable accuracy and a relatively small number of states with high deviations. Choosing the RMSE over the MAE gives a higher weight to the actual deviation between the full order model and its low order estimate.

3.4.2 Validation Procedure

For the model validation scheme the approximation error between the full order model and the surrogate is evaluated on a set of random parameter vectors $P = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m\}$ that are sampled from the parameter space \mathcal{P} , using the RMSE. The samples are taken from a multivariate uniform distribution $\mathcal{U}(\boldsymbol{\mu}_{min}, \boldsymbol{\mu}_{max})$. In the following, the parameter set P will also be referred to as the *validation points* or the *validation set*. In order to be repeatable

and comparable over all surrogate models, the sample is drawn using a pseudo random number generator with a common seed.

The number of evaluation points that are chosen for the validation has a significant influence on the performance and the accuracy of the procedure. Choosing too few points leads to a wrong error estimate, while choosing too many increases the computational effort without improving its accuracy. To answer the question for the appropriate size of the validation set, its influence on a number of models has been investigated and the results are shown in Figure 3.1. The evaluation is based on the realistic network model defined in section 2.4.3 and the surrogates for the evaluation are based on the three Cartesian refinements of the snapshot matrix with 121, 441 and 1681 elements taken from a parameter space spanned by the two defined demand patterns. Three refinements of the reduced model are then evaluated with reduced basis that contain 10, 40 and 120 degrees of freedom. On the y-axis the average

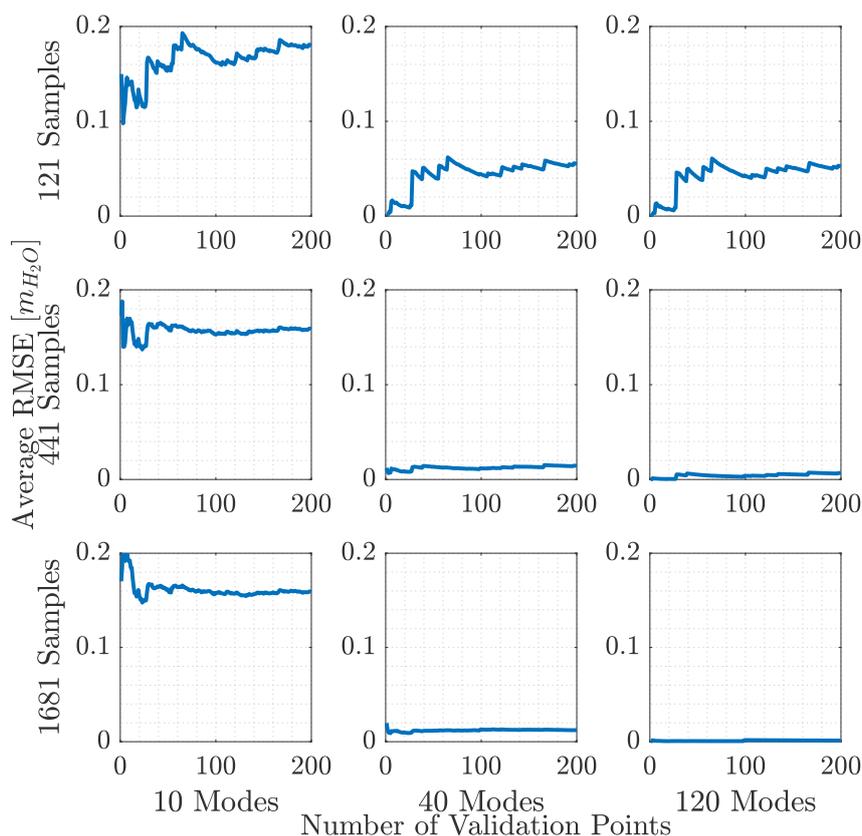


Figure 3.1: Approximation error convergence for 9 reduced order models as a function of the validation set size. The surrogates are based on three realizations for the Cartesian sampling of the snapshot matrix $n_{samples} = [121, 441, 1681]$ and with three different degrees of freedom $n_{modes} = [10, 40, 120]$.

RMSE is plotted as a function of the number of validation point on the x-axis. The figure shows the development of the of the error with the addition of further validation points. It is obvious that both, the sample size and the degrees of freedom in the model influence the convergence of the error plots. Disregarding the actual error value, it seems that in most cases the average RMSE starts to converge between 50 and 100 validation points. Based on this result the size of the validation set is fixed to 100 points.

3.5 Adaptive Parameter Sampling

The selection of simulation points in the parameter-space is a crucial task in sampling based reduced order models. In the application of interpolation methods they are used as support points at which the interpolation conditions hold and for the POD they define the snapshots from which the reduced basis is calculated. Both cases demand a sufficiently rich and detailed data set. For problems with few parameters simple structured or random sampling methods like Cartesian grids or Monte Carlo sampling are feasible approaches. However, with raising numbers of parameters these methods become computationally expensive as the number of sampling points grows exponentially with the dimension d of the parameter-space. Further, the use of such methods may result in either over- or under sampling of the domain in certain regions. Although the dimensional effect can be mitigated through the use of advanced sampling strategies, like sparse grid sampling and latin hypercube sampling, for dimensions $d > 10$ the sampling starts to becomes infeasible [Benner *et al.*, 2013]. In such cases more advanced and problem-aware adaptive sampling methods have to be applied.

One possible approach is the application of greedy sampling strategies. Given a reduced order, a parameter value is determined for which the approximation error is maximal. This parameter is then selected as new sampling point. The method has been introduced to reduced basis methods by Prud'Homme *et al.* [2002] and has been applied in the context of POD modelling [Bui-Thanh *et al.*, 2008; Haasdonk et Ohlberger, 2008; Benner *et al.*, 2015a] and rational interpolation methods [Druskin *et al.*, 2010]. Another approach, proposed by Bond et Daniel [2007] uses the local sensitivities of the state vector with respect to the parameters in order to judge if a change in parameters will result in a state that can not be represented with acceptable accuracy in the current reduced basis. State changes are modelled on the basis of a first-order Taylor series expansion for small changes in the parameter on each sampling point and the sensitivities are obtained through solving a sparse linear system. Other methods like the one presented by Borggaard *et al.* [2014], combine adaptive sampling strategies with basis interpolation for parametric reduced order models. Based on the discretization of the parameter-space

with the centroidal voronoi tessellation used by [Du et al. \[1999\]](#) and [Burkardt et al. \[2006\]](#), [Lombardi et al. \[2009\]](#) introduced an adaptive greedy sampling algorithm on the basis of the residual function.

The here proposed algorithm uses a greedy algorithm on the basis of the Delaunay triangulation of the parameter-space using the Leave-One-Out (LOO) errors. Given be the set of parameters $\mathbf{P} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_N\}$, on the parameter-space \mathcal{P} . The leave-one-out error $\epsilon_{LOO}(\boldsymbol{\mu}_i)$ in each point $i = 1, \dots, N$ is calculated from a reduced parameter set $\mathbf{P}^i = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{i-1}, \boldsymbol{\mu}_{i+1}, \dots, \boldsymbol{\mu}_N\}$, where the i -th element has been removed. Given a reduced order model

$$\hat{\mathbf{A}}(\boldsymbol{\mu}) \hat{\mathbf{x}}(\boldsymbol{\mu}) + \hat{\mathbf{B}}(\boldsymbol{\mu}) = \mathbf{0} \quad (3.36)$$

constructed on the parameter set \mathbf{P}^i , with the objective function

$$\mathbf{y}(\hat{\mathbf{x}}, \boldsymbol{\mu}) = \hat{\mathbf{C}}(\boldsymbol{\mu}) \hat{\mathbf{x}}(\boldsymbol{\mu}), \quad (3.37)$$

the objective function \mathbf{y}_i for the missing parameter $\boldsymbol{\mu}_i$ is estimated and the leave-one-out error is defined as

$$\epsilon_{LOO}(\boldsymbol{\mu}_i) = \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_{L_2}. \quad (3.38)$$

Using the Delaunay point set triangulation this allows to calculate the weighted LOO error $\epsilon_{LOO}(\mathbf{t}_i)$ over the vertices \mathbf{v} for each triangle $\mathbf{t}_i \in \mathbf{T}$, as

$$\epsilon_{LOO}(\mathbf{t}_i) = \frac{A(\mathbf{t}_i)}{3} \sum_{j=1}^3 \mathbf{t}_i(\mathbf{v}_j). \quad (3.39)$$

In the application to the hydraulic model the objective function $\mathbf{y}(\mathbf{x}, \boldsymbol{\mu})$ is given by the head-loss function $\Delta \mathbf{h}(\mathbf{q}(\boldsymbol{\mu}))$. The greedy approach selects the triangle with the biggest weighted error

$$\hat{\mathbf{t}} = \arg \max_{\mathbf{t}_i} (\epsilon_{LOO}(\mathbf{t}_i)), \quad (3.40)$$

from which the next parameter combination for sampling $\boldsymbol{\mu}_{N+1}$ is chosen as its weighted centroid

$$\boldsymbol{\mu}_{N+1} = \frac{1}{\epsilon_{LOO_{sum}}} \sum_{j=1}^3 \epsilon_{LOO_j} \boldsymbol{\mu}_j. \quad (3.41)$$

In two dimensions the final model-aware adaptive algorithm can be described as follows:

1. Create a set of parameters $\mathbf{P} = \{\boldsymbol{\mu}_i\}_{i=1}^k$ from the parameter space \mathcal{P} .
2. Calculate the set of leave-one-out errors $\epsilon_{LOO}(\boldsymbol{\mu}_i)$ for each parameter combination in the set \mathbf{P} .

3. Create the Delaunay point set triangulation \mathbf{T} on the parameter point set \mathbf{P} .
4. For each triangle i in \mathbf{T} , calculate the mean error weighed leave-one-out error $\epsilon_{LOO}(\mathbf{t}_i)$.
5. Add weighted centroid of the triangle with the biggest error $\epsilon_{LOO}(\mathbf{t}_i)$ to the parameter set \mathbf{P} .
6. Perform the high-fidelity simulation for the added parameter combination to the snapshot database.
7. If the error is bigger than the tolerance, restart from step 2.

The algorithm can be extended to higher dimensions without loss of generality.

In its practical application the performance of this algorithm can be influenced by a number of parameters. Some of the most important are the choice of the starting discretization, the error measure on which the LOO error is based and the reduced model used for its evaluation.

3.6 Evaluation on the medium size network

The objective of the case studies presented in this section is to apply the previously introduced reduced order modelling techniques on the medium size network depicted in Figure 2.4 and evaluate the results on their accuracy and performance. With respect to these two aspects, a focus will be set on investigating the influence of the parameters for each method. As detailed in Section 2.4, the network contains nearly 4000 elements in the head-flow formulation. Expressing the system in the more efficient loop-flow formulation from (2.27) already reduces the number of degrees of freedom to 348. The model is evaluated during the peak demand hour under steady state conditions. The consumption at the demand nodes is parametrized with a two-dimensional parameter-space. These two parameters are chosen to be two dimension-less demand multipliers $\boldsymbol{\mu} = \{m_1, m_2\}$, corresponding to the demand patterns defined in the model. They are applied to two groups of demand nodes in the entire network through the base-demand matrix \mathbf{K} . The evaluation of the model errors follows Section 3.4.1. and a general threshold for the acceptable model error is defined at a mean RMSE of $10^{-2} [m_{H_2O}]$ following the description in Section 3.4. The error threshold is chosen on the basis of the solver tolerance of the hydraulic equations. However, it has to be mentioned that this value is theoretical in nature, as the measurement accuracy in practical applications is more in the interval of $0.1 [m_{H_2O}]$ to $1 [m_{H_2O}]$. In effect, this allows for the use of less accurate and more efficient reduced models in such cases.

In the first study, the application of interpolation meta-models is investigated. The second study evaluates the application of the POD and the Galerkin projection to the system. The third study shows the implementation of the EIM in order to reduce the computational cost of the non-linear term in the residual function. A final study is presented that investigates the benefits of the adaptive sampling strategy.

The Cartesian discretisations of the parameter-space used throughout this chapter use an equidistant division for each parameter starting with 11 points and roughly doubling with each increase to 21 points, 41 points and so on. This approach allows to reuse sampling points from previous discretisations and helps to ease the computational effort.

3.6.1 Evaluating Interpolation

The interest of using interpolation and regression methods for the construction of a meta-model clearly lies in its easy implementation and fast evaluation. In order to evaluate the possible performance, a selection of interpolation methods are tested on the regular Cartesian samplings of the parameter space. Figure 3.2 shows the average RMSE for the simple linear, bi-linear, bi-cubic and radial basis function interpolation using a cubic element, together with the error threshold. The approximation error is given as a function of the resolution in the snapshot database.

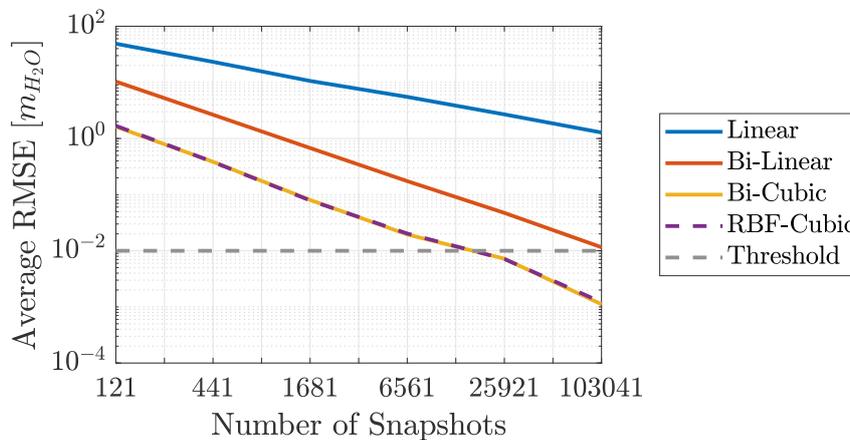


Figure 3.2: Average RMSE in the validation set as a function of the snapshot database size using linear, bi-linear, bi-cubic and RBF interpolation with a cubic element.

The linear interpolation applied in this case uses the three closest neighbouring points in the parameter-space as support. It is at the same time the fastest, but also the most inaccurate approximation. In comparison the bi-linear interpolation is able to reduce the average RMSE by an order of

magnitude, especially for higher amounts of snapshots. This increase in accuracy can be explained by the addition of the quadratic interdependency to the regression polynomial. A similar improvement can be observed with the application of the bi-cubic interpolation. The use of more complex spline elements for the interpolation was tested, however, the improvement over the bi-cubic interpolation was marginal and is not visualized here. Figure 3.2 also shows that the RBF interpolation on the basis of the cubic elements gives nearly the same result as the bi-cubic approach. Testing the interpolation with the most popular RBF elements shows that the cubic element gives the most accurate results.

While the bi-linear and bi-cubic interpolation are well suited for the use with data on a regular grid like the basic Cartesian discretizations, the RBF interpolation is also capable to interpolate data on an irregular grid, like the one created using the adaptive sampling of the parameter-space in Section 3.5. This benefit comes at a certain cost, as an additional preprocessing step is added with the fitting process for the RBF kernels.

In general it can be shown, as expected, that the accuracy of all interpolation methods increases and the average RMSE converges with an increasing number of sampling points. Especially the bi-cubic and the RBF interpolation with the cubic elements give an error that is usable in practice. However, at sample size of $\mathcal{O}(10^4)$ and $\mathcal{O}(10^5)$ for a two-dimensional parameter-space it is obvious that the extension of the approach to higher-dimensional applications would become infeasible from a computational point of view.

3.6.2 Evaluating POD

In this study the POD method is applied to the hydraulic network equations. The derivation of the reduced Galerkin system is described in Section 3.2. The section is divided into the application of the POD with a global and a local reduced basis. Both cases investigate the influence of the resolution in the Cartesian parameter-space samplings and the order of dimensional reduction in the reduced model on the approximation accuracy and improvement of computational time.

Global Basis

In the first step, the global reduced basis is constructed for each of the Cartesian samplings. Figure 3.3 shows the ordered, normalized singular values on the y-axis for each discretization (121, 441, 1681 and 6561 samples) taken from the SVD on the snapshot database. As the high-fidelity model has a number of 348 degrees of freedom, this is the maximum index for all singular value plots except the one with 121 samples. For the snapshot matrix with 121 high-fidelity simulations, a sharp drop in the singular values can be observed

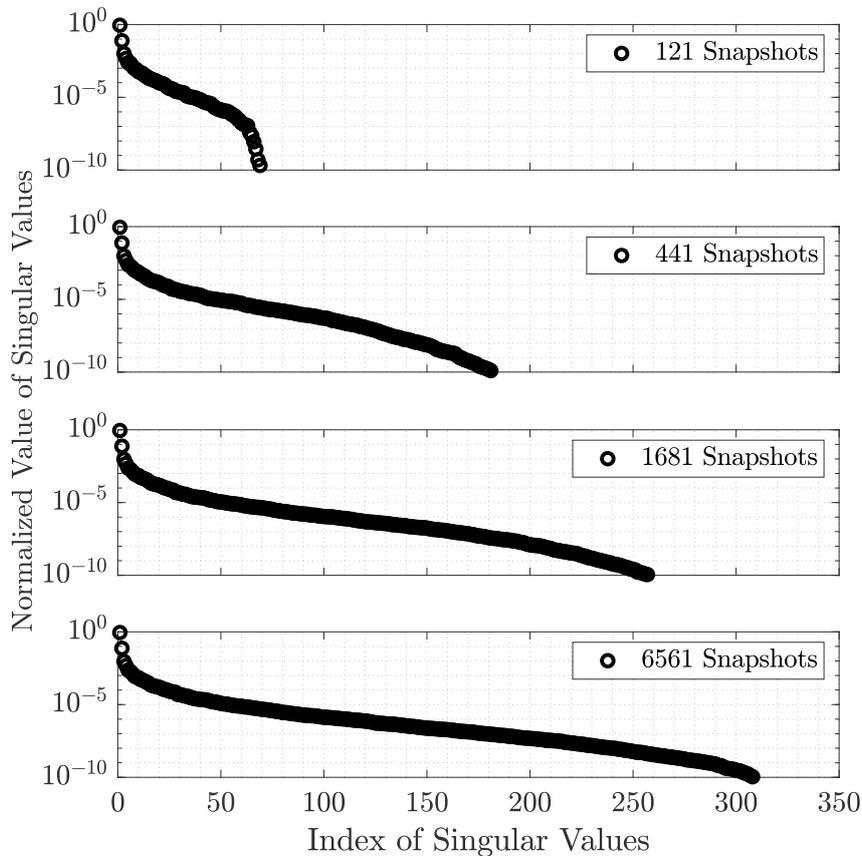


Figure 3.3: Normalized singular values for Cartesian samplings of the parameter-space \mathcal{P} with 121, 441, 1681 and 6561 high-fidelity simulations.

for elements higher than 70. This behaviour is not replicated for the more detailed discretizations, suggesting that this is due to a poor fidelity of the snapshot database. Figure 3.3 clearly shows that for all the more detailed samplings the singular values converge relatively slowly to zero. However, no sharp drop in the singular values can be observed. Further, with the addition of more discretization points to the snapshot database, the singular values decline more and more slowly. For the samplings with 25921 and 103041 no observable improvement could be achieved, so they are not included in Figure 3.3. The fact that for most of the samplings no sharp drop in the singular values is observable means that it is not possible to have a definite estimate for the maximum number of modes that should be used in the reduced model. It follows then that for the subsequent investigations the degrees of freedom in the reduced model are also treated as a parameter that has to be investigated.

The results of the basic evaluation in the global reduced basis and the global reduced model are depicted in Figure 3.4. It shows the average RMSE as a function of the number of the reduced degrees of freedom, evaluated on the validation set. While the full lines represent the error in the reduced

order model evaluated through the approximation error, the dashed lines show projection error, giving the error of the theoretically best approximation that is achievable with the given reduced basis. The threshold for the acceptable approximation error is represented by the gray, dashed line.

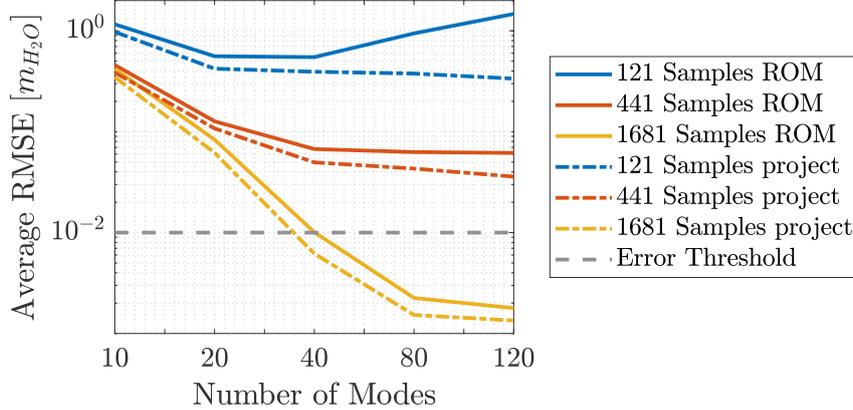


Figure 3.4: Average root mean square error for the approximation error and the projection error in the validation points as a function of the degrees of freedom in the global reduced order model and for different Cartesian sampled snapshot databases.

As the plot of the singular values already suggested, the reduced model with the 121 element snapshot matrix lacks the fidelity to approximate the full order model with any accuracy. This becomes apparent twofold, in the stagnating projection error and the error of the reduced order model that is not converging for higher degrees of freedom. For the reduced order models based on a more refined sampling, the projection error and the approximation error behave roughly in the same manner. As expected, the projection error acts as a lower bound for the approximation error, but the reduced order models based on the Galerkin projection stay within a reasonable margin to it. Adding further sampling points to the snapshot database helps to improve the accuracy of the reduced order model significantly. The effect of the additional data becomes especially apparent for the reduced models with degrees of freedom higher than 20. However, the effect seems to converge as the improvement for the accuracy between the database with 1681 and the one with 6561 elements does not represent the increase in computational effort in the off-line phase.

Having established, that the reduced order models on a global basis can approximate the high-fidelity model with acceptable accuracy, a closer look is cast on the benefit in computational effort. To answer this question and explain the results additional information from the solution process has to be reviewed. Figure 3.5 shows the average relative solution time of the reduced order model compared to the high-fidelity model as a function of the number of modes. While all of the observed models are computationally more efficient

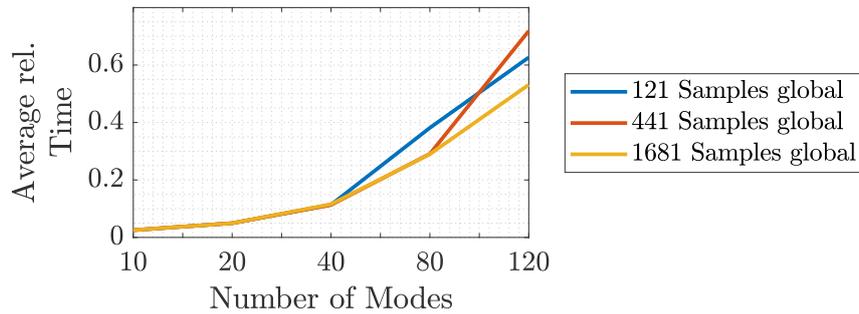


Figure 3.5: Average relative number of iterations for the solution of the global reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.

than the full order model, there is a big difference in the computation time with respect to the reduction in the degrees of freedom. The observed reduction in computational time is the same for the models with 10, 20 and 40 modes. For higher mode numbers the curves diverge. The number of function evaluations is closely related to the computational time. Figure 3.6 gives the average relative number of function evaluations, that have been performed by the non-linear solver. Comparing this plot to the average relative evaluation time the strong correlation is apparent. This is an expected result, as the main computational

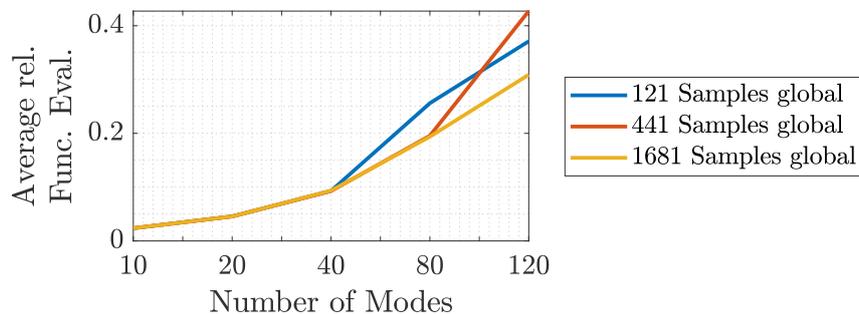


Figure 3.6: Average relative number of iterations for the solution of the global reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.

effort in the solution of a non-linear system of equations lays in the evaluation of the residual function. Most of these evaluations are linked with the numerical calculation of the Jacobian matrix. Due to the dimensional reduction in the hydraulic model, the number of elements in the Jacobian matrix is greatly reduced, making its evaluation far more efficient. The average relative number of iterations of the non-linear solution process is shown in Figure 3.7. Once again the results are the same for lower degrees of freedom. However, for higher degrees of freedom the reduced order models that are based on a more sparse snapshot database need more iterations to converge. This explains very well

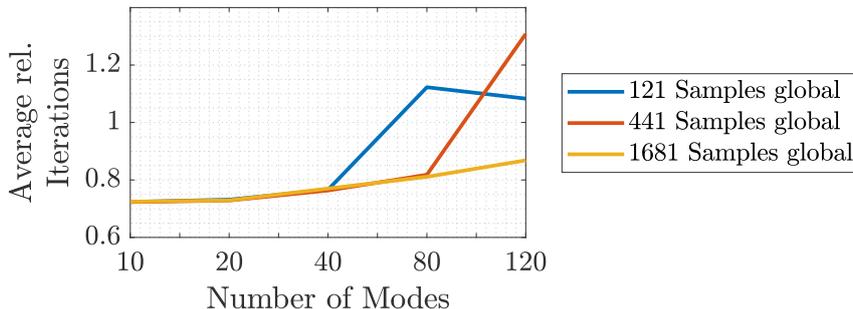


Figure 3.7: Average relative number of iterations for the solution of the global reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.

the difference in evaluation time for models with a higher number of modes. In contrast to the relative average time and the average relative number of function evaluations, the reduction in solver iterations is relatively low.

From the average RMSE plot in Figure 3.4 it can be concluded that the reduced model with 40 degrees of freedom on the 1681 snapshot database has an acceptable error. For this model the computational time has been reduced to only 10% of the high-fidelity model.

Local Basis

While the global POD creates one reduced basis that is applied on the whole parameter-space, the local POD creates a number of individual local reduced basis. This local basis is constructed using a selection of snapshots that are closest to the parameter that is approximated. The definition of distance can take a number of forms, as it may be defined either on the parameter or the solution space and use different definitions for the distance. For the following study, the distance is defined by an Euclidean norm on the parameter-space. The local reduced basis for a reduced order model with k degrees of freedom is created using a snapshot database containing the state vectors of the k closest parameter combinations.

In Figure 3.8 the average RMSE of the approximation error and the projection error is shown as a function of the number of modes in the reduced local model for a selection of Cartesian samplings of the parameter space. The projection error for the reduced model based on 121 evaluations of the full order model shows that the snapshot database does not have the fidelity to approximate the hydraulic state. The figure shows that it is possible to construct a model with 80 degrees of freedom on the 1681 snapshot database that satisfies the error constraint of $10^{-2}[m_{H_2O}]$, for the use of the local surrogate model in a hydraulic application. For the further discussion on the computational efficiency of the local reduced model, additional information on the

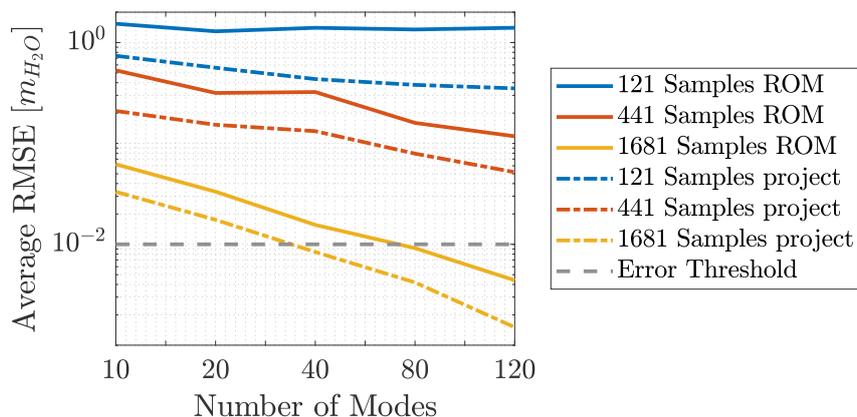


Figure 3.8: Average root mean square error for the reduced order model and the projection of the high-fidelity solution onto the reduced basis in the validation points as a function of the degrees of freedom in the local reduced order model and for different Cartesian sampled snapshot databases.

solver performance is presented in the following plots. In Figure 3.9 the average relative evaluation time for the a particular reduced order model compared to the high-fidelity model is depicted as a function of the systems reduced order. Once again, a significant increase for the computational time can be observed

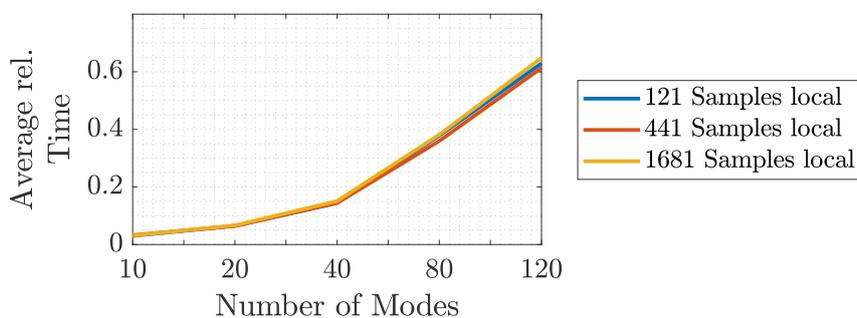


Figure 3.9: Average relative evaluation time for the solution of the local reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.

with the increases in the modelled degrees of freedom, proving the benefit of the dimensional reduction on the basis of a local model. Figure 3.10 shows the average relative number of function evaluations during the solution of the local reduced order model as a function of the degrees of freedom. As before, the strong correlation between the computational time and the number of function evaluations confirms that the evaluations of the residual function is responsible for the bulk of the computational effort. For the average relative number of iteration in Figure 3.11 it can be seen that the local reduced model stays within a 10% margin around the high-fidelity model.

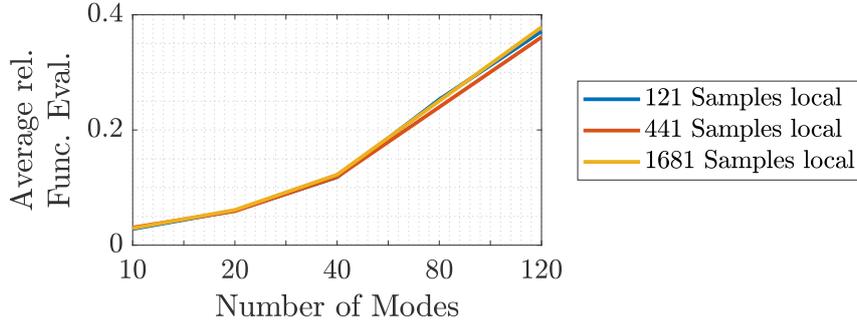


Figure 3.10: Average relative number of function evaluations for the solution of the local reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.

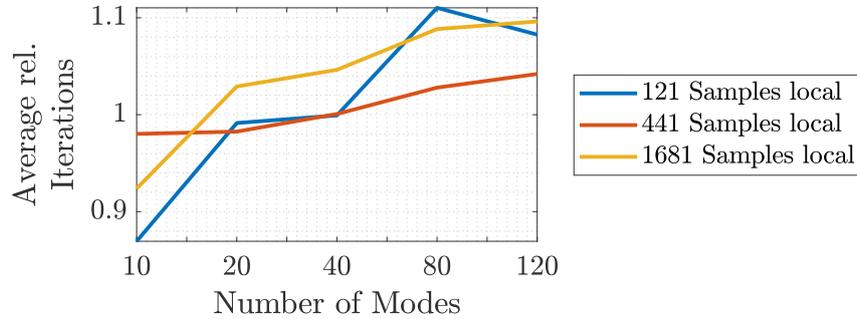


Figure 3.11: Average relative number of iterations for the solution of the local reduced order model in the validation points as a function of the degrees of freedom and for different Cartesian sampled snapshot databases.

In general the viability of the local reduced modal has been demonstrated in this section. However, the application of the method in the presented study is flawed in two points. First, due to the local construction of the snapshot database, the construction of the reduced basis has to be performed during the on-line step, adding to the computational effort mainly through the application of the SVD that has not been taken into account here. This could be improved through the definition of predefined sub-domains as proposed by [Amsallem et Haasdonk \[2016\]](#); [Haasdonk et Ohlberger \[2008\]](#) or [Maday et Stamm \[2013\]](#). Second, by selecting the same number of elements for the snapshot database as for the degrees of freedom, the normalized singular values of the higher modes are not significant and information that can be discarded as noise. This can be interpreted as the effective degree of freedoms being lower than the size of the reduced basis. A solution for this problem is to build the reduced basis from a database that contains more snapshots than degrees of freedom and ensure to contain at least k significant modes. The additional computational effort of this approach would be mitigated through the use of predefined sub-domains by making it an off-line step.

Comparing Global and Local POD

Having shown that both the global and the local approach to the construction of a reduced basis and a reduced order model can result in a viable model for the use in hydraulic applications, the objective of this paragraph compares their performance. Figure 3.12 shows the results for the accuracy of the global and the local reduced order models as a function of the model complexity and the resolution of the parameter-space. The general observations of the

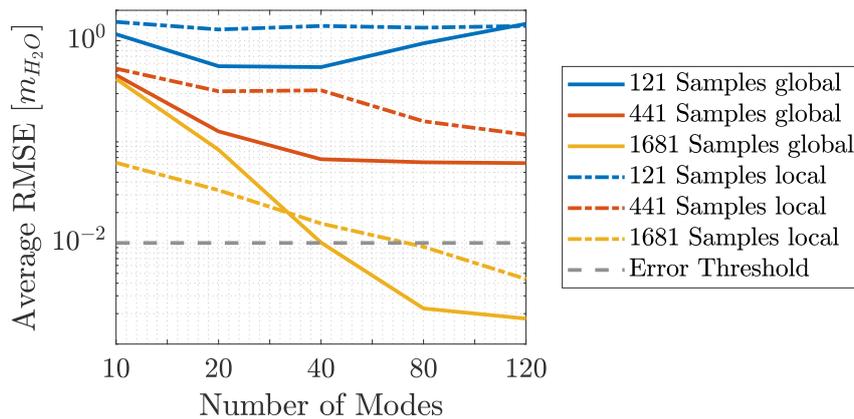


Figure 3.12: Average root mean square error for the reduced order model and the projection of the high-fidelity solution onto the reduced basis in the validation points as a function of the degrees of freedom in the local reduced order model and for different Cartesian sampled snapshot databases.

reduced order model on local basis are the same as for the use of a global basis, as the use of a more detailed sampling of the parameter space and higher degrees of freedom improve the model. But, although both approaches converge for a higher number of modes, the convergence rate of the global model is significantly better. For the very low order models with mode numbers between 10 and 40, modelled on the 1681 snapshot database, the local model has an advantage over the global model. However, the lowest order model with an error below the threshold is the global model with 40 modes modelled on the 1681 snapshot database.

Comparing the projection error of both methods in Figures 3.4 and 3.8, one can observe, that the difference between the approximation error and the projection error of the local approach is significantly bigger than for the global model. A possible explanation for this may be the inclusion of non-significant modes for the local model, that has been discussed previously.

For the computational time and the number of function evaluations, both models show a significant increase with the increases in the modelled degrees of freedom. However, no substantial improvement of the local model over the global model can be seen, despite the convergence issues of the global

model for higher mode numbers. Comparing the average relative number of iterations in Figures 3.11 and 3.7 a clear benefit for the global model can be identified. While the local model stays within a $\pm 10\%$ margin around the full order model, the global model has consistently lower iteration numbers for the model created on the 1681 element snapshot database.

In conclusion, in this study, the global approach gives a more accurate reduced model paired with a higher computational efficiency. However, as indicated previously, improvements to the local approach are possible and might change this assessment.

As a side note, the results for both, the global and the local reduced order model, have been obtained under the condition that the direct evaluation of the Jacobian matrix is not possible and it has to be obtained through numerical differentiation. As the main factor in the reduction of the computational effort has been identified as the reduced number of function evaluations during the numerical evaluation of the Jacobian matrix, this benefit does not hold with the same significance in cases where its direct evaluation is possible. While this assumption is not always the best interest for an efficient solution of the hydraulic system, there exist practical applications where the use of the numerical Jacobian is inevitable. One example is the use of the Darcy-Weisbach formulation for the head-loss function, as it is not continuously differentiable at all points.

3.6.3 Evaluating EIM

The application in the previous section has successfully shown, that the dimensional reduction in the state vector can be used to reduce the computational time. This reduction is mainly due to the reduced number of function evaluations in the calculation of the Jacobian matrix. In contrast, the focus of this section is on the reduction of the computational cost for each single function evaluation through interpolation methods. This study applies the Empirical Interpolation Method, introduced in Section 3.3 to the non-linear head-loss function, in order to speed up the evaluation of the function residuals.

As the EIM is a sampling based method, in a first step, the parameter-space is sampled by Cartesian grids of increasing resolutions in order to build a reduced basis in the head-loss domain. Figure 3.13 shows the normalized singular values for the three ensemble of 441, 1681 and 6561 snapshots. For the 441 snapshot matrix their value drops below 10^{-8} around mode 150 and modes higher than 700 can effectively be discarded as noise. For the more detailed samplings with the 1681 and the 6561 snapshot matrix the singular values drop below 10^{-8} for modes higher than 200. While for the 1681 snapshot database modes above 700 can be considered as noise, this threshold lays at 950 modes for the 6561 snapshot database. Similar results can be observed for more detailed samplings that are not presented here.

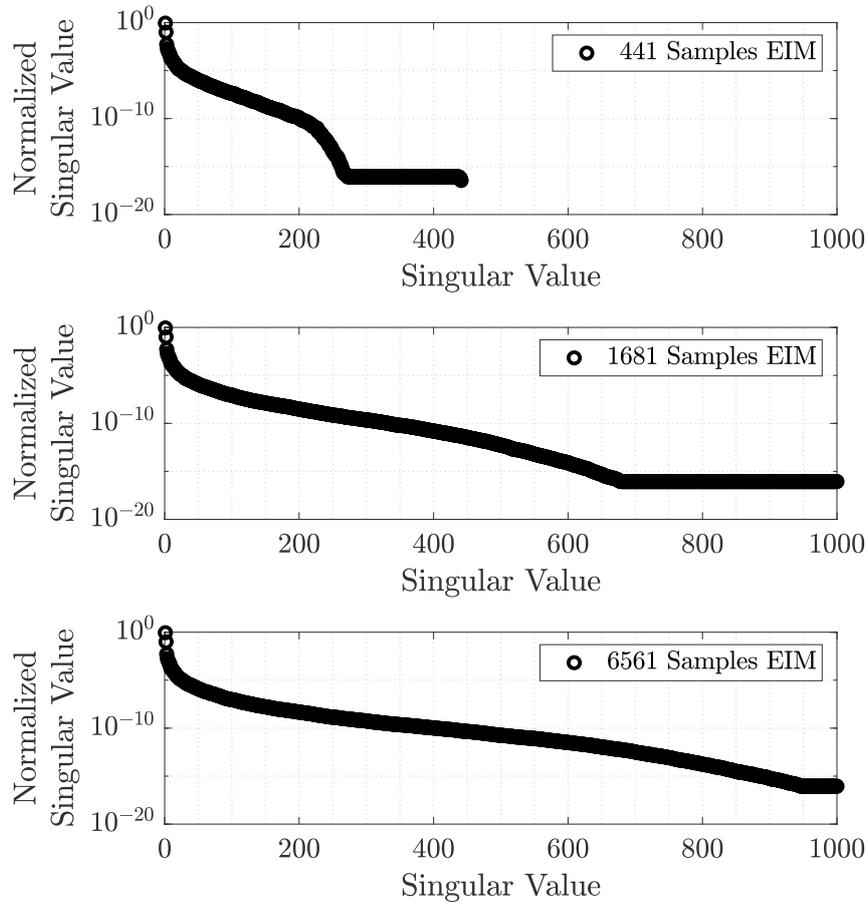


Figure 3.13: Normalized singular values for the reduced basis built on the 441, 1681 and 6561 snapshot database in the head-loss domain.

The evaluation of the interpolation models built on these reduced basis is visualized in Figure 3.14. It shows the approximation and the projection error of the EIM of the head-loss in the final states of the solution for the validation set. From the projection error of the 441 element discretization it becomes apparent, that this model does not contain enough information to approximate the non-linear function with any accuracy. The models on the more detailed sampling databases behave relatively similar to each other. All three models produce an acceptable approximation error for the use of more than 200 modes. This observation is in line with the results from the normalized singular values that drop below 10^{-8} around singular value 200. While the projection error suggests a slightly better performance for the bigger snapshot databases, this effect can only be seen in part for the interpolation results.

Attempts to solve the reduced order model using these interpolation models did not lead to satisfactory results, as the non-linear solver does not converge. Projecting the head-loss vectors taken during the solution steps of the reduced

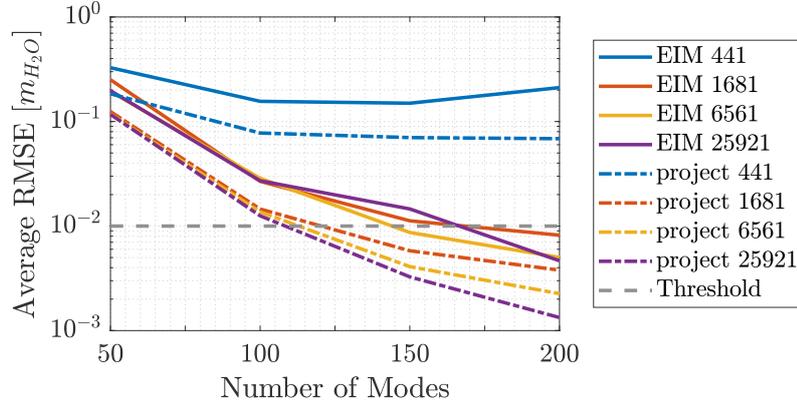


Figure 3.14: Projection and empirical interpolation of the head-loss function in the converged solution of the hydraulic system.

model without application of the interpolation method onto the reduced basis in Figure 3.15 shows that these states can not be approximated with any accuracy. This is easily explained by the fact that the reduced state vector has

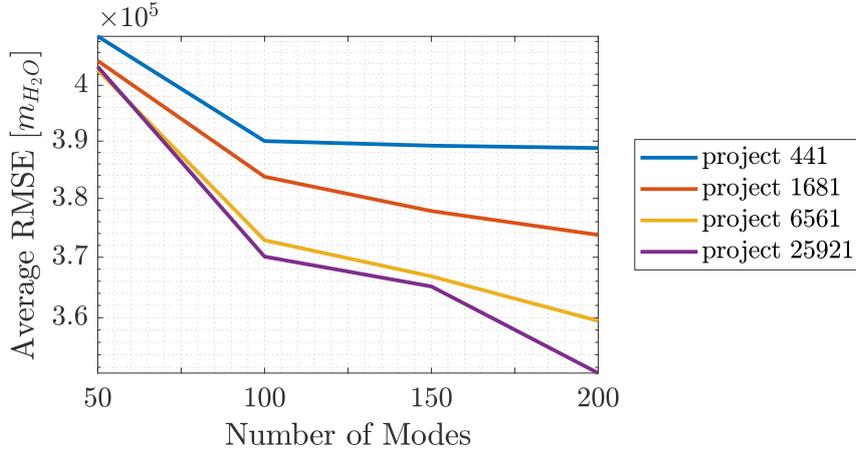


Figure 3.15: Projection of the head-loss function onto the reduced basis in the solver steps of the hydraulic system equations.

to be treated as a parameter during the solution process, which means that the parameter space for the head-loss database has to be extended. Adding the reduced state vector to the parameter-space would increase the dimension of the application to a size that would make the sampling problem intractable. As an alternative, [Fritzen et al. \[2018\]](#) proposed to add a reduced selection of realizations in the parameter-space, given by the reduced state vectors that are used during the solution process during sampling the snapshot database. The basic idea is, that the reduced solver for non-sampled parameters $\boldsymbol{\mu}$ follows a similar trajectory that lies within a reasonable neighbourhood of this extended

snapshot database.

Including the sampling in the reduced state vector in this way, increases the snapshot database for the sampling in 1681 parameter combinations to a total number of 36411 snapshots. In Figure 3.16 the normalized singular values for the SVD of this extended snapshot database are visualized. Compared to

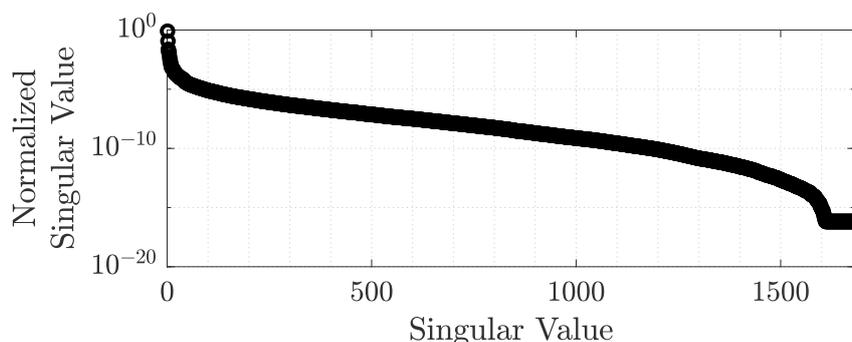


Figure 3.16: Ordered normalized singular values for the sampling of the extended snapshot database that includes the solver trajectory.

the previous samplings in the final state it is apparent, that the information content increases considerably, as the normalized singular values drop below 10^{-8} around the 900-th singular value compared to the 300-th in the previous attempt.

In order to validate this extended approach, the approximation error of the interpolation and the projection error on the reduced basis for the head-loss are evaluated for all the steps during the solution process. Figure 3.17 shows the results for the 1681 element discretization in the parameter-space of the demand multipliers and the basis built on the according 36411 element sampling matrix. The solutions of the validation set are projected on the extended reduced basis, together with the intermediate steps from the non-linear solver. It can be seen that, for an interpolation of the full head-loss function with acceptable accuracy, about 1000 modes are needed. This implies the evaluation of the original head-loss function in 1000 support points. The time threshold in Figure 3.17 indicates the number of modes that could maximally be used before the matrix vector multiplication in the EIM becomes computationally more demanding than the direct evaluation of the full head-loss function. The indicated time threshold is a conservative estimate, as the computational effort for the evaluation of the support points in the head-loss function is not included for its calculation.

Attempts to solve the reduced order model based on this refined interpolation of the head-loss function, despite the lack of savings in computational effort, still did not result on a converging solution. Following the residuals and the projection error of the head-loss function during the solution process shows that residuals become unstable after some initial steps and that the projec-

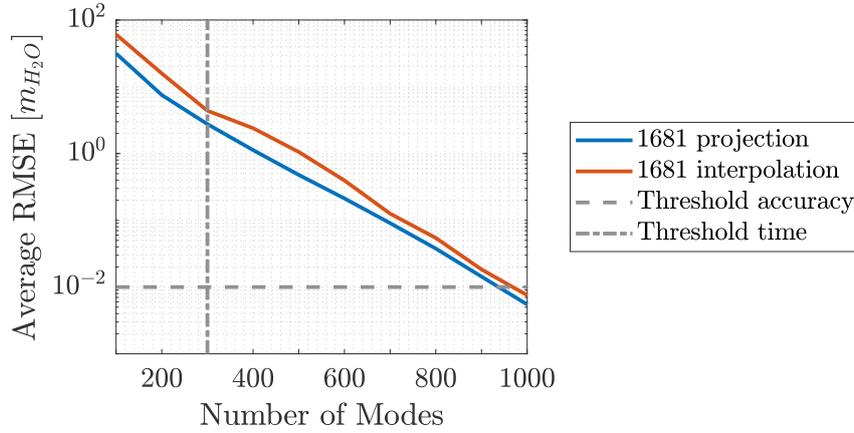


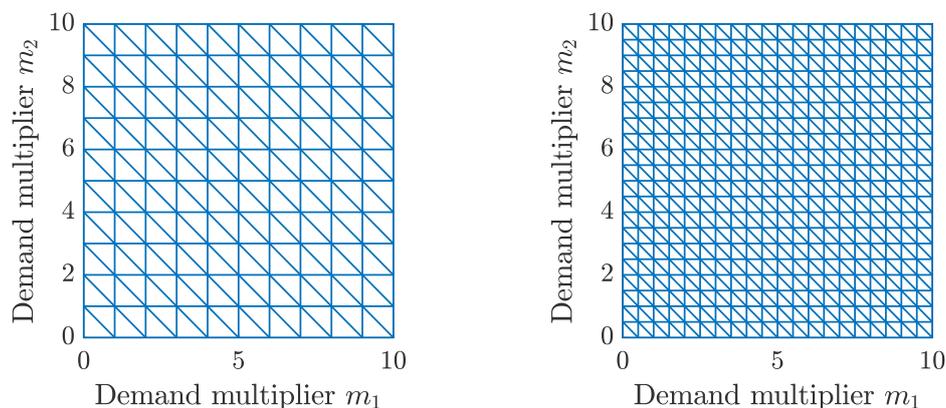
Figure 3.17: Projection of the head-loss function in the solver steps, using an extended snapshot matrix.

tion error increases continually. A possible interpretation of this observation is that, while the error of the non-linear function is acceptable, the solver is running out of the pre-sampled region where the interpolation of the head-loss function becomes increasingly inaccurate, leading to the diverging residuals. Attempts for a more stable solution method could be done by reformulating the reduced model as a constrained optimization problem that limits the decision values to the pre-sampled region. However, this approach would not improve the computational efficiency. Other strategies that could be followed in future applications, would be the inclusion of information on the derivative with respect to the reduced variables or the preconditioning of the sample data.

3.6.4 Evaluating Adaptive Sampling Strategy

The previous studies on the performance of reduced order models have been performed using Cartesian grid parameter-space discretization of increasing resolution. Objective of the study presented in this section is the investigation of the adaptive sampling strategy introduced in Section 3.5 and the evaluation of its influence on the performance and accuracy of the hydraulic reduced order model. The focus of this investigation is on the choices for the implementation details. As mentioned in the introduction of the adaptive sampling algorithm, the main factors in its application are given by the choice of the starting discretization, the model used in the estimation of the LOO errors and the model used for the reconstruction. The quality of the adaptive sampling is mainly evaluated through the average RMSE in the state vectors, but also in the evaluation of the triangulation of the parameter-space.

In Figure 3.18 two possible Cartesian starting discretizations for the two-dimensional parameter-space are shown. The grid in Figure 3.18a uses 121



(a) Cartesian starting grid discretization with 121 sampling points. (b) Cartesian starting grid discretization with 441 sampling points.

Figure 3.18: Delaunay starting triangulations for Cartesian grid discretizations of the parameter-space \mathcal{P} .

points and Figure 3.18b shows the 441 point grid. While it would be preferable to let the adaptive sampler choose as many points as possible, starting from a minimal sample of 4 elements, simulations have shown that the algorithm can ignore certain regions of the parameter-space.

The LOO errors are evaluated on the basis of the projection errors in the head-loss function. As discussed previously, the head-loss error is a convenient indicator for the quality of an approximation. Using the projection error over the approximation error greatly speeds up the calculation of the LOO error, as it is evaluated through vector-matrix multiplications and the time consuming evaluation of the non-linear reduced order system can be evaded. Further, as the previous applications have shown, it still provides a decent surrogate for the approximation error.

In the following the adaptive sampling strategy is discussed on the basis of a global and a local reduced basis, followed by the discussion on the use of global and local reduced order models on the adaptive snapshot databases.

Global Adaptive Sampling

Starting from the 121 element discretization in Figure 3.18a, Figure 3.19 shows the development of the average RMSE in the leave-one-out projection error using a global reduced basis with 80 modes and one with 120 modes. The sampling approach based on the 80 mode reduced basis manages to reduce the LOO error below $10^{-2} [m_{H_2O}]$. However, the error converges and is not reduced for the addition of sampling point above $N_{samples} = 200$. While the adaptive sampling on the 120 mode reduced basis can reduce the LOO error further than the 80 mode approach, converges as well and does not improve

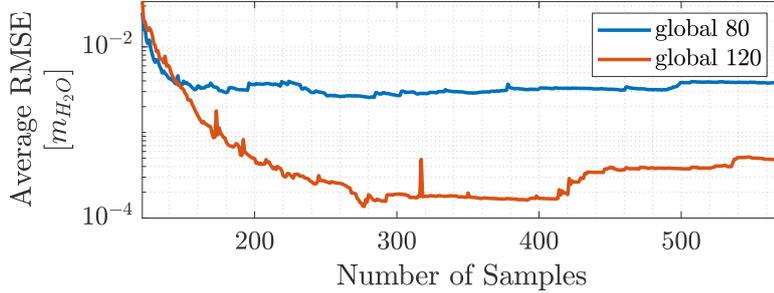
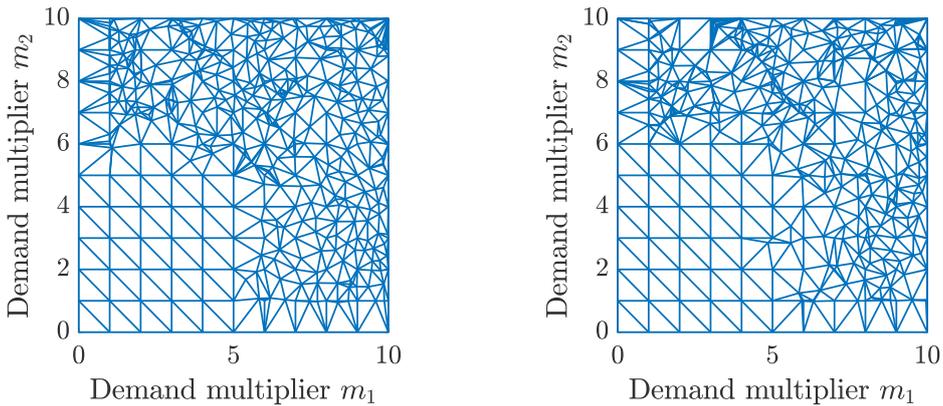


Figure 3.19: Average leave-one-out error development as a function of the snapshot database sample size for the adaptive sampling strategy with a global 80 mode and a global 120 mode reduced basis.

for the addition of sampling points over $N_{samples} = 300$.

In Figure 3.20 the triangulation of the adaptive sampling method are shown for both, the 80 mode and the 120 mode global basis. Both triangulations show



(a) 441 element triangulation using a 80 mode global reduced basis. (b) 441 element triangulation using a 120 mode global reduced basis.

Figure 3.20: Triangulation of the global adaptive sampling strategy starting from a 121 element Cartesian grid.

a similar characteristic, as the sampling algorithm mainly refines the grid for higher values of the demand multipliers. In contrast, the region with demand multipliers $m_1, m_2 < 5$ the initial discretization has barely been improved at all.

A possible explanation for this behaviour is that the reduced basis is constructed in a way that it allows a comparably good representation of the states for lower demand multipliers, leading the algorithm to favour sampling points with higher demand multipliers. This result suggests that the use of local basis may improve the performance of the adaptive sampling algorithm, as it

prevents the reduced basis to represent only parts of the parameter space.

Figure 3.21 shows the result of the model validation based on the set of random validation points. The approximation error and the projection error of the adaptive sampling using the 80 mode and the 120 mode reduced basis are compared to the Cartesian grid as a benchmark. All of the snapshot databases use the same number of 441 samples in order to be comparable. It

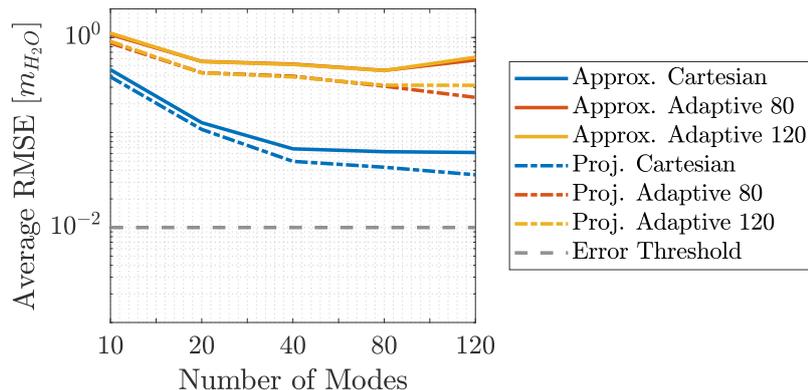


Figure 3.21: Average RMSE approximation and projection error on the validation set as a function of the size of the reduced model for Cartesian sampling and the 80 mode and 120 mode adaptive sampling strategy.

is obvious that the adaptive sampling approach used with a global basis is not able to improve the results of the reduced model compared to the Cartesian benchmark. This bad performance of the projection error suggests that this is mainly a problem caused by the reduced basis. The difference between the two global samplings in both, the approximation error and the projection error is marginal.

Local Adaptive Sampling

The global adaptive sampling strategy has been shown to under-sample the region of lower demand multipliers and to stagnate in the LOO error for sample sizes over $N_{samples} = 300$. In order to improve on this result the global reduced basis is replaced by local basis that more closely represent the current point based on the samples in its vicinity. As for the global approach the starting discretisation is chosen as the 121 element Cartesian grid and the LOO error is evaluated on the basis of the RMSE projection error. Figure 3.22 shows the development of the LOO error as a function of the size of the snapshot database. Comparing this figure to the development of the global sampling approach in Figure 3.19 two observations can be made. First, the mean error for the local approach is much higher for a small number of samples and second,

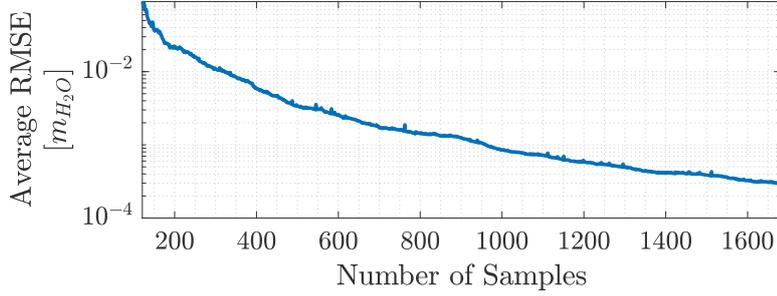


Figure 3.22: Average leave-one-out error development as a function of the snapshot database sample size for the adaptive sampling strategy with a local 10 mode reduced basis.

the mean error is reduced continually, even for the addition of samples for a sample database size well over $N_{samples} = 300$.

The according triangulations are depicted in Figure 3.23 for 441 samples and for 1681 samples. For both triangulations regions with a higher and a

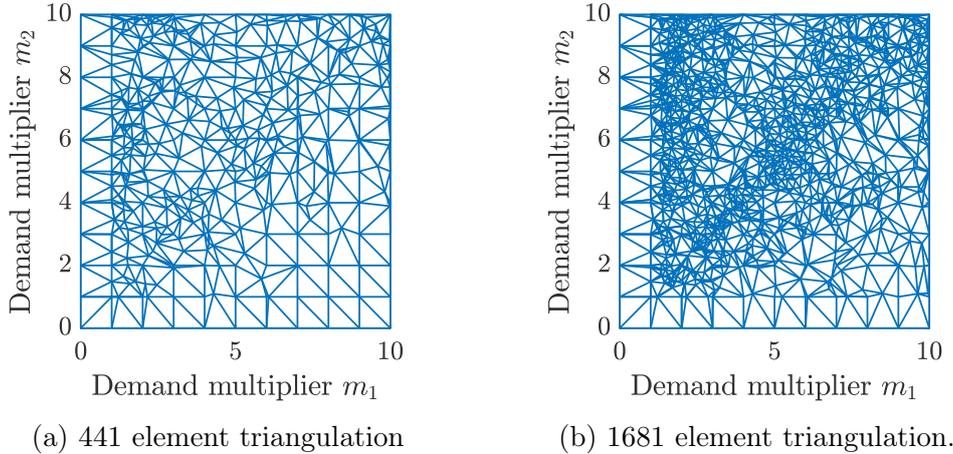


Figure 3.23: Triangulation of adaptive sampling strategy with a local 10 mode reduced basis starting from a 121 element Cartesian grid.

lower sample density can be identified. However, compared to the triangulation of the global approach, there is no region that is completely ignored by the algorithm.

Looking at the evaluation of the errors on the validation set the improved quality of the selected sampling point is apparent. Figure 3.24 shows the approximation error of local reduced models on the adaptive sampling based on a reduced order model with 80 degrees of freedom, compared to equivalent Cartesian samplings. For the 441 element samplings the adaptive approach performs better than the Cartesian sampling independent of the chosen model. How-

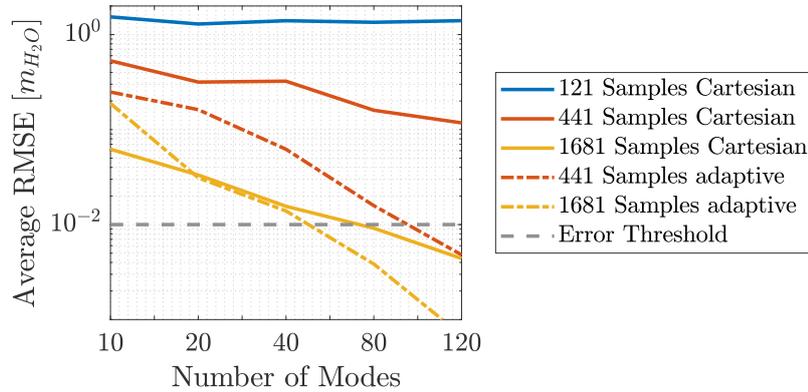


Figure 3.24: Average root mean square error for the reduced order model in the validation points as a function of the degrees of freedom, calculated on a Cartesian and adaptive snapshot database using a 80 mode reduced order model.

ever, for models with a higher number of degrees of freedom this performance benefit increases significantly. Similar results can be observed for the local adaptive 1681 element sampling. Although the model performs nearly on the same level as the 441 element sampling, with an increased amount of degrees of freedom it performs considerably better. In comparison with the equivalent Cartesian model the adaptive approach cannot compete for small mode numbers, but for degrees of freedom above 40 the improvement is significant.

The effect of an increased number of 120 modes for the local reduced basis in the adaptive sampling is visualized in Figure 3.25. This change influences

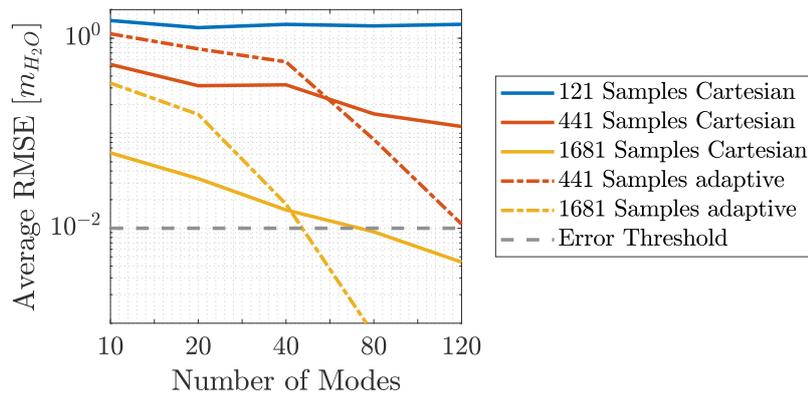


Figure 3.25: Average root mean square error for the reduced order model in the validation points as a function of the degrees of freedom, calculated on a Cartesian and adaptive snapshot database using a 120 mode reduced order model.

the accuracy in two ways. On the one hand, for a low number of degrees of freedom the reduced models on the adaptive sampling perform worse than the Cartesian equivalents. On the other hand, they perform significantly better for a higher number of degrees of freedom.

An observation that has been made in the application of the POD to the hydraulic model in Section 3.6.2 is, that the global reduced model performed better in the estimation of the full hydraulic state. In order to investigate if this effect holds true for the adaptive sampling approach, a global reduced model was applied to the local adaptive sampling. The results in the validations set is shown in Figure 3.26 for the 80 mode adaptive sampling. Based on the average

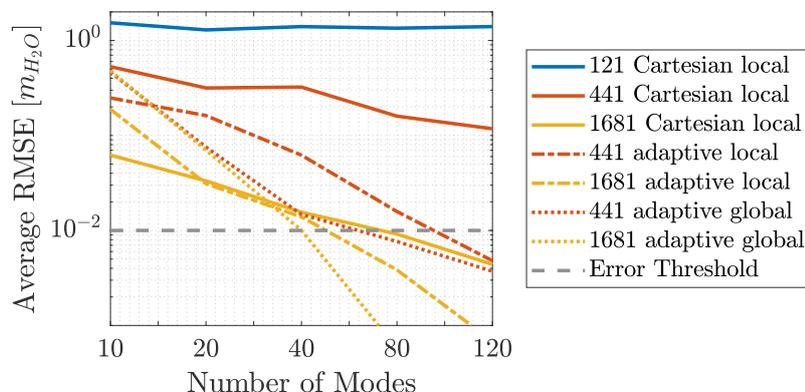


Figure 3.26: Average root mean square error for a global and a local reduced order model in the validation points as a function of the degrees of freedom, calculated on a local adaptive snapshot database using a 80 mode reduced order model and compared to a Cartesian sampling.

RMSE the previous findings can also be confirmed for the adaptive sampling, as the accuracy of the local model is better for lower mode numbers and worse for higher mode numbers.

The global approach to the adaptive sampling shows to be limited in its application, as well as its performance. Due to the fact that the addition of a new sampling point influences the global reduced basis, the leave-one-out error estimation has to be re-evaluated for all of the parameter sampling point. In effect the computational burden increases with each addition of a sampling point. Although this could be mitigated in part by the use of the projection error the approach becomes infeasible for more detailed discretizations. Further, it has been shown that the LOO error stagnates above 300 samples, suggesting that the further addition of sampling points does not improve the reduced basis created from it. Using a local reduced basis during the adaptive sampling phase greatly improves the computational effort, as the evaluation of the leave-one-out error can be limited to the samples in the direct vicinity of the new sample. The accuracy of reduced models created from the local

sampling show to be significantly better than the equivalent Cartesian discretizations especially for models that use a similar number in the reduced model as for the sampling process. Finally, it could also be shown that the use of a global reduced model on a locally sampled snapshot database performs significantly better for models that are sufficiently accurate to substitute the full order model.

3.7 Conclusion

This section introduces advanced, sampling based, deterministic meta-modelling techniques for the application in the hydraulic modelling framework. Section 3.2 introduces the proper orthogonal decomposition for the dimensional reduction of stationary, parametric systems and demonstrates its application to formulate the reduced hydraulic system equations through the Galerkin projection in Section 3.2.2. Reduced models in the state equations, as well as a reduced formulation of the Jacobian matrix are developed for both, the head-flow and the loop-flow formulation. This is followed by a short review on the use of global and local reduced basis in Section 3.2.3. Section 3.3 introduces the empirical interpolation method as a hyper-reduction approach for the non-linear residual function, for the more efficient solution of the hydraulic system. As the quality of sampling based model reduction techniques is inherently dependent on the fidelity of the parameter space discretization, Section 3.5 discusses the use of adaptive sampling techniques in the literature, followed by the introduction of an adaptive, sequential sampling method, that uses a greedy algorithm on the Delaunay triangulation of the discretized parameter-space to select the centroid of the biggest error triangle as a subsequent sampling point, based on the leave-one-out projection error.

Section 3.6 applies the previously introduced meta-modelling techniques to the medium sized network model, in order to validate their performance based on the validation procedure introduced in Section 3.4.

In a first step the application of interpolation and regression based methods are evaluated. For these reduced order models it can be concluded, that models of sufficient accuracy can be formulated using the bi-cubic interpolation for data on a regular grid or the cubic radial basis function interpolation for irregular grids. While the approaches are easy to implement and time efficient in the evaluation, it takes a copious amount of pre-simulated data for the interpolation approach to reach this level of accuracy, making the methods infeasible for the application, especially in cases with higher dimensional parameter-spaces.

The application of global and local projection-based reduced order modelling techniques to the hydraulic network model is investigated in Section 3.6.2. Results for the global reduced order model are encouraging as they

admit the construction of an accurate surrogate model that only uses about 10 – 20% of the degrees of freedom compared to the high-fidelity model. Following in suit with this dimensional reduction of the hydraulic system, is a reduction in computational time of up to 90%, depending on the choice of evaluation for the Jacobian matrix. While the local reduced order model can achieve similar accuracy and performance as the global approach, it does not achieve the same degree of dimensional reduction. However, the construction method used for the creation of the local reduced basis puts the approach at a disadvantage. Limiting the number of snapshots to the models degrees of freedom, does not result in the construction of an efficient basis. This can be concluded from the SVD of the local snapshot database that contains a small number of insignificant singular values. Two suggestions to improve the local approach are thus the addition of local parameter combinations to the local snapshot matrix, until all of the local modes are significant and the use of pre-calculated local basis for a number of sub-domains in the parameter-space.

Following the dimensional reduction of the hydraulic problem, the sampling based empirical interpolation method is applied as a hyper-reduction approach to increase the computational efficiency in the evaluation of the non-linear elements. However, the analysis of the SVD on the sampling database shows a slow convergence in the singular values suggesting that the use of a high number of modes might be necessary for an accurate approximation. Evaluation of the interpolation method shows that it is possible to approximate the non-linear function within appropriate accuracy, giving a projection error below $10^{-2}[m_{H_2O}]$ in the validation set. Yet, the method needs around 1000 support points to interpolate the full vector of about 2000 values, which does not offer any increase in computational efficiency. Further, the attempt to solve the reduced system based on the EIM did not converge. Based on the analysis of the application, a number of suggestions for the future improvement of the approach can be made.

As the main reason for the convergence problem seems to be linked to inaccuracies in the evaluation of the Jacobian matrix, that leads the solver trajectory out of the pre-sampled region in the reduced state vector, two conceivable improvements are the addition of derivative information to the interpolation method and the reformulation of the reduced hydraulic model as a constrained optimization problem

$$\begin{aligned} & \underset{\hat{\mathbf{x}}}{\text{minimize}} && ||\hat{\mathbf{r}}(\hat{\mathbf{x}}, \boldsymbol{\mu})||_2 \\ & \text{subject to} && \{\hat{x}|(1 - \delta) \min \hat{x} \preceq \hat{x} \preceq (1 + \delta) \max \hat{x}\} \end{aligned} \quad (3.42)$$

where the state vector is limited to the sampled region with a certain margin through the inequality constraints. While this could assure the validity of the interpolation basis and lead to an appropriate approximation, this approach will most certainly not result in a performance gain for the evaluation of the reduced system.

Further improvement of the computational efficiency for the EIM can be achieved through the reduction in the evaluated support points. The use of pre-conditioning methods is used with benefit in numerous applications like optimization and eigenvalue problems. Applying these methods in the construction of the reduced EIM basis could improve the evaluation efficiency to a viable degree.

The final application in Section 3.6.4 evaluates the benefit of the adaptive sequential sampling approach to the medium size hydraulic model. The sampling algorithm is evaluated on the basis of a global and a local reduced basis. In general, the use of the projection error for the evaluation of the leave-one-out error is far more efficient, as it does not include the solution of a non-linear system of equations like the evaluation of the approximation error and the previous applications for the POD and the EIM have shown the projection error to be a good surrogate for it. While the global approach converges relatively quick in the average RMSE of the LOO error, the quality of the sampling grid is impractical for the application of a reduced model. The most likely explanation is, that the global reduced basis is well adapted for a region of the model and additional points that are put in the regions of insufficient approximation are not used in the basis construction. This indicates that the use of a local sampling basis might improve the results. Applying the local adaptive sampling method proves that the existence of such a local optimum is no issue. Although the local approach converges slowly in the average RMSE, compared to the global method, the quality of the triangulation is much better and the reduced models built on it are accurate enough for the use as a surrogate surpassing the Cartesian benchmark samplings. Further, the local sampling approach is more efficient than the global approach. As the addition of a new sampling point potentially changes the reduced basis, the LOO error for the global model has to be evaluated in all of the sampling points for each added sample. As the local basis is built on the samples in its proximity, the recalculation of the LOO errors is limited to the points in the vicinity of the new parameter combination.

For the construction of reduced order models on the adaptively sampled database two observations can be made. First, the reduced models seem to perform better if they use a similar number of degrees of freedom as the reduced basis that was used in the sampling process. Second, the global reduced model gave better results than the local one. However, as explained for the application of the POD this is most probably due to the problem in the construction of the local reduced model.

Choosing the parameters for a reduced order model always involves a compromise between performance and accuracy, as well as the computational effort for the off-line and on-line procedures. For the model developed in this chapter, the error threshold is well defined as $\mathcal{O}(10^{-2}) [m_{H_2O}]$. From this it can be concluded that the reduced model built from the adaptively and locally sam-

pled 1681 sample snapshot database performs exceptionally well for a reduced order of 60 degrees of freedom. If the focus is shifted to a lower number of 441 samples, an accurate reduced order model can be constructed using a reduced number of 100 degrees of freedom.

Chapter 4

Uncertainty Quantification

Motivation: *Quantification and propagation of parameter uncertainties in the hydraulic model.*

Principal Elements:

- *Classical methods in uncertainty quantification*
- *Intrusive polynomial chaos expansion*
- *Non-intrusive polynomial chaos expansion*

Conclusions:

- *Intrusive polynomial chaos expansion evaluates directly the coefficients of the stochastic expansion from the stochastic system of equations that has to be formulated.*
- *Non-intrusive polynomial chaos expansion can directly use the deterministic equations and is easily adaptable.*
- *Non-intrusive spectral projection is more efficient than the Monte Carlo method in the use of collocation points.*
- *The polynomial chaos expansion is comparable to the Monte Carlo capable to capture non-linear effects.*

Central part of Uncertainty Analysis is the propagation of errors and uncertainties by means of the mathematical model. To do so, a multitude of algorithms are available that have been tested and proven in numerous large scale applications like weather and climate models, hydrology, biology and the design of nuclear reactors [Smith \[2013\]](#). It is possible to classify the majority of these methods as one of three groups. The perturbation or sensitivity methods, the sampling methods and the spectral methods.

- **Perturbation Methods:** These methods calculate the moments for the distribution of the quantity of interest directly from the system equations by means of a truncated Taylor expansion. Typically, the expansions employed are limited to first- or second-order expansions. This limits their accuracy for highly non-linear models [[Cacuci et al., 2005](#)].

- **Sampling Methods:** With Monte Carlo Simulations as one of the most prominent representatives for this group, sampling methods are often applied for the propagation of uncertainties in non-linear models. Although, in general, implementation of the method is a straightforward task, it is computationally expensive, as the variance of the mean estimate converges with a rate of $1/\sqrt{M}$, where M is the number of simulations [Xiu, 2010; Fishman, 2013].
- **Stochastic Spectral Methods:** The objective of the polynomial chaos expansion is to calculate an approximation for random variables in a mathematical model. This truncated series uses orthogonal polynomials with increasing frequency as a basis. Utilizing the smoothness of these polynomials leads to an efficient convergence behavior [Xiu, 2010; Smith, 2013].

In reality, some methods are not strictly limited to one of these groups. For example, the non-intrusive spectral projection (NISP) is a spectral method that uses a sampling approach for the calculation of its coefficients [Le Maître et Knio, 2010; Smith, 2013].

Section 4.1 introduces classical approaches to the formulation of stochastic systems. In Section 4.2 the polynomial chaos expansion (PCE) is introduced and the stochastic hydraulic system is derived using intrusive and non-intrusive projection methods. A number of use cases are presented in Section 4.3 that analyse the influence of parameter uncertainties on the hydraulic state vector and the performance of the PCE compared to the state of the art methods.

4.1 Classical Uncertainty Quantification

Two of the most prominent methods in the formulation of stochastic systems and uncertainty quantification are stochastic moment approximation methods like the perturbation method and stochastic collocation methods like Monte Carlo simulation. Both approaches are introduced in this section and discussed with regard to their application to the hydraulic model.

4.1.1 Perturbation Methods

Perturbation methods treat the random part of the solution as perturbation around the mean value. These perturbations are generally assumed to be small, to guaranty accurate results. As a result, standard deviations are required to be small. The First Order Second Moment method models the uncertain parameters symmetrical about their nominal value, which in case of an uncertain

demand would be \bar{d}_i and \bar{r}_i for an uncertain roughness. The perturbed parameters can then be represented by the generalized vector

$$\mathbf{P} = \bar{\mathbf{p}} + \delta\mathbf{P} = [\bar{d}_1 + \delta D_1, \dots, \bar{d}_{nd} + \delta D_{nd}, \bar{r}_1 + \delta R_1, \dots, \bar{r}_{np} + \delta R_{np}]^T. \quad (4.1)$$

Here, $\bar{\mathbf{p}}$ is taken as the expected value of the parameters and $\delta\mathbf{P}$ is a random perturbation or uncertainty on \mathbf{P} . To propagate the parameter perturbation the QoI \mathbf{X} is developed in a Taylor expansion as a function of the perturbed parameters \mathbf{P} :

$$\mathbf{X} = \mathbf{x}(\mathbf{P}) = \mathbf{x}(\bar{\mathbf{p}} + \delta\mathbf{P}) \approx \mathbf{x}(\bar{\mathbf{p}}) + \mathbf{S}\delta\mathbf{P}. \quad (4.2)$$

where \mathbf{S} is the sensitivity matrix evaluated at the mean value $\bar{\mathbf{p}}$ of the random variable. For this first order expansion the expected value of the QoI can be calculated by

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = \mathbf{x}(\bar{\mathbf{p}}) \quad (4.3)$$

Using the Taylor expansion of the random parameter in (4.2) the variance can be determined as:

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^T] = \mathbb{E}[\mathbf{S}\delta\mathbf{P}\delta\mathbf{P}^T\mathbf{S}^T] = \mathbf{S}\boldsymbol{\Sigma}_{\mathbf{P}}\mathbf{S}^T. \quad (4.4)$$

where $\boldsymbol{\Sigma}_{\mathbf{P}}$ denotes the covariance matrix of the parameters [Cacuci et al. \[2005\]](#). The application of the FOSM to the hydraulic model of a water distribution network is especially efficient, as there exists the formulation for the direct evaluation of the sensitivity matrix presented in [Section 2.1.3](#).

Using a first order Taylor approximation limits the method to the estimation of the second moment. If quadratic terms are also taken into account a full second order sixth moment approach would be possible [\[Kriegesmann, 2012\]](#). However, practical applications are usually limited to incomplete second order approaches like the second order third moment method [\[Hong et al., 1999\]](#).

4.1.2 Monte Carlo

Stochastic collocation methods like Monte Carlo simulations (MCS) use random samples in order to obtain a stochastic approximation for the mean and variance of random variables. In uncertainty quantification applications it is used to repeatedly evaluate the deterministic system equations for a random sample of the uncertain parameter with the objective to obtain an approximate representation of the PDF of the QoIs. Monte Carlo methods are very popular due to their straightforward implementation of the general procedure. In the first step, random samples are generated from the parameter space $\mathbf{P}^{(i)} = (p_1^{(i)}, \dots, p_k^{(i)})^T, i = 1, \dots, M$ according to their respective distributions. This step makes heavy use of random number generation algorithms like the ones described in [Section 3.2](#). In the second step the deterministic system is evaluated for each sample $i = 1, \dots, M$ from the parameter space \mathbf{P} to obtain

the solution ensemble $\mathbb{X} = \{x^{(1)} \dots x^{(M)}\}$. In the last step the solution ensemble is used to evaluate the solution statistics defined in Section 3.1 where the mean is approximated by the sample average

$$\mathbb{E}[X] \approx \hat{\mu}_X = \frac{1}{M} \sum_{i=1}^M x^{(i)} \quad (4.5)$$

and the sample variance as

$$\mathbb{E}[(X - \mu_X)^2] \approx \hat{\sigma}_X^2 = \frac{1}{M-1} \sum_{i=1}^M (x^{(i)} - \hat{\mu}_X)^2 \quad (4.6)$$

The Monte Carlo method is supported by two basic statistical principles: The Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). The Law of Large Numbers states that, if the samples are independent and identically distributed (i.i.d.), the sample average $\hat{\mu}_X$ will converge to the true mean in the limit of $M \rightarrow \infty$. This also holds for the sample variance $\hat{\sigma}_X^2$ and higher moments. Although the LLN guarantees the convergence of the MCS it does not evaluate the accuracy of the approximation. To do so the CLT has to be applied. Under the condition that the sample size justifies the LLN and the solution ensemble is i.i.d., the Central Limit Theorem states that the sample distribution of the sample average converges to a Gaussian distribution $\mathcal{N}(\mathbb{E}[X], (\sigma_X^2)/M)$, with a standard deviation of σ_X/\sqrt{M} and σ_X as the standard deviation of the true solution. This relation justifies the concept that the MCS converges proportional to the inverse of the square root of the sample size. It is obvious that the MCS can be easily generalized to more complex and even high dimensional applications, but due to its slow convergence with $1/\sqrt{M}$ it is prone to suffer from the curse of dimensionality [Xiu, 2010].

Pseudo random sampling methods are often used to make stochastic collocation methods more efficient. This is achieved through the faster convergence in the stochastic moments for the sampled distribution. Popular choices are low-discrepancy sampling methods like the Halton sequence or the Hammersley point set [Halton, 1964; Hammersley, 1960; Wong *et al.*, 1997]. Looking at applications with multivariate input parameter-spaces experimental design based on the Latin Hypercube Sampling (LHS) is a popular approach [McKay *et al.*, 1979]. The LHS is a stratified Monte Carlo sampling method. Generating M samples for a k -dimensional input parameter $\mathbf{p} = [p_1, \dots, p_k]$, the domain of the multivariate distribution is divided into M sub-domains of equal probability $1/M$. For each sub-domain a parameter is selected randomly, on the basis of the probability density of the interval. The components of the M parameter vectors $\mathbf{p}^{(i)}$ are then combined at random resulting in the sample vector $\mathbf{P} = [\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(M)}]$. The hypercube created from this procedure is made up of N^k sub-domains or cells.

4.2 Polynomial Chaos Expansion

The idea of the polynomial chaos expansion is to create a stochastic series expansion as a function of a basic random variable, using an orthogonal polynomial basis. This basic random variable, often also called the germ distribution is closely related to the choice of the polynomial basis functions.

If \mathbf{Z} is a vector of independent and identically distributed random variables on Ω , then a scalar random variable $X : \Omega \rightarrow \mathbb{R}$ can be represented by the infinite expansion

$$x(\mathbf{Z}) = \sum_{k=0}^{\infty} x_k \Psi_k(\mathbf{Z}), \quad (4.7)$$

where x is a function of the random variable \mathbf{Z} and Ψ_k are the orthogonal polynomials. In practical applications, a truncated expansion with a total number of $N + 1$ elements is used to approximate random variables

$$x_N(\mathbf{Z}) = \sum_{k=0}^N x_k \Psi_k(\mathbf{Z}). \quad (4.8)$$

The number of terms in this finite polynomial chaos expansion is dependent on the dimension of the parameter space n and the order of the expansion l . It is calculated as:

$$N + 1 = \frac{(n + l)!}{n!l!}. \quad (4.9)$$

Polynomial chaos expansion as introduced by [Wiener \[1938\]](#) uses orthogonal Hermite polynomials to model arbitrarily distributed random variables based on a Gaussian distribution. [Xiu et Karniadakis \[2002\]](#) generalized the method to the use of a wider variety of random processes by introducing broader classes of polynomials defined by the Askey scheme [Xiu \[2010\]](#). [Table 4.1](#) shows the link between the probability distribution of the random variable Z and the set of orthogonal polynomials that are used as polynomial basis.

Two of the most important probability distributions in engineering applications are given by the uniform distribution and the normal distribution. [Figures 4.1 and 4.2](#) the first 5 elements are illustrated for the Legendre polynomials and the Hermite polynomials.

The probability distribution function F_Z and the orthogonal polynomial basis are linked through the weighted inner product. For each choice of polynomials $\Psi(z)_k$ a weighted inner product $L^2(\Omega)$ is defined as

$$\langle \Psi_i, \Psi_j \rangle = \int \Psi(z)_i \Psi(z)_j dF_Z(z) = \delta_{ij} \|\Psi_i\|_{L^2_p}. \quad (4.10)$$

The first step in any application of the polynomial chaos expansion is the projection of the known random input parameters P on the chosen polynomial

	Distribution of Z	gPC basis polynomials	Support
Continuous	Gaussian	Hermite	$(-\infty, \infty)$
	Gamma	Laguerre	$[0, \infty)$
	Beta	Jacobi	$[a, b]$
	Uniform	Legendre	$[a, b]$
Discrete	Poisson	Charlier	$\{0, 1, \dots, N\}$
	Binomial	Krawtchouk	$\{0, 1, \dots, N\}$
	Negative Binomial	Meixner	$\{0, 1, \dots, N\}$
	Hypergeometric	Hahn	$\{0, 1, \dots, N\}$

Table 4.1: Correspondence between the type of generalized polynomial chaos and their underlying random variables [Xiu, 2010].

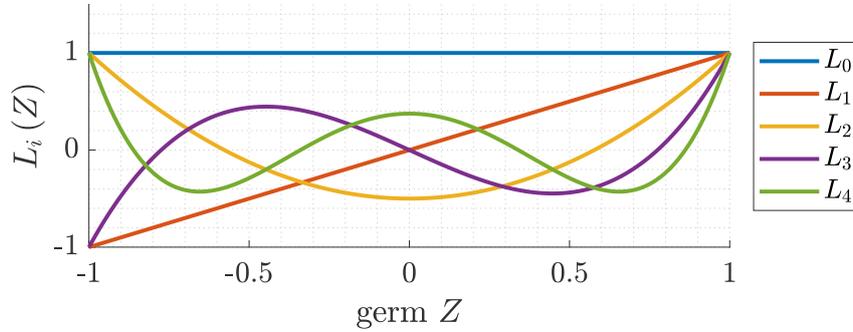


Figure 4.1: First five Legendre polynomials as a function of the uniform germ distribution $Z \sim \mathcal{U}(-1, 1)$.

basis. The coefficients P_N of their truncated PC expansion

$$p_N(\mathbf{Z}) = \sum_{k=0}^N p_k \Psi_k(\mathbf{Z}) \quad (4.11)$$

are calculated using the definition in (4.10)

$$p_k = \frac{\langle P, \Psi_k \rangle}{\langle \Psi_k, \Psi_k \rangle}. \quad (4.12)$$

The existence and convergence of this projection follows from the classical approximation theory.

In practice there exist two ways to perform the projection called the strong and the weak approximation. The strong approximation applies in cases where the parameter is explicitly known as a function P of the basis random variable Z . For that case the evaluation of (4.12) can be done straightforward by the integral

$$p_k = \frac{1}{\|\Psi_k\|_{L^2_\rho}} \int_{\Omega} P(z) \Psi_k(z) f_Z(z) dz \quad (4.13)$$

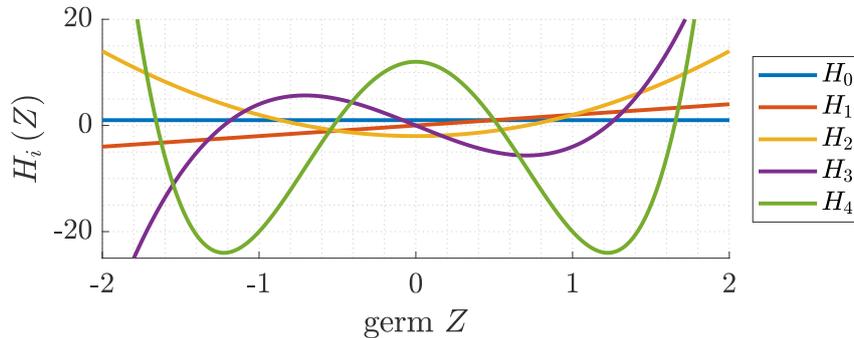


Figure 4.2: First five Hermite polynomials as a function of the normal germ distribution $X \sim \mathcal{N}(0, 1)$.

Such a case is for example given by a random variable $P \sim \mathcal{N}(\mu, \sigma)$, which is defined as $P(Z) = \mu + \sigma Z$. Usually the direct definition of the random parameter is not known. Instead the parameter is characterized by its probability distribution function $F_P(p) = \mathbb{P}(P \leq p)$. For such cases a weak approximation is performed. In contrast to the strong approximation a convergence of the weak approximation is not guaranteed. Instead it converges in probability. With $F_P : I_P \rightarrow [0, 1]$ and $F_Z : I_Z \rightarrow [0, 1]$ mapping the random variables P and Z to a uniform distribution in $[0, 1]$, (4.13) may be rewritten. Assume that $U = F_P(P) = F_Z(Z)$, the random variables may be expressed as $P = F_P^{-1}(U)$ and $Z = F_Z^{-1}(U)$ respectively. This allows to give $P = F_P^{-1}(F_Z(Z))$ as a function of Z , which results in

$$p_k = \frac{1}{\|\Psi_k\|_{L^2_p}} \int_{\Omega} F_P^{-1}(F_Z(z)) \Psi_k(z) f_Z(z) dz \quad (4.14)$$

For multivariate random variables the inversion of the probability density function is generally not possible. This means that multivariate input parametrizations usually use the tensorization of univariate random variables

$$p(\mathbf{Z}) = \prod_{i=1}^n p(Z_i). \quad (4.15)$$

However, this way of modelling demands for the random parameters to be independent. As (4.9) shows, the use of a multivariate input parameter-space dramatically increases the size for the finite PCE. Especially the addition of high dimensional interdependencies in the input parameters increase the number of terms in the expansion, many of which can have small or close to zero coefficients. This makes the approach computationally expensive for intrusive and non-intrusive methods alike. Blatman et Sudret [2011] address this problem through the use of a sparse polynomial chaos expansion that eliminates not significant elements. They introduce a truncation strategy that reduces

the expansion complexity based on hyperbolic index sets and an adaptive algorithm based on Least Angle Regression (LAR) in order to detect the significant PCE coefficients.

Evaluation of the coefficients for the quantities of interest can be performed in an intrusive or a non-intrusive manner. The intrusive approach formulates an extended stochastic system in the coefficients of the PCE which allows for the direct evaluation without any need to sample the initial system, whereas the non-intrusive methods evaluate the original deterministic system in order to approximate the coefficients.

4.2.1 Intrusive stochastic hydraulic model

Due to the direct numerical evaluation of the stochastic expansion of the quantities of interest, an intrusive stochastic model may have considerable benefits compared to stochastic collocation methods like Monte Carlo and non-intrusive spectral projection based on the repeated evaluation of the deterministic system. With this in mind, this section presents the basic steps for the development of the intrusive stochastic hydraulic model.

In applying the intrusive PCE to the hydraulic equations all uncertain parameters \mathbf{d} and \mathbf{r} are replaced by the polynomial series expansion of the order $N + 1$, following (4.12).

$$\begin{aligned}\mathbf{d}_N(\mathbf{Z}) &= \sum_{k=0}^N \mathbf{d}_k \Psi_k(\mathbf{Z}) \\ \mathbf{r}_N(\mathbf{Z}) &= \sum_{k=0}^N \mathbf{r}_k \Psi_k(\mathbf{Z})\end{aligned}$$

The uncertain hydraulic state vectors \mathbf{q} and \mathbf{h} are then replaced by their expansions

$$\begin{aligned}\mathbf{q}_N(\mathbf{Z}) &= \sum_{k=0}^N \mathbf{q}_k \Psi_k(\mathbf{Z}) \\ \mathbf{h}_N(\mathbf{Z}) &= \sum_{k=0}^N \mathbf{h}_k \Psi_k(\mathbf{Z})\end{aligned}$$

with the unknown coefficients \mathbf{q}_k and \mathbf{h}_k . Replacing these expansions in (2.13) and (2.14) gives the stochastic system of equations:

$$\begin{aligned}\mathbf{A}\mathbf{q}_N + \mathbf{d}_N &= \mathbf{0} \\ \Delta\mathbf{h}(\mathbf{r}_N, \mathbf{q}_N) - \mathbf{A}^T\mathbf{h}_N - \mathbf{A}_f^T\mathbf{h}_f &= \mathbf{0}\end{aligned}\tag{4.16}$$

Projecting this approximation onto the $N + 1$ polynomial basis functions Ψ_k leads to an augmented system of equations which contains $N + 1$ times the number of equations as the original system.

$$\begin{aligned} \langle \mathbf{A}\mathbf{q}_N + \mathbf{d}_N, \Psi_k \rangle &= \mathbf{0} \\ \langle \Delta\mathbf{h}(\mathbf{r}_N, \mathbf{q}_N) - \mathbf{A}^T\mathbf{h}_N - \mathbf{A}_f^T\mathbf{h}_f, \Psi_k \rangle &= \mathbf{0} \end{aligned} \quad (4.17)$$

The formulation of such a system of equations allows to directly evaluate the coefficients for the expansion of the QoIs. However, the direct evaluation of (4.17) is not possible as the head-loss function $\Delta\mathbf{h}$ is a non-polynomial function. The treatment of this term is further discussed in the Section 4.2.3.

4.2.2 Non-Intrusive Spectral Projection

Determining the coefficients of the polynomial chaos expansion for the hydraulic state vector x , using the non-intrusive approach can be done in a number of different ways. Smith [2013] divides them into collocation methods and discrete projection methods. One thing all of these methods have in common is, that they use a set of realizations from the deterministic model and can be used in combination with existing, deterministic models or software solutions.

The general approach for the non-intrusive collocation can be described in two steps. First, based on a set of M collocation points $\{Z^m\}_{m=1}^M$ from the germ distribution, an ensemble of solutions $X = [x(p(Z^1)), \dots, x(p(Z^M))]$ is calculated for the set of parameters $\{p(Z^m)\}_{m=1}^M$, using the deterministic system. Second, the coefficients for the polynomial series expansion

$$x(p(Z^m)) = x^K(p(Z^m)) \quad (4.18)$$

are estimated using linear regression, K indicating the polynomial expansion order. For a general orthogonal basis polynomial this leads to the matrix system:

$$\begin{bmatrix} \Psi_0(Z^1) & \dots & \Psi_N(Z^1) \\ \vdots & \ddots & \vdots \\ \Psi_0(Z^M) & \dots & \Psi_N(Z^M) \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} x(p(Z^1)) \\ \vdots \\ x(p(Z^M)) \end{bmatrix}. \quad (4.19)$$

Given that this system is well conditioned and the system is not under-determined, the system can be solved in a least squares sense through the inversion of the matrix on the left hand side of (4.19).

The set of collocation points $\{Z^m\}_{m=1}^M$ can be created using either stochastic methods like Monte Carlo sampling or deterministic methods like sparse grid sampling. Further, detail on the efficient placement of collocation points is given by Xiu [2010] or Smith [2013].

4.2.3 Pseudo-Spectral Head-Loss Function

In the development of an intrusive stochastic formulation of the hydraulic system equations, the treatment of the non-polynomial head-loss function is one of the most complicated tasks.

One possibility is to replace the head-loss function by a Taylor approximation, which gives it polynomial form and is thus easily treatable as a polynomial non-linearity following [Le Maître et Knio \[2010\]](#). However, the approximation of the head-loss function through the Taylor expansion is limited and has to be performed at a minimum for the third order if non-linear effects are supposed to be captured.

A more accurate approximation of the head-loss function, that has proven to closely approximate the Hazen-Williams as well as the Darcy-Weisbach equation is given in (2.9) by [Pecci et al. \[2017\]](#). Starting from this quadratic approximation and replacing the sign function in the quadratic term by the product $q|q|$ a stochastic approximation of the head-loss function can be formulated as:

$$\Delta h(q(Z)) = aq(Z) + b|q|(Z)q(Z). \quad (4.20)$$

This formulation is still not polynomial as it uses the Absolute value function, but following [Le Maître et Knio \[2010\]](#) the spectral coefficients of the absolute value function are defined through the spectral product:

$$|q|(Z)|q|(Z) = q^2(Z). \quad (4.21)$$

Using the Galerkin product, a set of non-linear equations can be formulated for the expansion coefficients of $|q|(Z)$

$$\begin{bmatrix} \sum_{j=0}^P C_{j00}|q|_j & \cdots & \sum_{j=0}^P C_{jP0}|q|_j \\ \vdots & \ddots & \vdots \\ \sum_{j=0}^P C_{j0P}|q|_j & \cdots & \sum_{j=0}^P C_{jPP}|q|_j \end{bmatrix} \begin{bmatrix} |q|_0 \\ \vdots \\ |q|_P \end{bmatrix} = \begin{bmatrix} q_0^2 \\ \vdots \\ q_P^2 \end{bmatrix}, \quad (4.22)$$

with the coefficients C_{ijk} calculated as

$$C_{ijk} = \frac{\langle \Psi_i, \Psi_j, \Psi_k \rangle}{\langle \Psi_k \bar{\Psi}_k \rangle}. \quad (4.23)$$

The system can be solved using a non-linear standard solver. Once the coefficients for the stochastic expansion of the absolute value function are determined, the calculation of the coefficients for the stochastic head-loss function is straight forward and it can be replaced in (4.17).

Figure 4.3 shows the pseudo spectral approximation of the absolute value function based on the Hermite polynomial for a flow rate that is normally distributed as $q \sim \mathcal{N}(0, 1)$. Due to the lack of differentiability in the absolute value function, the Hermite polynomials have a problem approximating it around the mean value.

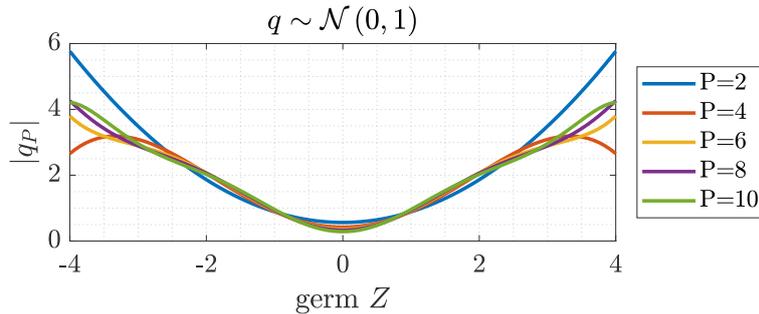


Figure 4.3: Pseudo-spectral approximation of the absolute value function of a flow variable distributed as $q \sim \mathcal{N}(0, 1)$ for different orders P of Hermite polynomials.

With the coefficients of the absolute value function, the pseudo-spectral projection of the non-linear term in the spectral head-loss function can be evaluated. The results are shown in Figure 4.4. Two observations can be

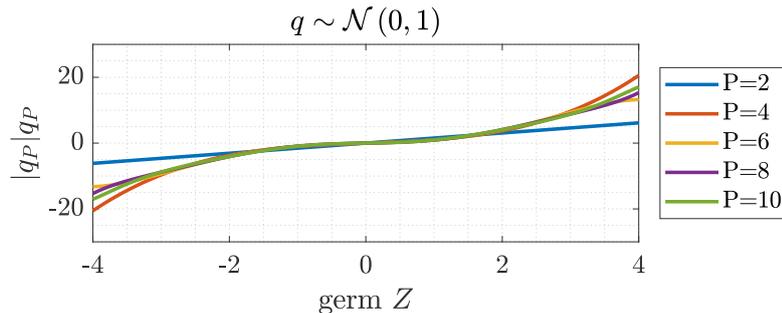


Figure 4.4: Pseudo-spectral approximation of the quadratic term in the head-loss function of a flow variable distributed as $q \sim \mathcal{N}(0, 1)$ for different orders P of Hermite polynomials.

made in this case. First, the error that is introduced by the lack of differentiability in the absolute value is not significant as the multiplication with the flow smooths over the region around the mean value. Second, while the anti-symmetric approximation around the mean value is reasonably accurate, the approximation gets inaccurate for big Z values.

Looking at a random flow variable that is distributed as $q \sim \mathcal{N}(1, 0.5)$ in Figure 4.5, shows the same inaccuracy for the approximation of the non-differentiable part in the absolute value function. In Figure 4.6 the same smoothing effect can be observed for the expansion of the non-linear head-loss term.

This pseudo-spectral approximation of the head-loss function seems promising for the application in an intrusive stochastic hydraulic system. One issue however, is the calculation of the coefficients. As the head-loss is a function of

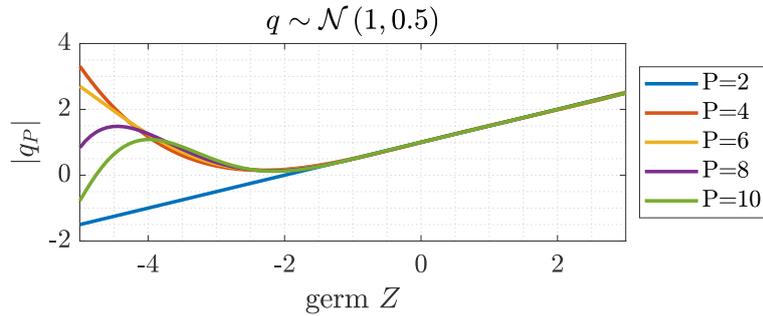


Figure 4.5: Pseudo-spectral approximation of the absolute value function of a flow variable distributed as $q \sim \mathcal{N}(1, 0.5)$ for different orders P of Hermite polynomials.

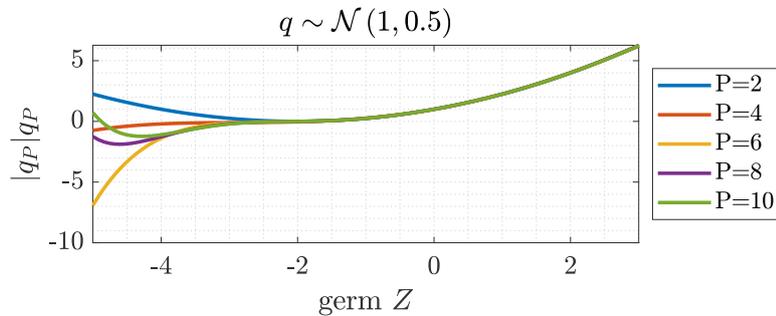


Figure 4.6: Pseudo-spectral approximation of the quadratic term in the head-loss function of a flow variable distributed as $q \sim \mathcal{N}(1, 0.5)$ for different orders P of Hermite polynomials.

the stochastic state variable, the expansion coefficients have to be calculated repeatedly for each element of the flow rate vector during the evaluation of the stochastic system. This poses a considerable computational effort, especially as the solution of the non-linear system (4.22) often depends on a good choice for the initial value.

4.3 Applications

In this section three use cases are presented that have been designed with specific objectives. The first application is mainly used for the validation of the implementation for the polynomial chaos expansion. The tree shaped graph of the network guarantees good results for the FOSM, as the flow rates are determined through a set of linear equations. In the second use case the application to the pressure driven model is investigated. The introduction of the pressure-outflow relation introduces highly non-linear behaviour to the hydraulic system in order to challenge the polynomial chaos expansion. The third

application is used to challenge the PCE through the use of a six dimensional parameter-space on the demand driven hydraulic model.

4.3.1 Illustrative Model

The illustrative model is used for the first implementation and validation of both, the intrusive and non-intrusive PCE algorithms. The topology of the tree shaped network graph is shown in Figure 2.2. The tree structure is convenient, as the flow rates are calculated by a determined set of linear equations in this case, which facilitates the validation process. However, the application can be extended to more complex use cases with looped networks without any changes. The parameters of the mathematical model have been introduced in Section 2.1 and are given by the incidence matrix \mathbf{A} , the demand vector \mathbf{d} , the friction vector \mathbf{r} and the fixed head potential vector \mathbf{h}_f . The quantities of interest are the flow rate vector \mathbf{q} and the head vector \mathbf{h} .

Scenario

The case presented in this scenario introduces a system with a single uncertain parameter. This parameter is defined by the demand multiplier at node 10, which is the end point of one of the branches in the network depicted in Figure 2.2. The parameter uncertainty is modelled as Gaussian random variables $\mathcal{N} \sim (1, 0.3)$. This way the mean of the demand flow rate, which is defined as the base demand times the demand multiplier, has its mean at the base demand and is highly unlikely to become negative. A total number of $N_{samples} = 1e5$ samples are generated using Latin Hypercube Sampling in order to ensure an error of approximately $e = 0.1\%$ for the Monte Carlo simulation. The generated distribution is shown in Figure 4.7

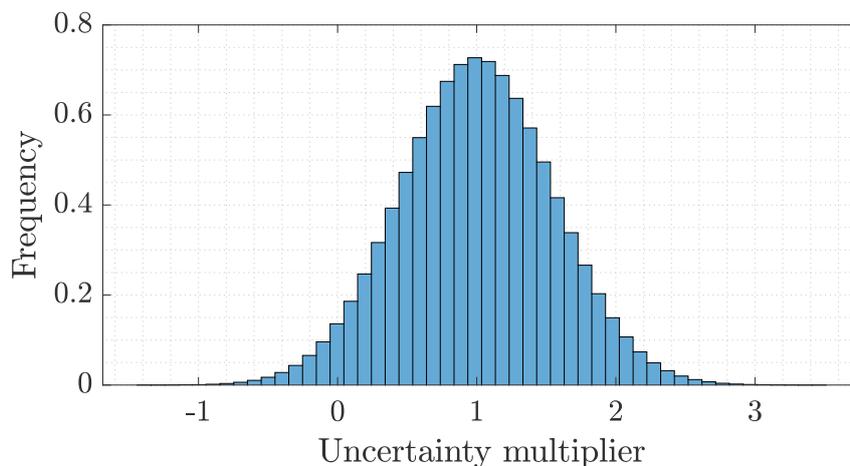


Figure 4.7: Parameter uncertainties for demand $Z \sim \mathcal{N}(1, 0.3)$.

Results

This section compares the results from the Monte Carlo simulation, first order perturbation method and the PCE. The results for the quantities of interest are illustrated by the flow rate through pipe 5 and the head at node 5. These elements are representative for the results in the rest of the network.

Figure 4.8 shows the resulting probability distribution for the flow rate through link 5. The abscissa gives the flow rate q_5 in l/s and the ordinate quantifies its probability $P(q_5)$. The result of the Monte Carlo simulation uses the full sample size of $N_{mc} = 1e5$ evaluations and is given by the histogram. The yellow curve gives the result from the First Order Second Moment method, which calculates the mean and the variance and is approximated as a Gaussian distribution. Finally the red curve shows the result of a first order PC expansion, which has been evaluated using a sample size of $N_{pce} = 1e2$. All three methods are in good agreement for the quantification of uncertainty in the flow rate q_5 . Information on the convergence behavior of the Monte

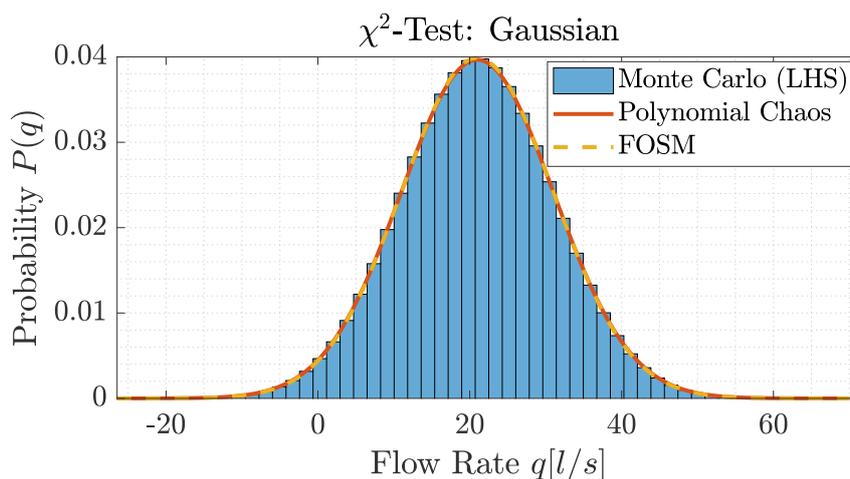


Figure 4.8: Estimated probability density function for the flow rate through link 5 using FOSM, MCS and a 1st-order PCE.

Carlo method and the polynomial chaos expansion is given in Figure 4.9 for the mean estimated value and in Figure 4.10 for the variance. The graphs show the estimated mean and standard deviation as a function of the number of samples that have been used for the evaluation. A logarithmic scale is used for the abscissa.

Figure 4.11 shows the results for the head h_5 in m_{H_2O} at node 5 on the left hand side and is again completed by the convergence plots in mean and standard deviation on the right hand side. For the FOSM method the estimate probability density function is once again chosen to be Gaussian. However, in this case the approximation is not as accurate as the results from both the Monte Carlo simulations and the fourth order PCE that give a non-symmetric

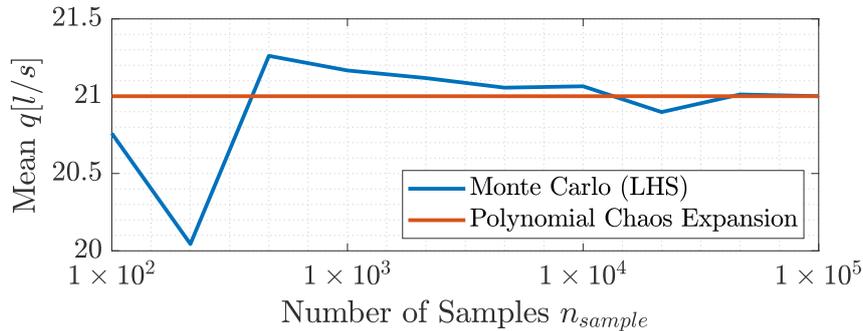


Figure 4.9: Convergence of mean estimated flow rate through link 5 using MCS and a 1st-order PCE.

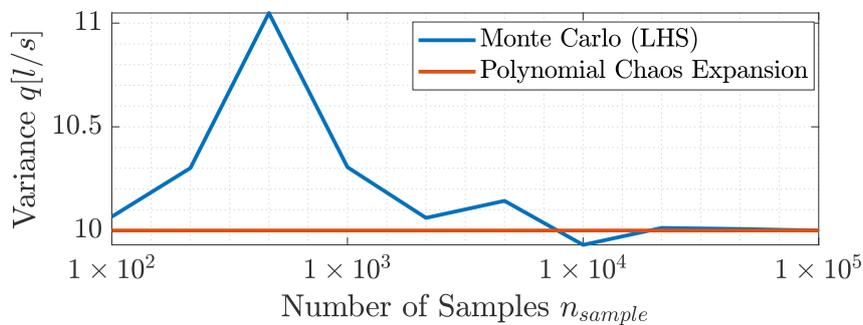


Figure 4.10: Convergence of estimated standard deviation for the flow rate through link 5 using MCS and a 1st-order PCE.

distribution that is skewed in the direction of lower heads.

Taking a close look at the methods used for quantifying the uncertainties and the structure of the model they have been applied to, it is possible to explain the difference in performance shown by Figures 4.8 and 4.11. In the special case of the hydraulic model defined in Section 2.4.1 the flow rates can be calculated by the set of linear equations in (2.13) due to the tree structure of the network. From this it follows for one that the FOSM method uses a first order or linear approximation of the random variable for the calculation of the first and second moment. Similar to this it can be shown that for a Gauss-Hermite PCE any Gaussian distribution can be described exactly by a first order expansion with the coefficient u_0 as mean value and u_1 as standard deviation. In conclusion it can be argued that both methods give the exact result in this special case and the result is confirmed by the Monte Carlo simulation.

As introduced in Section 2.1.1 the head is strongly influenced by the non-linear head-loss function. That explains in part the different results in Figure 4.11. Once again the Monte Carlo simulation with the full sample size $N_{mc} = 1e5$ gives the result for validation. The FOSM method produces a Gaussian distribution that is based on the sensitivity of the system and is not able to

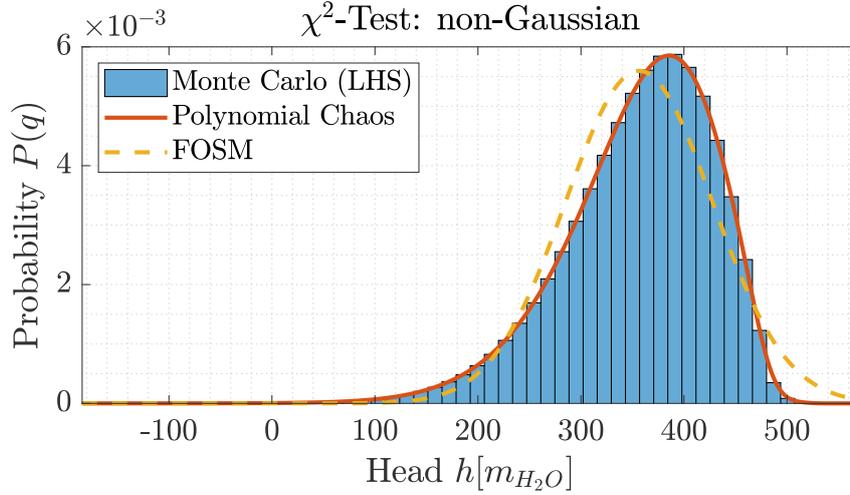


Figure 4.11: Estimated probability density function for the head at node 5 using FOSM, MCS and a 8th-order PCE.

capture the asymmetry of the probability distribution. Figure 4.11 shows a good fit in the central part even for a substantial variance in the demand. Major problems in the description are limited to the distribution tails. In contrast, a PCE of the order four is capable to give an accurate description of the head probability distribution that is in agreement with the Monte Carlo simulation. For the convergence in both the mean value in Figure 4.12 and the variance in Figure 4.13 similar results can be observed as for the flow rate. While the PCE converges already for a few hundred samples, the MC method needs more than $N_{mc} = 1e4$ samples to achieve similar precision.

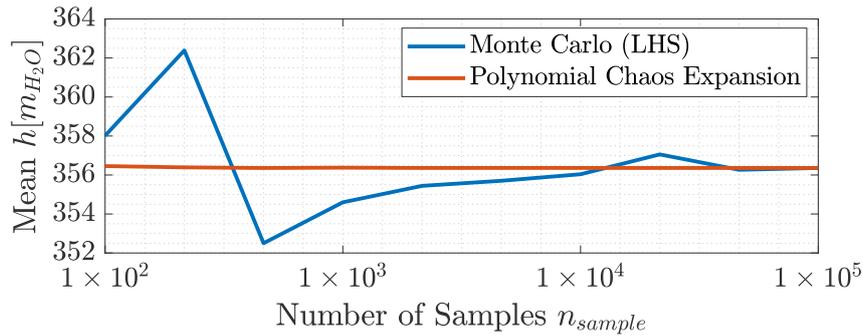


Figure 4.12: Convergence of estimated mean value for the head at node 5 using MCS and a 8th-order PCE

Discussion

Monte Carlo versus Non-Intrusive Spectral Projection: Looking at the fact that both the Monte Carlo simulation and the non-intrusive spectral projection

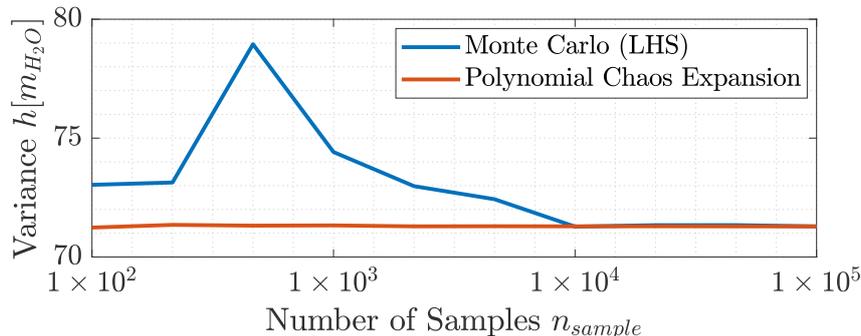


Figure 4.13: Convergence of estimated standard deviation for the head at node 5 using MCS and a 8th-order PCE.

	Galerkin		Non-Intrusive Spectral Projection			
	2 nd -order Taylor		5 Samples		10 Samples	
	q_5	h_5	q_5	h_5	q_5	h_5
u_0	21	845.2393	21	845.2393	21	844.8645
u_1	10	-80.6156	10	-80.6156	10	-80.2202
u_2	0	-15.0044	1e-8	-15.0044	3e-13	-15.7426
u_3	0	-3e-18	1e-9	0.4011	2e-13	0.5413
u_4	0	0	1e-11	-0.0202	2e-15	-0.1009

Table 4.2: Calculated coefficients for the polynomial chaos expansion using the intrusive Galerkin projection and the non-intrusive matrix inversion approaches for the flow rate through pipe 5 q_5 and the head at node 5 h_5 .

may be classified as sampling methods one may ask why the application of the polynomial chaos expansion is beneficial. From literature the answer to this question lays in the fact that PC methods use the smoothness of the orthogonal basis polynomials and in effect have a superior convergence behavior [Smith \[2013\]](#). This is illustrated in the convergence plots for the mean and standard deviation in Figures 4.9 and 4.10, as well as Figures 4.12 and 4.13. For a low dimensional problem as the one discussed in this scenario a very small number of points is sufficient to get a good estimation of the PCE coefficients, while the Monte Carlo method with pseudo-random LHS needs more than $1e4$ samples to give comparable accuracy.

Intrusive versus Non-Intrusive Methods: As introduced in Section 4.2 there exist two basic approaches to calculate the expansion coefficients with the intrusive and non-intrusive methods. For the application of the PCE in this thesis both have been tested and the results are shown in Table 4.2. The coefficients for the Galerkin approach have been obtained by using a second order Taylor expansion for approximating the non-linear head-loss function. For the non-intrusive matrix inversion, the coefficients have been calculated using sample sizes of 5 and 10. The table shows that the coefficients u_0 to u_2 calculated

by the Galerkin projection are in good agreement with the NISP calculations. Coefficients of an order higher than the expansion of the head-loss function are effectively zero, suggesting that in future applications the expansion order should match that of the PCE. The coefficients from the NISP results no major difference can be identified between the calculations based on 5 and 10 samples underlining the efficiency of the stochastic collocation approach. However, under the circumstances of the example it is probable that the 10 sample NISP is more accurate due to the higher number of collocation points and with respect to the second order approximation of the head-loss function for the Galerkin projection. Applying the intrusive approach is challenging since the newly created set of equations changes for the addition of new, uncertain input parameter or with a change in the expansion order. This means it is not easily adaptable to new network models. The adaptation of non-intrusive methods to higher order expansions and a bigger parameter space on the other hand is relatively easy. This makes the non-intrusive approach more flexible. Further, the non-intrusive method does not require the approximation of the non polynomial head loss term in the system equations, which means that it contains the correct non-linear behavior. In conclusion, it can be stated that in small applications like the one presented in this article the matrix inversion is the better approach due to the flexible application and efficient sampling.

Evaluating the Expansion Order: An important task for any application of an expansion approach is the evaluation of accuracy for the chosen development order. Since it is not possible to do so a priori this section shows the measures that have been taken based on the estimated coefficients. In a first iteration, the expansion order is chosen due to experience. Based on the evaluation, it has to be adapted. The appropriate expansion order depends on factors like the non-linear properties of the modeled system and the desired accuracy for the application. Similar to other examples from polynomial approximation theory, it is assumed that the expansion converges to the true solution and that the theoretical infinite series may be represented by a truncated series of order N . From this it follows that coefficient values of higher order polynomials should be small and go to zero. As can be seen in Table 4.2 the coefficients decrease with order of the expansion and approach zero for higher orders, which justifies the series truncation.

4.3.2 Strong Non-Linearities

In order to evaluate the polynomial chaos expansions capability to capture highly non-linear effects in a model, this use case applies it to the pressure driven model of the small looped network shown in Figure 2.3. The network contains 9 consumer nodes, 12 links and 3 reservoirs.

Scenario

The one-dimensional parameter-space for the uncertain parameter is given by a demand multiplier. This demand multiplier is applied to the base demand, defined on a subset of the free nodes. The quantities of interest are the flow rates at the pipes and the nodal heads. The system is modelled using the demand driven and the pressure driven paradigm in order to investigate the influence on the QoIs. The pressure outflow relation is modelled by the classical Wagner function.

Results

Figure 4.14 shows the estimated probability density function for the flow rate through one of the pipes in the network loop. The non-intrusive PCE uses a polynomial of the order 12 and a sample size of 100 realizations, while the Monte Carlo simulation is based on an ensemble of $1e4$ samples. The upper plot shows the result for the demand driven approach, whereas the lower plot gives the result of the pressure driven model. The results show that PDF

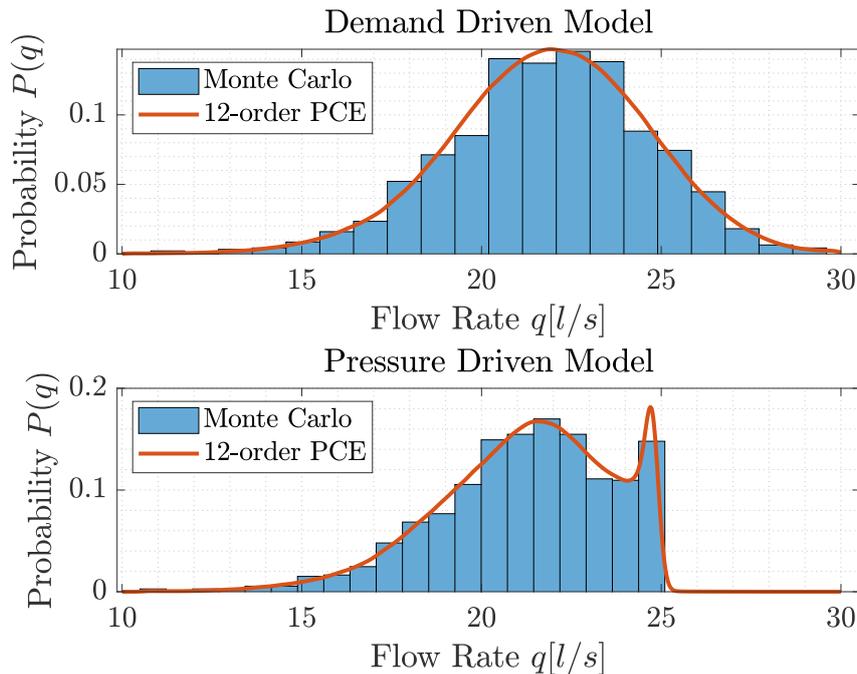


Figure 4.14: Estimated probability density for the flow rate through the pipe 2 in the small network loop, using Monte Carlo simulation and a 12th order PCE on the DDM and PDM approach.

based on the DDM is given by a slightly skewed, bell shaped distribution. In contrast, both distributions generated from the PDM clearly show a second peak and an abrupt cut-off for higher flow rates.

Looking at the results for the nodal pressure in the looped network part in Figure 4.15, similar observations can be made. While the head distribution

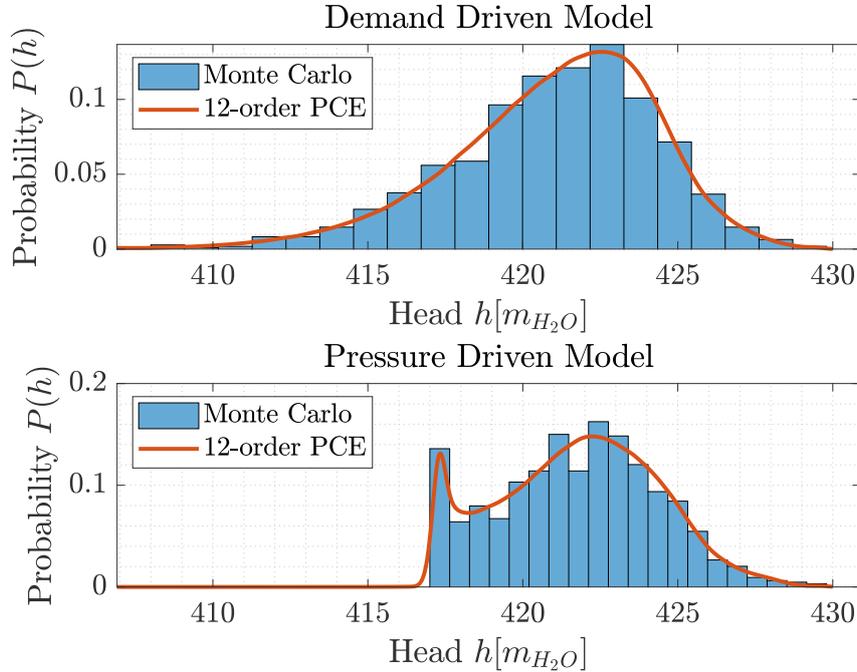


Figure 4.15: Estimated probability density for the head at node 7 in the network loop of the small model, using Monte Carlo simulation and a 12th order PCE on the DDM and PDM approach.

based on the DDM is clearly skewed in the direction of lower heads, the PDM once again introduces a second peak to the distribution and the cut-off to lower heads.

Discussion

The presented results show the benefit of the Monte Carlo simulation and the polynomial chaos expansion over perturbation methods like the first order second moment and second order third moment methods. Both the MC and PCE are able to capture the second peak in the distribution introduced by the PDM. Comparing the performance of the two approaches, the PCE shows to use the data more efficiently as it uses a much lower number of simulations to produce results comparable to the MC simulation.

Looking at the results from a hydraulic point of view, the significance of combining the PDM and the uncertainty propagation is apparent. The cut-off in the flow rate and the head distribution can be easily explained by the inclusion of the pressure-outflow relation introduced in Section 2.2. With rising demand or outflow at a consumption node, higher pressures are necessary.

While the shift of the main peak of the flow distribution in Figure 4.14 can be explained by the reduced outflow for $h_m \leq h \leq h_s$ in the Wagner function (2.32), the outflow is reduced to zero for insufficient pressure. This effect also creates the second peak, which agglomerates the states of higher demands that cannot be served to the maximum possible outflow. At the lower end of the flow PDF no changes can be observed, since the POR does not act in this region.

The effect of the POR can also be observed in the probability distribution of the head in the looped part of the network. While the DDM produces a distribution with a long tail in the direction of lower heads, the distribution from the PDM has a lower limit given by the minimum service pressure defined in the Wagner function. This, in turn, generates the second peak for the cumulated lower head cases.

4.3.3 High Dimensional Parameter-Space

In this use case the scenario is extended with respect to the network size as well as in the dimension of the parameter space. First, the subject network is defined by a realistic model that has been supplied by VEDIF through the ResiWater project. This network contains 2,175 pipes, 1,822 nodes and one reservoir in the highly looped region. Its topology is illustrated in Figure 4.16. Second, parameter uncertainties are defined by a 12 dimensional model in demand flow rates and pipe roughness.

Scenario

For the scenario, the network is divided into the 6 different regions using the k-means algorithm. The regions are illustrated in Figure 4.16. For each of the regions an uncertain demand multiplier is applied to the nodes with their respective base demand and roughness multiplier is applied to the design roughness of each pipe taken from the network model. This ensures that the mean value of the uncertain parameters is close to that of the deterministic model. All of the 12 parameters are independent and normally distributed random variables. The demand multipliers are defined as $\mathcal{N} \sim (2, 0.6)$ in order to peak demand period flow rates and avoid negative demands in the model. The roughnesses are chosen as $\mathcal{N} \sim (2, 0.3)$ in order to simulate the ageing process of the pipes, as the mean roughness is expected to rise with age and it is unlikely to be lower than the design value. The parameter space has been chosen to be based on Gaussian distributions to make the results comparable to similar studies; however, using the PCE the design of more complex parameter spaces with multiple different probability distributions is possible without the loss of generality. Samples from the 12 dimensional distribution are created using a Latin Hypercube Sampler. In order to ensure a very low estimated

Cluster Assignments and Centroids

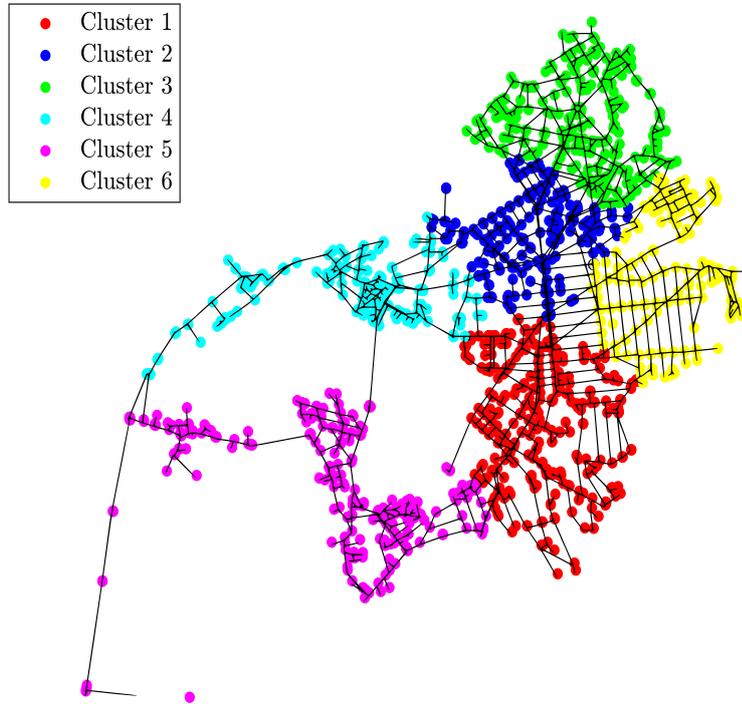


Figure 4.16: Clusters for parameter uncertainties in the medium size network.

error of $e = 0.1\%$ for the Monte Carlo simulation a maximum sample size of $N_{samples} = 1e5$ has been generated. Figure 4.17 shows the realizations for one of the demand multipliers and one of the roughness multipliers. The hydraulic system is modelled using the demand driven approach.

Results

As for the illustrative example the results are presented by the probability density function, which is tested for normality using the Pearson χ^2 -test and the convergence plots in mean and standard deviation. Since it has been shown previously that the first order perturbation method is inadequate for the propagation of non-linear effects and the application to the network is expected to have non-linear effects due to the looped structure of the network, the results are limited to the Monte Carlo simulation and the PCE and the use of the non-intrusive PCE method. For the Monte Carlo simulation represented by the histogram the full sample size of $N_{mc} = 1e5$ has been used. In comparison, the PCE with the NISP approach has been developed on the basis of a total number of $N_{pce} = 5e2$ evaluations of the full hydraulic system in order to calculate the coefficients for a 6th order expansion. The kernel density estimation uses

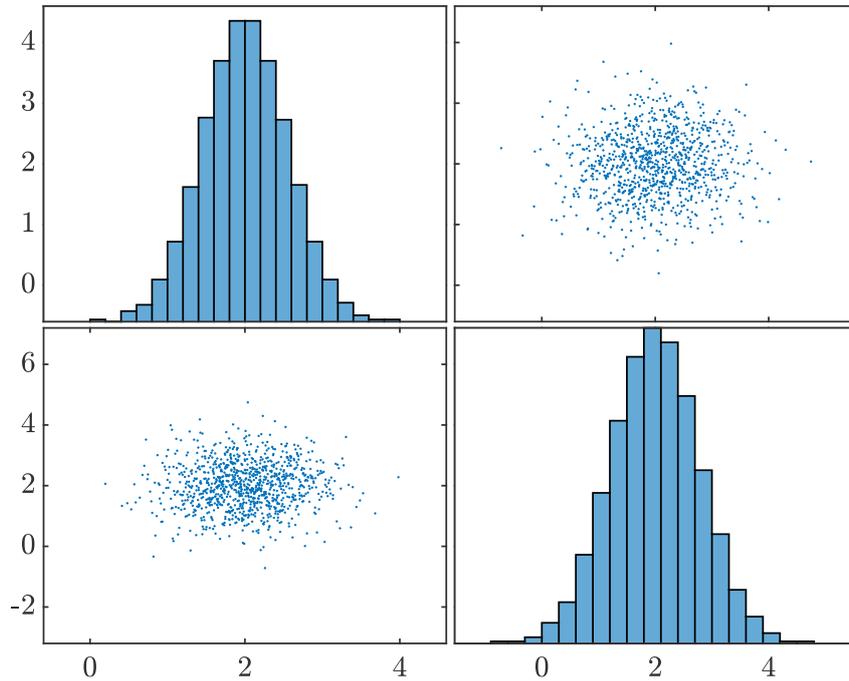


Figure 4.17: Exemplary parameter uncertainties for demand multiplier $Z_{1,\dots,6} \sim \mathcal{N}(2, 0.6)$ and roughness multiplier $Z_{7,\dots,12} \sim \mathcal{N}(2, 0.3)$.

$N_{kde} = 1e5$ evaluations of the resulting meta-model, which is computationally insignificant.

Figure 4.18 shows the estimated probability distribution of the flow rate through one of the pipes in the highly looped section of the network that is located next to the reservoir. In the χ^2 -test it is confirmed that the flow rate is distributed normally. The location of the pipe suggests that the demand flows from the all over the network are collected at this location which explains the high flow rates and the Gaussian distribution.

For the convergence it can be seen that the PCE is able to give a very accurate estimate in mean (Figure 4.19) and variance (Figure 4.20) of the flow, even for small sample sized, while the Monte Carlo simulation starts converging after a number of $1e4$ samples. This fast convergence of the PCE could also be observed for the flow rate in the illustrative model and can be explained by the linear behaviour that is easily approximated by a second order PCE.

In Figure 4.21 the flow is illustrated for one of the pipes in the highly looped area with a medium distance from the reservoir. The distribution is slightly skewed. This is confirmed by the rejection of the null hypothesis in the χ^2 -test. For the mean in Figure 4.22 and especially the variance in Figure 4.23 it can be seen, that convergence for the PCE is reached with about $1e3$ samples. Although the flow rate distribution in this part of the network is clearly non-

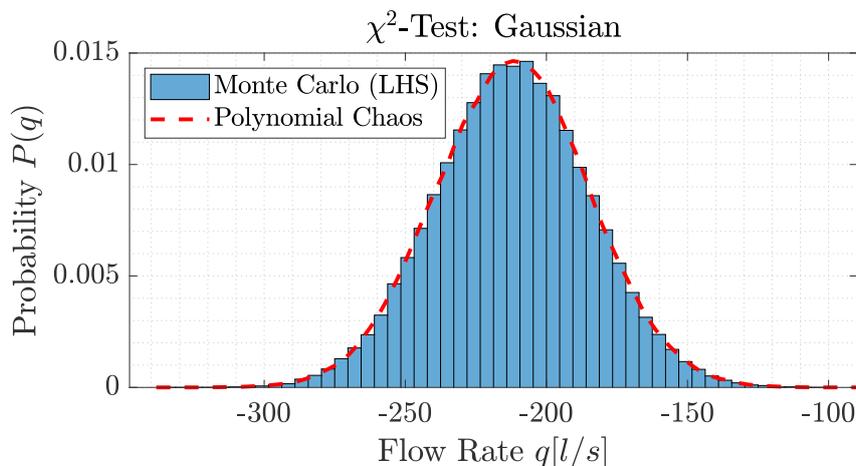


Figure 4.18: Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6th-order PCE.

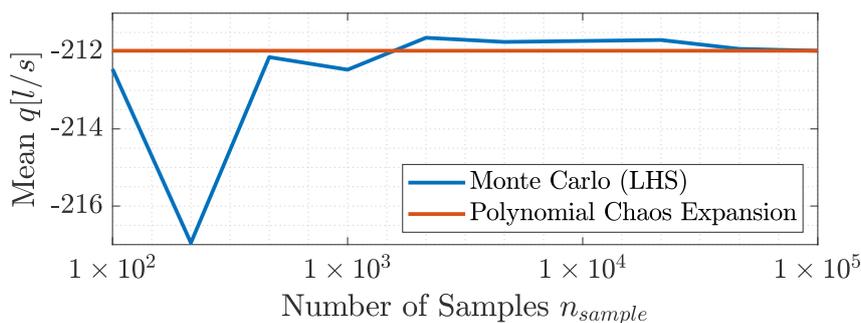


Figure 4.19: Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

Gaussian, in the same part of the network also contains pipes with similar characteristics where the flow rates were classified as normally distributed.

Figure 4.24 shows the probability distribution for the head at one of the nodes in the looped region of the network. In contrast to the flow rate distribution, which can be linear or non-linear at different positions in the network, this result is very characteristic for the head distributions all over the network. The form of the distribution can likely be explained from the head-loss function that is approximately quadratically dependent on the flow rate. For a normally distributed flow rate this leads to χ^2 distribution of the head-loss and the $1 - \chi^2$ characteristic in the remaining head.

The convergence in mean head illustrated in Figure 4.25 and head variance depicted in Figure 4.26, once again shows that the PCE stabilizes around

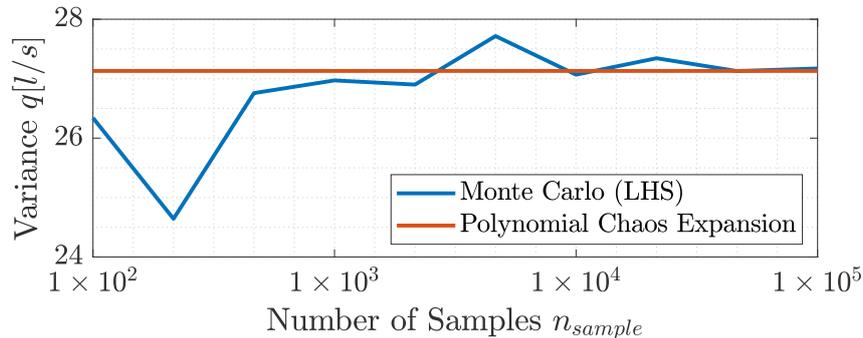


Figure 4.20: Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

$n_{samples} = 1e3$ while the Monte Carlo simulation needs more than $1e4$ samples. As before for it can be seen that for non-Gaussian distributions the PCE needs considerably more samples than for close to Gaussian distributions.

Discussion

In the Flow rates few generalizations can be made from the results of the high dimensional scenario. For one it can be seen that flow rates close to the reservoir follow a Gaussian distribution. This is most likely due to the effect of demand agglomeration. As the network is supplied by a single reservoir all demands have to flow through these pipes. In contrast, for arbitrary pipes that are part of the loop structure in the network it is not possible to give any sort of generalization. There results show some flow rate distributions that are classified as Gaussian and others as non-Gaussian by the χ^2 -test with no apparent difference to the topological position. However, the results from the illustrative network example has some added relevance, as it describes the behavior of the tree structures in the network.

For the head on the other hand, it can be seen that the distribution generally has a characteristic form similar to the one shown in Figure 4.24. As stated before, this form can be directly explained by the close to quadratic relation in the head-loss formula which leads to a χ^2 distribution. Similar results have been reported for a study by Piller *et al.* [2003] who introduced a stochastic consumer model based on a binomial distribution and propagate the uncertainties using MCS and the FOSM. In contrast, Hwang *et al.* [2017] report that nodal heads behave more linearly for looped network. This difference in results might be explained by the topology of the network and the fact that distribution networks in North America are over-sized, compared to their European counterparts. This would lead to laminar flow and in turn the

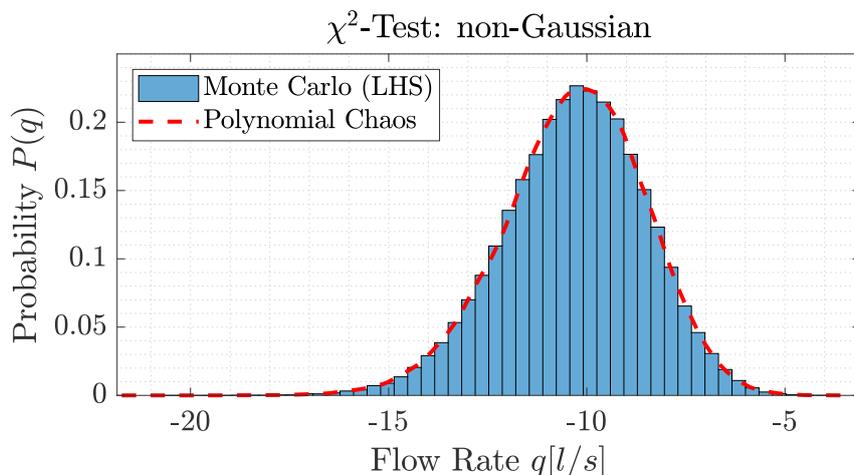


Figure 4.21: Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

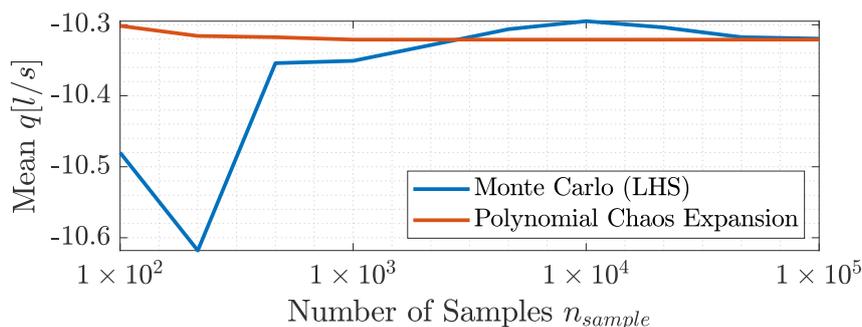


Figure 4.22: Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

head-loss function would become linear. In that case a normally distributed probability density function for the head would be a logical conclusion.

For the application of the PCE with the high dimensional parameter space a number of issues have to be mentioned. First, for an efficient evaluation of the Monte Carlo approach the samples have been generated using Latin Hypercube sampling. While this makes the MCS more efficient, this benefit is not limited to the method as the PCE collocation approach profits from this in the same way. Second, Section 4.2.2 shows that the coefficients are estimated based on the linear regression (4.19). The calculation of the coefficients involves the inversion of a $M \times N$ matrix, where M is the number of collocation points and N is the number of coefficients in the meta-model. This matrix inversion can become very demanding or even prohibitive for memory resources in cases of huge sample sizes or a very big number of coefficients. This is not the case for

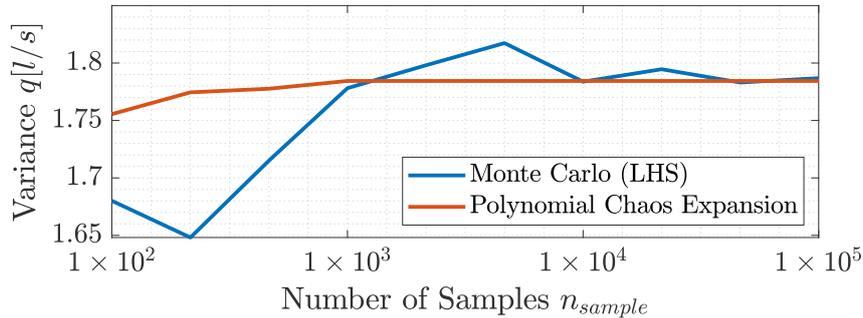


Figure 4.23: Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

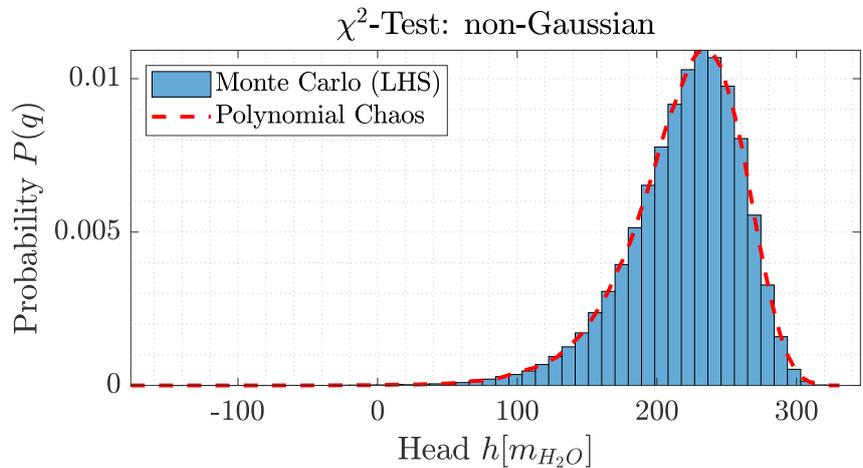


Figure 4.24: Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

the intrusive PCE approach. However, as stated in [Le Maître et Knio \[2010\]](#) the development of special solvers is necessary for such applications.

4.4 Conclusion

The objective of this chapter is the introduction of sophisticated propagation methods for uncertain parameters to the hydraulic equations.

Section 4.1 gives an overview on current state of the art methods in uncertainty analysis.

In Section 4.2 the polynomial chaos expansion is introduced as a novel approach for the propagation of parameter uncertainties in non-linear models. For the intrusive application of the PCE, the derivation of the stochastic hydraulic equations is presented on the basis of the Galerkin projection, together

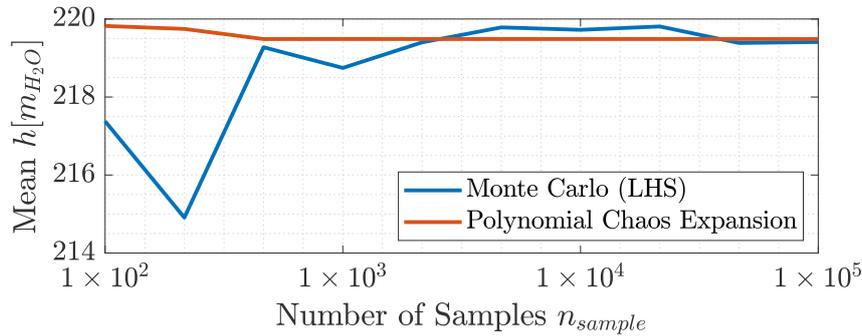


Figure 4.25: Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

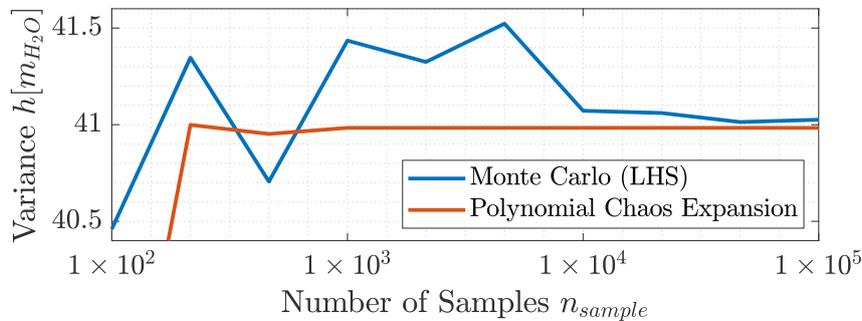


Figure 4.26: Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

with the derivation of a pseudo-spectral formulation for the non-polynomial head-loss function. This is followed by a discussion on the application of the non-intrusive spectral projection, which permits the use of the PCE as the basis of deterministic legacy code.

Section 4.3 applies the PCE to a selection of challenging use cases in order to evaluate its performance to traditional methods. The application of intrusive and non-intrusive Polynomial Chaos expansion methods to the hydraulic models has shown that it is capable to capture non-linearities in the quantities of interest with comparable accuracy to Monte Carlo simulations. But, in comparison to these classical Monte Carlo simulations the Polynomial Chaos expansion is by far computationally more efficient. In the application of the intrusive Galerkin projection, the reformulation of the system equations for the calculation of the expansion coefficients together with the necessary approximation of the non-polynomial head-loss function has been found to be challenging. Especially since this new set of equations is only applicable to one specific instance of the parameter space. The application of non-intrusive

methods has proven to be more generic. It allows for an easy adaptation to new input parameters and deals with non-polynomial elements directly through the sampling of parameter space.

The application of the PCE to the highly non-linear PDM model particularly demonstrates the benefit of the method over the perturbation methods. Based on a relatively small sample size, the method is able to capture the second peak in the probability distribution and the hard cut-off introduced by the pressure outflow relation.

In the application to high dimensional parameter spaces experience has shown, that the PCE profits from efficient sampling strategies in the way as the MCS. However, in the fitting of the meta-model current implementations have to calculate the inverse of a matrix whose size is determined by the number samples times the number of elements in the polynomial expansion. Due to this, system memory becomes a limiting factor in practical applications.

Applying the first order second moment method to the hydraulic equations has been shown to be limited since it is not possible to capture non-linear effects. But, due to the direct formulation of the sensitivities and highly efficient calculation of the covariance matrix, this method may still be relevant in cases with high dimensional input spaces and concentrated variance in the uncertain parameters.

For all the use cases, it is assumed that the product from the number of nodal consumers and the probability of consuming is greater than 5 so that the normal distribution approximation holds. However, the framework of the generalized polynomial chaos expansion also allows to model parameter uncertainties that follow distributions like the Binomial and the Poisson distribution which should be tested in the future.

While a large part of the evaluations in this section are based on Matlab code that was created during this thesis, some software solutions, that offer applications for the PCE, have been evaluated. First, there exists a NISP toolbox for the scientific calculation environment SciLab. However, this toolbox is not actively supported and incompatible with the current version. At the ETH Zurich the reliability toolbox UQLab is developed. While some parts of the package are open source, the core of the application is closed off and licensed. The software Dakota developed by Sandia National Laboratories gives a broad platform for parameter study based methods like optimization and uncertainty quantification. Applications can be implemented either through script based coupling of Dakota and the computational model or directly through the open C++ code. The open source project OpenTURNS has not been investigated further, but it seems to be supported by an active community.

Chapter 5

Efficient Uncertainty Propagation

Motivation: *Combined application of deterministic and stochastic meta models for the efficient propagation of parameter uncertainties.*

Principal Elements:

- *Projection based reduced order models*
- *Non-intrusive polynomial chaos expansion*

Conclusions:

- *Combined application brings massive gains in computational efficiency.*
- *Small sample sizes in the snapshot database limit the attainable accuracy of the reduced model.*
- *Degrees of freedom in the reduced order model can be seen as design parameters for the attainable accuracy.*

With the successful development of the reduced order model for the hydraulic systems in Chapters 3 and the encouraging results from the spectral uncertainty quantification in Chapter 4, the next logical step is the combined application of the deterministic and the stochastic meta models. To do so, this chapter applies the efficient uncertainty propagation to one of the model networks. The results of the uncertainty quantification based on the reduced order model is evaluated with respect to its accuracy and the benefits in computational cost and compared to more conventional approaches.

5.1 Realistic Network Model

Subject for the combined validation of the efficient uncertainty propagation is the realistic network model from the Paris region, introduced in Section 2.4.

5.1.1 Scenario

The scenario is defined similar to the use case discussed in the development of the reduced order model in Chapter 3. It is based on the loop-flow formulation of the demand-driven hydraulic equations. As such the quantities of interest are given by the state description in the loop-flow vector \mathbf{q}_c .

The two-dimensional parameter-space for the parameter uncertainties is defined in the demand-multipliers using a bivariate normal distribution. They are applied globally to the demand nodes in the complete network through the selection matrix \mathbf{K} . The reduced model is constructed using the same parameter-space as the parameter uncertainties. In the application it is important to ensure that the samples from the parameter distribution is confined to the pre-sampled region in the reduced model, otherwise inaccuracies could be introduced due to extrapolation.

The main interest in contrast to the studies in Chapter 4 is used of the reduced order model for the evaluation of the stochastic collocation points. As it has been discussed for the application of the reduced order model, two of the most important factors influencing its accuracy and performance are given by the models degrees of freedom and the refinement of the snapshot database. These two factors and their influence on the probability distribution in the QoIs will be investigated in the following sections.

5.1.2 Results

Figure 5.1 shows the histogram for the conventional validation procedure using a full scale Monte Carlo sampling with a total number of $n_{MC} = 10^4$ samples on the high-fidelity model. In comparison the kernel density estimation is illustrated for a 12th order polynomial chaos expansion is shown. The PCE only uses a tenth of the full scale evaluations, with $n_{PCE} = 10^3$. While the PCE is able to capture the non-linear behaviour in the distribution, a certain amount of smoothing is apparent. This is also expressed in the difference for the mean values in the MC and PCE distributions.

The results for the application of the reduced order model for the propagation of the parameter uncertainties is shown in Figures 5.2 and 5.3. Both models used in the estimation of the probability density function are constructed on the basis of the adaptively sampled, 1681 element snapshot database.

Figure 5.2 illustrates the estimate for the probability density function using a reduced order model with a total number of 10 degrees of freedom. It is obvious, that the application of this low order approximation to the hydraulic model is not at all capable to capture the complex structure of the PDF shown in Figure 5.1. However, it does shows to be accurate in some of the main probability measures like the mean, variance and the confidence intervals in the tails of the distribution.

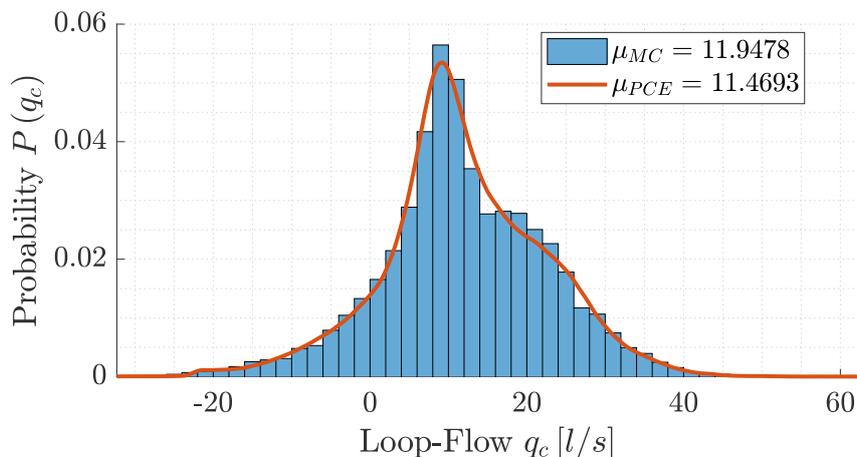


Figure 5.1: MC and PCE estimates of the probability density function for the loop flow, using the high fidelity model.

Increasing the degrees of freedom for the reduced order model clearly increases the model's capability to capture the intrinsic non-linear effects. The estimated probability density function for the loop-flow in Figure 5.3 is based on a reduced order model with a total number of 40 degrees of freedom. While a considerable amount of smoothing is visible compared to the full order model in Figure 5.1, the model behaviour is approximated much more closely.

Figure 5.4 shows the relative computational effort as a function of the complexity for the reduced order model in its degrees of freedom. The baseline in all cases is given by the computational effort for the full scale Monte Carlo simulation using the high fidelity-model. The figure contains the blue line that visualizes the direct reduction in computational effort due to the performance gain in the evaluation of the reduced order model. For the range of models analysed in this section, this benefit can be quantified by a 40% to 50% reduction. The second line visualizes the reduction achieved in the combined application of the deterministic and stochastic meta model, which is in the range of 90% to 95%.

In the context of the results presented in the construction of the reduced order model in Chapter 3 it has to be mentioned that during these calculations the Jacobian matrix was evaluated through the direct formulation given in (3.21) as opposed to numerical differentiation.

On the influence of the fidelity of the snapshot database pretty similar observations could be made to the influence of the model complexity. For the use of a sufficiently big snapshot database, which in the context of this study is the case for the adaptive sampling with more than 441 elements, no bigger limitations of the model could be observed, if a sufficient number of modes are used. However, for the use of a small snapshot database, like the basic 121 element Cartesian sampling introduced in Chapter 3, the achievable accuracy

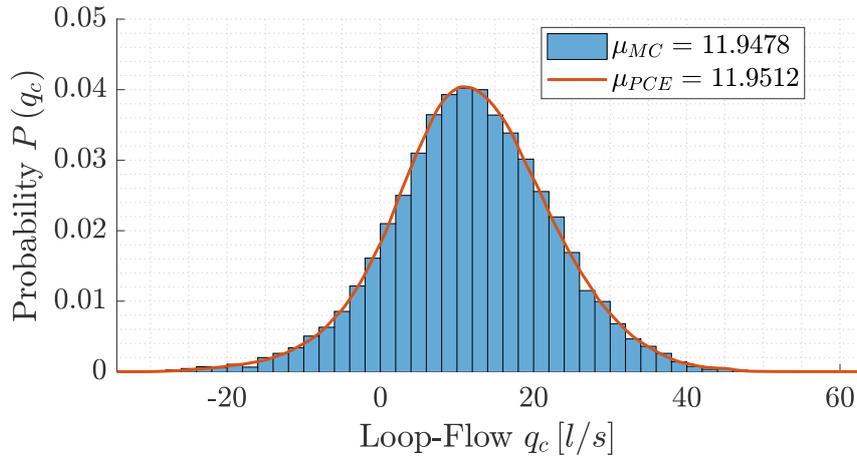


Figure 5.2: MC and PCE estimates of the probability density function for the loop flow, using the reduced 10 mode model and the 1681 element adaptive sampling.

in the probability density function is limited to that comparable to a 10 degrees of freedom model, even if higher numbers of modes are used.

5.1.3 Discussion

Comparing the results from the random sampling based on the full model to the results of the reduced order model, shows that the reduced model gives a good approximation of the system. The results of the full scale validation reproduce the experience from the spectral propagation of uncertainties in Chapter 4, namely the faster convergence of the spectral approximation compared to the Monte Carlo simulation for a greatly reduced number of stochastic collocation points. This also translates into a reduction in computational effort.

With respect to this, the main focus of the study is on the influence of the reduced order model and the effect of the discretization in the parameter-space. For the effect of the degrees of freedom in the reduced order model the results of the study illustrate strikingly how, with the reduction of fidelity, the estimate for the probability density function of the loop-flow approaches more and more that of a linear system. Looking at the influence of the sampling in the parameter-space, similar behaviour could be reproduced for the reduction of sampling points in the snapshot matrix. The most obvious interpretation of this behaviour, with respect to the model complexity, is that the first POD mode can be seen as the normal or mean state in the loop-flow vector and the addition of further degrees of freedom models the non-linear modes for a better approximation of the high fidelity system. This interpretation is also in line with the effect of the sparsely sampled parameter-space, where the inclusion of the higher modes does not improve the results due to the lack of these

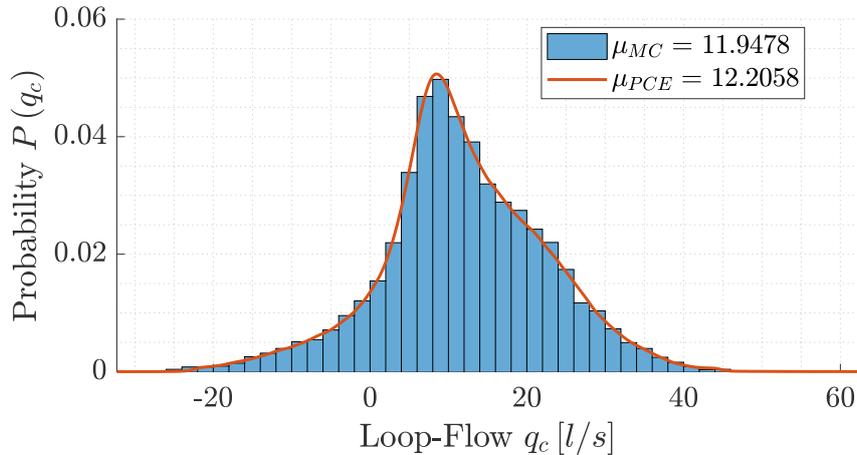


Figure 5.3: MC and PCE estimates of the probability density function for the loop flow, using the reduced 40 mode model and the 1681 element adaptive sampling.

higher modes in the snapshot matrix. However, although the full non-linear behaviour is not reproduced for the low order or sparsely sampled models, this does not discard the use of these models. An interesting observation that has been made is, that these models produce good results with respect to mean values, variance and the description of the distribution tails with the confidence interval.

While all these effects shown in the chapter on reduced order modelling could be reproduced in the context of the uncertainty propagation, the stochastic sampling seems to be more forgiving with respect to the model accuracy. A possible explanation for this observed behaviour is, that the parameter uncertainties are normally distributed around the centre of the parameter-space. The inaccuracies in the reduced order model are especially high for lower values of the demand multipliers. These errors are mitigated by the fact that these regions are not sampled as densely for the creation of the stochastic surrogate model.

5.1.4 Conclusion

This chapter successfully shows the improved benefit of the combined application for the reduced order model and the polynomial chaos expansion in the uncertainty propagation framework. While the separate benefits for each methods could be confirmed once more, their combination proves to be far superior and to greatly reduced the computational effort.

However, in the context of practical applications a number of consideration have to made. The massive gain in computational performance during the on-line step is very attractive for the use with parametrised models. Further, the

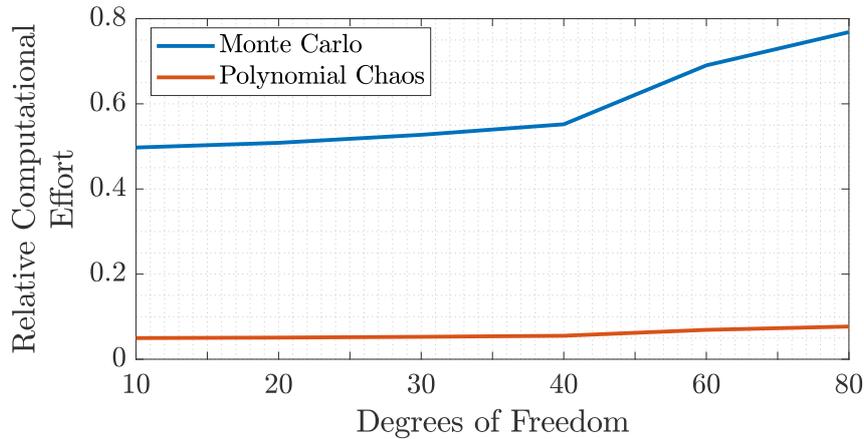


Figure 5.4: Relative computational effort for the reduced order model MC sampling and the 12th order PCE.

observed mitigation of the approximation errors in the reduced order model allows for the use of more time efficient models with a lower number of degrees of freedom. Some aspects of the application that have to be addressed during the model creation are the lack of an a priori error estimation. As a result it has to be seen as an iterative process where a deliberate choice of the model parameters has to be made. The development of such a priori error estimators is part of the current development and would greatly increase the efficiency during the application of these methods.

In conclusion the choice of the model parameters is closely tied to the required level of detail. If the main interest is to describe the resulting distribution in the first and second moment or the confidence intervals, the use of a low order model on a sparsely sampled parameter-space leads to a highly efficient propagation of the uncertainties. In the case of more detailed applications, for example in the context of a Bayesian framework, a more complex reduced model can offer a sizeable reduction in computational effort while creating a close approximation of the true probability distribution for the quantity of interest.

Conclusion

Work

- In Chapter 1 the basic mathematical tools that are used throughout the thesis are introduced in topical order. First, a short introduction is given for graph theory. This is relevant in the hydraulic modelling process for water distribution networks, the hierarchical reduction of these models and the Delaunay triangulation. The adjoint method is presented as a convenient tool for the determination of the system derivatives in Chapter 2. The SVD is introduced as one possible method to obtain a suitable orthogonal basis for the projection based model reduction methods used in Chapter 3 and the Delaunay triangulation builds the corner stone of the adaptive sampling strategy. Finally, for the use in Chapter 4 some basic definitions for statistics are introduced together with the framework of orthogonal polynomials for their use in the polynomial chaos expansion.
- Chapter 2 deals with the hydraulic network model in general. Section 2.1 introduces the demand-driven modelling paradigm. This contains the presentation of the friction induced head-loss, the formulation of the static hydraulic equations and the formulation of the sensitivity matrix. This is followed by the presentation of the pressure-driven modelling paradigm in Section 2.2, with the introduction of the pressure-outflow relationship and the formulation of the altered hydraulic equations. The capabilities of both the demand-driven and pressure-driven modelling paradigms for modelling deficient networks are discussed in Section 2.3. The chapter closes with the presentation of the graphs for the example networks that are used throughout the thesis.
- In Chapter 3 a categorization of common reduced order modelling techniques is presented with interpolation based methods, projection based methods and hierarchical models. This is followed by the definition of a common validation procedure on the basis of the approximation error and the projection error in Section 3.4. For the application to the hydraulic network models, the interpolation based and projection based

methods are introduced in further detail. Section 3.1 introduces classical response surface interpolation models and interpolation models based on the radial basis functions. For the projection based methods Section 3.2 concentrates on the introduction of the SVD based proper orthogonal decomposition. Further, the reduced hydraulic model equations are derived for the head-flow and the loop-flow equations, together with the projected form for the direct evaluation of the Jacobian matrix. Finally it introduces the use of global and local reduced basis. Section 3.3 introduces the application of the empirical interpolation method for the application in the hydraulic equations, followed by the introduction of an adaptive projection error based sampling strategy in Section 3.5.

- Chapter 4 deals with the propagation of parameter uncertainties using mathematical models. It first gives a characterization of the basic propagation approaches with the perturbation methods, sampling or stochastic collocation methods and the stochastic spectral methods. In Section 4.1 the perturbation method and the Monte Carlo simulation are presented in more detail followed by the introduction of the polynomial chaos expansion in Section 4.2. The section continues to discuss the formulation of the stochastic spectral equations for the hydraulic model.
- Chapter 5 evaluates the combined application of reduced order modelling techniques and spectral uncertainty propagation on the example of a realistic water distribution network with a two dimensional parameter-space.

Findings & Contributions

- In Chapter 2 a new approach to the derivation of the hydraulic sensitivities through the use of the adjoint method is presented, in contrast to the direct formulation reach through the differentiation of the hydraulic equations. The new way to derive the hydraulic sensitivities through the use of the adjoint method gives an alternative to the conventional direct formulation reach through the differentiation of the hydraulic equations.
- Chapter 3 shows for the first time an application of the projection based POD method to the hydraulic equations of a water distribution network. It compares the accuracy and computational efficiency of interpolation based and projection based methods for the application in a realistic network model. Further, a new projection error based adaptive sampling strategy is introduced. For an even more efficient evaluation of the hydraulic system the empirical interpolation methods is applied for the first time in the hydraulic modelling framework to decrease the computational effort for the non-linear residual function.

- It shows, that both the interpolation based and the projection based reduced order modelling approach can produce a meta-model of sufficient accuracy in the approximation error. However, convergence in the interpolation model comes at the price of a high number of evaluations in the high-fidelity model. This makes its use limited, as higher dimensional applications are computationally prohibitive. The POD is much more efficient in this perspective as it keeps the physical information from the hydraulic equations through the Galerkin projection. This makes it more efficient from a sampling point of view. Further, the possible dimensional reduction in the systems degrees of freedom results in an increased performance for solving the non-linear set of equations. Depending on the fact if the Jacobian matrix is evaluated by its direct formulation or numerically the computational time is reduced by 50 – 90%.
- Even though the pure application of the POD results in such a speed up, the main computational effort lays in the evaluation of the non-linear residual function. The application of hyper reduction techniques like the empirical interpolation method to the non-linear head-loss function is supposed to greatly reduce this factor. However, the experience in this thesis is, that the parameter-space of the head-loss function is too complex for the application of sampling based methods, as the reduced system state vector has to be added to it, making the problem intractable. While the approximation and projection errors could be brought to an acceptable level, there is no gain in computational efficiency and the solver for the reduced system on basis of the EIM does not converge, most probably due to the solver stepping out of the region that was pre-sampled during the off-line step.
- The application of the adaptive sampling strategy shows that choosing the sampling points based on the leave-one-out error greatly improves the result of the reduced order model. It could also be shown that the use of the projection error and a local basis in the sampling strategy greatly increases its performance, as the calculation of the projection is very efficient and the formulation of a local basis limits the number of points influenced by a new sample. Comparing the use of global and local basis, the applications in this thesis have shown the global approach to be more efficient. However, as the number of significant modes created from the local basis is usually lower than the number of degrees of freedom, this is more a sign that the construction of the local basis is inefficient.
- Chapter 4 uses for the first time a spectral series expansion method for the propagation of uncertainties in the context of the hydraulic water distribution network model. Further, a full formulation of the intrusive stochastic hydraulic equations is derived. Finally, a new spectral for-

mulation of the non-polynomial head-loss function is derived using the pseudo-spectral projection.

- Comparing the Monte Carlo simulation to the non-intrusive spectral projection, while they are both stochastic collocation methods, the PCE uses the samples more efficient and can use the smoothness of the orthogonal basis functions in order to converge much faster. Analysing the results of the intrusive and the non-intrusive approach, they are both in good agreement. However, the application of the intrusive formulation is not easily adaptable to changes in the parameter-space or in the expansion order. This makes the non-intrusive approach much more versatile, as it allows for the use of legacy code and the easy adaptation of the spectral expansion. Determining the appropriate expansion order of the PCE is an iterative process. As currently there exists no a priori error estimation, the suitability of a selected PCE is commonly evaluated a posteriori. This is usually performed as a post processing step, through the evaluation of the convergence in the PCE coefficients.
- In the context of an application with strongly non-linear effects like the PDM model, the PCE manages to capture the non-linear effect on the PDF of the QoIs, in contrast to the FOSM. The results are comparable to those of the MC simulation and it uses less collocation points in the process. However, compared to the MC simulation, the use of the continuous polynomials has a problem in approximating the hard cut-off in the head and flow distributions, which can be explained by the Gibbs phenomenon.
- For the application in the context of a 12-dimensional parameter-space similar observations could be made. In order to alleviate the computational burden Latin Hypercube sampling was used for the sample creation. Both the MC and the PCE profit from this method. Analysing large scale applications the non-intrusive calculation of the expansion coefficients may become computationally prohibitive, as it performs a matrix inversion. LHS and least angle regression models are ways to mitigate the problem, however in cases where this is insufficient the use of an intrusive model may be the last resort. However, the use of such a model may necessitate the development and use of special solvers.
- The application proves the potential of the combined application of a deterministic and a stochastic meta-model.
- As a conclusion, it could be shown that the number of degrees of freedom chosen in the reduced order model is one of the main factors for the accuracy of the resulting probability density function, together with the refinement of the snapshot matrix. For models built on a small sample

of snapshots, the reachable accuracy is limited even with the inclusion of a high number of degrees of freedom in the reduced model, as the reduced basis does not contain enough information. For a sufficiently dense sampled parameter-space, the models degrees of freedom can be used to recreate the distribution of the high-fidelity model with varying accuracy. For most of these approximations to the PDF there is a good approximation in the first two stochastic moments, as well as the confidence intervals. This means that, depending on the application, the required model complexity may be even lower than what has been deemed a suitable approximation in the context of this thesis. A general effect that has also been observed is, that the nature of the stochastic application in the stochastic collocation process strongly mitigates the errors of the reduced order model especially in the region of lower demand multipliers.

Future Work

- The discussion on the deficient networks shows that in the future, further effort has to be put in the development of better hydraulic models, as for current models there exist numerous extreme scenarios for which they fail to produce reliable results.
- Although the main application of the reduced order modelling techniques was successful, there remain some questions that have to be studied further in the future. This concerns for one the application of the framework to higher dimensional problems. Applications of the reduced model in this thesis were limited to two dimensions, mainly to facilitate the validation and analysis process. However, real life applications like optimization or uncertainty quantification generally use a much higher number of parameters. In those frameworks the scalability of these methods has to be evaluated. The second focus lays on the application of the EIM. While the application of the methods was not met with success in this thesis, there exist a number of approaches that can improve its performance. One possibility could be the reformulation of the hydraulic problem in an optimization framework. This way it would be ensured that the solver does not leave the pre-sampled region. A second possibility is the inclusion of the derivatives into the interpolation method, that could help the solver to converge. A third approach is the use of a pre-conditioner to make the problem more suitable for the numerical solution process. Finally, as the construction of the local reduced basis has been shown to lack in fidelity, the selection process for the local snapshot matrix has to be adapted to give a more accurate basis.
- One of the main problems in the application of the polynomial chaos

expansion is the lack of an a priori error estimate. The availability of such an error measure would greatly improve the use of the PCE in practical applications. This means that finding a suitable model may require a number of iteration. Further, even though the method is more efficient in the use of stochastic collocation points than the Monte Carlo simulation, the application to high-dimensional problems will become computationally infeasible. This may at best be mitigated through the application of the intrusive stochastic model, as in this case the size of the system of equations may increase and become hard to solve standard solution methods.

- While the test application was limited to a case where the Jacobian matrix was evaluated using the direct formulation, applying it in the context of the numerical Jacobian may further improve the computational benefit from around 93 – 95% to about 99% or even more. Further improvement of the efficient propagation of uncertainties is directly influenced by the reduced order modelling and the polynomial chaos expansion. From this it can be concluded that the possible improvements to these methods also benefit the combined application. A second path, that has not been followed in this thesis, is the combination of the reduced order model with the intrusive spectral projection. Formulating the stochastic equations of a big deterministic model may lead to a massive system of equations. The size of such a stochastic system may greatly benefit from the reduced number of degrees of freedom given by a projection based reduced hydraulic model.

Bibliography

- AMSALLEM, David, CORTIAL, Julien, CARLBERG, Kevin et FARHAT, Charbel, 2009. A method for interpolating on manifolds structural dynamics reduced-order models. *International journal for numerical methods in engineering*, 80(9):1241–1258.
- AMSALLEM, David et FARHAT, Charbel, 2008. Interpolation method for adapting reduced-order models and application to aeroelasticity. *AIAA journal*, 46(7):1803–1813.
- AMSALLEM, David et HAASDONK, Bernard, 2016. Pebl-rom: Projection-error based local reduced-order models. *Advanced Modeling and Simulation in Engineering Sciences*, 3(1):6.
- ANTOULAS, Athanasios C, 2005. *Approximation of large-scale dynamical systems*, tome 6. SIAM.
- ASKEY, Richard et WILSON, James Arthur, 1985. *Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials*, tome 319. American Mathematical Soc.
- BALAJEWICZ, Maciej, AMSALLEM, David et FARHAT, Charbel, 2016. Projection-based model reduction for contact problems. *International Journal for Numerical Methods in Engineering*, 106(8):644–663.
- BAUR, Ulrike, BEATTIE, Christopher, BENNER, Peter et GUGERCIN, Serkan, 2011. Interpolatory projection methods for parameterized model reduction. *SIAM Journal on Scientific Computing*, 33(5):2489–2518.
- BECK, Stephen et COLLINS, Richard, 2008. Moody diagram. Online; accessed January 22, 2019.
URL https://commons.wikimedia.org/wiki/File:Moody_EN.svg
- BENNER, Peter et BREITEN, Tobias, 2011. On h2-model reduction of linear parameter-varying systems. *PAMM*, 11(1):805–806.

- BENNER, Peter, FENG, Lihong, LI, Suzhou et ZHANG, Yongjin, 2015a. Reduced-order modeling and rom-based optimization of batch chromatography. Dans *Numerical Mathematics and Advanced Applications-ENUMATH 2013*, pages 427–435. Springer.
- BENNER, Peter, GUGERCIN, Serkan et WILLCOX, Karen, 2013. A survey of model reduction methods for parametric systems.
- BENNER, Peter, GUGERCIN, Serkan et WILLCOX, Karen, 2015b. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531.
- BHAVE, Pramod, 1981. Node flow analysis distribution systems. *Transportation engineering journal of ASCE*, 107(4):457–467.
- BIRKHOFF, Garrett, 1963. A variational principle for nonlinear networks. *Quart. Appl. Math*, 21:160–162.
- BLATMAN, Géraud et SUDRET, Bruno, 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367.
- BLOKKER, EJM et VAN DER SCHEE, W, 2006. Simulation of water demands provides insight. *Water Supply and Drainage for Buildings, CIB W*, 62.
- BOND, Bradley et DANIEL, Luca, 2007. A piecewise-linear moment-matching approach to parameterized model-order reduction for highly nonlinear systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 26(12):2116–2129.
- BORGGGAARD, Jeff, POND, Kevin et ZIETSMAN, Lizette, 2014. Parametric reduced order models using adaptive sampling and interpolation. *IFAC Proceedings Volumes*, 47(3):7773–7778.
- BOULOS, Paul, Lansy Kevin et KARNEY, Bryan, 2004. *Comprehensive Water Distribution Systems Analysis Handbook For Engineers and Planners*. MWH Soft Inc.
- BRAUN, Mathias, BERNARD, Thomas, PILLER, Olivier et SEDEHIZADE, Fereshte, 2014. 24-hours demand forecasting based on sarima and support vector machines. *Procedia Engineering*, 89:926–933.
- BRAUN, Mathias, PILLER, Olivier, DEUERLEIN, Jochen et MORTAZAVI, Iraj, 2017. Limitations of demand-and pressure-driven modeling for large deficient networks. *Drinking Water Engineering and Science*, 10(2):93.

BIBLIOGRAPHY

- BUCHBERGER, Steven et WELLS, Greg, 1996. Intensity, duration, and frequency of residential water demands. *Journal of Water Resources Planning and Management*, 122(1):11–19.
- BUI-THANH, Tan, WILLCOX, Karen et GHATTAS, Omar, 2008. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288.
- BURKARDT, John, GUNZBURGER, Max et LEE, Hyung-Chun, 2006. Centroidal voronoi tessellation-based reduced-order modeling of complex systems. *SIAM Journal on Scientific Computing*, 28(2):459–484.
- BUSH, Cheryl et UBER, James, 1998. Sampling design methods for water distribution model calibration. *Journal of Water Resources Planning and Management*, 124(6):334–344.
- CACUCI, Dan, IONESCU-BUJOR, Mihaela et NAVON, Ionel Michael, 2005. *Sensitivity and uncertainty analysis, Volume II: Applications to large-scale systems*, tome 2. CRC press.
- CARLBERG, Kevin, TUMINARO, Ray et BOGGS, Paul, 2015. Preserving lagrangian structure in nonlinear model reduction with application to structural dynamics. *SIAM Journal on Scientific Computing*, 37(2):B153–B184.
- CARPENTIER, Pierre et COHEN, Guy, 1993. Applied mathematics in water supply network management. *Automatica*, 29(5):1215–1250.
- CARPENTIER, Pierre, COHEN, Guy et HAMAM, Y., 1985. Water network equilibrium, variational formulation and comparison of numerical algorithms. *EURO VII*.
- CHATURANTABUT, Saifon et SORENSEN, Danny, 2010. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764.
- CHATURANTABUT, Saifon et SORENSEN, Danny C, 2009. Discrete empirical interpolation for nonlinear model reduction. Dans *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 4316–4321. IEEE.
- CHENEY, Elliott Ward, 1966. Introduction to approximation theory.
- CHERRY, Colin, 1951. Cxvii. some general theorems for non-linear systems possessing reactance. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 42(333):1161–1177.

- CHEUNG, PB, VAN ZYL, JE et REIS, LFR, 2005. Extension of epanet for pressure driven demand modeling in water distribution system. *Computing and Control for the Water Industry*, 1:311–316.
- CHIHARA, Theodore, 2011. *An introduction to orthogonal polynomials*. Courier Corporation.
- CHUNG, T., 2010. *Computational fluid dynamics*. Cambridge university press.
- COLEBROOK, Cyril, BLENCH, T., CHATLEY, H., ESSEX, E., FINNIECOME, J., LACEY, G., WILLIAMSON, J. et MACDONALD, G., 1939. Correspondence. Turbulent flow in pipes, with particular reference to the transition region between the smooth and rough pipe laws.(includes plates). *Journal of the Institution of Civil engineers*, 12(8):393–422.
- COLLINS, M. ; Cooper, L. ; Helgason R. ; Kennington J. ; LeBlanc L., 1978. Solving the pipe network analysis problem using optimization techniques. *Management Science*, 24:747–760.
- CROSS, Hardy, 1936. Analysis of flow in networks of conduits or conductors. Rapport technique, University of Illinois at Urbana Champaign, College of Engineering. Engineering Experiment Station.
- CROUS, Pedro, VAN ZYL, Jakobus et ROODT, Yuko, 2012. The potential of graphical processing units to solve hydraulic network equations. *Journal of Hydroinformatics*, 14(3):603–612.
- DE STURLER, Eric, GUGERCIN, Serkan, KILMER, Misha E, CHATURANTABUT, Saifon, BEATTIE, Christopher et O’CONNELL, Meghan, 2015. Nonlinear parametric inversion using interpolatory model reduction. *SIAM Journal on Scientific Computing*, 37(3):B495–B517.
- DEGROOTE, Joris, VIERENDEELS, Jan et WILLCOX, Karen, 2010. Interpolation among reduced-order matrices to obtain parameterized models for design, optimization and probabilistic analysis. *International Journal for Numerical Methods in Fluids*, 63(2):207–230.
- DELAUNAY, Boris, 1934. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7(793-800):1–2.
- DEUERLEIN, Jochen, 2008. Decomposition model of a general water supply network graph. *Journal of Hydraulic Engineering*, 134(6):822–832.
- DEUERLEIN, Jochen, SIMPSON, Angus et MONTALVO, Idel, 2012a. Preprocessing of water distribution systems to assess connectivity and solvability in the presence of flow control devices. *World Environmental and Water Resources Congress 2012: Crossing Boundaries*, pages 3237–3247.

BIBLIOGRAPHY

- DEUERLEIN, Jochen, SIMPSON, Angus *et al.*, 2012b. A modified global gradient algorithm based on topological decomposition. Dans *WDSA 2012: 14th Water Distribution Systems Analysis Conference, 24-27 September 2012 in Adelaide, South Australia*, page 384. Engineers Australia.
- DRUSKIN, Vladimir, LIEBERMAN, Chad et ZASLAVSKY, Mikhail, 2010. On adaptive choice of shifts in rational krylov subspace reduction of evolutionary problems. *SIAM Journal on Scientific Computing*, 32(5):2485–2496.
- DU, Qiang, FABER, Vance et GUNZBURGER, Max, 1999. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM review*, 41(4):637–676.
- EFTANG, Jens, PATERA, Anthony et RØNQUIST, Einar, 2010. An "hp" certified reduced basis method for parametrized elliptic partial differential equations. *SIAM Journal on Scientific Computing*, 32(6):3170–3200.
- EGGENER, Charles et POLKOWSKI, Lawrence, 1976. Network models and the impact of modeling assumptions. *Journal-American Water Works Association*, 68(4):189–196.
- ELHAY, Sylvan, PILLER, Olivier, DEUERLEIN, Jochen et SIMPSON, Angus, 2015. A robust, rapidly convergent method that solves the water distribution equations for pressure-dependent models. *Journal of Water Resources Planning and Management*, 142(2):04015047.
- ELHAY, Sylvan et SIMPSON, Angus, 2011. Dealing with zero flows in solving the nonlinear equations for water distribution systems. *Journal of Hydraulic Engineering*, 137(10):1216–1224.
- ELHAY, Sylvan, SIMPSON, Angus, DEUERLEIN, Jochen, ALEXANDER, Bradley et SCHILDERS, Wil, 2014. Reformulated co-tree flows method competitive with the global gradient algorithm for solving water distribution system equations. *Journal of Water Resources Planning and Management*, 140(12):04014040.
- ESTEP, Donald, 2002. *Practical analysis in one variable*. Springer Science & Business Media.
- FISHMAN, George, 2013. *Monte Carlo: Concepts, algorithms, and applications*. Springer Science & Business Media.
- FRITZEN, Felix, HAASDONK, Bernard, RYCKELYNCK, David et SCHÖPS, Sebastian, 2018. An algorithmic comparison of the hyper-reduction and the discrete empirical interpolation method for a nonlinear thermal problem. *Mathematical and computational applications*, 23(1):8.

- FUJIWARA, Okitsugu et GANESHARAJAH, Tharmarajah, 1993. Reliability assessment of water supply systems with storage and distribution networks. *Water Resources Research*, 29(8):2917–2924.
- GALBALLY, David, FIDKOWSKI, Krzysztof, WILLCOX, Karen et GHATTAS, Omar, 2010. Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *International journal for numerical methods in engineering*, 81(12):1581–1608.
- GENTLE, James E, 2012. *Numerical linear algebra for applications in statistics*. Springer Science & Business Media.
- GERBRANDS, Jan, 1981. On the relationships between SVD, KLT and PCA. *Pattern recognition*, 14(1-6):375–381.
- GILBERT, Denis, MORTAZAVI, Iraj, PILLER, Olivier et UNG, Hervé, 2017. Low dimensional modeling of double t-junctions in water distribution networks using kriging interpolation and delaunay triangulation. *Pacific Journal of Mathematics for Industry*, 9(1):2.
- GODSIL, Chris et ROYLE, Gordon, 2013. *Algebraic graph theory*, tome 207. Springer Science & Business Media.
- GOREV, N., KODZHESPIROVA, I. et SIVAKUMAR, P., 2016. Nonunique steady states in water distribution networks with flow control valves. *Journal of Hydraulic Engineering*, 142(9):04016029.
- HAASDONK, Bernard, DIHLMANN, Markus et OHLBERGER, Mario, 2010. A training set and multiple bases generation approach for parametrized model reduction based on adaptive grids in parameter space/simtech, university of stuttgart. 2010 (2010-28). *Preprint*.
- HAASDONK, Bernard et OHLBERGER, Mario, 2008. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(2):277–302.
- HALTON, John, 1964. Radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702.
- HAMBERG, Dan et SHAMIR, Uri, 1988a. Schematic models for distribution systems design. I.: Combination concept. *Journal of Water Resources Planning and Management*, 114(2):129–140.
- HAMBERG, Dan et SHAMIR, Uri, 1988b. Schematic models for distribution systems design, II. Continuum approach. *Journal of Water Resources Planning and Management*, 114(2):141–162.

BIBLIOGRAPHY

- HAMMERSLEY, John, 1960. Monte carlo methods for solving multivariable problems. *Annals of the New York Academy of Sciences*, 86(3):844–874.
- HERRERA, Manuel, TORGO, Luís, IZQUIERDO, Joaquín et PÉREZ-GARCÍA, Rafael, 2010. Predictive models for forecasting hourly urban water demand. *Journal of hydrology*, 387(1):141–150.
- HONG, Y., XING, J. et WANG, J., 1999. A second-order third-moment method for calculating the reliability of fatigue. *International Journal of Pressure Vessels and Piping*, 76(8):567–570.
- HWANG, Hwee, LANSEY, Kevin et JUNG, Donghwi, 2017. Accuracy of first-order second-moment approximation for uncertainty analysis of water distribution systems. *Journal of Water Resources Planning and Management*, 144(2):04017087.
- JAUMOILLÉ, Elodie, PILLER, Olivier et VAN ZYL, Jakobus, 2007. A hydraulic model for water distribution systems incorporating both inertia and leakage. Dans *Water Management Challenges in Global Change (CCWI2007 and SUWM2007 Conference)*, Leicester, GBR, pages 3–5.
- KAPELAN, Zoran, SAVIC, Dragan et WALTERS, Godfrey, 2007. Calibration of water distribution hydraulic models using a bayesian-type procedure. *Journal of Hydraulic Engineering*, 133(8):927–936.
- KRIEGESMANN, Benedikt, 2012. *Probabilistic design of thin-walled fiber composite structures*. Thèse de doctorat, Technische Informationsbibliothek und Universitätsbibliothek Hannover (TIB).
- LALL, Sanjay, KRYSL, Petr et MARSDEN, Jerrold, 2003. Structure-preserving model reduction for mechanical systems. *Physica D: Nonlinear Phenomena*, 184(1-4):304–318.
- LANCZOS, Cornelius, 1997. *Linear differential operators*, tome 18. SIAM.
- LANSEY, KE, EL-SHORBAGY, W, AHMED, I, ARAUJO, J et HAAN, CT, 2001. Calibration assessment and data collection for water distribution networks. *Journal of Hydraulic Engineering*, 127(4):270–279.
- LANSEY, Kevin, 1997. Uncertainty in water distribution network modeling. *Journal of Contemporary Water Research and Education*, 103(1):5.
- LAWSON, Charles, 1972. Generation of a triangular grid with applications to contour plotting. *Jet Propul. Lab. Techn. Memo*, 299:2.
- LE MAÎTRE, Olivier et KNIO, Omar M, 2010. *Spectral methods for uncertainty quantification: With applications to computational fluid dynamics*. Springer Science & Business Media.

- LEE, Der-Tsai et SCHACHTER, Bruce, 1980. Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242.
- LIESEN, Jörg et STRAKOS, Zdenek, 2013. *Krylov subspace methods: Principles and analysis*. Oxford University Press.
- LOHMANN, Boris et EID, Rudy, 2007. Efficient order reduction of parametric and nonlinear models by superposition of locally reduced models. Dans *Methoden und Anwendungen der Regelungstechnik. Erlangen-Münchener Workshops*, pages 27–36.
- LOMBARDI, Edoardo, BERGMANN, Michel, CAMARRI, Simone et IOLLO, Angelo, 2009. *Low-order models: Optimal sampling and linearized control strategies*. Thèse de doctorat, INRIA.
- LU, Zhiming et VESSELINOV, Velimir V, 2015. Analytical sensitivity analysis of transient groundwater flow in a bounded model domain using the adjoint method. *Water Resources Research*, 51(7):5060–5080.
- MADAY, Yvon et STAMM, Benjamin, 2013. Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces. *SIAM Journal on Scientific Computing*, 35(6):A2417–A2441.
- MARCHUK, Guri I, 2013. *Adjoint equations and analysis of complex systems*, tome 295. Springer Science & Business Media.
- MATALA, Anna, 2008. Sample size requirement for monte carlo simulations using latin hypercube sampling. *Helsinki University of Technology, Department of Engineering Physics and Mathematics, Systems Analysis Laboratory*.
- MCKAY, Michael, BECKMAN, Richard et CONOVER, William, 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.
- MOODY, Lewis F, 1944. Friction factors for pipe flow. *Trans. AMSE*, 66:671–684.
- NIKURADSE, Johann, 1933. *Strömungsgesetze in rauhen Röhren: Mit 13 Zahlentafeln*. VDI-Verlag.
- PALUSZCZYSZYN, Daniel, SKWORCOW, Piotr et ULANICKI, Bogumil, 2013. Online simplification of water distribution network models for optimal scheduling. *Journal of Hydroinformatics*, 15(3):652–665.

BIBLIOGRAPHY

- PASHA, MFK et LANSEY, K, 2010. Effect of parameter uncertainty on water quality predictions in distribution systems-case study. *Journal of hydroinformatics*, 12(1):1–21.
- PECCI, Filippo, ABRAHAM, Edo et STOIANOV, Ivan, 2017. Quadratic head loss approximations for optimisation problems in water supply networks. *Journal of Hydroinformatics*, 19(4):493–506.
- PEHERSTORFER, Benjamin, BUTNARU, Daniel, WILLCOX, Karen et BUNGARTZ, Hans-Joachim, 2014. Localized discrete empirical interpolation method. *SIAM Journal on Scientific Computing*, 36(1):A168–A192.
- PERELMAN, Lina, HOUSH, Mashor et OSTFELD, Avi, 2013. Robust optimization for water distribution systems least cost design. *Water Resources Research*, 49(10):6795–6809.
- PERELMAN, Lina, MASLIA, Morris L, OSTFELD, Avi et SAUTNER, Jason, 2008. Using aggregation/skeletonization network models for water quality simulations in epidemiologic studies. *Journal (American Water Works Association)*, 100(6):122–133.
- PILLER, Oliver, 1995. *Modelisation du fonctionnement d'un reseau - Analyse hydraulique et choix des mesures pour l'estimation de parametres*. Thèse de doctorat, L'Université de Bordeaux I.
- PILLER, Olivier, BREMOND, Bernard et POULTON, Matthew, 2003. Least action principles appropriate to pressure driven models of pipe networks. pages 19–30.
- PILLER, Olivier, ELHAY, Sylvan, DEUERLEIN, Jochen et SIMPSON, Angus R, 2016. Local sensitivity of pressure-driven modeling and demand-driven modeling steady-state solutions to variations in parameters. *Journal of Water Resources Planning and Management*, page 04016074.
- PILLER, Olivier, GILBERT, Denis et VAN ZYL, Jakobus, 2010. Dual calibration for coupled flow and transport models of water distribution systems. *Water Distribution Systems Analysis*, pages 722–731.
- PILLER, Olivier et VAN ZYL, Jakobus, 2007. A unified framework for pressure-driven network analysis. *Proc. of Computer and Control in Water Industry, Water Management Challenges in Global Changes, Supplementary Proceedings*, pages 25–30.
- PILLER, Olivier et VAN ZYL, Jakobus, 2009. Pressure-driven analysis of network sections supplied via high-lying nodes. *Proceedings of the Computing and Control in the Water Industry, Integrating water systems*, Taylor & Francis Group, London.

- PRANDTL, L, 1961. Neuere Ergebnisse der Turbulenzforschung, Z. VDI 77 (1933) 105-114. Nachdruck: Ges. Abh., pages. 819-845.
- PRUD'HOMME, Christophe, ROVAS, Dimitrios, VEROY, Karen, MACHIELS, Luc, MADAY, Yvon, PATERA, Anthony et TURINICI, Gabriel, 2002. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80.
- RAZAVI, Saman et GUPTA, Hoshin, 2015. What do we mean by sensitivity analysis? the need for comprehensive characterization of “global” sensitivity in earth and environmental systems models. *Water Resources Research*, 51(5):3070–3092.
- RESIWATER, 2019. <http://www.resiwater.eu/>. Online; accessed January 22, 2019.
- SAVIC, Dragan, KAPELAN, Zoran et JONKERGOUW, Philip, 2009. Quo vadis water distribution model calibration? *Urban Water Journal*, 6(1):3–22.
- SCHWALLER, Julie et VAN ZYL, Jakobus, 2014. Implications of the known pressure-response of individual leaks for whole distribution systems. *Procedia Engineering*, 70:1513–1517.
- SHAMOS, Michael Ian et HOEY, Dan, 1975. Closest-point problems. Dans *Foundations of Computer Science, 1975., 16th Annual Symposium on*, pages 151–162. IEEE.
- SIROVICH, Lawrence, 1987. Turbulence and the dynamics of coherent structures. I. Coherent structures. *Quarterly of applied mathematics*, 45(3):561–571.
- SMITH, Ralph C, 2013. *Uncertainty quantification: Theory, implementation, and applications*, tome 12. SIAM.
- STEWART, GW, 2000. The decompositional approach to matrix computation. *Computing in Science & Engineering*, 2(1):50–59.
- TANYIMBOH, Tiku T et TEMPLEMAN, Andrew B, 2004. A new nodal out-flow function for water distribution networks. Dans *Proceedings of the 4th International Conf. on Eng. Computational Technology*. Civil-Comp Press, Stirling, UK, page 12.
- TODINI, Ezio et PILATI, S., 1988. A gradient method for the solution of looped pipe networks, computer applications in water supply.
- TRÜEB, Ernst, 1961. Druckverlustberechnung mit hilfe der formel von prandtlbrook. *Schweizerische Zeitschrift für Vermessung, Kulturtechnik und Photogrammetrie*, 59(5):142–170.

BIBLIOGRAPHY

- TUCCIARELLI, Tullio, CRIMINISI, Antonio et TERMINI, Donatella, 1999. Leak analysis in pipeline systems by means of optimal valve regulation. *Journal of Hydraulic Engineering*, 125(3):277–285.
- ULANICKI, Bogumil, ZEHNPFUND, Andreas et MARTINEZ, Fernando, 1996. Simplification of water distribution network models. Dans *Proc., 2nd Int. Conf. on Hydroinformatics*, pages 493–500. Balkema Rotterdam, Netherlands.
- VAN BLOEMEN WAANDERS, B., HAMMOND, G., SHADID, J., COLLIS, S. et MURRAY, R., 2005. A comparison of Navier–Stokes and network models to predict chemical transport in municipal water distribution systems. Dans *Impacts of Global Climate Change*, pages 1–10.
- VAN ZYL, Jakobus, CASSA, Amanda, SAVIC, Dragan, KAPELAN, Zoran et BUTLER, David, 2011. Linking the power and FAVAD equations for modelling the effect of pressure on leakage. *Eleventh International Conference on Computing and Control for Water Industry (CCWI), University of Exeter, United Kingdom*.
- VAN ZYL, JE et CLAYTON, CRI, 2007. The effect of pressure on leakage in water distribution systems. Dans *Proceedings of the Institution of Civil Engineers-Water Management*, tome 160, pages 109–114. Thomas Telford Ltd.
- WAGNER, Janet, SHAMIR, Uri et MARKS, David, 1988. Water distribution reliability: Simulation methods. *Journal of Water Resources Planning and Management*, 114(3):276–294.
- WALSKI, Thomas, CHASE, Donald et SAVIC, Dragan, 2001. Water distribution modeling.
- WALSKI, Thomas, CHASE, Donald, SAVIC, Dragan, GRAYMAN, Walter, BECKWITH, Stephen et KOELLE, Edmundo, 2003. Advanced water distribution modeling and management.
- WIENER, Norbert, 1938. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936.
- WONG, Tien-Tsin, LUK, Wai-Shing et HENG, Pheng-Ann, 1997. Sampling with Hammersley and Halton points. *Journal of graphics tools*, 2(2):9–24.
- XIU, Dongbin, 2010. *Numerical methods for stochastic computations: A spectral method approach*. Princeton university press.

- XIU, Dongbin et KARNIADAKIS, George E., 2002. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644.
- ZHENG, Feifei, SIMPSON, Angus, ZECCHIN, Aaron et DEUERLEIN, Jochen W., 2013. A graph decomposition-based approach for water distribution network optimization. *Water Resources Research*, 49(4):2093–2109.