Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

9-1-2022

Uncertainty Assessment in Hydrological Modelling Using Polynomial Chaos Expansion

Rahul Narula
University of Windsor

Uncertainty Assessment in Hydrological Modelling Using Polynomial Chaos Expansion

By

Rahul Narula

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

2022

Uncertainty Assessment in Hydrological Modeling Using Polynomial Chaos Expansion

By

Rahul Narula

APPROVED BY:

_____

D. Ting

Department of Mechanical, Automotive and Materials Engineering

_____

R. Ruparathna

Department of Civil and Environmental Engineering

_____

T. Bolisetti, Advisor

Department of Civil and Environmental Engineering

July 28, 2022

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Hydrological models are powerful tools that simulate the natural hydrological cycle and natural processes like surface runoff, groundwater flow, and evapotranspiration, which are needed to be understood and quantified for a wide range of applications like water resource management, climate change impact assessment, flood studies and water quality assessment. Errors and uncertainties are bound to creep into the modelling process because of various reasons like incomplete understanding and representation of the natural phenomena, model errors, approximation errors, and parameter uncertainties. This study aims to efficiently quantify the parameter uncertainty by making use of surrogate modeling techniques. There is an inherent trade-off between model complexity and the parameter uncertainty i.e., parameter uncertainty usually increases if complex hydrological model with high model accuracy is employed, but the required computational effort would increase significantly. Considering this tradeoff, one lumped, conceptual model (HYMOD) and one semi-distributed, process-based model (SWAT) for a small (179 km^2) and mid-sized (2318 km^2) watershed are considered for this study. Hydrological modelling processes are frequently hampered by computationally costly simulations. Consequently, modellers can opt a surrogate model which is a machine learning model that approximates another model but requires less computational effort. This thesis uses Polynomial Chaos Expansion (PCE) method of surrogacy which represents an accurate approximation of the model as the sum of carefully selected polynomials, each separately weighted. The aim of this study is to use Polynomial Chaos Expansion to 1) Improve the calibration and uncertainty assessment procedure for HYMOD, and 2) Quantify parameter uncertainties in SWAT

Keywords: Hydrological models, SWAT, parameter uncertainty, surrogate model, Polynomial Chaos Expansion (PCE)

DEDICATION

To my beloved parents, brother and to God.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIES   Affine Invariant Ensemble Sampler

ANN   Artificial Neural Network

CI   Confidence Interval

CREAMS   Chemicals, Runoff, and Erosion from Agricultural Management Systems

CRR   Conceptual Rainfall-Runoff

DEM   Digital Elevation Model

DHI   Danish Hydraulic Institute

DREAM   Differential Evolution Adaptive Metropolis

EPIC   Environmental Policy Integrated Climate

ESWAT   Enhanced Soil and Water Assessment Tool

FAST   Fourier Amplitude Sensitivity Test

FEFLOW   Finite element sub-surface flow and transport simulation system

GLEAMS   Groundwater Loading Effects of Agricultural Management Systems

GLUE   Generalized Likelihood Uncertainty Estimation

GSA   Global Sensitivity Analysis

HRU   hydrologic Response Unit

HYMOD   Hydrological Model

IHDM   Institute of Hydrology Distributed Model

LAR   Least Angle Regression

LHS   Latin Hypercube Sampling

LOO   Leave-One-Out

LULC	Land Use Land Cover
MAP	Maximum a Posteriori
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MOAT	Morris One-at-a-time Screening
MODFLOW	Modular finite difference ground-water flow model
MOPEX	Model Parameter Estimation Experiment
NED	National Elevation Dataset
NLCD	National Land Cover Database
NOAA	National Oceanic and Atmospheric Administration
NRCS	Natural Resources Conservation Service
NSE	Nash-Sutcliffe Efficiency
OLS	Ordinary Least Squares
PBIAS	Percentage Bias
PCE	Polynomial Chaos Expansion
RSM	Response Surface Modelling
RSR	RMSE-observations Standard deviation Ratio
SA	Sensitivity Analysis
SD	Standard Deviation
SVM	Support Vector Machines
SWAT	Soil and Water Assessment Tool
SWAT-CUP	Soil and Water Assessment Tool Calibration and Uncertainty Programs
SWAT-G	Gridded Soil and Water Assessment Tool

SWRRB	Simulator for Water Resources in Rural Basins
UQ	Uncertainty Quantification
USDA	United States Department of Agriculture
WRF	Weather Research and Forecasting

CHAPTER 1: INTRODUCTION

1.1 Background

Extreme events are occurring more frequently as a result of climate change, which calls for increased readiness to reduce risk and prevent catastrophic events (Xu, 1999). In the field of hydrology, designers, researchers, and decision-makers have been working to create more reliable daily (or sub-daily) probabilistic forecasting methods. Hydrological modelling is crucial for managing flood risks, allocating water resources, and operating and planning water infrastructure (Fowler et al., 2007). There are several different kinds of hydrological models for predicting flow. There are two commonly employed types of hydrological models: physically based models and conceptual models. Based on spatial variability, physically based models can be categorised into three groups: lumped, semi-distributed, and fully distributed models (Devia et al., 2015). Given that the catchment is viewed as a single entity and that there is no spatial variability within the watershed, the lumped model is also known as a conceptual model. The watershed is divided into sub-basins by the semi-distributed model, and each sub-basin has its own set of parameter values. The catchment is divided into grids in the fully distributed model, which is the most complex because each grid has unique properties and processes. Although fully distributed models appear to be the best, as the model gets more complex, more input data is needed. The model becomes more uncertain as there are more parameters added (Moges et al.,2021). On the other hand, physical laws are taken into account by conceptual models, but they are greatly simplified. A conceptual model is an accurate depiction of a hydrologic system that includes the modeler's knowledge of the pertinent physical, chemical, and hydrologic conditions. Conceptual models of rainfall-to-runoff generation are used to predict the size of streams by simulating internal variables like soil moisture using a variety of response functions (Jajarmizadeh et al., 2012).

1.1.1 Physically Based Models

A system of hydrological and/or hydraulic processes that simulates the catchment's reaction to precipitation events is known as a physically-based model. There are some simplifications to the actual hydrological processes in physically based models. The type of physically-based model will determine how simplified the model is (Beven, 1989). The most simplified models are lumped, and fully distributed models contain the most information about the actual catchment response. Physically based models must first be calibrated in order to offer valid and trustworthy predictions (Fatichi et al., 2016). The number of parameters that must be calibrated can vary from less than ten to hundreds, depending on the type of model.

To cut down on computational time and requirements, sensitivity analysis can be performed first. Non-sensitive parameters discovered through the sensitivity analysis can be disregarded during the calibration process. Local and global sensitivity analyses are two of the various types of sensitivity analysis. When determining local sensitivity, all parameters are fixed, and only one parameter is changed at a time near the value of interest to see how the output changes as a result (White and Chaubey, 2005). Local sensitivity analysis is simple, quick, which is a trade-off for its low accuracy. When computational resources are scarce, they are useful (Karkee and Steward, 2010; López-Cruz et al., 2012). The parameters that have the greatest impact on the model output are identified by global sensitivity analysis, which considers output change in relation to changes in all parameters throughout the entire parameter space (Dos Santos and Lu, 2015; López-Cruz et al., 2012; Scire et al., 2001). Global sensitivity is becoming more widespread due to the quick advancement of high-performance computing technology and the fact that it yields more accurate results than local sensitivity.

Previous research has looked into various methods for enhancing physically based models. The first option is to develop new hydrological models or to enhance the simulation of hydrological and hydraulic processes within the model (Ehteram et al., 2018; Farzin et al., 2018). The second option is to improve the calibration algorithms in order to get more precise parameter values that can replicate the observed outflow (Singh et al., 2013; Yang et al., 2008). Investigating the trade-off between accuracy and simplicity is the third option (Herman et al., 2013; Vos et al., 2010). Finding hydrological models with a structure as straightforward as lumped models but capable of producing outcomes as precise as fully distributed models when the data is available is the aim. Because there is a lack of data and fully distributed models take a long time to set up and calibrate, lumped and semi-distributed models are currently used more frequently than fully distributed models. With sufficient data, both lumped and semi-distributed models could deliver respectable predictions. There isn't a single model that works well across all watersheds. Hydrological model improvement is still a very active area of research. This thesis employs the use of SWAT hydrological model which is a semi-distributed, physically-based model.

1.1.2 Conceptual Model

The conceptual models are developed based on the development of relationships between input and output, without the complete explicit knowledge of the physical processes. Researchers and decision-makers can now gather more frequent and accurate data thanks to the advancement of technology (Montáns et al., 2019). Conceptual models have been widely researched and used in hydrological modelling as data availability and computational power have increased (Jothiprakash and Kote, 2011; Solomatine and Ostfeld, 2008; Taormina and Chau, 2015). The quantity and quality of the data have a significant impact on conceptual models. A specific percentage of the data must be used to train a conceptual model before the remaining data is used for validation and

testing when forecasting with the model. About 70% is typically used for training, with the remaining 30% being split equally between validation and testing (Wu and Chau, 2010; Zeroual et al., 2016). Depending on the amount of available data, the percentage may change. Conceptual models must have input data to function. The input variables that will be used in the model must be carefully chosen based on the conceptual model and the output that is required. This thesis employs the use of HYMOD which is a lumped, conceptual hydrological model.

1.1.3 Uncertainty Analysis

There are various types of uncertainties in all hydrological models, whether they are physically based, data-driven, or hybrid. First, a significant source of uncertainty is the model parameters (Moges et al., 2021). The parameter values may not accurately reflect the real values even after the model calibration is completed (Beven and Binley, 1992). The input data or observations are a further source of uncertainty. A further source of uncertainty is the model's structure. Different mathematical representations of hydrological relationships are used by various hydrological models and there may be various modelling approaches even for the same process in the same model (Talebizadeh et al., 2010; Tolessa et al., 2015). The analysis and measurement of parameter uncertainty have been the subject of numerous studies. Only the most sensitive parameters discovered through sensitivity analysis are typically used for the analysis of parameter uncertainty in order to minimise the computational requirement. Generalized likelihood uncertainty estimation (GLUE), Monte Carlo (MC), and bootstrap sampling are the commonly employed methods for assessing parameter uncertainty and producing probabilistic predictions (Li et al., 2009; Wu and Liu, 2012; Zhang et al., 2016). Finding algorithms for uncertainty quantification that can cut down on the amount of computational time and resources needed while maintaining the accuracy of the probabilistic prediction is a top priority.

1.1.4 Surrogate Models

Studies on water resources and decision-making processes heavily rely on hydrological models, which simulate abstract representations of physically based systems using mathematical concepts and language. Hydrological models can be used to solve a variety of issues, such as prediction, optimization, sensitivity analysis, and uncertainty analysis. In order to improve simulation models' fidelity to the real-world system, there are additional issues like model calibration and model parameter sensitivity analysis. In the context of modelling, fidelity refers to the level of realism in a simulation model. According to Keating et al. (2010), Mugunthan et al. (2005), and Zhang et al. (2009), modern simulation models are typically computationally intensive because they accurately represent detailed scientific knowledge about real-world systems. These simulation models must be run thousands of times for many model-based engineering analyses, which makes them prohibitively expensive to run. A second level of abstraction called "surrogate modelling" is concerned with creating and using "surrogates" of the "original" simulation models that are easier to run.

There are many different types of substitute models that can be effectively used in place of simulation models. Response surface modelling and lower-fidelity modelling are two broad families that fall under the broad category of surrogate modelling. Data-driven function approximation techniques are used by response surface surrogates to empirically approximation the model response. The objective of a surrogate modelling technique (response surface surrogates or lower-fidelity physically based surrogates) is to approximatively represent the response(s) of an original simulation model, which is typically computationally intensive, for different values of explanatory variables of interest. The terms "response surface" and "response landscape" refer to the surface that represents the model response with respect to the variables of interest (which is

typically a nonlinear hyper-plane). Different response surfaces must be fitted to each model response of interest for the majority of response surface surrogate modelling techniques (or each function aggregating multiple model responses). One technique that makes an exception and can fit multiple model responses is neural network technology. In contrast, one lower-fidelity surrogate model can typically approximate several important model responses because lower-fidelity surrogates still have some physically based characteristics of the original model. Making better use of the available, typically constrained, computational budget is the primary driver behind the development of surrogate modelling strategies.

1.1.5 Polynomial Chaos Expansion (PCE)

Recently, Polynomial Chaos Expansion (PCE) demonstrated the potential to quantify parameter uncertainties in an effective and efficient way (Fan et al., 2016, 2014; Wang et al., 2015). PCE has been employed for uncertainty quantification (UQ) in a wide range of fields (e.g., in solid mechanics, fluid flows, thermal sciences, etc.). Since the current PCE method depends on observations to quantify the propagation of parameter uncertainties within a model, it is unable to make hydrological forecasts under uncertainty. Furthermore, since most hydrological models' parameters interact with one another, independent model parameters are required for PCE to be able to quantify parameter uncertainties.

The probabilistic technique called PCE projects the model's output onto the basis of orthogonal stochastic polynomials in the input data. The stochastic projection offers a compact and practical representation of the input-dependent output variability of the model. In other words, a surrogate model approximates a computationally expensive model. It mimics the behaviour of the original model and replicates the underlying physics. PCE can be used to efficiently and accurately perform parameterization, sensitivity analysis, and uncertainty quantification. A PCE represents the model

as a sum of carefully chosen polynomials each individually weighted to give an accurate approximation. Advantages of PCE lie in the fact that it is a non-intrusive method in the sense that it is based on the runs of the computational model, similar to Monte Carlo simulation. Additionally, PCE is suited to a high-performance computation. PCE also has disadvantages like its need for a rigorous validation. The PCE is highly efficient compared to Monte Carlo simulations by at least 2-3 orders of magnitude (Sudret, 2007). This thesis uses the meta-modeling approach of surrogate modelling using polynomial chaos expansions (PCE) to approximate the model output.

In the past, a PCE coupled with the Least Angle Regression method has not been linked with either HYMOD or SWAT. Applications of a surrogate model like PCE with a complex semi-distributed model has only been explored in a handful of studies. These are the research needs that this thesis aims to address.

1.2 Objectives of the Study

The objective of this study is to examine Polynomial Chaos Expansion surrogate model's ability in capturing the behaviour of the conceptual hydrological model HYMOD and the physically-based SWAT model. This thesis aims to demonstrate the applicability of PCE meta-models to quantify the parameter uncertainty in SWAT models. Additionally, investigating the applicability of PCE to create a generalized uncertainty quantification (UQ) framework for HYMOD, which can be used to optimize any UQ algorithm which is demonstrated by Markov Chain Monte Carlo (MCMC) in this study.

1.3 Structure of the Thesis

This thesis is organized into 5 chapters. Chapter 1 introduces the background information about hydrological models, both conceptual and physical, uncertainty assessment followed by surrogate models and polynomial chaos expansion (PCE). Chapter 2 presents the literature review on

hydrological model, both HYMOD and SWAT models, uncertainty assessment, surrogate models and finally PCE. Chapter 3 presents the data on study area and input data necessary for the current study followed by methodology for PCE, hydrological model and uncertainty analysis. Results and discussions on HYMOD, SWAT model and their surrogate modeling and uncertainty assessment and sensitivity analysis using PCE is presented in Chapter 4. Subsequent comments and conclusions along with suggestions for future work are presented in the final Chapter 5.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews the previous literature to provide an insight into hydrological models, uncertainty assessment, surrogate model and polynomial chaos expansion (PCE).

2.1 Hydrological Model

A hydrologic model is a simplified representation of a real-world system (e.g., surface water, soil water, or groundwater) that can be used to help understand, predict, and manage water resources. They conceptualise and aggregate the complex, spatially distributed, and highly interconnected water, energy, and vegetation dynamics in a watershed using relatively basic mathematical equations (Vrugt et al., 2005). Hydrological models serve as tools to estimate the quantities of water movement in different processes in the natural hydrological cycle. Estimating the availability of water resources is a direct example of a technical challenge brought about by a lack of historical data as well as issues related to the measurement of river discharge. To this extent, the models use a combination of various laws of physics like law of conservation of mass and law of conservation of energy and continuity equation.

Natural hydrological processes are greatly heterogenous and have a high degree of spatial and temporal variance. To simulate the heterogeneity, the models can adopt a huge number of small-sized spatial units and can consider infinitesimally small time steps but doing so will greatly heighten the computational power required for the simulation. Hence it is essential to choose a hydrological model after giving due consideration to the available computational power and the degree of heterogeneity in the model simulation. For this purpose, a wide variety of models are available, for instance, lumped (HYMOD), semi-distributed (SWAT) and fully distributed (WRF) hydrological models exist based on the spatial discretization. The models can also be classified into empirical (SVM, ANN), conceptual (HYMOD, NAM) and physical (SWAT, MIKE-SHE)

based on the complexity of natural hydrological process descriptions. According to Maier and Dandy (1997), Daniell (1991) was the first person to introduce ANNs to the field of hydrological modelling. Daniell used an ANN to anticipate monthly water intake and estimate the occurrence of floods at the time. ANNs have been applied to many different aspects of water resource management, such as time-series prediction for rainfall data (French et al., 1992), rainfall-runoff mechanisms (Minns and Hall, 1996; Shamseldin, 1997), and representing soil and water processes, such as soil moisture (Altenford, 1992).

On the other hand, physically based models don't use parameterized equations to reflect these processes; rather, they use mathematical-physics equations that are based on real-world motion and quantum physics, hydrodynamics, and heat and mass transfer (Beven, 2001). For instance, the Institute of Hydrology Distributed Model (IHDM) (Beven et al., 1987) and MIKE SHE (Refsgaard and Storm, 1995) use physically based laws to solve different components within the model. These physically based laws include the St. Venant equations of channel flow and the 3D finite difference groundwater flow governed by the Darcian law, respectively. Distributed models have the capability to take into consideration, the spatial variation of each and every parameter and variable that exists within the catchment. Distributed models are typically physically based and discretize the basin into a network of grid cells or a large number of elements. These models then solve the parametric mathematical equation for each of the grid cells independently (Beven, 2002). Within the MIKE SHE model, the finite difference/grid mesh method is used to represent the surface and sub-surface flow equations within a regular spatial grid. This method can be found within the framework of the MIKE SHE model (DHI-WE, 2005 and Beven, 2002).

2.1.1 SWAT Model

For modelling hydrological, sediment, nutrient, bacterial, and tile drainage processes, the USDA Agricultural Research Service created the semi-distributed, process-based SWAT model. Instead of simulating a single storm, SWAT simulates hydrological processes over time. It is frequently employed to comprehend the hydrology of various areas of interest. SWAT model is run on sub-daily, daily, monthly and yearly time scales. Extension of the SWRRB model is SWAT. Later additions to this advanced model allowed it to simulate chemicals, runoff, erosion, and groundwater using CREAMS, GLEAMS, and EPIC. Modules for reservoir and pond storage were added to test their effects on water flow. Regression equations from USGS were imported along with nutrient transport and loading components from SWMM. Before the SWAT2012 hydrological model was used in this study, there were a number of uses for earlier, more basic SWAT models, some of which are listed below: For sub-daily water quality studies, ESWAT had an automatic calibration routine (Griensven and Bauwens, 2001); SWAT-G had enhanced transpiration mechanics (Eckhardt et al., 2002); and SWIM included crucial hydrological processes at both smaller and larger (> 10,000 sq. km) basin levels (Krysanova et al., 2005). The improved SWAT2012 used in this study was enhanced from the prior basic versions of the model.

The most sensitive parameters for various water budget components, pollutants, and nutrients were reviewed along with calibration and validation techniques with a thorough explanation of the steps that should be taken for calibration and uncertainty analysis by Arnold et al. (2012). In most cases, SWAT is utilised in the analysis of large river basins, covering an area of close to thousands of square kilometres, provided that there is a high level of congruence between the simulated and actual data (Jayakrishnan et al., 2005). On the other hand, the SWAT model is also capable of

being validated and applied at the scale of smaller watersheds (Arnold et al., 1996; Arnold et al., 1999; Arnold and Williams, 1987)

2.1.2 HYMOD

A lumped model is that hydrological model where the entire watershed is represented using a single spatial unit. The runoff is calculated at the sub-watershed scale in a conceptual rainfall-runoff (CRR) model on the basis of governing hydrologic phenomena without taking into account the subtle processes but providing adequately optimum hydrologic response of the watershed system. The conceptual model HYMOD (Boyle 2001) has been used in the current study's catchment hydrologic simulation. The model is widely used for a variety of applications and across a wide range of hydrologic regimes (Chilkoti et al. 2017; Sikorska et al. 2014; Formetta et al. 2011; Gharari et al. 2013; Montanari 2005; Wagener et al. 2001). HYMOD is a strong, continuous-process simulation-capable runoff generation model that is relatively easy to use. The model was selected primarily due to its broad acceptance, small number of parameters, and adequate process representation that takes into account both slow and fast responses. A nonlinear soil moisture component is included in the model and is computed using the Moore loss model (Moore, 1985). The five model parameters are the watershed storage capacity Cmax, the watershed's imperviousness percentage α, the recession coefficients for fast and slow flows/ base flow and Rq and Rs, respectively, and the degree of spatial variability of soil moisture capacity within a watershed Bexp. Boyle (2001) is a source that can be used to get more information about the model configuration.

2.2 Uncertainty Assessment

Uncertainty can be defined as the state of limited knowledge/data where it is not possible to perfectly describe an existing state or future outcomes. Uncertainty can be either

aleatoric/statistical or epistemic/systematic. Uncertainty decomposition refers to the breakdown of total uncertainty into the contributions of induvial sources. In this study, the total parameter uncertainty is broken down into contribution of induvial parameter expressed in Sobol indices, using PCE. Input uncertainty, uncertainty regarding the climate model, uncertainty regarding the simplification of the underlying physical processes, uncertainty regarding the output, uncertainty regarding the hydrological model, and uncertainty regarding the parameters all contribute to the overall uncertainty in the process of modelling climate change (Beven, 2016; Athira and Sudheer, 2015). Because of the sources, there is always some degree of uncertainty in the modelling process; as a result, identifying those sources and then attempting to quantify them is one of the most difficult challenges that hydrologic modellers must overcome (Beven and Freer, 2001). Several studies on the effects of climate change have been carried out, with the goal of identifying and quantifying the various factors that contribute to the uncertainty in the results. These factors include the choice of climate model (Arnell, 2011; Chen et al., 2011; Dobler et al., 2012; Karlsson et al., 2016; Her et al., 2019); the hydrological model (Arnell, 2011; Karlsson et al., 2016; Wang et al., 2020; Tarek et al., 2021), bias correction methods (Aryal et al., 2019; Wang et al., 2020), land-use scenarios (Karlsson et al., 2016), hydrological parameters (Chilkoti, 2019), RCP scenarios (Wilby and Harris, 2006; Vetter et al., 2017; Chegwidden et al., 2019; Stojkovic et al., 2020) among many other factors like internal variability (Deser at al., 2012).

Defining uncertainty in hydrologic models is intensively crucial for a variety of water resources applications, including environmental protection, drought management, water utility operations, reservoir operation, and sustainable water resource management (Fan et al., 2012). Series of techniques for assessing the uncertainty in hydrologic projections have been developed previously, the most notable ones are: generalized likelihood uncertainty estimation (GLUE) (Beven and

Binley, 1992), Monte Carlo simulation (Ang and Tang, 1984), Bayesian recursive estimation (Thiemann et al., 2001), Metropolis algorithms – Shuffled complex (Vrugt et al., 2003) and differential evolution adaptive metropolis (DREAM) (Vrugt et al., 2009), Dual state estimation using Ensemble Kalman Filter (Moradkhani et al., 2005), Simultaneous Optimization and Data Assimilation (Vrugt et al., 2005). MC is the most extensively utilised of these because it offers the most degree of flexibility for uncertainty propagation. Stable estimates of the model output distribution can be easily obtained with a sufficiently large number of parameter samples (Mishra, 2009). However, the primary disadvantage of these methodologies like the MC simulation and GLUE is that it requires a large number of model simulations to obtain appropriate estimates of the output statistics, which is a difficulty for models with a high computing demand (Y. Liu and Gupta, 2007; Tran and Kim, 2019).

2.3 Surrogate Model

Highly heterogeneous physical characteristics and processes govern how groundwater flows. Many groundwater management issues require complex, fully distributed models that can accommodate fields for the hydraulic properties and boundary conditions that vary in time and space in order to capture such heterogeneity. Fully distributed groundwater models typically solved using a finite difference approximation, like that implemented in MODFLOW (Harbaugh, 2005) or a finite element method, like that used by FEFLOW (Diersch, 2005), have a tendency to include more physical processes, increase numerical resolution, and expand the model domain. However, more complicated conceptual models require more parameters and require longer model runs.

Long runtimes prevent the use of models in many-run applications like inverse modelling, uncertainty analysis, sensitivity analysis, and integrated modelling (where groundwater flow

models are coupled with models of different processes). Additionally, slow runtimes prevent models from being used in real-time, which is essential for applications like decision support. The "curse of dimensionality" is also manifested in uncertainty analysis, sensitivity analysis, or calibration where the number of samples needed to cover the parameter space grows exponentially with the number of model parameters. In order to reduce runtime and make many model runs computationally tractable, numerical resolution must be decreased or physical processes must be disregarded when a model's runtime increases. Complex models may be accelerated using surrogate models without compromising precision or level of detail.

Surrogate models are less expensive computationally and are used to approximate the main features of a complex model. They are also known as metamodels (Blanning, 1975), reduced models (Willcox and Peraire, 2002), model emulators (O'Hagan, 2006), proxy models (Bieker et al., 2007), lower fidelity models (Robinson et al., 2008), and model emulators (O'Hagan, 2006). In addition to increasing computational efficiency, there are other benefits to using a surrogate model (Razavi et al., 2012a). When numerical instability is reduced, calibration and uncertainty analysis are made easier (Doherty and Christensen, 2011). A complex model's irrelevant parameters and insensitive outputs may become apparent during the emulator building process (Young and Ratto, 2011). In order to analyse model simplification and the ways that models simplify reality, substitutes may be used as didactic tools (Watson et al., 2013). They can also be used to reduce the ill conditioning of a conjugate gradient optimizer by using eigenvector approximations (Vuik et al., 1999) or to smooth an objective function surface, enabling the use of gradient-based, nonlinear programming methods for optimization problems (Hemker et al., 2008; Kavetski and Kuczera, 2007). By simulating and calibrating different model structures simultaneously (Matott and Rabideau, 2008) or by including data and physical processes at various

scales (Weinan and Engquist, 2003), the increase in computational efficiency makes it possible to investigate the structural model uncertainty. Additionally, in interactive decision support environments, substitutes with sufficiently short runtimes have been used (Roach and Tidwell, 2009). In order to increase accuracy, surrogates have been used in "complementary" modelling, which involves fitting a simple model to the residual of a complex model (Demissie et al., 2009; Xu et al., 2012).

Utilizing a parallel computing approach is one way for compensating for the increased time necessary to perform uncertainty quantification. However, the disadvantage is that the requirements for computer hardware configuration result in high costs (Cintra and de Campos Velho, 2018). Possible solutions to the problem include: 1) utilising a more efficient sampling approach, such as the Latin hypercube method; 2) replacing the computationally demanding model with a faster surrogate model (Hu et al., 2019). The primary goal of a surrogate model is to produce results that are almost identical to those of the original model, to quantify uncertainty more rapidly, and to evaluate the model's sensitivity effortlessly (Blatman and Sudret, 2010; Xiu and Karniadakis, 2002).

To bridge this issue and combine both solutions, surrogate modelling based on the polynomial chaos expansion (PCE) theory (Wiener, 1938; Xiu and Karniadakis, 2002) has received much attention in the literature as a highly efficient method to quantify uncertainty.

2.4 Polynomial Chaos Expansion (PCE)

In the recent past, studies focused on the propagation of parameter uncertainty into prediction uncertainty within a hydrological model demonstrated that the use of PCE can significantly reduce the computing costs associated with simulating a large number of ensemble runs (Fan et al., 2015; Ghaith and Li, 2020; Hu et al., 2019; Tran and Kim, 2019; Wang et al., 2015, 2017).

PCE, which originated with Wiener's homogeneous chaos theory (Wiener, 1938) and was generalised later on utilising Wiener-Askey polynomial chaos (Xiu and Karniadakis, 2002), provides an orthogonal polynomial-based spectral expression for any statistical distribution. It can be used to expand any random process of the second order (a process having finite variance across all time), which is applicable to most physical processes (Xiu and Karniadakis, 2002). PCE approximates the dynamic model by allowing the model output to have a fully probabilistic distribution. By sampling the random variable in PCE, statistical information (such as the mean, variance, covariance, and skewness) can be determined. By the virtue of being a surrogate model, PCE offers an efficient way of sampling without running the original model, which is appealing particularly to computationally expensive models (Hu et al., 2019).

The efficiency of the surrogate model is dependent on the estimation of PCE coefficients from the response of an original model at design points in the input space. The two main non-intrusive methods used for the estimation of PCE coefficients are: regression and projection (Sudret, 2008). The projection approach can be formulated as a numerical integration problem using quadrature or sparse-grid methods, and the regression method uses least square regression to minimise the mean square error between the surrogate and original model outputs (Sudret, 2008). While it is worthwhile to investigate the utility of the generalised PCE approach for assessing the uncertainty associated with hydrological predictions caused by uncertain parameters, it has certain drawbacks which need to be addressed. Given the enormous number of model evaluations required, both techniques are ineffective at optimising a large number of PCE coefficients. The number of PCE coefficients exponentially increase with the increase in the number of uncertain inputs and the polynomial order (Blatman and Sudret, 2010; B. Liu et al., 2014; Razavi et al., 2012). This

"generalized" PCE necessitates an absurdly huge number of model assessments, significantly limiting engineering applications (Sargsyan et al., 2014).

To address this issue, techniques for reducing the PCE coefficients have been developed such as least angle regression (LAR) (Blatman and Sudret, 2008), sparse collocation (Shi et al., 1996), and Bayesian compressive sensing (Sargsyan et al., 2014). Among them, LAR has gained recent attention due to its demonstrated ability to achieve large computational advantages over original PCE (Xiao et al., 2021; Xie et al., 2017). In general, the goal of LAR is to estimate just the coefficients for the significant PCE base components and to set the coefficients for the nonessential elements to zero (Blatman and Sudret, 2008). Hence, LAR permits the fitting of high-order polynomials to nonlinear complex models without significantly raising the computing cost of the surrogate model development. Even though the efficacy of LAR has been established, very few studies have focused on the use of LAR-PCE as a tool to quantify uncertainty, while reliably capturing the input-output relationship of a hydrological model.

2.5 Summary

This chapter enumerates various literary sources on hydrological models, uncertainty assessment, surrogate models, polynomial chaos expansion. SWAT 2012, a popular semi-distributed model and HYMOD, a popular conceptualized model is chosen in this study to emulate the watershed characteristics and obtain the streamflow. This chapter also provides an insight into uncertainty assessment and various techniques, especially the recently popular surrogate modelling techniques. It highlights the importance of data-driven models in today's world of ever-increasing data and complexity of models. PCE serves as an excellent example of surrogate models as it requires minimal training data and computational power to simulate numerical models. In comparison with the other PCE meta-models available, LAR-PCE or Sparse PCE has been

recognized to simulate numerical models based on natural processes most efficiently. In the subsequent chapters, application and advantages of PCE at various stages of hydrological model development would be discussed.

CHAPTER 3: DATA AND METHODOLOGY

Chapter 3 provides a detailed explanation of the data and methodology used in the current study. This chapter summarizes the study areas, geospatial data and meteorological and hydrometric input data required for the hydrological models followed by the methodology for Polynomial Chaos Expansion (PCE), hydrological modeling and Uncertainty Analysis.

3.1 Study Area

Two study areas are chosen to find out the effects of varying climatic, land use and area profiles on their SWAT calibration as well as uncertainty quantification results. These watersheds of varying sizes, namely Chehalis River watershed (large watershed, greater than 2000 sq. km) and Indian Creek watershed (small watershed, smaller than 200 sq. km) with different climatic profiles, as depicted in Figure 3-1, were chosen for this thesis.

a) Indian Creek watershed
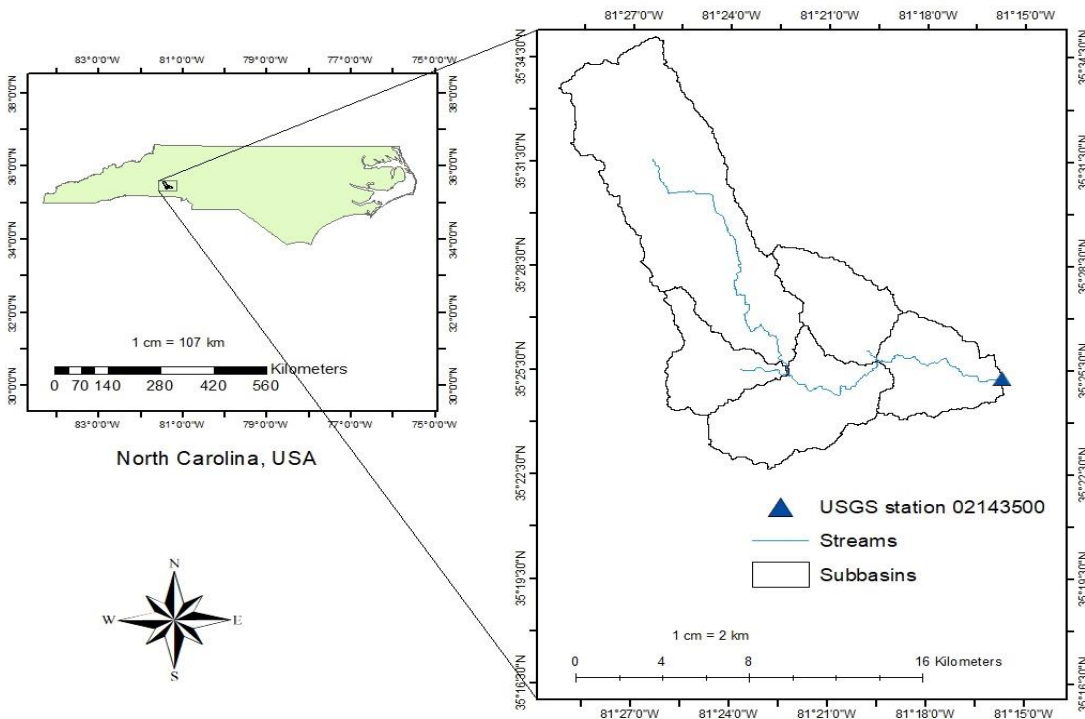
b) Chehalis River watershed
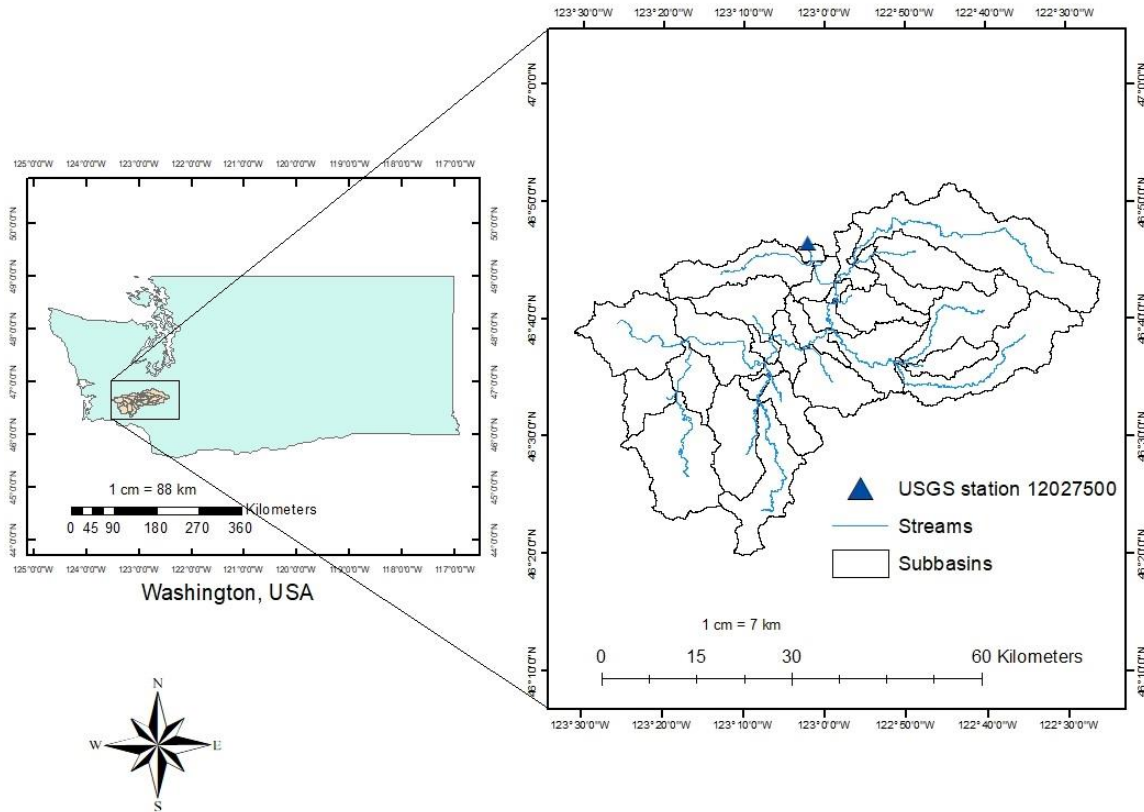
Figure 3-1: Watersheds considered for the study (a) Indian Creek watershed, North Carolina, and (b) Chehalis River watershed, Washington, USA

Indian Creek Watershed drains an area of 179 sq. km and is located in the Inner Piedmont belt of the Western Piedmont of North Carolina. The study area is spread over three counties, namely Catawba, Lincoln and Gaston Counties. The Indian Creek stream channel has a constant flow for roughly 32 km while descending approximately 100 m in elevation from its point of inception, with an average slope of 2.5 m/km (Daniel et al., 1997). The predominant land use in the watershed is agriculture (46%) followed by forest (41%), urban (12%) and the remaining is water bodies (Herrmann, 2008). The climate profile of the Indian Creek watershed is humid subtropical with

long, hot, humid summers and mild winters. The study area has a mean annual precipitation over the watershed is 1270 mm. The coldest month is January with 4.4° C mean monthly temperature while July is the hottest month with a mean monthly temperature of 25.7° C. NC State Climate Office Statistics from 1970-2000 indicates that drought has been intense around the watershed, particularly the Catawba county.

The Chehalis River Watershed has a size of 2318 sq. km and is located in Northwest Washington originating from Willapa Hills region's surface runoff. The study area is spread out over five counties namely Wakhiakum, Cowlitz, Pacific, Lewis and Thurston. The local residents and governments of the Chehalis River watershed have worked together with Federal, State, and Tribal authorities through the Chehalis Basin Partnership to produce a long-term watershed management plan as a result of the demand for water resources (Gendaszek, 2011). Chehalis river main stream runs through two large cities namely Chehalis and Centralia. Land use in the Upper Chehalis basin is predominantly forested with 82% forests, 12% agricultural lands and around 6% urban areas using the NLCD 2001 land use classification. The watershed consists of underlying unconsolidated alluvial and glacio-fluvial aquifers, confined by hydrogeologic units consisting of fine alluvial sediments with low permeability (Gendaszek, 2011). Approximately 30% of the total precipitation converts into surface runoff, 30% percolates underground to recharge groundwater, 35% is lost to evapotranspiration and around 5% can be attributed to interception losses (Gendaszek et al., 2018). The climate profile of Chehalis River watershed is temperate Mediterranean with a mean annual precipitation over the watershed is 1205 mm. The coldest month is December with 5° C mean monthly temperature while August is the hottest month with a mean monthly temperature of 19° C. In regards to Chehalis River Watershed, Department of Ecology for the State of Washington states that the trend of bigger, more frequent winter flooding occurrences will continue as a result

of climate change and the droughts are anticipated to occur more frequently throughout the basin, and hotter and dryer summers.

3.2 Input Data

A brief overview of the geo-spatial and hydrometric and meteorological data used in the current study is present in the following sections.

3.2.1 GIS Data

This study uses both HYMOD and SWAT hydrological models. While HYMOD is a conceptual model and does not require any geo-spatial inputs, SWAT is a physically-based model and requires GIS spatial inputs to enable the process-based simulations. These spatial inputs include a Digital Elevation Model (DEM), land use land cover (LULC) layer, soil layer, and watershed boundary shapefile. The DEM used for both the study areas of this thesis was a standard USGS 30 m resolution DEM which is a part of the National Elevation Dataset (NED), accessible through the USGS TNM download website (https://apps.nationalmap.gov/downloader/#/). Additionally, the required soil layers were obtained from Natural Resources Conservation Service (NRCS) website (https://websoilsurvey.sc.egov.usda.gov/). This study performs hydrologic modelling over the study areas over a period of (1993-2001), National Land Cover Database 2001 (NLCD) from the data gateway website of USDA (https://datagateway.nrcs.usda.gov/) was obtained and used as the LULC layer. The watershed shapefile for performing GIS operations for SWAT modeling is obtained directly for the Model Parameter Estimation Experiment (MOPEX) obtained from National Oceanic and Atmospheric Administration (NOAA) website (https://hydrology.nws.noaa.gov/pub/).

3.2.2 Meteorological Data and Flow Data

This thesis uses the Model Parameter Estimation Experiment, or MOPEX dataset for the meteorological and hydrometric data. It is an international project comprising a massive and comprehensive database of historical weather conditions and land surface characteristics for a variety of hydrologic basins located all over the world. Data from 438 catchments in the United States as well as data from several other catchments located all over the world are available as part of MOPEX. The data set for the United States includes daily time series of mean aerial values for a variety of climatic variables, including maximum temperature, minimum temperature, potential evaporation, and precipitation, amongst others. There is also a record of the daily streamflow available for each of the basins. For the model development and subsequent calibration and validation, the dataset from two MOPEX basins, namely Indian Creek watershed and Chehalis River watershed were used in this thesis.

HYMOD requires daily precipitation and daily potential evapotranspiration (PET) data along with daily time series of outlet discharge data for both the study areas. All of this data is directly available form the MOPEX dataset. SWAT model requires mainly daily precipitation and maximum, minimum temperature data for the model simulations which are also available directly from the MOPEX dataset. SWAT also requires additional information like wind velocity, humidity and solar radiation in order to represent the physical processes of the hydrological cycle, but these data are supplemented using the data from SWAT weather generator in-built into the SWAT program itself.

3.3 Hydrological Model Setup

Hydrological model is an extremely useful tool at a modeler's disposal which simulates the natural hydrological processes. This section gives a brief explanation on the operational framework and subsequently, the methodologies of both HYMOD and SWAT models.

3.3.1 HYMOD

The HYMOD model of rainfall and runoff is based on conceptualized physical processes (Figure 3-2) (Moore, 1985). Three steps can be used to divide the modelling process. To begin with, the amount of infiltration and runoff generated is determined using the excess infiltration method. After subtracting evapotranspiration and infiltration, runoff is defined as the remaining water. The input variable for this step is evapotranspiration, and the infiltration is calculated using the soil infiltration capacity, which is determined by the two parameters Cmax and Bexp as given in Table 3-1. When addressing the spatial distribution of water storage, Bexp is used to determine the maximum storage capacity, or Cmax. Second, using a coefficient, surface runoff and base flow are created from the runoff (excess water). The surface runoff is computed by three consecutive, identical quick reservoirs with a travel time of Rq, and the base flow is computed by a slow reservoir with a travel time of Rs. The total of the discharges from the quick and slow reservoirs is then used to calculate the discharge's schematic provides an overview of the modelling procedure.
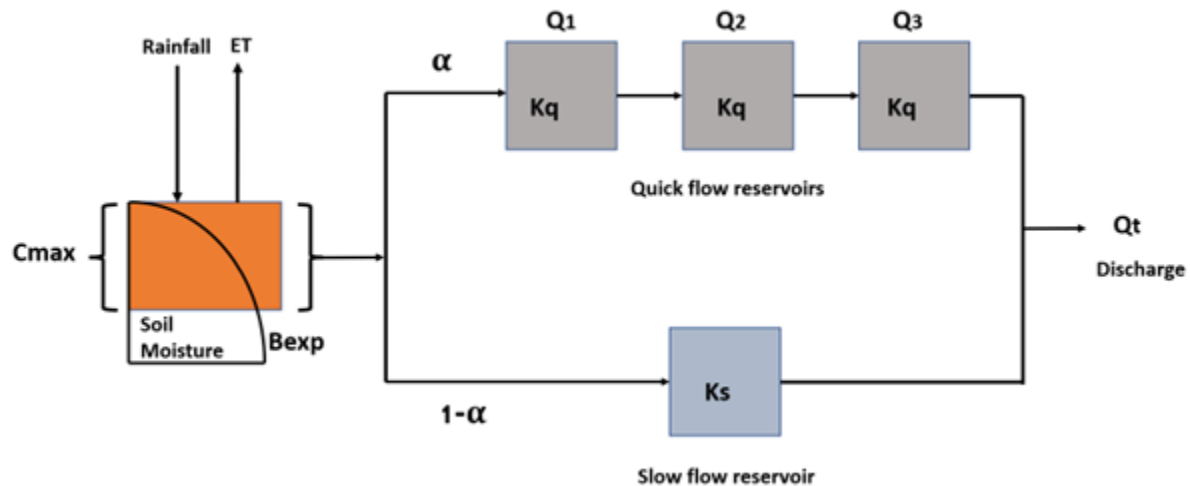
Figure 3-2: HYMOD schematic (Quan et al., 2015)

As previously mentioned, there are five parameters in the HYMOD model. The first three parameters – Cmax, Bexp, and α are used to calculate the generated runoff, whereas Rq and Rs are used in the routing process to calculate the discharge at the catchment outlet. The five parameters are not explicitly measured in the field because this is a lumped rainfall runoff model; rather, their ranges have been established in prior research (Quan et al., 2015; Vrugt et al., 2008) as shown in Table 3-1. The parameters' distribution is assumed to be uniform in this study (Quan et al., 2015; Vrugt et al., 2008).

Table 3-1: HYMOD parameter description and ranges

Parameter	Unit	Description	Lower Bound	Upper Bound
Cmax	mm	Maximum soil moisture	1	500
Bexp	-	Exponential parameter in soil routing	0	2
α	-	Partitioning factor	0.01	0.99
Rs	day	Travel time of slow tank	0	0.1
Rq	day	Travel time of quick reservoirs	0.1	1

3.3.2 SWAT Model

In order to set up a SWAT model, streamlines are first burned onto the DEM to ensure the accuracy of the automatic delineation. The outlet points are then selected based on the desired flow gauges. Third, the process of delineation is conducted based on the selected outlets. To determine the hydrologic response units (HRUs) for each subcatchment, the study area is classified according to land-use, soil characteristics, and slope. Fifth, the model receives all weather data as a time-series table. Lastly, model parameters are estimated based on slope, land use, soil information, and weather conditions. SWAT is now prepared for calibration and validation, preferably after a warmup period to prevent initialization errors. There are a variety of methods for calibrating models. SWAT-CUP is one of the most popular toolkits due to its versatility in calibration techniques and its ease of use.

SWAT-CUP contains three automatic calibration algorithms: SUFI, GLUE, and Para-Sol. In this study, SUFI is used to calibrate models for multiple reasons. SUFI is the quickest algorithm because it employs the LHS method to cover the parameter ranges. SUFI is dependent on running the model multiple times with progressively narrower parameter ranges. In addition, previous research has demonstrated that SUFI performs slightly better than the other two algorithms (Khatun et al., 2018; Singh et al., 2013). After defining the uniform distribution of each parameter, the number of iterations, and the objective function, calibration can be performed. The parameter ranges utilised in this study for both hydrological models are based on previous research and are presented in

Table 3-2 (Xie et al., 2020). In order to maintain a relatively consistent spatial distribution, the values of spatially distinct parameters are modified by multiplying the original value by a ratio. Curve Number (CN2.mgt) is an example of spatial parameters, so it is denoted as "R" for relative

change. Other catchment-wide parameters with fixed values are modified by randomly selecting a

value from the corresponding uniform distribution. These parameters are noted as "V" for re-

positioning change. The calibration's objective is to maximise the NSE value. After executing the

automatic calibration tool, the optimal parameter set is selected, and the model can be re-executed

in SWAT-CUP for validation or in SWAT for additional simulation.

Table 3-2: SWAT Model Parameters

Parameter	Change	Min.	Max.	Parameter Description
v__ALPHA_BF.gw	Replace	0	1	Baseflow alpha factor (days)
v__ALPHA_BNK.rte	Replace	0	1	Baseflow alpha factor for bank storage (days)
r__CN2.mgt	Ratio	-0.35	0.35	SCS runoff curve number
v__ESCO.hru	Replace	0	1	Soil evaporation compensation factor
v__SFTMP.bsn	Replace	-5	5	Snowfall temperature
r__SOL_AWC(..).sol	Ratio	-0.3	0.3	Soil available water content
v__GWQMN.gw	Replace	0	5000	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)
v__RCHRG_DP.gw	Replace	0	1	Deep aquifer percolation fraction
v__CH_L2.rte	Replace	-0.05	500	Length of main channel
v__SNO50COV.bsn	Replace	0	0.9	Snow water equivalent that corresponds to 50% snow cover
v__GW_REVAP.gw	Replace	0.02	0.2	Groundwater "revap" coefficient
v__REVAPMN.gw	Replace	0	500	Threshold depth of water in the shallow aquifer for "revap" to occur (mm)
r__SOL_K(..).sol	Ratio	-0.3	0.3	Saturated hydraulic conductivity
r__SOL_BD(..).sol	Ratio	-0.5	0.5	Soil moist bulk density

During the calibration procedure, SWAT-CUP is able to generate a global sensitivity analysis report. The sensitivity report identifies the most sensitive parameters that should serve as the focus of uncertainty analysis. The report on sensitivity may be presented as scatter plots or in a statistical format (p-test and t-test). If the scatter plot is distributed uniformly along the range, the corresponding parameter is not sensitive. If the scatter points exhibit a distinct trend, then this parameter is sensitive. To avoid human bias, the statistical format is preferred over the scatter plot format. If the p-value is less than 0.05, which also indicates that the t-test values are high, the parameter is deemed sensitive (Khatun et al., 2018). When two parameters have identical p-values, a t-test can be used to distinguish between them. In this study, the five most significant parameters in two different hydrological models are identified by this method and are subjected to further uncertainty analysis.

3.4 Polynomial Chaos Expansion (PCE)

The generalized PCE and the LAR-PCE were compared by examining the ability to build a suitable surrogate model with a small training dataset, the degree of accuracy reflecting uncertainty in streamflow prediction, and the degree of improvement in efficiency of two surrogate models compared to the original hydrological models described in section 3.3. PCEs of degree 3 and several different experimental designs were considered with values ranging between N = 100 – 2000 for the generalized PCE, and N = 100 – 500 for the LAR-PCE based on previous literature (Dwelle et al., 2019; Hampton and Doostan, 2014; Torre et al., 2018). The Leave-one-out (LOO) error values (section 3.4.3) for the PCE meta-models were compared and the best performing models from each generalized and LAR-PCE were considered for further analysis.

After obtaining best performing generalized and LAR-PCE models, they were statistically analyzed by comparing with the SWAT model MC runs, utilizing performance characteristics such

as MUE, MUAE of mean, standard deviation, skewness, and kurtosis, along with the NSE, PBIAS, and RSR values. The most effective and accurate PCE model out of the two was then chosen for backward MCMC and sensitivity analysis of the SWAT models. This section details the methodologies of two types of surrogate PCE model namely fully polynomial chaos – Ordinary Least Square (OLS) regression and sparse polynomial chaos – LAR, followed by the leave-one-out error estimation.

3.4.1 Full Polynomial Chaos – Ordinary Least Square (OLS) Regression

In general, a model's output is a function of its input fields. As a result, the output can be expressed as a nonlinear function of the collection of random variables used to represent the stochasticity of the input (Huang et al., 2007). Typically, the polynomial chaos (PC) method is used to express the evolution of uncertainty in dynamical systems with randomly generated inputs. As first proposed by Wiener (1938), it was suggested that a second-order random process (which is applicable to the vast majority of physical processes) be expressed in terms of orthogonal polynomials. In the process, Hermite polynomials were used to decompose the model stochastic process into Gaussian random variables. However, the convergence of the Hermite polynomial expansion is not optimal for non-Gaussian random input variables (e.g., Beta and uniform) (Xiu and Karniadakis, 2002). Hence, Xiu and Karniadakis (2002) proposed a generalized form of PCE, in which according to the type of random input, the polynomials can be chosen from the Wiener-Askey family of polynomials (Table 3-3). They suggested that if the polynomial basis is chosen based on the distribution of the random inputs, an "optimal choice" is made.

Table 3-3: Correlation between random variables and Wiener-Askey polynomial chaos (Xiu and Karniadakis, 2002)

	Random Variable	Wiener-Askey chaos	Support
Continuous	Gaussian	Hermite-Chaos	$(-\infty, \infty)$
	Gamma	Laguerre-Chaos	$[0, \infty)$
	Beta	Jacobi-Chaos	$[a, b]$
	Uniform	Legendre-Chaos	$[a, b]$
Discrete	Poisson	Charlier-Chaos	$\{0, 1, 2,... \}$
	Binomial	Krawtchouk-Chaos	$\{0, 1,...,N\}$
	Negative-Binomial	Meixner-Chaos	$\{0, 1, 2,... \}$
	hypergeometric	Hahn-Chaos	$\{0, 1,...,N\}$

Given a dynamic model $Y = M(X)$, where X is an input vector comprising of N uncertain parameters and Y is the desired output response (e.g., simulated streamflow). The hydrological simulation model M converts the input X to the output Y; additionally, the model response Y can be estimated using the PCE meta-model M^{PC} composed of a set of polynomial bases (Tran and Kim, 2021):

$$Y = M(X) \approx M^{PC}(X) = \sum_{i=0}^{\infty} a_i \Psi_i(X)$$

(3-1)

where $\Psi_i(X)$ denotes corresponding multivariate polynomials in terms of N uncertain parameters; i is a multi-index identifying the components of the multivariate polynomials; and a_i denotes unknown PCE coefficients. In Equation (3-1), the multivariate polynomials Ψ_i are constructed as the tensor product of the univariate orthogonal polynomials $\Psi_{\alpha_j}^{(j)}(X)$ with the degree α_j:

$$\Psi_i(\boldsymbol{X}) = \prod_{j=1}^{N} \Psi_{\alpha_j}^{(j)}(\boldsymbol{X}_j) \qquad (3\text{-}2)$$

Approximations are made in practise using a finite summation in a finite dimensional space. This is accomplished by truncating Equation (3-2) to obtain a low-order PCE (Sudret, 2008) as expressed in Equation (3-3),

$$Y = \boldsymbol{M}(\boldsymbol{X}) \approx \boldsymbol{M}^{PC}(\boldsymbol{X}) = \sum_{i=0}^{P-1} a_i \Psi_i(\boldsymbol{X}) \qquad (3\text{-}3)$$

where P denotes the number of PCE coefficients (i.e., the number of polynomial expansion basis terms), which is determined by the number of parameters N, and the degree of the polynomial p as:

$$P = \frac{(N+p)!}{N!\,p!} \qquad (3\text{-}4)$$

For polynomial chaos methods, there are three essential steps. They are the determination of the orthogonal polynomial basis of model input parameters, the calculation of the coefficients of polynomial chaos, and the estimation of the accuracy of polynomial chaos approximations. For the determination of orthogonal polynomial basis, according to the Wiener–Askey polynomial chaos, there exist different optimal polynomials for various probability density functions (Table 3-3) such as normalised Legendre (respectively Hermite) polynomials being associated with uniform (respectively Gaussian) probability density functions. In this study, it is assumed that the model parameters are uniformly distributed, and the Legendre polynomial chaos is selected.

In the process of calculating coefficients of polynomial expansion, intrusive and non-intrusive approaches can be utilised (Xiu, 2010). The intrusive method must modify the code of the original model, whereas the non-intrusive method does not. Thus, the non-intrusive method is applicable

to the study of most of the models. The non-intrusive method consists of two primary methods: regression and projection (Blatman and Sudret, 2011). In this work, the regression-based nonintrusive method ordinary least square regression (OLS) is chosen.

For the collection of multivariate polynomials $\Psi_i(\boldsymbol{X})$, the following step is to compute the PCE coefficients a, using OLS. The OLS method seeks to identify the PCE coefficients that minimize the mean-square error of the surrogate model's approximation of the model response (3-5).

$$a = \arg min_{a \in \mathbb{R}^P} \mathbb{E}\left[\left(Y - \sum_{i=0}^{P-1} a_i \Psi_i(\boldsymbol{X})\right)^2\right] \tag{3-5}$$

The number of PCE coefficients that must be estimated is P, which can be determined using the Equation (3-4). Given an experimental design consisting of a collection of N sets of random parameters, $\boldsymbol{X} = \{X^{(1)}, X^{(2)}, \dots, X^{(N)}\}$ followed by the corresponding model response $\boldsymbol{Y} = \{\boldsymbol{M}(X^{(1)}), \boldsymbol{M}(X^{(2)}), \dots, \boldsymbol{M}(X^{(N)})\}$, the estimated PCE coefficients are provided by:

$$a = \arg min_{a \in \mathbb{R}^P} \frac{1}{N} \sum_{k=1}^{N} \left(Y^{(k)} - \sum_{i=0}^{P-1} a_i \Psi_i(X^{(k)})\right)^2 \tag{3-6}$$

The OLS solution for Equation (3-6) is:

$$a = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{Y} \tag{3-7}$$

where $\boldsymbol{A} = \{A_{ij} = \Psi_j X^{(i)}, i = 1, \dots, P; j = 1, \dots, card\ A\}$ consists of the values of all basis polynomials in the experimental design points and is referred to as the experimental matrix.

3.4.2 Sparse Polynomial Chaos – Least Angle Regression (LAR)

A complementary strategy for favouring sparsity in high dimension is to directly modify the least-square minimization problem in Equation (3-5) by adding a penalty term of the form $\lambda ||a||_1$:

$$a = \arg\min_{a \in \mathbb{R}^P} \mathbb{E}\left[\left(Y - \sum_{i=0}^{P-1} a_i \Psi_i(X)\right)^2\right] + \lambda||a||_1 \qquad (3\text{-}8)$$

where λ is a non-negative constant and the regularisation term $||a||_1 = \sum_{i \in A} |a_i|$ forces the minimization to favour sparse solutions with a low rank.

The primary difference between LAR and OLS is the number of PCE coefficients, which is fewer in LAR. Specifically, LAR identifies only the multivariate polynomials $\Psi_i(X)$ that have the greatest influence on the model response, while discarding all other polynomial terms. Estimates are made for the selected weighty PCE coefficients, while other insignificant coefficients are set to zero. Based on the sparse set of PCE terms, a surrogate model is constructed and can be described in Equation (3-9):

$$Y = M(X) \approx M^{PC}(X) = \sum_{i=0}^{S-1} a_i{}^s \Psi_i{}^s(X) \qquad (3\text{-}9)$$

where $\Psi_i{}^s(X) = \{\Psi_0{}^s(X), \dots, \Psi_{S-1}{}^s(X)\}$ represent the set of significant polynomials, a_i are the set of corresponding coefficients, and S are the number of PCE terms that are preserved.

3.4.3 Leave-one-out (LOO) error estimation

After the polynomial coefficients have been computed, the polynomial chaos-based model with the required degree of precision is successfully established through error estimation. In this work, the leave-one-out (LOO) method for error estimation is chosen. Cross-validation is used to create the leave-one-out (LOO) cross-validation error ε_{LOO}. It involves constructing N meta-models $Y^{PC/i}$, each based on a reduced experimental design $X^{(i)} = \{X^{(j)}, j = 1, \dots, N, j \neq i\}$, and comparing each model's prediction for the excluded point $X^{(i)}$ to its actual value $Y^{(i)}$. The LOO cross-validation error can be represented as follows:

$$\varepsilon_{LOO} = \frac{\sum_{i=1}^{N}\left(\boldsymbol{M}(X^{(i)}) - \boldsymbol{M}^{PC\backslash i}(X^{(i)})\right)^2}{\sum_{i=1}^{N}(\boldsymbol{M}(X^{(i)}) - \mu_Y)^2} \tag{3-10}$$

where μ_Y represents the sample mean of the experimental design's response. In practical application, when the results of a least-squares minimization are available, it is not necessary to explicitly calculate N distinct meta-models, as Equation (3-10) can be reduced to:

$$\varepsilon_{LOO} = \frac{\sum_{i=1}^{N}\left(\dfrac{\boldsymbol{M}(X^{(i)}) - \boldsymbol{M}^{PC}(X^{(i)})}{1 - h_i}\right)^2}{\sum_{i=1}^{N}(\boldsymbol{M}(X^{(i)}) - \mu_Y)^2} \tag{3-11}$$

where h_i represents the diagonal term of the matrix $\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$. The definition of information matrix $\boldsymbol{A}$ is given in Equation (3-7).

3.5 Uncertainty Analysis

Three different types of approaches to uncertainty analysis are used in this study: the forward approach, the backward approach, and sensitivity analysis using the PCE metamodels. The forward method focuses primarily on sampling from parameter distributions and quantifying their effects on model outputs. The backward method focuses on determining parameter ranges or distributions based on the propagation of prediction errors in reverse. The primary objective of the forward approach is to identify the optimal model outcomes, whereas the backward approach serves primarily as a monitoring tool and an early warning system for extreme events. Uncertainty analysis can make use of global sensitivity analysis (GSA). A sensitivity analysis can be performed on a hydrological model to identify the parameters that have the greatest impact on the output responses and to determine the parameter ranges, so that the uncertainty of the most sensitive parameters can be evaluated.

3.5.1 SWAT Uncertainty Quantification

After performing the automatic calibration and sensitivity analysis of the SWAT models using SWAT-CUP, five most sensitive parameters are chosen to perform the Monte Carlo (MC), generalized-PCE, and LAR-PCE analyses. To perform MC, the selected parameters are varied randomly within their physical range for 2,500 realisations to determine the period of uncertainty. To perform PCE, however, the selected parameters are assumed to be independent and transformed to standard normal distributions. Then, for each parameter, collocation points are selected from the normal distribution. The SWAT model is then executed with all possible parameter value combinations at the collocation points, and one output value is obtained for each combination at each time step. For each parameter combination at each time step, it is then possible to establish a linear equation in which the PCE coefficients serve as the unknowns. At this step, LAR differentiates from the generalized PCE approach by considering only the most significant and sparse set of meaningful coefficients. The PCE coefficients can be obtained by solving the system of linear equations. Using the obtained PCE coefficients, the PCE surrogate model for the specified time period is constructed.

Sensitivity analysis (SA) is utilised to determine the extent to which each parameter contributes to the output uncertainty and permits the identification of crucial parameters that govern model behaviour. In general, SA can be divided into two categories: local and global. The local SA computes the changes to the simulation model by modifying one parameter while holding the remaining parameters constant. However, it is frequently incapable of producing meaningful outcomes (Saltelli et al., 2004). The global SA, on the other hand, investigates the model changes by adjusting all parameters at once. The Fourier amplitude sensitivity test (FAST) (Cukier et al., 1973), the Morris one-at-a-time screening (MOAT), the Sobol' sensitivity indices, and the

response surface methodology (RSM) are some of the commonly used global SA techniques in the literature. In this study, we used Sobol's sensitivity indices to perform a global sensitivity analysis.

As a ratio of partial and total variances, the Sobol' indices can be calculated as given by:

$$S_{i_1\dots i_s} = \frac{D_{i_1\dots i_s}}{D} \tag{3-15}$$

Where D is the total variance of model $\boldsymbol{M}$. The first-order indices can be expressed in the case of a single input variable as follows:

$$S_i = \frac{D_i}{D} \tag{3-16}$$

The total Sobol' indices indicate the total contribution of an input variable to the output variance. These are determined by adding up all the indices pertaining to that variable:

$$S_i^T = \sum S_{i_1\dots i_s} \tag{3-17}$$

Finding the relative importance of each input variable in relation to the model's output also makes it possible to identify the variables that can be discounted.

This makes it possible to fix those variables to deterministic values after performing the Sobol' analysis. There are numerous techniques that can be used to determine the values of the partial and total variances. Realizing that a PCE approximates the Sobol decomposition of a model into polynomials is one of them.

The partial variances can also be expressed using the same notation used in the total variance estimation by expressing the total variance in terms of the PCE coefficients:

$$\sigma(M_v{}^{PC}x_v)^2 = \sum_{i=1\ldots N^v} a_i{}^2 \qquad (3\text{-}18)$$

Where $M_v{}^{PC}$ is the v th PCE term and a is the i th PCE coefficient, and v stands for the index set of the PCE expansion terms that only depend on the subset of input variables x_v. It is possible to obtain the Sobol' indices directly from post-processing of the PCE coefficients by calculating the ratio between Eqs. (3-20) and (3-9), which makes it possible to obtain all indices essentially for free.

The first-order indices will have values between 0 and 1, since they allot a portion of the total variance to each individual model parameter. Since they also include the interaction terms, if any, the total indices will always be greater than or equal to the first-order indices. Repeating these among various total indices is possible (interaction between two parameters increases value of two total indices). Therefore, the sums will be:

$$\sum_{i=1}^{N} S_i \leq 1 \leq \sum_{i=1}^{N} S_i^T \qquad (3\text{-}19)$$

These uncertainty quantification techniques are used to find the parametric uncertainties of the SWAT models for Indian Creek and Chehalis River watersheds.

3.5.2 HYMOD Uncertainty Framework using MCMC

Once HYMOD is set up and ran, the surrogate model calculates a mathematical relationship between the input HYMOD parameter values and the streamflow output using it's output database. The UQLab tool was used to select and create polynomial chaos expansion (PCE) metamodels for this study (Sudret, 2008). Due to its widespread use in uncertainty quantification for streamflow estimation and hydrological applications, this type of metamodel was chosen.

The next step is to figure out how to build the PCE surrogate by calculating the values of the polynomial coefficients Eq. (3-9). There are numerous ways to approach this issue, including projection techniques (like Gaussian quadrature), conventional least-square techniques, subspace pursuit, Bayesian compression, or least angle regression, the latter of which is the approach taken in this study. In order to promote sparsity in high-dimensional problems, the least angle regression modifies the traditional least square minimization by including a penalty term (Blatman and Sudret, 2011). Additionally, compared to other methods, least angle regression offers a quick computation time. The leave-one-out (LOO) cross-validation error is the metric used while training the PCE model for HYMOD. By creating multiple PCE metamodels, each with a smaller training set, and evaluating each model's prediction on the excluded point, this metric prevents the PCE model from being overfit to the training data. This might make it easier for the PCE model to generalise to new data.

An experimental design of 500 HYMOD parameter sets, within their physical parameter ranges was considered for building the LAR-PCE metamodel with a degree of 3. The experimental design and degree of the polynomials were based on previous PCE studies conducted with HYMOD (Hu et al., 2019; Wang et al., 2015). The uniform distribution over a specific (prior) range was assumed to govern the prior distributions for the uncertain parameters. Latin hypercube sampling (LHS) was used due to its efficiency to determine parameter sets from their distribution.

The goal of MCMC is to create an invariant distribution that is as close as possible to the target posterior distribution by building a Markov chain over the prior distribution. The transition probability $K(x^{t+1}|x^t)$, which determines the likelihood of moving from the current step t at each iteration to the following step at a time t+1 when the condition in Eq. (3-1) is satisfied, is a crucial part of MCMC.

$$\pi(x^{t+1}|D)K(x^t|x^{t+1}) = \pi(x^t|D)K(x^{t+1}|x^t) \tag{3-20}$$

Where D is the measured data and $\pi(x)$ represents our fundamental understanding of the hydrological watersheds and is the prior distribution of the input parameters x at time t.

This requirement guarantees the chain's reversibility, or, alternatively, that the probability of moving from x^t to x^{t+1} is equal to the probability of moving from x^{t+1} to x^t. It is possible to demonstrate in Eq. that the resulting distribution is equal to the posterior by integrating Eq. (3-1) over the support of the parameters R.

$$(\pi(x^{t+1}|D) = \int_R \pi(x^t|D)\, K(x^t|x^{t+1}))K(x^{t+1}|x^t)dx^t \tag{3-21}$$

Afterwards, a post-processing on the MCMC samples can be done to calculate the posterior distribution moments (mean, standard deviation, etc.), or fit the distribution to the closest defined distribution shape (normal, lognormal, gamma, etc.).

The Metropolis-Hastings algorithm, Adaptive Metropolis algorithm, and Hamiltonian Monte Carlo algorithm are some of the most popular MCMC algorithms. The affine invariant ensemble algorithm (AIES) (Goodman and Weare, 2010) is more resistant to sampling from posterior distributions that show strong correlation between its parameters, according to Goodman and Weare (2010) and Wagner et al., (2019). As a result, unlike the recommended classical MCMC algorithms, AIES is able to sample from both types of distributions with or without correlation without explicitly requiring the target distribution to undergo an affine transformation. As a result, in this study, samples from the posterior distribution are taken using AIES. It is important to note that AIES requires more computation than other MCMC algorithms because it is slower. Before reaching the posterior distribution, MCMC algorithms typically require hundreds of thousands of

model evaluations. This is why the introduction of PCE is beneficial in the context of calibration of hydrological models.

By using the MCMC's sampling option from the posterior, each calibrated parameter in *x* receives a random distribution of potential values rather than a single value. However, in the context of this study we are looking for a single set of parameter values to use in practise that maximises the posterior distribution. With this method, known as Maximum A Posteriori (MAP), the *x* value that maximises the posterior distribution is reported in Eq.

$$x_{MAP} = argmax_x \pi(x|D)$$
(3-22)

In this study, we present the MAP values for each HYMOD parameter along with the MCMC-discovered random distributions for each parameter.

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 General

Chapter 4 offers organized results and their subsequent discussion. This chapter covers the results pertaining to HYMOD and SWAT model separately. For HYMOD, Markov Chan Monte Carlo algorithm is used for parameterization and subsequent calibration followed by an enhanced calibration achieved through the use of a surrogate model, PCE. For SWAT model, in-built SUFI2 was used to calibrate and validate followed by using PCE to quantify the parameter uncertainty.

4.2 HYMOD

This section provides description of the results pertaining to HYMOD modeling and it includes parameterization and the results of using surrogate model for improved calibration. Table 3-1 provides the parameter settings for the MCMC UQ analysis. Along with other climatological inputs, MOPEX streamflow data that have been observed are used for both the study areas. The AIES sampler employs 100 parallel chains altogether. The total number of generated samples from the posterior distribution is 20,000 samples, with each chain consisting of 200 steps. We postprocess the other half of the samples and discard the first 5,000 samples, or 25% of the total. Since the first samples in the chain typically have lower quality before the sampler starts to converge, this burn-in period is a common practise in MCMC. Finally, the point estimates or MAPs (x_{MAP}) are presented.

4.2.1 Indian Creek Watershed

After building the PCE surrogate with the help of the parameters, it was discovered that the PCE model had a low enough LOO value of 0.02 to support sensitivity analysis and UQ applications.

It is important to note that since we are not using the surrogate for point-wise predictions but rather as a tool to speed up MCMC analysis, the PCE model is not supposed to be very accurate. As we will ultimately validate the MCMC results using real computer code, capturing the trend of the data could therefore be sufficient for the surrogate. All things considered, this metric increases the trustworthiness of the PCE surrogate model when it is used in place of the HYMOD model in MCMC analysis.

We postprocess the posterior samples by eliminating the initial 25% burn-in samples, and then in Table 4-1 we present summary statistics of the calibrated parameters. The statistics consist of the mean value, standard deviation, the 95% confidence interval (CI), which includes lower and upper bounds, and lastly the MAP estimate of Eq . (3-20). The small standard deviation and the close confidence intervals in Table 4-1 indicate that the inverse UQ results once more show a significant reduction in parametric uncertainty in the two model parameters (Rs and alpha). Figure 4-1 shows that the major issue in the HYMOD model for Indian Creek watershed is the prediction of peak flows. While the MCMC does well in capturing the baseflow scenarios, the extreme flood scenarios are not simulated well.

Table 4-1: Posterior marginals for the Indian Creek watershed

| Parameter	x_{MAP}	SD	(0.05-0.95) Quant.	Type
Cmax	3.30E+02	35	(2.8e+02 - 3.9e+02)	Model
Bexp	0.77	0.16	(0.54 - 1)	Model
alpha	0.14	0.042	(0.1 - 0.23)	Model
Rs	0.0037	0.0031	(0.00049 - 0.0096)	Model
Rq	0.31	0.16	(0.23 - 0.84)	Model
sigma2	23	1.6	(21 - 26)	Discrepancy
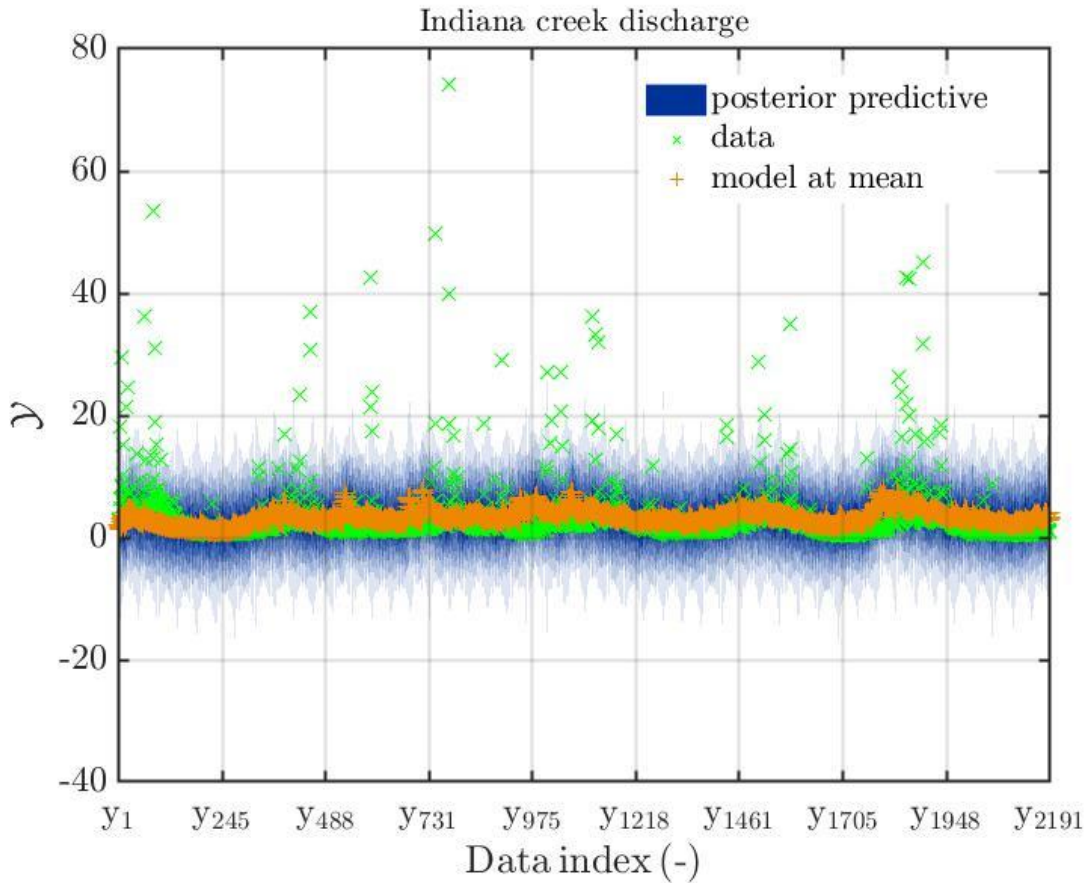
Figure 4-1: 95% confidence interval band streamflow graph for Indian Creek watershed

Similar to the uncertainty captured by the 95 PPU, Figure 4-2 shows a plot of the marginal (also known as univariate) posterior distribution of the five model parameters (residual discrepancy is excluded since it is not a model parameter). Three parameters exhibit a skewed distribution, while the posteriors of the two model parameters (Cmax, Bexp) resemble a normal distribution. Additionally, four out of the five parameters' distribution ranges where the Maximum a Posterei (MAP) estimate falls indicate high accuracy of the MCMC sampling in capturing the posterior marginal. Based on Figure 4-2, it can be seen that MCMC does a great job in narrowing down the uncertainty band for all five parameters.
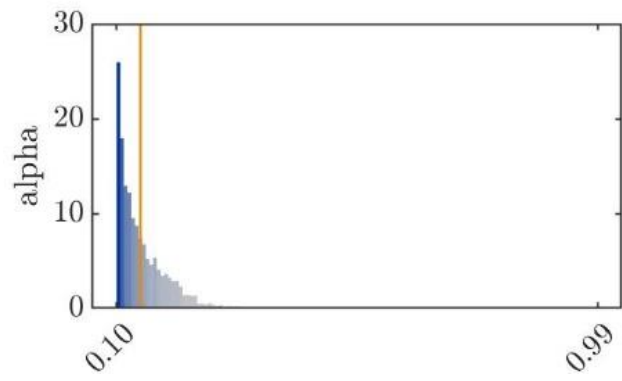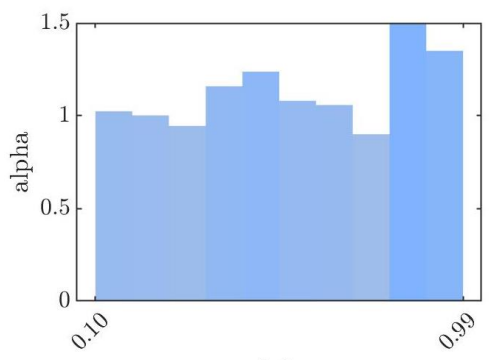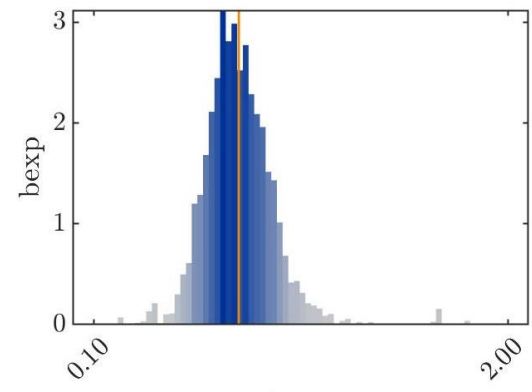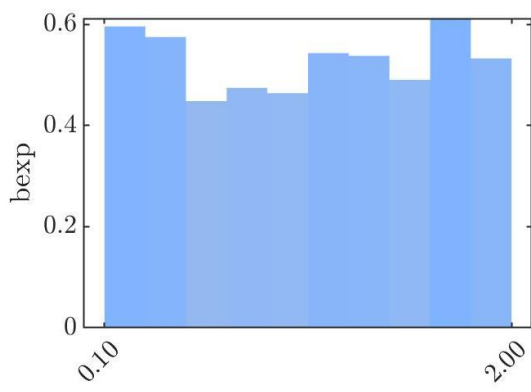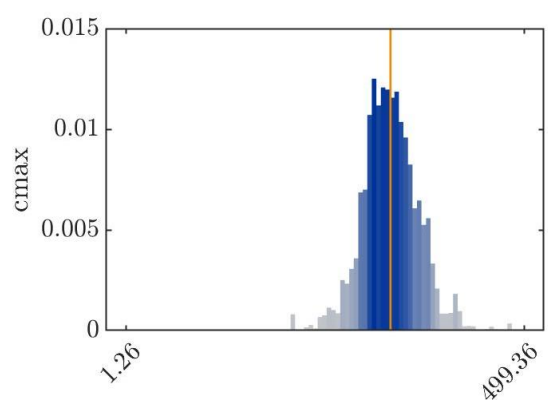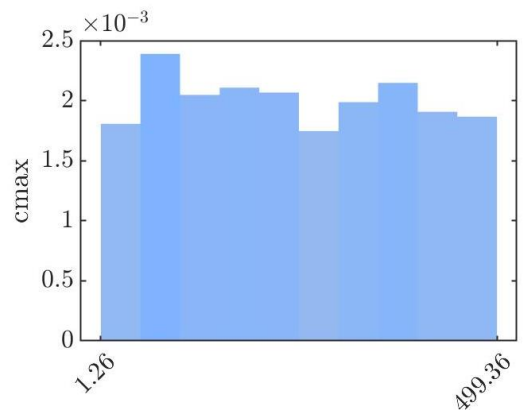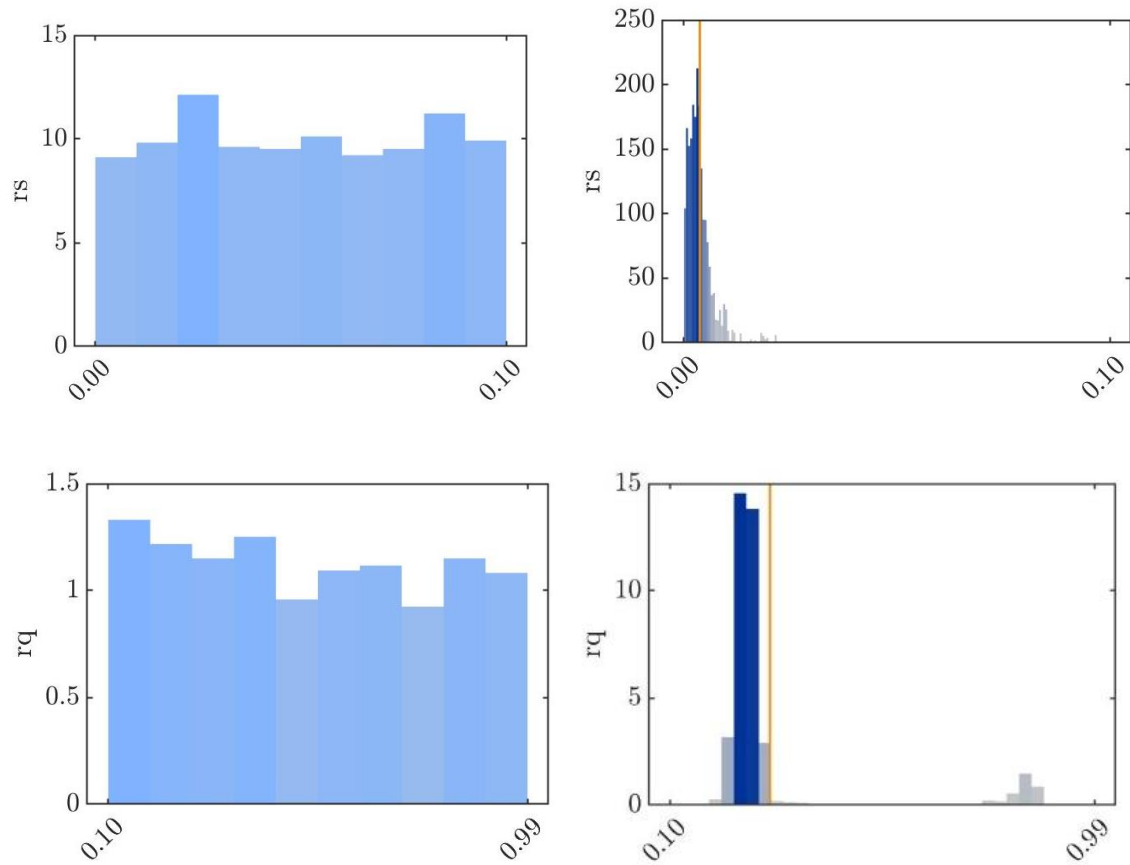
Figure 4-2: Prior (left) and posterior (right) parameter distribution for the MCMC analysis

When sampling posteriors with a high degree of parameter correlation, the AIES sampler is advantageous. The correlation matrix between the five parameters is displayed in Table 4-2. Next, there is not a strong (generally >-0.5) negative correlation between the model parameters Rq and alpha. The largest positive correlation between travel times for quick reservoirs and slow tanks can be attributed to the Indian Creek watershed's predominantly well-drained subsoil.

Table 4-2: Correlation matrix of HYMOD parameters for Indian Creek watershed

	Cmax	Bexp	alpha	Rs	Rq
Cmax	1	-0.37	-0.082	0.2	0.33
Bexp	-0.37	1	0.17	-0.21	-0.29
alpha	-0.082	0.17	1	0.074	-0.31
Rs	0.2	-0.21	0.074	1	0.61
Rq	0.33	-0.29	-0.31	0.61	1

4.2.2 Chehalis River Watershed

It was found that the PCE model had a low enough LOO value of 0.03 after using the Table 3-1 parameters to build the PCE surrogate to support sensitivity analysis and UQ applications. After removing the initial 25% burn-in samples from the posterior samples, we present summary statistics for the calibrated parameters in Table 4-3. The 95 percent confidence interval (CI), which includes lower and upper bounds, the mean value, standard deviation, and finally the MAP estimate of Eq. (3-20) comprises the statistics. Table 4-3's tight confidence intervals and small standard deviation show that the inverse UQ results once more demonstrate a sizable reduction in parametric uncertainty in the two model parameters (Rs and alpha).

Table 4-3: Posterior marginals for the Chehalis River watershed

Parameter	Mean	Std	(0.05-0.95) Quant.	Type
Cmax	4.10E+02	60	(3e+02-4.9e+02))	Model
Bexp	0.64	0.18	(0.38-0.98)	Model
alpha	0.14	0.058	(0.1-0.21)	Model
Rs	0.0061	0.0054	(0.0024-0.01)	Model
Rq	0.37	0.29	(0.18-0.88)	Model
sigma2	3.00E+02	4.2	(2.9e+02-3e+02)	Discrepancy

Figure 4-3 shows that the HYMOD model for Chehalis River watershed does a better job at representing the peak flow scenarios than the Indian Creek Model. However, in a broader context, peak flow simulation is still an issue in this case, along with a slight over-prediction of baseflows. A plot of the posterior distribution of the five model parameters is shown in Figure 4-4. One model parameter, Bexp, has a posterior that resembles a normal distribution, while all other parameters, with the exception of that one, have skewed distributions. We observe that two parameters (alpha, Rs) are tightly confined within a small uncertainty band. Furthermore, the MAP estimate falls within the distribution ranges of four out of the five parameters, indicating high accuracy of the MCMC sampling in capturing the posterior marginal. It's interesting to note that in neither of the two scenarios has MAP passed through Rq.
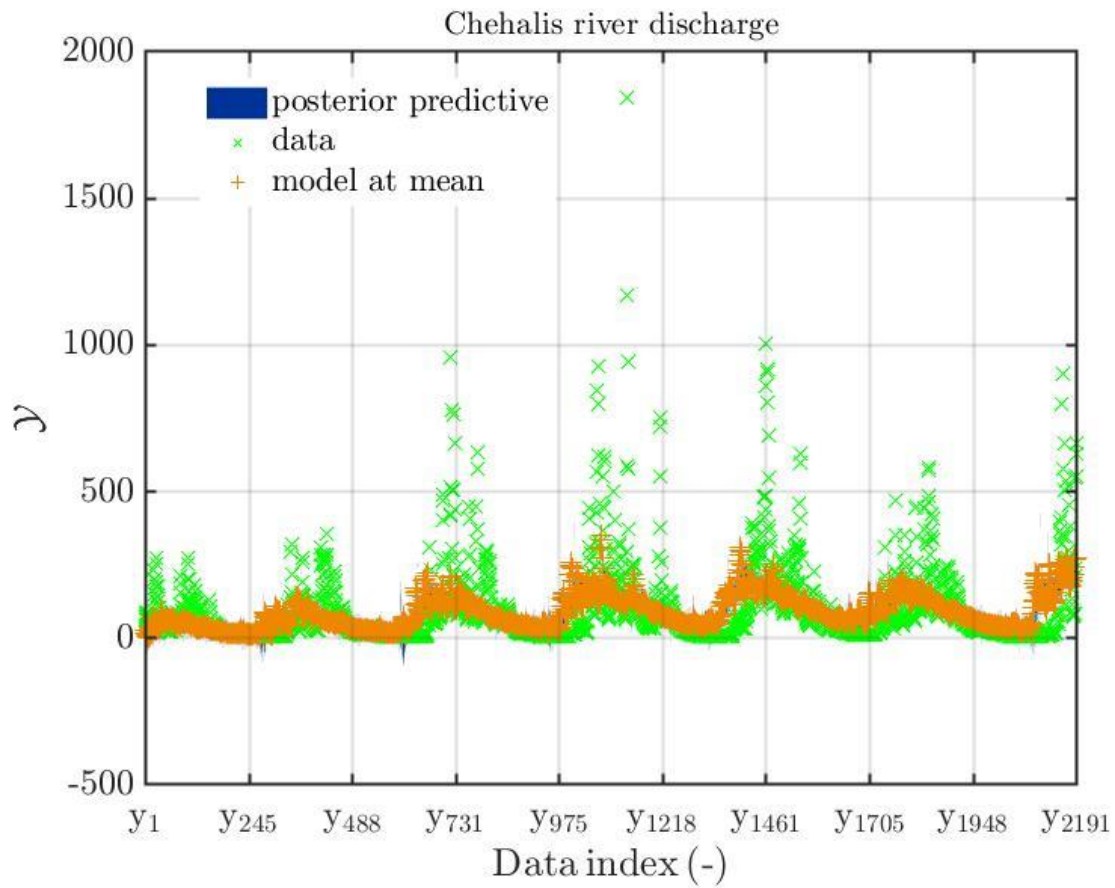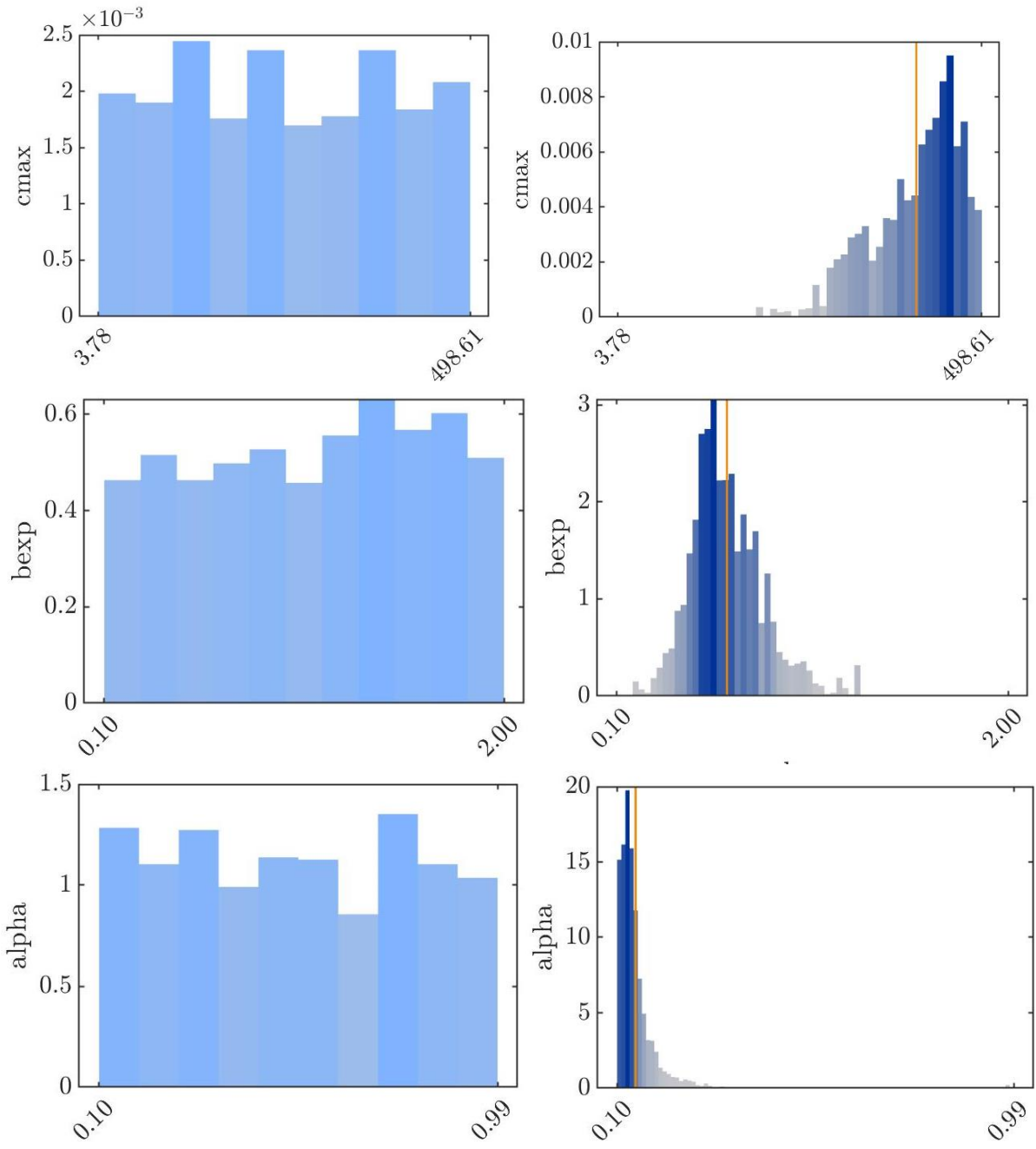
Figure 4-3: 95% confidence interval band streamflow graph for Chehalis River watershed
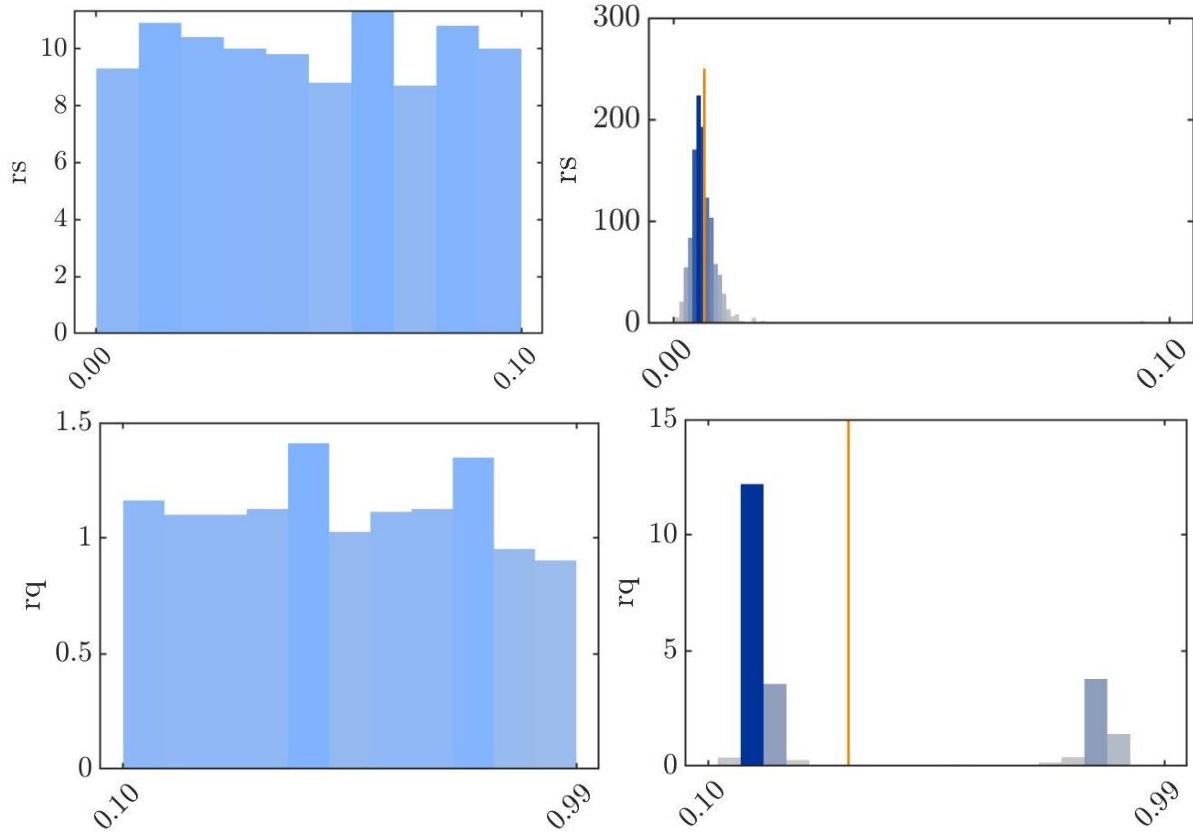
Figure 4-4: Prior (left) and posterior (right) parameter distribution for the MCMC analysis

Table 4-4 shows the correlation matrix between the five parameters. Next, the model's soil water storage (Cmax) and time of flow in the slow tank have a strong negative correlation (-0.54). (Rs). This can be explained by the large alluvial floodplain that makes up the Chehalis River watershed, which has a propensity to absorb and hold moisture. The partitioning factor for quick and slow tanks (alpha) and the time of flow in slow tanks have a strong positive correlation, which makes sense given that the former can affect the latter.

Table 4-4: Correlation matrix of HYMOD parameters for Chehalis River watershed

|	Cmax	Bexp	alpha	Rs	Rq
Cmax	1	-0.35	0.038	-0.051	-0.54
Bexp	-0.35	1	0.23	0.15	0.26
alpha	0.038	0.23	1	0.74	-0.29
Rs	-0.051	0.15	0.74	1	0.017
Rq	-0.54	0.26	-0.29	0.017	1

4.3 SWAT model

This section presents results of parameter sensitivity and calibration of the SWAT model along with uncertainty assessment using PCE, followed by sensitivity analysis using PCE.

4.3.1 SWAT Calibration and Parameter Sensitivity

A total of 14 parameters were calibrated for the SWAT model over the historical period from 1993 to 1998 for both the SWAT models. Four parameters were updated by a specific ratio to maintain spatial consistency during the automatic calibration process, while the remaining parameters were replaced using values drawn at random from uniform distributions over their respective physical ranges as shown in . The observed vs simulated streamflows are displayed in Figure 4-5 for both the watersheds, as simulated by SWAT.

Observed Vs Simulated Streamflow

a)

Observed Vs Simulated Streamflow
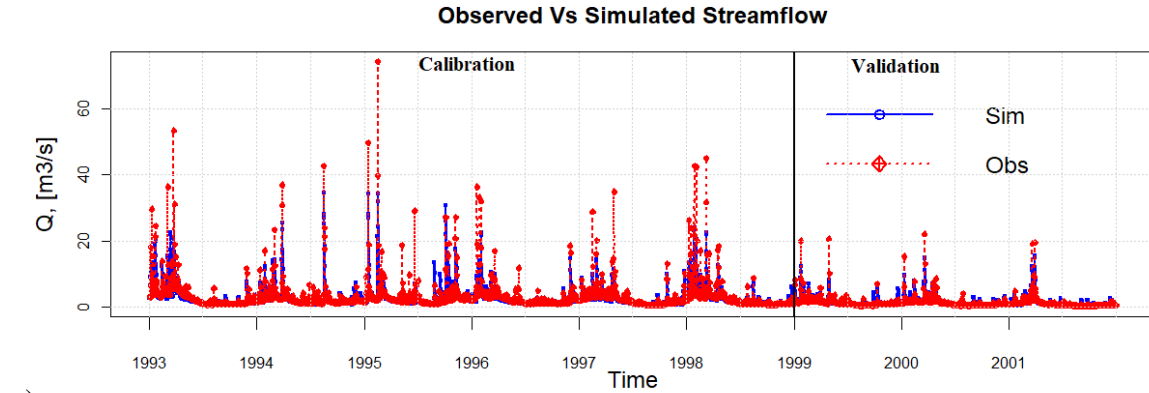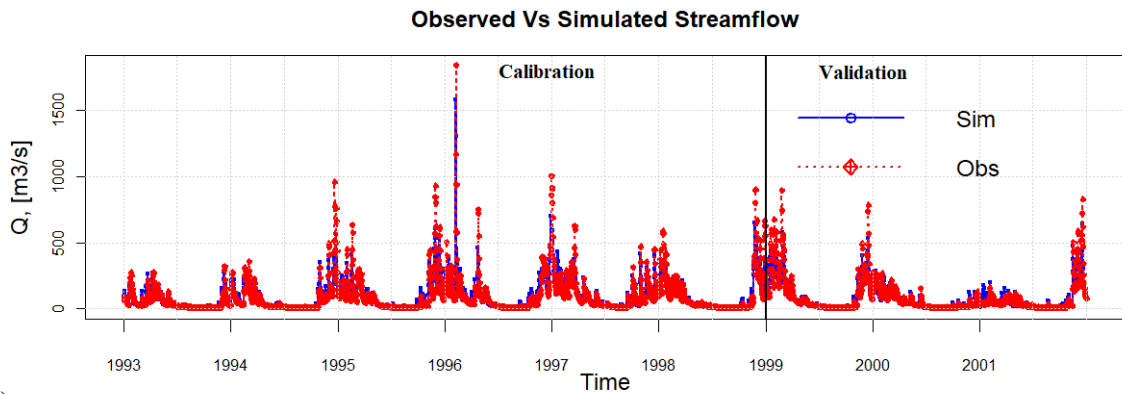
b)

Figure 4-5: Observed vs Simulated flow for calibration and validation periods for a) Indian Creek and b) Chehalis River

The fitted values are extracted and displayed in Table 4-6 for Indian Creek watershed and Table 4-7 for the Chehalis River watershed after two automatic calibration processes with 2,000 iterations each were run. The SWAT-CUP calibration process produces the following results as described in Table 4-5.

Table 4-5: Calibration and Validation results

Watershed	Calibration (1993-1998)	Validation (1999-2001)
Indian Creek Watershed	0.64	0.61
Chehalis River Watershed	0.77	0.82

Table 4-6: Indian Creek Calibrated Values

Parameter Name	Best Fit	t-Stat	P-Value
R__CN2.mgt	0.31	-3.01	0.00
R__SOL_AWC(..).sol	0.11	2.79	0.01
R__SOL_BD(..).sol	0.43	2.17	0.03
V__ALPHA_BNK.rte	0.99	0.88	0.38
V__GWQMN.gw	971.61	-0.81	0.42
V__ESCO.hru	0.45	-0.74	0.46
R__REVAPMN.gw	336.20	-0.56	0.57
V__SNO50COV.bsn	0.55	0.33	0.74
R__SOL_K(..).sol	-0.24	0.30	0.76
V__SFTMP.bsn	-0.92	-0.26	0.79
R__GW_REVAP.gw	0.09	-0.12	0.91
V__RCHRG_DP.gw	0.72	-0.06	0.95
V__CH_L2.rte	96.11	0.01	0.99
V__ALPHA_BF.gw	0.04	0.00	1.00

Table 4-7: Chehalis River Calibrated Values

Parameter Name	Best Fit	t-Stat	P-Value
R__CN2.mgt	-0.33	-70.46	0.00
R__SOL_BD(..).sol	0.24	40.48	0.00
V__GWQMN.gw	51.25	-24.80	0.00
V__RCHRG_DP.gw	0.25	-19.75	0.00
V__SFTMP.bsn	4.51	17.22	0.00
R__SOL_K(..).sol	0.30	11.26	0.00
R__SOL_AWC(..).sol	-0.08	8.51	0.00
V__GW_REVAP.gw	0.18	-6.07	0.00
V__ALPHA_BF.gw	0.58	4.81	0.00
V__ESCO.hru	0.78	0.89	0.37
V__REVAPMN.gw	144.93	-0.67	0.50
V__SNO50COV.bsn	0.56	-0.50	0.62
V__ALPHA_BNK.rte	0.50	-0.23	0.81
V__CH_L2.rte	154.43	0.10	0.92

The results suggest that SWAT can reliably predict daily flow for the study area (Li et al., 2010; Zhang et al., 2016). It is important to note that the models' performance was also validated for three years, which did not change significantly from the calibration period, further proving SWAT's capability to simulate both Indian Creek and Chehalis River catchments.

Table 4-6 and Table 4-7 also include the results of the sensitivity analysis as the t- and p-values. Table 4-6 demonstrates 6 sensitive parameters for the Indian Creek watershed model, and Table 4-6 gives 10 sensitive parameters for the Chehalis River watershed based on the p-significance value's level (p-value less than 0.05). According to the t-stat, CN is the parameter that is most sensitive in both watersheds. The remaining sensitive parameters can be divided into three groups. The first group discusses channel properties, which are directly measurable through stream and watershed surveys and weren't regarded as uncertain parameters because they can be measured in this way. The second group of parameters includes those related to soil moisture and density, which are the most sensitive ones because they significantly affect the infiltration, runoff, and evapotranspiration processes. The third set of sensitive parameters relates to baseflow, which establishes how much infiltrated water contributes to streamflow and how long it takes for water to leave the watershed to reach the outlet. This study evaluated how five parameters' uncertainties spread over time. The five parameters are composed of CN and two each from the second and third groups. To represent the evaporation process and the degree of infiltration, respectively, the soil evaporation compensation factor (ESCO) and available water content (SOL_AWC) were chosen from the second group for both the watersheds. From the third group, shallow aquifer's threshold depth for return flow to occur (GWQMN) was chosen for both watersheds, whereas the fifth parameter considered was, threshold of water in the shallow aquifer for deep percolation to occur (REVAPMN) for Indian Creek watershed, and deep aquifer's percolation fraction (RCHRG_DP) for the Chehalis River watershed.

4.3.2 Uncertainty Assessment – PCE

This section provides an insight into the results of experimental design selection followed by the accuracy and uncertainty quantification of OLS-PCE Model and LAR-PCE Model, respectively.

4.3.2.1 Selection of Experimental Design

We examine how the size of the experimental design affects the precision of surrogate models built by OLS-PCE and LAR-PCE, thereby offering recommendations for selecting the right size of experimental design and proving superiority of LAR-PCE's over OLS-PCE. As mentioned in section 3, several PCE models were built with experimental designs ranging from 100-2000 for OLS-PCE and 100-500 for LAR-PCE. It can be observed from Figure 4-6, that the Leave one-out (LOO) error estimate becomes constant after a certain number of runs. The accuracy achieved by LAR-PCE for both the SWAT models has a faster convergence than the OLS-PCE models for similar results.
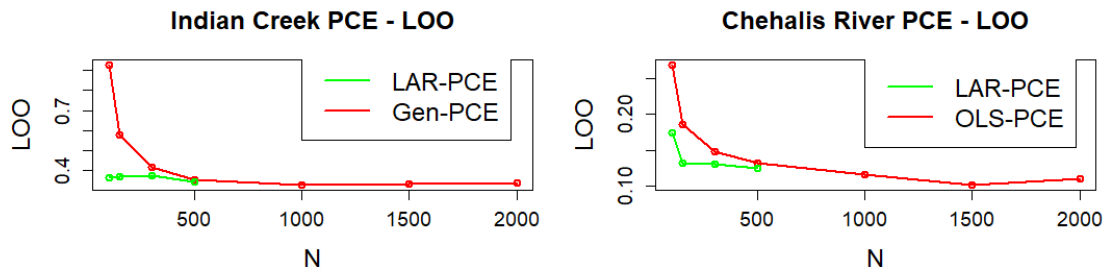

Figure 4-6: Leave-one-out error estimates of LAR-PCE and OLS-PCE for Indian Creek and Chehalis River watersheds for different runs (N)

The best experimental designs without compromising on both computational efficiency and accuracy were considered from Figure 4-6 upon visual inspection; OLS-PCE with an experimental design of 1000 was suggested for Indian Creek, and 1500 for Chehalis River watershed. An

experimental design of 500 was considered for LAR-PCE for both the watersheds, as it provided the least LOO values based on previous literature (Dwelle et al., 2019; Hampton and Doostan, 2014; Torre et al., 2018).

4.3.2.2 OLS-PCE Model – Accuracy and Uncertainty Quantification

OLS-PCE surrogate models were created for the Indian Creek Watershed and the Chehalis River Watershed with an experimental design of 1000 and 1500 runs, respectively. A total of 56 coefficients in the polynomials representing SWAT were estimated through linear regression by running SWAT on the number of sensitive parameter sets given by the experimental design. The surrogate models then created were tested for their ability to simulate the SWAT outputs by running them through realizations of 2,500 random parameter sets.

The comparison of the mean streamflow values obtained using the OLS-PCE and MC simulation methods is shown in Figure 4-7 and Figure 4-8. They show that the mean values obtained through OLS-PCE and the outcomes of the MC simulation are nearly identical. This means that in order to reflect the temporal variations for the streamflow, the OLS-PCE can typically replace the hydrologic model (i.e., SWAT).
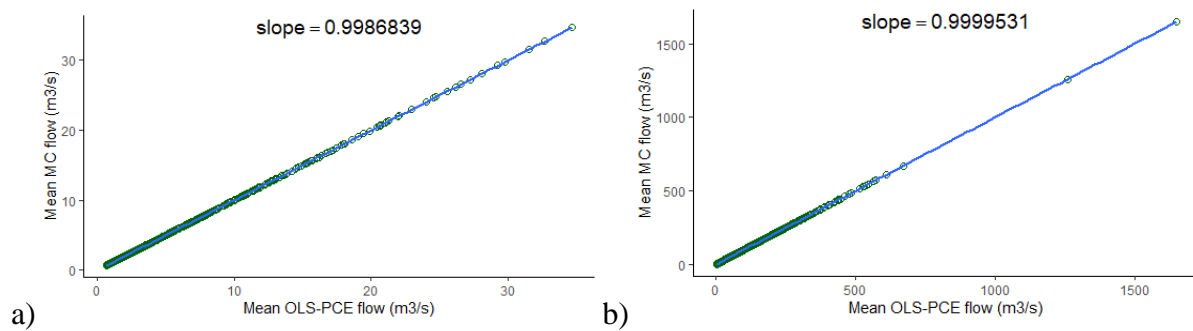
MC vs OLS-PCE Mean Streamflow

a)

MC vs OLS-PCE Mean Streamflow
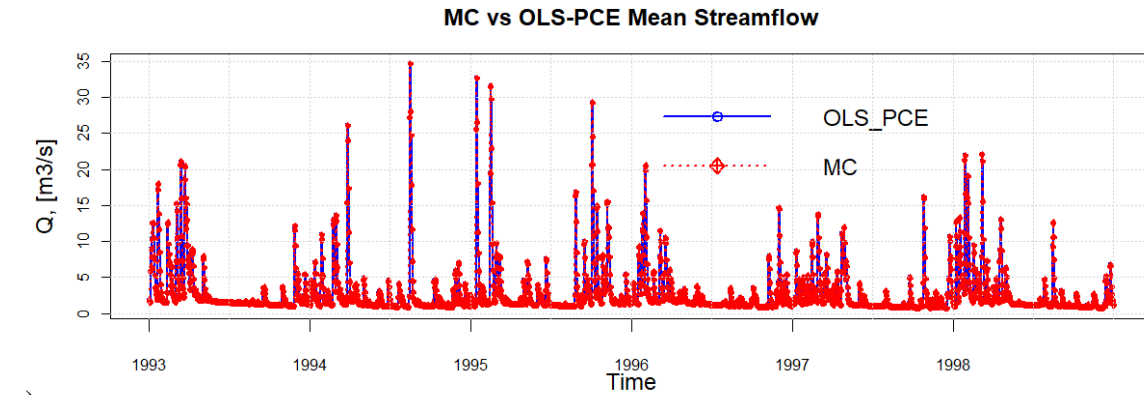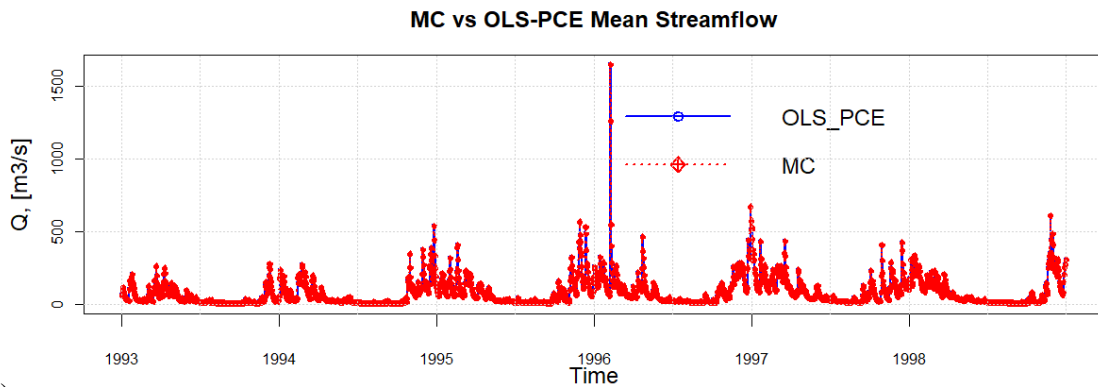
b)

Figure 4-7: A comparison of mean values of streamflow obtained through MC and OLS-PCE simulations for a) Indian Creek Watershed and b) Chehalis River Watershed

a)

b)

Figure 4-8: Scatter plots of mean values of streamflow of MC vs OLS-PCE simulation for a) Indian Creek Watershed and b) Chehalis River Watershed

The standard deviation of the streamflow at each time step as determined by the OLS-PCE and

MC simulation methods are compared in Figure 4-9 and Figure 4-10, respectively. They propose

59

that at both low and high uncertain flows, the standard deviation of OLS-PCE and MC simulation

be the same. Thus, it can be said that both the means and variances of the OLS-PCE results would

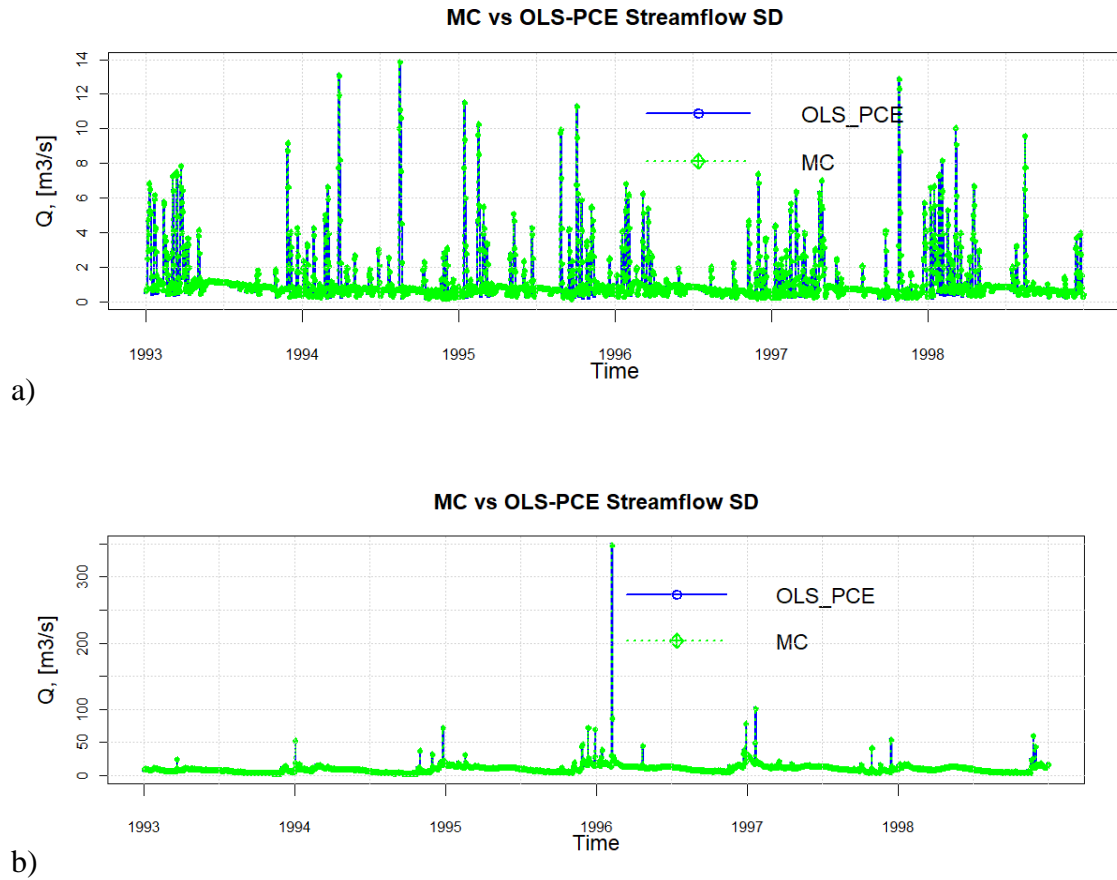fit well with the outcomes of the MC simulation.

a)

b)

Figure 4-9: A comparison of standard deviation of streamflow obtained through MC and OLS-
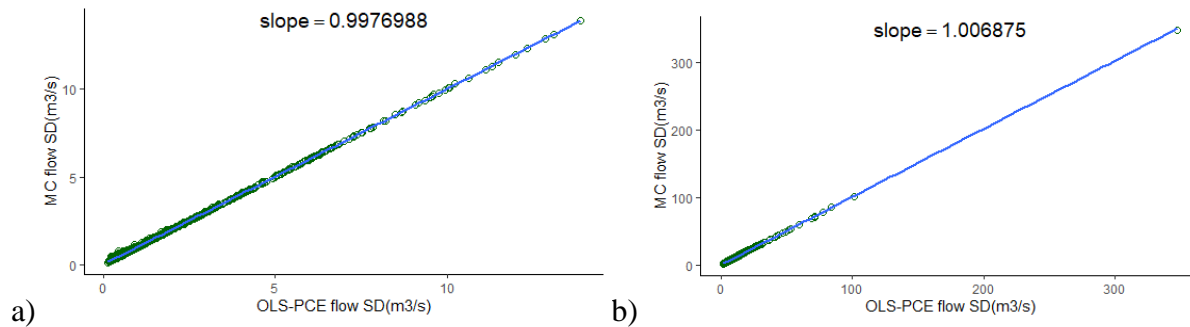PCE simulations for a) Indian Creek Watershed and b) Chehalis River Watershed

Figure 4-10: Scatter plots of standard deviation of streamflow values of MC vs OLS-PCE simulation for a) Indian Creek Watershed and b) Chehalis River Watershed

The in-depth statistical characteristics would be examined at some particular time periods in order to further compare the accuracy between the results of OLS-PCE and MC simulation. Table 4-8 lists the estimates of statistical characteristics, such as mean, standard deviation, kurtosis, and skewness. Low, medium, and high streamflow levels are present during those particular times. In general, the probability density distributions obtained by OLS-PCE and those obtained by MC simulation would be similar in low flows in both the watersheds, as shown in Table 4-8. However, the OLS-PCE results are a bit more skewed, and the probability density function is heavier at the outliers than the MC counterpart – which has a higher peak, represented by a higher skewness and kurtosis during the high flow day in the Indian Creek Watershed. Generally, the OLS-PCE simulations have a steeper probability density function than the MC simulations for the Indian Creek Watershed. The probability density function for OLS-PCE represents the heavily tailed streamflow metric of the Chehalis River Watershed during the high and median flow day. However, the OLS-PCE probability density functions for both the watersheds are adequately able to simulate the MC runs, with Chehalis River Watershed being a bit better to replicate.

Table 4-8: Comparison of OLS-PCE and MC simulation results at specific times

Indian Creek Watershed	OLS-PCE	MC	OLS-PCE	MC	OLS-PCE	MC	OLS-PCE	MC
	Mean		SD		Kurtosis		Skewness	
High Flow (Day - 435)	3.28	3.28	0.99	1.05	2.48	1.95	0.75	0.56
Median Flow (Day - 157)	1.56	1.57	1.16	1.18	1.63	1.75	0.03	0.08
Low Flow (Day - 177)	1.48	1.49	1.1	1.13	1.64	1.76	0.04	0.11
Chehalis River Watershed	OLS-PCE	MC	OLS-PCE	MC	OLS-PCE	MC	OLS-PCE	MC
	Mean		SD		Kurtosis		Skewness	
High Flow (Day - 128)	88.47	88.52	8.77	8.99	2.9	3.06	-0.21	-0.24
Median Flow (Day - 137)	21.42	21.47	7.96	8.17	2.36	2.58	-0.21	-0.16
Low Flow (Day - 510)	10.37	10.37	7.78	7.87	1.86	1.83	0.22	0.2

The histograms of the OLS-PCE and MC simulation results at the chosen time periods are displayed in Figure 4-11. Additionally, it suggests that Chehalis River watershed might be better

represented by OLS-PCE when compared to Indian Creek watershed based on visual correlation between the histograms.

a) Indian Creek Watershed
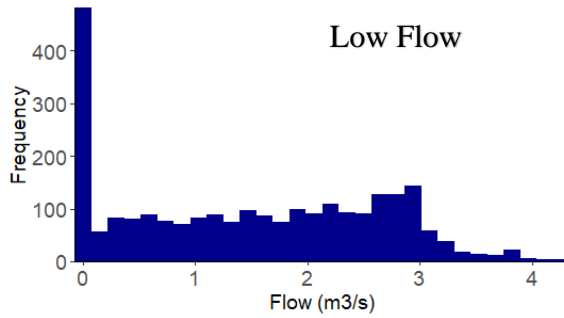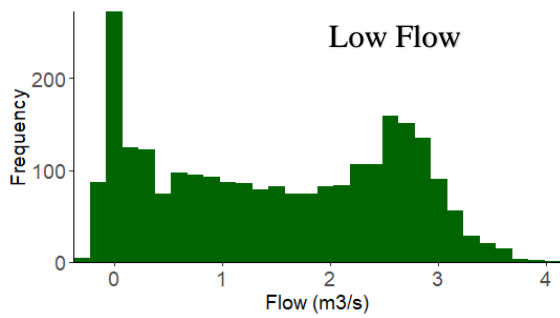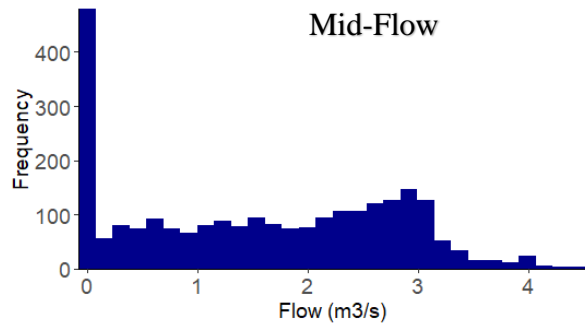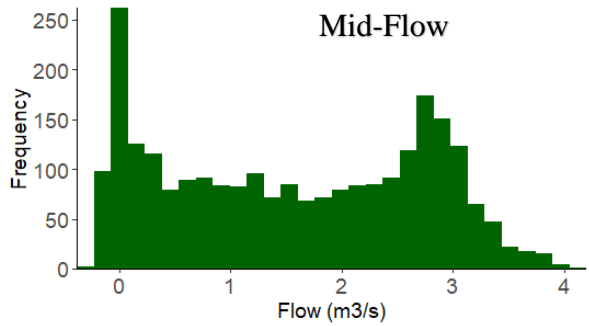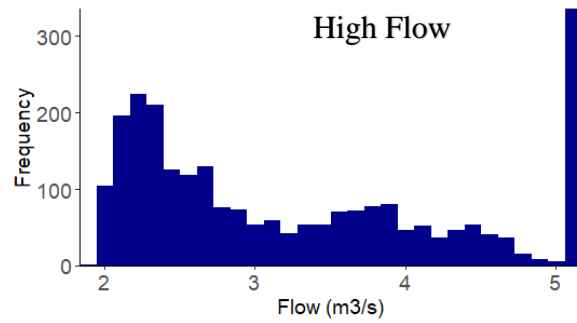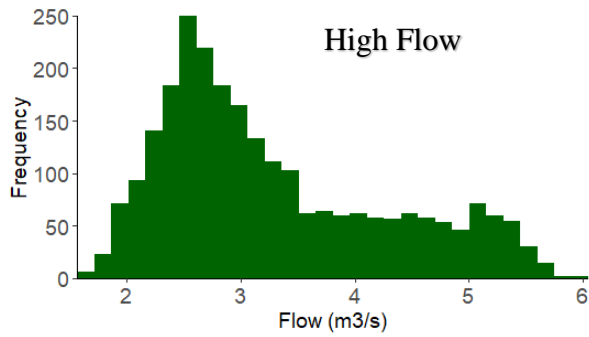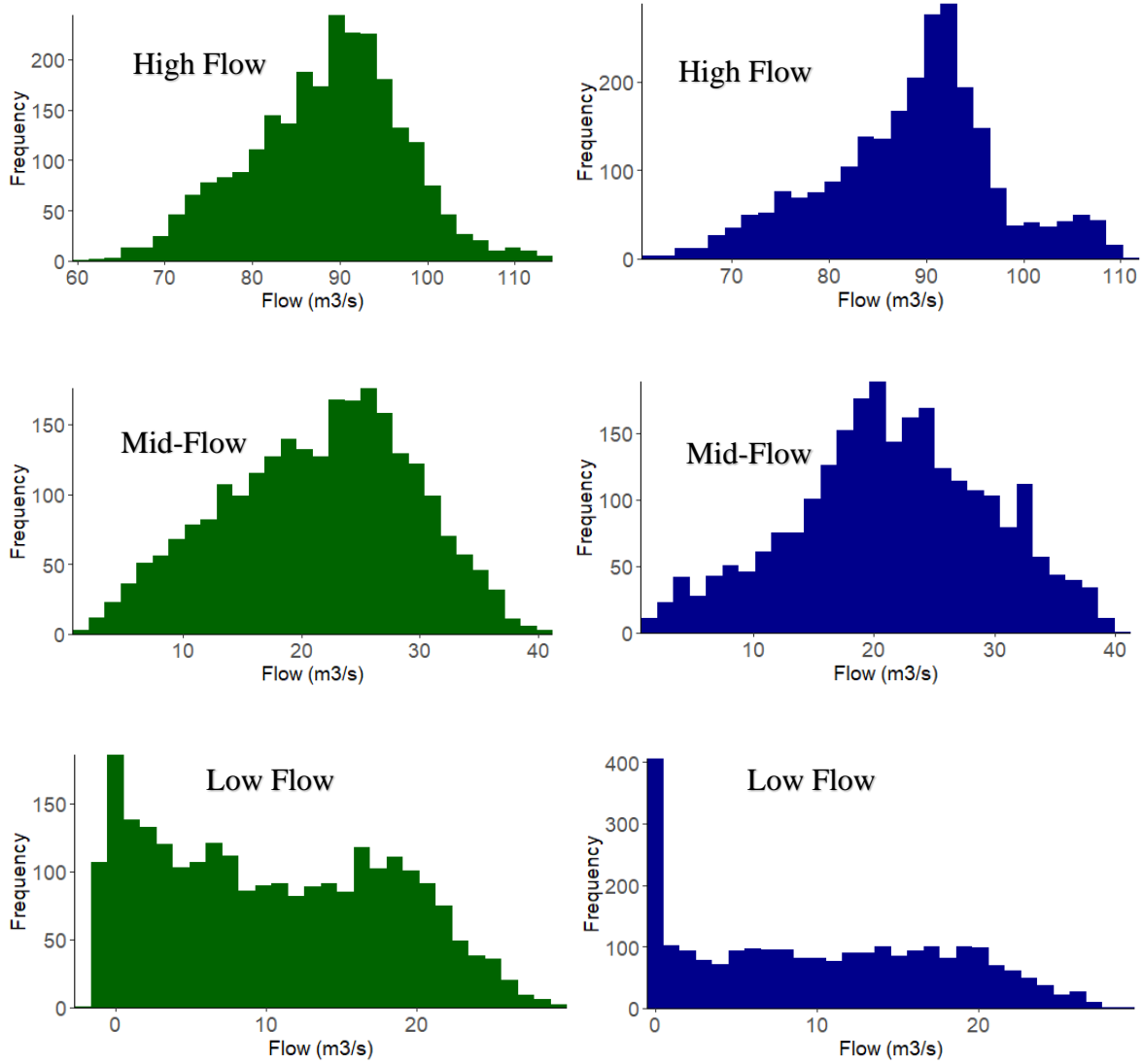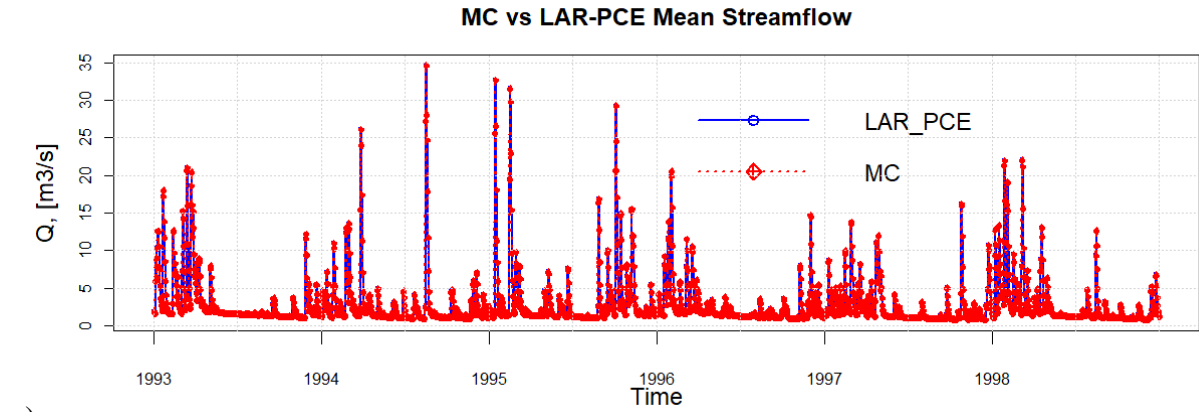
b) Chehalis River Watershed

Figure 4-11: Histograms comparing OLS-PCE and MC simulations for a) Indian Creek watershed and b) Chehalis River watershed

4.3.2.3 LAR-PCE Model – Accuracy and Uncertainty Quantification

With an experimental design of 500 runs, LAR-PCE surrogate models were developed for both the Indian Creek Watershed and the Chehalis River Watershed. A total of 26 important coefficients were truncated out of the total 56 coefficients by the LAR-PCE method. The polynomial coefficients were then calculated by regressing the values of the polynomials to the known SWAT outputs, which is run over a random set of sensitive parameters. Then, using realisations of 2,500 random parameter sets, the surrogate models were evaluated for their capability to simulate the SWAT outputs.

In Figure 4-12 and Figure 4-13, similar with the OLS-PCE simulations, the mean streamflow values obtained by the LAR-PCE and MC simulation methods are compared. They also demonstrate that there is almost no difference between the mean values obtained using LAR-PCE and the results of the MC simulation. This means that in most cases, the LAR-PCE can replace the hydrologic model in order to reflect the temporal variations for the streamflow (i.e., SWAT).
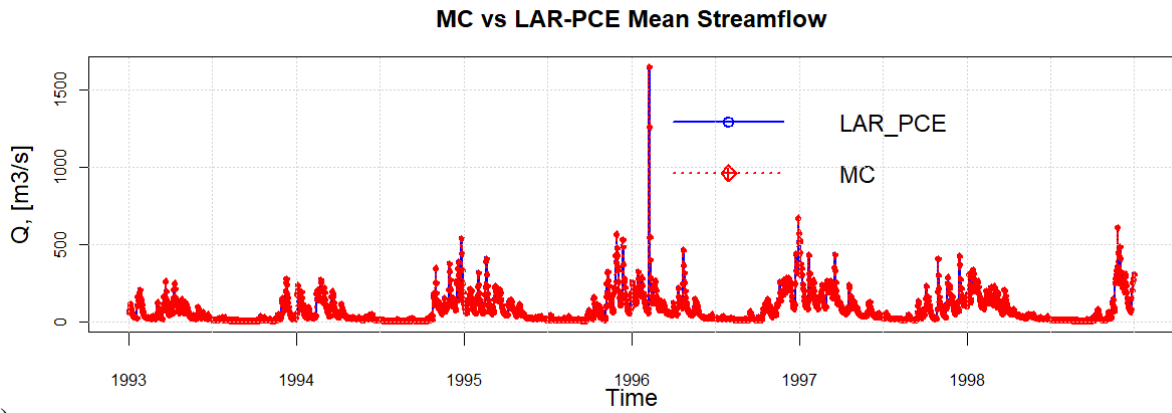
Figure 4-12: A comparison of mean values of streamflow obtained through MC and LAR-PCE simulations for a) Indian Creek Watershed and b) Chehalis River Watershed

Figure 4-13: Scatter plots of mean values of streamflow of MC vs LAR-PCE simulation for a) Indian Creek Watershed and b) Chehalis River Watershed

Figure 4-14 and Figure 4-15 show that the values of standard deviation for LAR-PCE and both the SWAT MC models are an order of magnitude closer than those of OLS-PCE. There is no variance between the two simulations, with LAR-PCE reflecting the model runs even at high and low flow uncertain conditions. As a result, the LAR- PCE does a good job of describing the degree of uncertainty in the SWAT predictions.
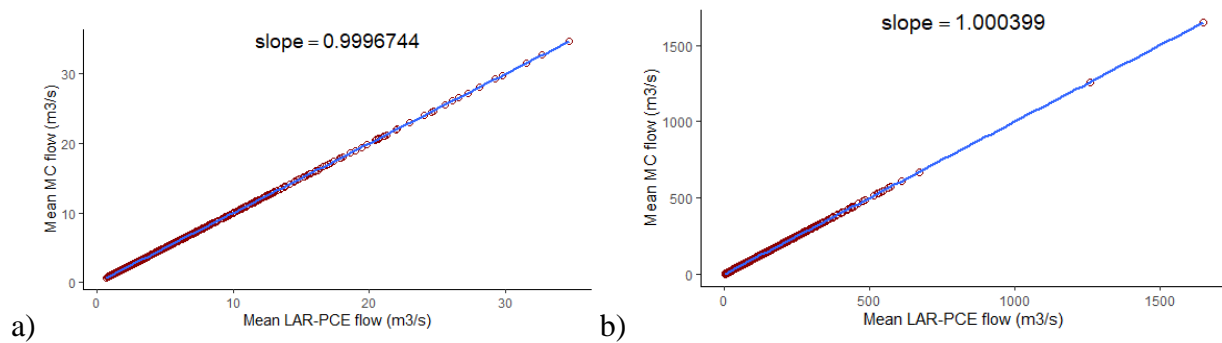

a)

b)
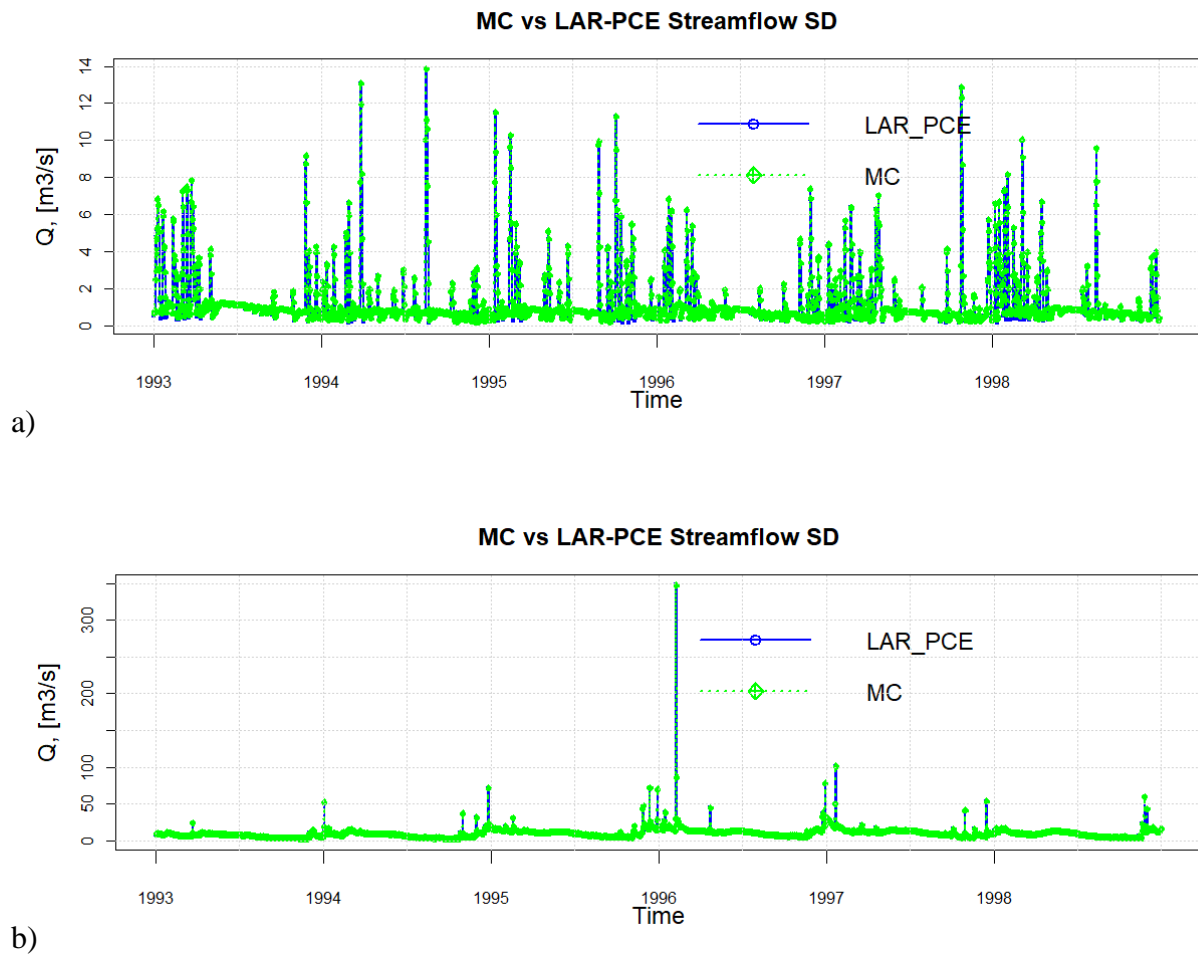
Figure 4-14: A comparison of standard deviation of streamflow obtained through MC and LAR-PCE simulations for a) Indian Creek Watershed and b) Chehalis River Watershed
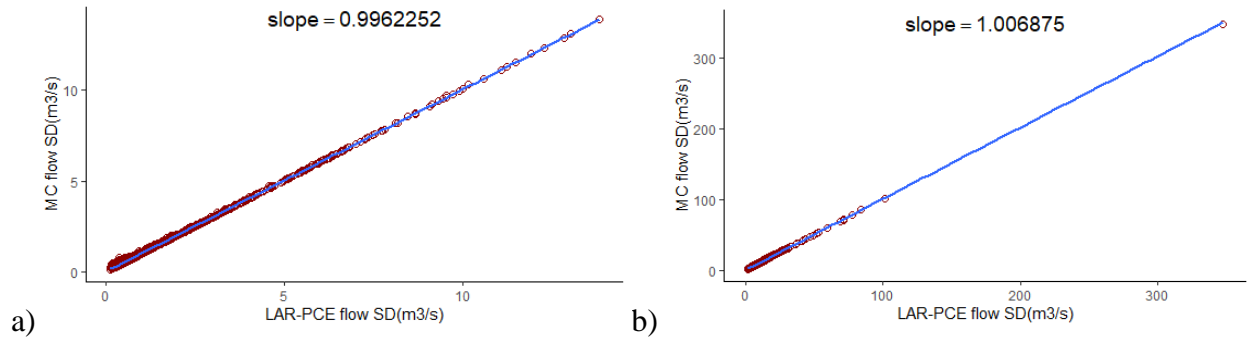
Figure 4-15: Scatter plots of standard deviation of streamflow values of MC vs LAR-PCE simulation for a) Indian Creek Watershed and b) Chehalis River Watershed

The statistic characteristics at some particular time periods would be thoroughly examined in order to compare the accuracy between the results of the LAR-PCE and MC simulations in more detail. The results of the LAR-PCE and MC simulations are shown in Table 4-9 along with their respective means, standard deviations, kurtosis, and skewness values. They suggest that the probability density distributions obtained by LAR-PCE and MC simulation would be fairly similar.

Table 4-9: Comparison of LAR-PCE and MC simulation results at specific times

Indian Creek Watershed	LAR-PCE	MC	LAR-PCE	MC	LAR-PCE	MC	LAR-PCE	MC
	Mean		SD		Kurtosis		Skewness	
High Flow (Day - 435)	3.27	3.28	1	1.05	2.47	1.95	0.83	0.56
Median Flow (Day - 157)	1.56	1.57	1.15	1.18	1.6	1.75	0	0.08
Low Flow (Day - 177)	1.49	1.49	1.1	1.13	1.6	1.76	0.01	0.11
Chehalis River Watershed	LAR-PCE	MC	LAR-PCE	MC	LAR-PCE	MC	LAR-PCE	MC
	Mean		SD		Kurtosis		Skewness	
High Flow (Day - 128)	88.42	88.52	8.59	8.99	2.83	3.06	-0.2	-0.24
Median Flow (Day - 137)	21.39	21.47	7.88	8.17	2.44	2.58	-0.28	-0.16
Low Flow (Day - 510)	10.35	10.37	7.72	7.87	1.81	1.83	0.19	0.2

However, as can be seen from Figure 4-16, there are certain peaks in the Indian Creek probability density function, which are not present in the LAR-PCE simulation. This can be corroborated by the fact that there is some difference in the kurtosis values, which can be seen as the tails of the

LAR-PCE probability density function being heavier than it's MC counterpart, making it well spread out. That being said, the errors are minimal in the practical context. The LAR-PCE model for the Chehalis River watershed definitely does a better job at giving a similar probability density function as the MC simulations, with closer SD, kurtosis, and skewness values.

a)

b)
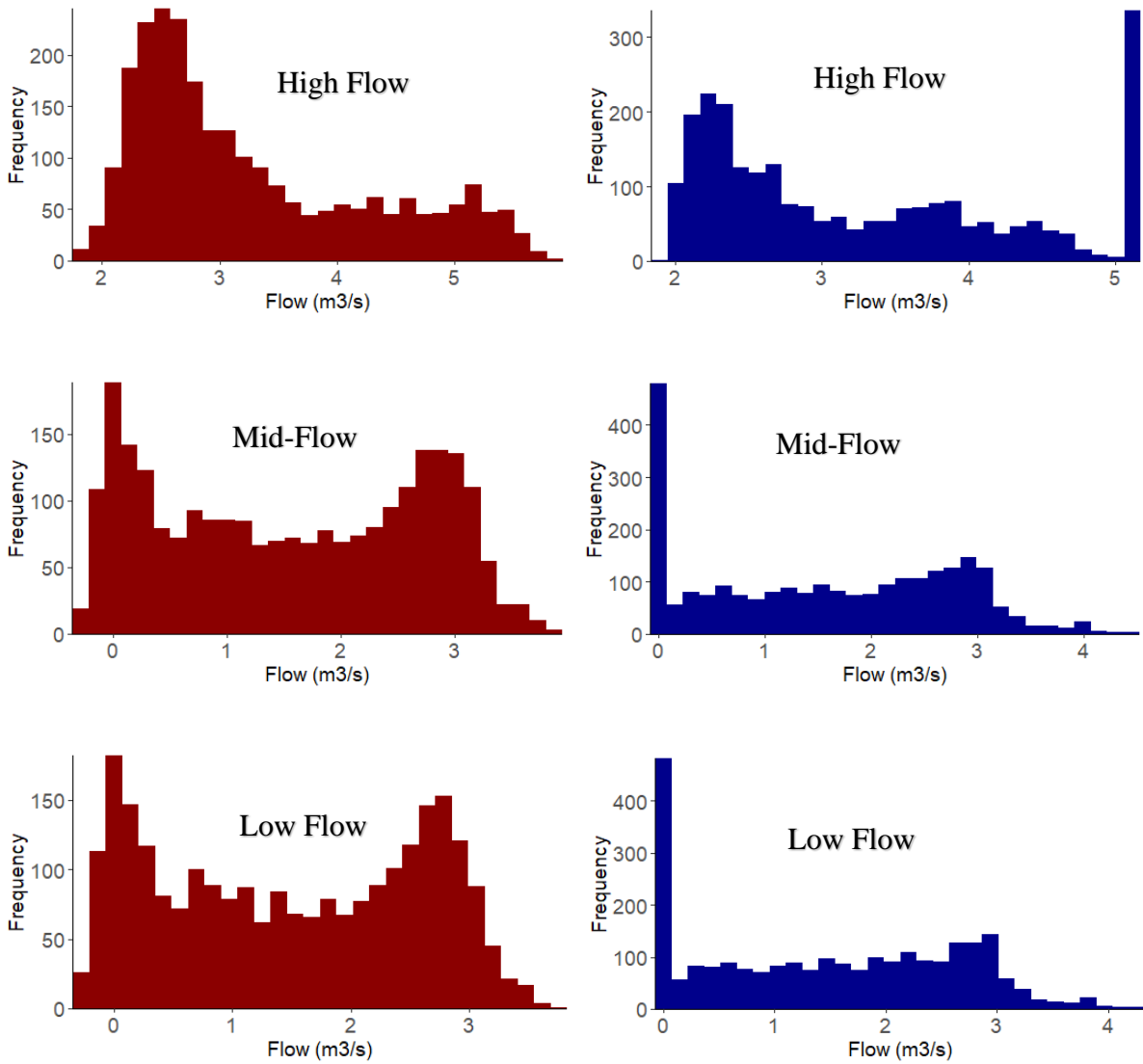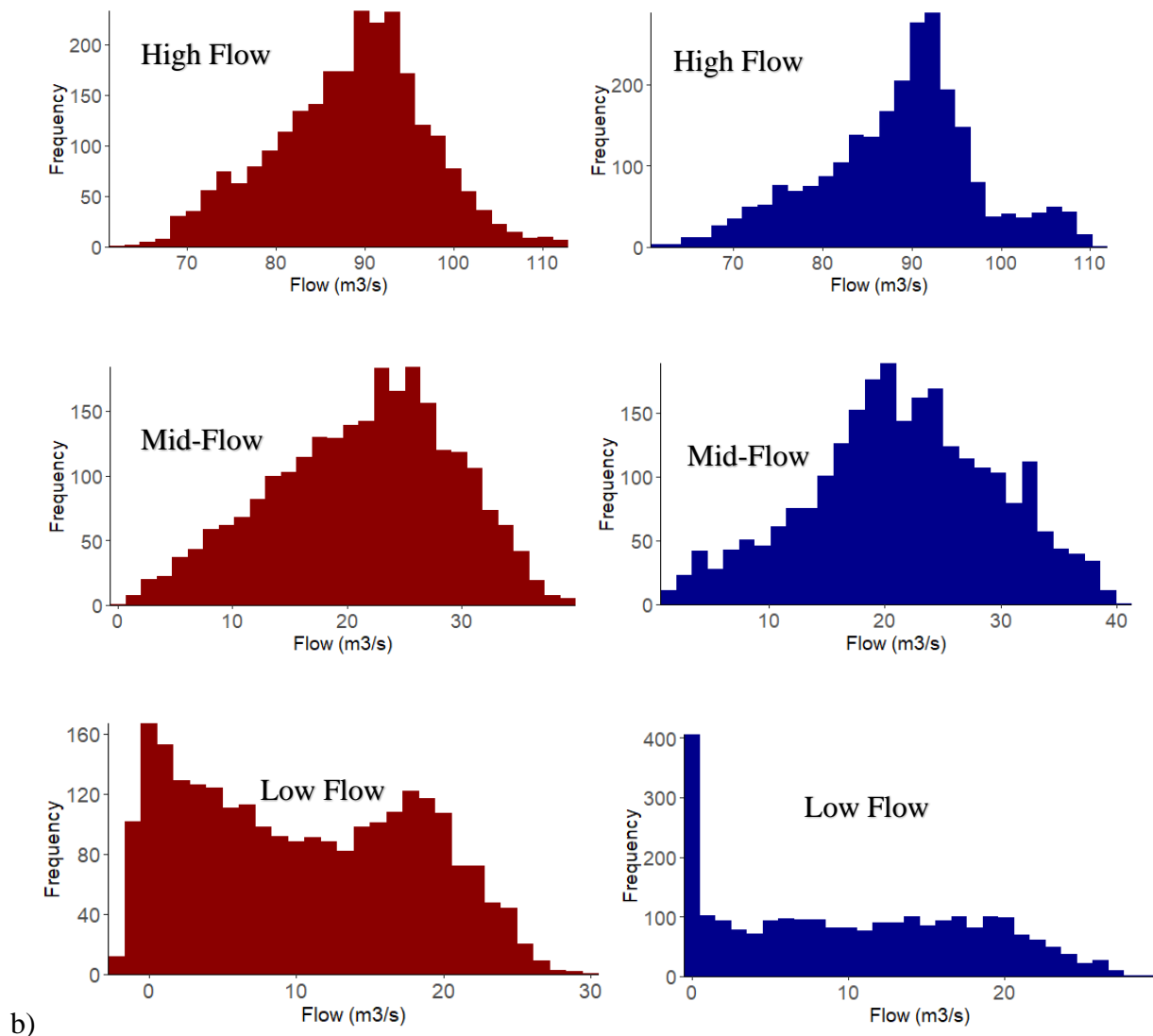
Figure 4-16: Histograms comparing LAR-PCE and MC simulations for a) Indian Creek watershed and b) Chehalis River watershed

4.3.3 Sensitivity Analysis using PCE

As described in the section 3.5.1, Sobol indices of any model can be directly found from the PCE coefficients built from the model. In this section, Sobol indices calculated from the OLS-PCE and LAR-PCE are used to find out the contribution of variability of each sensitive parameter to the variability of streamflow over low, median, and high flow days.

The first order indices in Figure 4-17 show a dominant dependence (>95%) of flow on the curve number (CN2) for the Indian Creek watershed. The results are consistent with both PCE metamodels, giving similar Sobol indices, representing that the curve number is a major factor in streamflow variability. On the other hand, for the Chehalis River watershed, the PCE models calculate that GWQMN, which is the threshold of water in deep aquifer, and RCHRG_DP, which is the deep aquifer percolation fraction, are the most sensitive parameters during high flow days. The flow has a major dependence on these two parameters, and their interactions with the other parameters, based on high total Sobol indices.

High flow day:

a) Sobol Indices based on OLS-PCE for Indian Creek
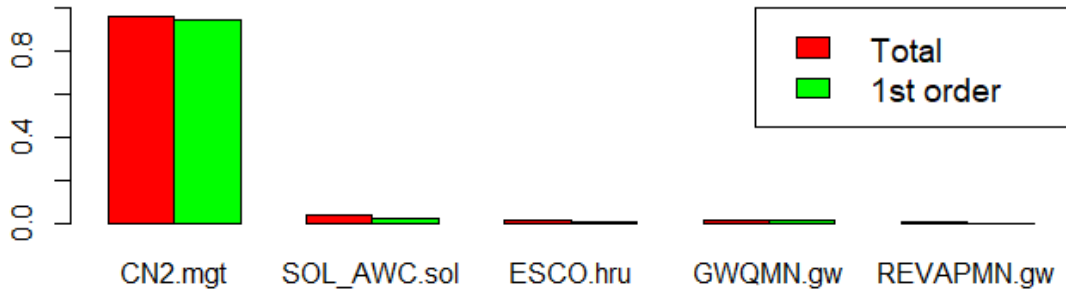
b) Sobol Indices based on LAR-PCE for Indian Creek
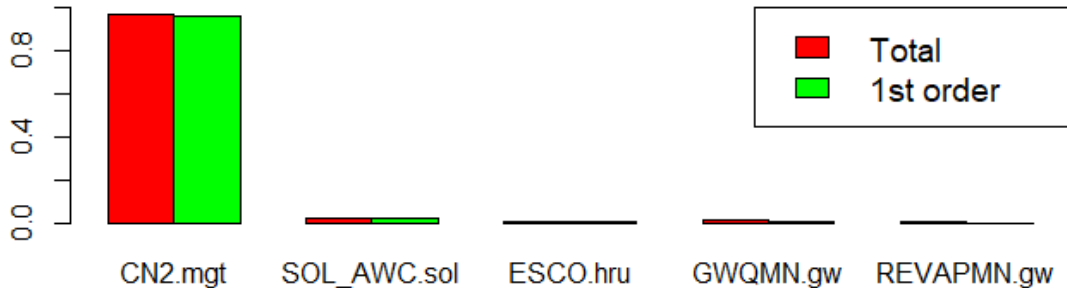
c) Sobol Indices based on OLS-PCE for Chehalis River

d) Sobol Indices based on LAR-PCE for Chehalis River
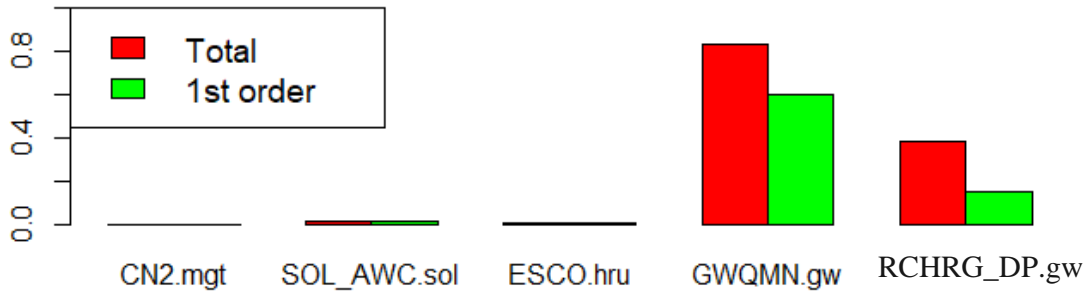
Figure 4-17: Sobol Indices corresponding to high flow days

The flow simulations on median flow day shown in Figure 4-18 give almost a similar pattern of dependencies of streamflow on the curve number for Indiana creek river; and GWQMN and RCHRG_DP for Chehalis River watershed. It can be concluded that the change of flow values from high to median, does not affect the sensitivity of parameters for both the watersheds.

Median flow day:

a) Sobol Indices based on OLS-PCE for Indian Creek
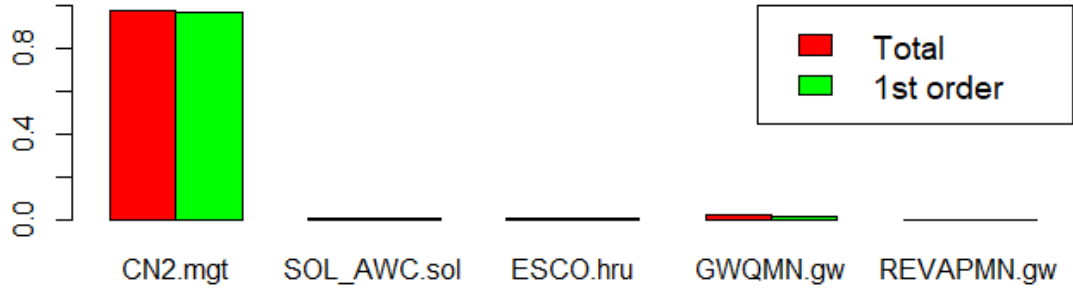
b) Sobol Indices based on LAR-PCE for Indian Creek
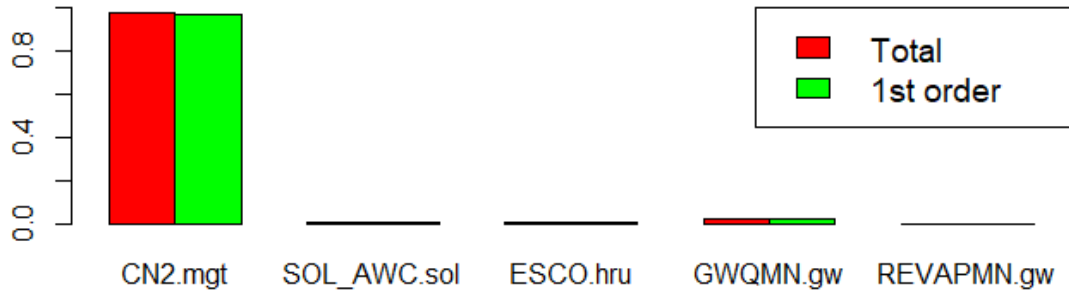
c) Sobol Indices based on OLS-PCE for Chehalis River
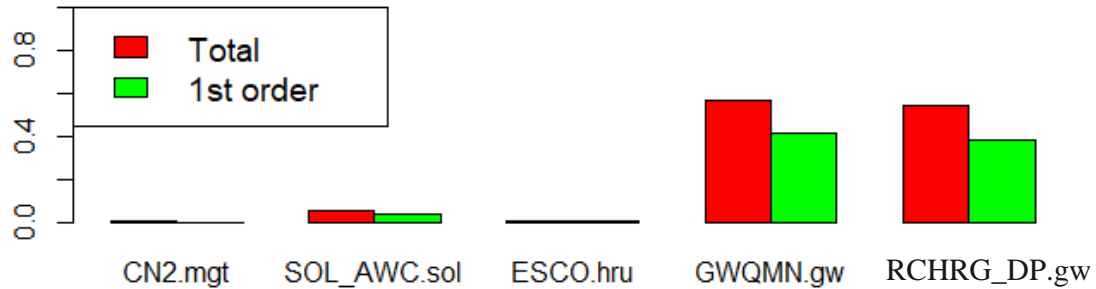
d) Sobol Indices based on LAR-PCE for Chehalis River
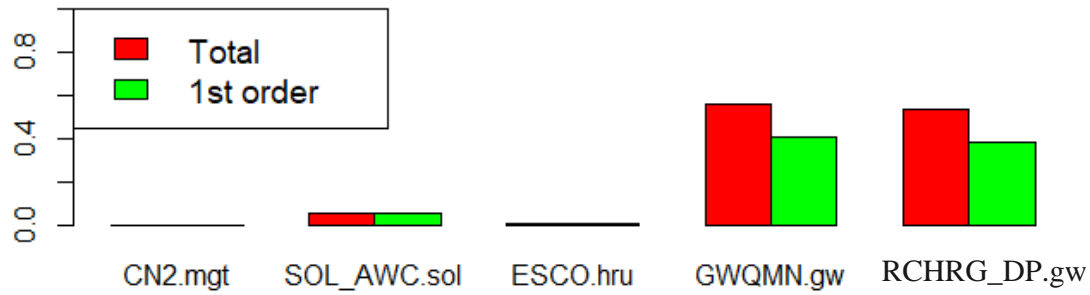
Figure 4-18: Sobol Indices corresponding to median flow days

In all flow conditions, curve number has the most impact on the streamflow of Indian Creek watershed. While it can be observed that the deep aquifer percolation fraction has the most significance amongst all parameters, during low flow days at the Chehalis River watershed, indicated in Figure 4-19.

Low flow day:

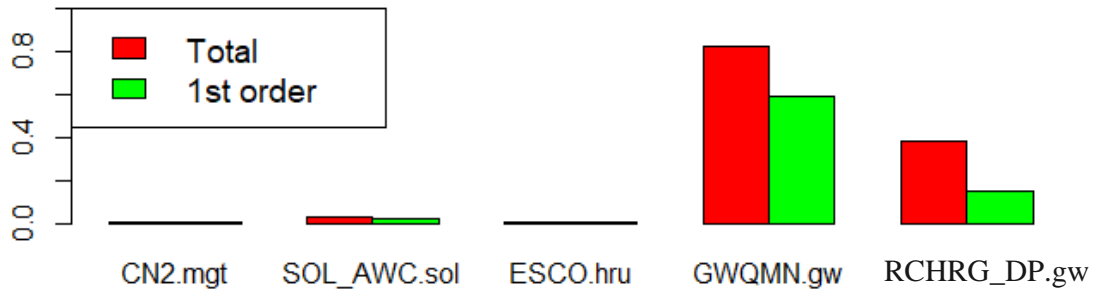a) Sobol Indices based on OLS-PCE for Indian Creek
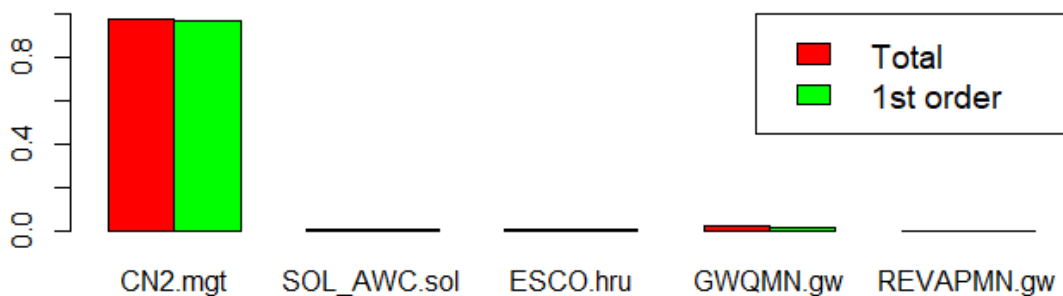
b) Sobol Indices based on LAR-PCE for Indian Creek
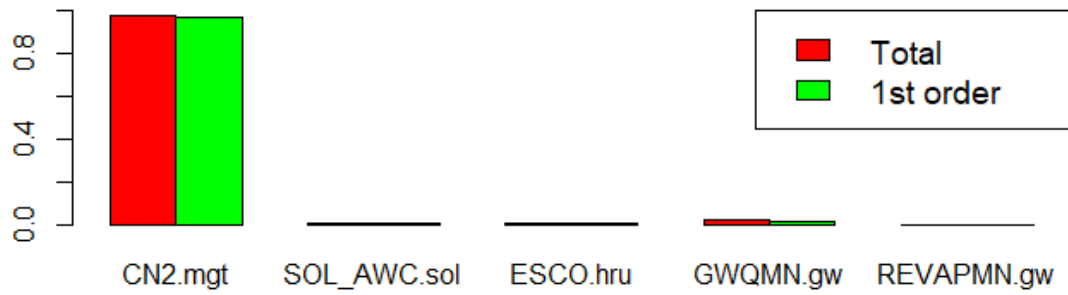
c) Sobol Indices based on OLS-PCE for Chehalis River
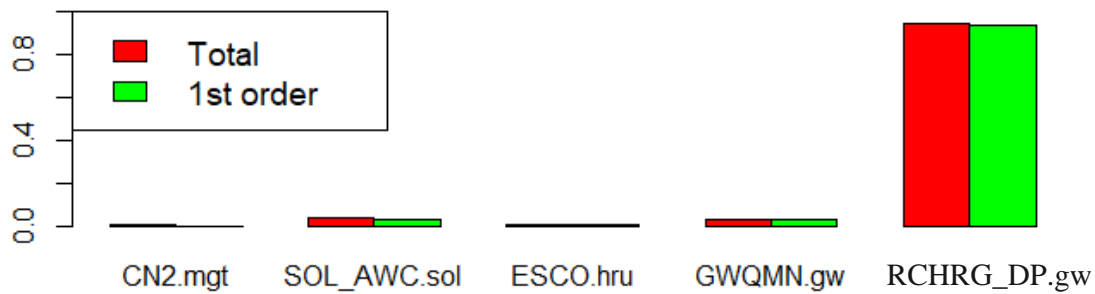
d) Sobol Indices based on LAR-PCE for Chehalis River
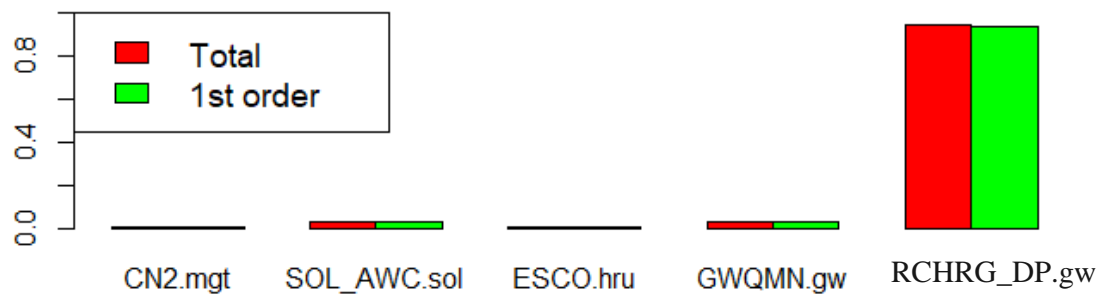
Figure 4-19: Sobol Indices corresponding to low flow days

It can be concluded that Polynomial Chaos Expansion (PCE) were efficiently able to simulate SWAT and HYMOD models. Prior and posterior distributions for HYMOD parameters were achieved by quantifying the uncertainty of it's surrogate LAR-PCE model. Obtaining the posterior

distributions of HYMOD parameters helped narrow down their range over which they are uncertain. LAR-PCE and OLS-PCE were rigorously validated and used to find out the sensitive parameters for SWAT models. It was found that curve number (CN2.mgt) is the major contributor towards uncertainty for the Indian creek watershed, irrespective of the quantity of flow. While for the Chehalis River Watershed, the major contribution is from the groundwater parameters.

CHAPTER 5: CONCLUSIONS AND

RECOMMENDATIONS

5.1 Conclusions

Hydrologic models conceptualise and combine the intricate, spatially distributed, and intricately entwined water, energy, and vegetation processes in a watershed into comparatively straightforward mathematical equations in order to simulate the rainfall-runoff processes. The model parameters' significant uncertainties, which have a big impact on hydrologic forecasts, are a significant result of process conceptualization. This study put forth a comprehensive framework for PCE-based hydrologic models to quantify uncertainty.

PCE was used with HYMOD, to simulate the hydrological model responses by a polynomial, for express calibration by running PCE tens of thousands of times for MCMC analysis, rather than running HYMOD again and again. The uncertainties and prior distribution associated with the HYMOD parameters were calculated in express pace by utilizing the PCE surrogate.

In order to efficiently build a surrogate model and quickly quantify its uncertainty for hydrological predictions, this study combined PCE with LAR. To the author's best knowledge, unlike prior studies this study is one of the first ones to use LAR-PCE with a semi-distributed model like SWAT.

As a result of learning and remembering only the most important polynomial basis terms, LAR produces a sparse set of PCE coefficients that can be more easily estimated. Comparing the performance of a surrogate model created using ordinary least square regression to the benefits of LAR-PCE (OLS). In order to quantify the uncertainties of hydrological models of a small and

medium scale watershed, the surrogate SWAT models OLS-PCE and LAR-PCE were created. The PCE models were then utilized to directly obtain Sobol indices for the sensitive SWAT parameters.

The key findings of this study are as follows:

1. Polynomial Chaos Expansion is an effective technique to surrogate a model, particularly if too many model runs are required or the model requires a long time to run.

2. Once a PCE surrogate model is built for a certain model, the Sobol indices of the input parameters of the model are automatically calculated.

3. LAR-PCE performs better than OLS-PCE using a fewer number of runs to calculate its coefficients, providing the same or a better magnitude of accuracy.

4. The application of PCE has been highlighted in this study to improve hydrological models, from their calibration, to forward uncertainty assessment by MC, or backward uncertainty assessment by MCMC, or sensitivity analysis by Sobol indices.

The contribution made by this thesis is to apply LAR-PCE to a complex semi-distributed hydrological model (SWAT) to quantify parameter uncertainty, which is not yet performed according to the author's best knowledge. A framework was proposed to effectively calibrate and quantify the uncertainty of a conceptual hydrological model, HYMOD using Markov Chain Monte Carlo. The framework can be later implemented to other complex models, which needs more data and has more parameter uncertainty.

5.2 Recommendations and Future Work

As the model complexity increases, it is anticipated that PCE's advantage in terms of computational efficiency will become more significantly advantageous. The PCE should be evaluated in the

future for additional intricate hydrological models and for a greater variety of watersheds. Applying PCE directly to field data could also be a solution to real-time forecasting and prediction scenarios. One of the limitations that PCE faces is that the parameters are assumed to be independent. Future studies are necessary in this field to address these issues.

REFERENCES

Altenford, C. T., 1992. Using a neural network for soil moisture predictions. ASAE Microfiche No. 92-3557. St. Joseph, MI: ASAE.

Ang, A. H.-S., and Tang, W. H. 1984. Probability Concepts in Engineering Planning and Design, Vol. 2 Decision, Risk, and Reliability, Illustrated Edition. John Wiley and Sons Inc., New York

Arnell, N.W., 2011. Uncertainty in the relationship between climate forcing and hydrological response in UK catchments. Hydrological Earth System Sciences15, 897–912. https://doi.org/10.5194/hess15-897-2011

Arnold, J G, and P M Allen. 1996. Estimating hydrologic budgets for three Illinois watersheds. Journal of Hydrology 176, 57-77.

Arnold, J. G., Neitsch, S. L. and Williams, J. R., 1999. Soil and Water Assessment Tool, User's Manual Version 98.1. Temple Texas:: USDA-ARS Grassland Soil and Water Research.

Arnold, J. G. and Williams, J. R., 1987. Validation of SWRRB: simulator for water resources in rural basins. Journal of Water Resources Planning and Management. ASCE, 113(2), pp. 243-256.

Aryal, A., Shrestha, S. and Babel, M.S. 2019. Quantifying the sources of uncertainty in an ensemble of hydrological climate-impact projections. Theoretical and Applied Climatology, **135,** 193–209. https://doi.org/10.1007/s00704-017-2359-3

Athira, P., Sudheer, K.P., 2015. A method to reduce the computational requirement while assessing uncertainty of complex hydrological models. Stochastic Environment Research and Risk Assessment, 29, 847–859. https://doi.org/10.1007/s00477-014-0958-4

Beven, K., and Binley, A. 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3), 279–298. https://doi.org/10.1002/HYP.3360060305

Beven, K., 2019. How to make advances in hydrological modelling. *Hydrology Research* 1 December 2019; 50 (6): 1481–1494. doi: https://doi.org/10.2166/nh.2019.134

Beven, K., 2016. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrological Sciences Journal. 61, 1652–1665. https://doi.org/10.1080/02626667.2015.1031761

Beven, K., and Freer, J. 2001. "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology." Journal of hydrology, 249(1), 11–29.

Beven, K. (1989). Changing ideas in hydrology—the case of physically-based models. Journal of hydrology, 105(1-2), 157-172.

Beven, K., 2001. On hypothesis testing in hydrology. Hydrological Processes, 15 (9), 1655-1657.

Beven, K. J., Calver, A. and Morris, E. M., 1987. The Institute of Hydrology Distributed Model, Institute of Hydrology, Report No. 98. Wallingford.

Bieker, H. P., Slupphaug, O., and T. A. Johansen. "Real-Time Production Optimization of Oil and Gas Production Systems: A Technology Survey." SPE Production and Operation 22 (2007): 382–391.

Blanning, R. W. 1975, The construction and implementation of metamodels, Simulation, 24(6), 177–184.

Blatman, G., and Sudret, B. 2008. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. Comptes Rendus Mécanique, 336(6), 518–523. https://doi.org/10.1016/J.CRME.2008.02.013

Blatman, G., and Sudret, B. 2010. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. Probabilistic Engineering Mechanics, 25(2), 183–197. https://doi.org/10.1016/J.PROBENGMECH.2009.10.003

Blatman, G., and Sudret, B. 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. Journal of Computational Physics, 230(6), 2345–2367. https://doi.org/10.1016/J.JCP.2010.12.021

Boyle, D. P. 2001. "Multicriteria Calibration of Hydrologic Models." PhD Thesis, The University of Arizona.

Charles C. Daniel, Douglas G. Smith, and Jo Leslie Eimers, 1997, Hydrogeology and simulation of ground-water flow in the thick regolith-fractured crystalline rock aquifer system of Indian Creek basin, North Carolina: U.S. Geological Survey Water Supply Paper 2341-C.

Chegwidden, O.S., Nijssen, B., Rupp, D.E., Arnold, J.R., Clark, M.P., Hamman, J.J., Kao, S.-C., Mao, Y., Mizukami, N., Mote, P. W., Pan, M., Pytlak, E. and Xiao, M. 2019. How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates. Earth's Future, 7, 623–637.

Chen, J. Eeden, C., and Zidek, J. 2010. Uncertainty and the conditional variance. Statistics and Probability Letters. 80. 1764-1770. 10.1016/j.spl.2010.07.021.

Chilkoti, V. 2019. Impacts of climate change on hydropower generation and developing adaptation measures through hydrologic modeling and multi-objective optimization, Electronic Theses and Dissertations.7691. https://scholar.uwindsor.ca/etd/7691

Chilkoti, V., Bolisetti, T., and Balachandar, R. 2017. Climate change impact assessment on hydropower generation using multi-model climate ensemble. Renewable Energy, 109, 510517.

Cintra, R. S., and De Campos Velho, H. F. 2018. Data Assimilation by Artificial Neural Networks for an Atmospheric General Circulation Model: Conventional Observation. Advanced Applications for Artificial Neural Networks. doi: 10.5772/intechopen.70791

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., and Schaibly, J. H. 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. The Journal of Chemical Physics, 59(8), 3873–3878.

Daniell, T. M., 1991. Neural networks - Applications in hydrology and water resources engineering. Proceedings, International Hydrology and Water Symposium, 3, 797-802. National Conference Publication 91/22, Institute of Engineering, Australia, Barton, ACT, Australia.

De Vos, N. J., T. H. M. Rientjes, and H. V. Gupta. 2010. "Diagnostic evaluation of conceptual rainfall: Runoff models." Hydrological Processes 24 (20): 2840–2850. https://doi.org/10.1002/hyp.7698.

Demissie, Y. K., A. J. Valocchi, B. S. Minsker, and B. A. Bailey. 2009, Integrating a calibrated groundwater flow model with error-correcting data-driven models to improve predictions, Journal of Hydrology, 364(3), 257–271.

Deser, C., Phillips, A., Bourdette, V., and Teng, H. 2012. Uncertainty in climate change projections: the role of internal variability. Climate Dynamics, 38(3-4), 527–546.

Devia, G. K., Ganasri, B. P., and Dwarakish, G. S. (2015). A review on hydrological models. Aquatic procedia, 4, 1001-1007.

DHI-WE, 2005. MIKE SHE Water Movement: User manual, Danish Hydraulic Institute – Water and Environment, Hørsholm, Denmark.

Dobler, C., Hagemann, S., Wilby, R. L., and Stötter, J. 2012. Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed. Hydrology and Earth System Sciences, 16(11), 4343–4360.

Doherty, J., and S. Christensen. 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, Water Resources Research, 47, W12534, doi:10.1029/2011WR010763.

Dos Santos, R.F., Lu, C.-T., 2015. Geography Markup Language (GML). Springer International Publishing, pp. 1–6. https://doi.org/10.1007/978-3319-23519-6_480-2

Diersch, H. 2005. FEFLOW finite element subsurface flow and transport simulation system, inReference Manual, 292 pp., WASY GmbHInst. Water Resources Planning Syst. Res., Berlin

Dwelle, M. C., Kim, J., Sargsyan, K., and Ivanov, V. Y. 2019. Streamflow, stomata, and soil pits: Sources of inference for complex models with fast, robust uncertainty quantification. Advances in Water Resources, 125, 13–31. https://doi.org/10.1016/J.ADVWATRES.2019.01.002

Ehteram, M., F. B. Othman, Z. M. Yaseen, H. A. Afan, M. F. Allawi, M. B. A. Malek, A. N. Ahmed, S. Shahid, V. P. Singh, and A. El-Shafie. 2018. Improving the Muskingum flood routing

method using a hybrid of particle swarm optimization and bat algorithm. Water 10 (6): 1–21. https://doi.org/10.3390/w10060807.

Eckhardt, K., Haverkamp, S., Fohrer, N., Frede, H.-G., 2002b. SWAT-G, a version of SWAT99.2 modified for application to low mountain range catchments. Physics and Chemistry of the Earth, in press.

Fan, Y., Huang, G., and Veawab, A. 2012. A generalized fuzzy linear programming approach for environmental management problem under uncertainty. Journal of the Air and Waste Management Association , 62(1), 72–86. https://doi.org/10.1080/10473289.2011.628901

Fan, Y. R., G. H. Huang, B. W. Baetz, Y. P. Li, K. Huang, Z. Li, X. Chen, and L. H. Xiong. 2016. "Parameter uncertainty and temporal dynamics of sensitivity for hydrologic models: A hybrid sequential data assimilation and probabilistic collocation method." Environmental Modelling Software. 86 (Oct): 30–49. https://doi.org/10.1016/j.envsoft.2016.09.012.

Fan, Y. R., Huang, W., Huang, G. H., Huang, K., and Zhou, X. 2014. A PCM-based stochastic hydrological model for uncertainty quantification in watershed systems. Stochastic Environmental Research and Risk Assessment, 29(3), 915–927. https://doi.org/10.1007/s00477-014-0954-8

Farzin, S., V. P. Singh, H. Karami, N. Farahani, M. Ehteram, O. Kisi, M. F. Allawi, N. S. Mohd, and A. El-Shafie. 2018. "Flood routing in river reaches using a three-parameter Muskingum model coupled with an improved bat algorithm." Water 10 (9): 1130. https://doi.org/10.3390/w10091130.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., ... and Tarboton, D. 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology, 537, 45-60.

Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., and Rigon, R. 2011. "The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: models of flow generation and propagation/routing." Geoscientific Model Development, 4(4), 943–955.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology: A Journal of the Royal Meteorological Society, 27(12), 1547-1578.

French, M., Krajewski, W. F. and Cuykendall, R. R., 1992. Rainfall forecasting in space and time using a neural network. Journal of Hydrology, 137, 1-31.

Gendaszek, A.S., and Welch, W.B., 2018. Water budget of the upper Chehalis River Basin, southwestern Washington: U.S. Geological Survey Scientific Investigations Report 2018-5084, 17 p., https://doi.org/10.3133/sir20185084.

Gendaszek, A.S., 2011. Hydrogeologic framework and groundwater/surface-water interactions of the Chehalis River basin, Washington: U.S. Geological Survey Scientific Investigations Report 2011-5160, 42 p.

Box, G. E. P. 1976 Science and Statistics, Journal of the American Statistical Association, 71:356, 791-799, DOI: 10.1080/01621459.1976.10480949

Ghaith, M., and Li, Z. 2020. Propagation of parameter uncertainty in SWAT: A probabilistic forecasting method based on polynomial chaos expansion and machine learning. Journal of Hydrology, 586, 124854. https://doi.org/10.1016/J.JHYDROL.2020.124854

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G. 2013. An approach to identify time consistent model parameters: sub-period calibration. Hydrology and Earth System Sciences, 17(1), 149–16

Harbaugh, A. W. 2005. MODFLOW-2005, the US geological survey modular ground-water model: The ground-water flow process, U.S. Geological Survey Technological Methods, 6-A16, 1–99

Hemker, T., K. R. Fowler, M. W. Farthing, and O. von Stryk. 2008, A mixed-integer simulation-based optimization approach with surrogate functions in water resources management, Optical Engineering, 9(4), 341–360.

Her, Y., Yoo, SH., Cho, J. et al. 2019. Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. Scientific Reports **9,** 4974. https://doi-org.ledproxy2.uwindsor.ca/10.1038/s41598-019-41334-7

Herman, J.D., Reed, P.M., Wagener, T., 2013. Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. Water Resources Research 49, 1400–1414. https://doi.org/10.1002/wrcr.20124

Hu, J., Chen, S., Behrangi, A., and Yuan, H. 2019. Parametric uncertainty assessment in hydrological modeling using the generalized polynomial chaos expansion. Journal of Hydrology, 579, 124158. https://doi.org/10.1016/J.JHYDROL.2019.124158

Huang, S., Mahadevan, S., and Rebba, R. 2007. Collocation-based stochastic finite element analysis for random field problems. Probabilistic Engineering Mechanics, 22(2), 194–205. https://doi.org/10.1016/J.PROBENGMECH.2006.11.004

Jajarmizadeh, M., Harun, S., and Salarpour, M. (2012). A review on theoretical consideration and types of models in hydrology. Journal of Environmental Science and Technology, 5(5), 249-261.

Jayakrishnan, R, R Srinivasan, C Santhi, and J. G Arnold. 2005. "Advances in the application of the SWAT model for waters resources management." Hydological Processes, 19: 749-762.

Arnold J. G., Moriasi D. N., Gassman P. W., Abbaspour K. C., White M. J., Srinivasan R., Santhi C., Harmel R. D., van Griensven A., Van Liew M. W., Kannan N., Jha M. K., 2012. SWAT: Model use, Calibration, and Validation. Transactions of the ASABE. 55(4): 1491-1508. (doi: 10.13031/2013.42256)

Jie Chen, François P. Brissette, Robert Leconte, Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, Journal of Hydrology, Volume 401, Issues 3–4, 2011, Pages 190-202, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2011.02.020.

Jothiprakash, V., and A. S. Kote. 2011. "Improving the performance of data-driven techniques through data pre-processing for modelling daily reservoir inflow." Hydrological Sciences Journal 56 (1): 168–186. https://doi.org/10.1080/02626667.2010.546358.

Karkee, M., Steward, B.L., 2010. Local and global sensitivity analysis of a tractor and single axle grain cart dynamic system model. Biosystems Engineeering 106, 352366. https://doi.org/10.1016/j.biosystemseng.2010.04.006

Karlsson, I.B., Sonnenborg, T.O., Refsgaard, J.C., Trolle, D., Børgesen, C.D., Olesen, J.E., Jeppesen, E., Jensen, K.H., 2016. Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change. Journal of Hydrology 535, 301–317. https://doi.org/10.1016/j.jhydrol.2016.01.069

Kavetski, D., and G. Kuczera, 2007, Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration, Water Resources Research,43, W03411, doi:10.1029/2006WR005195.

Keating, E. H., J. Doherty, J. A. Vrugt, and Q. J. Kang, 2010, Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality, Water Resources Research,46, W10517, doi:10.1029/2009WR008584.

Khatun, S., Sahana, M., Jain, S. K., and Jain, N. 2018. Simulation of surface runoff using semi distributed hydrological model for a part of Satluj Basin: parameterization and global sensitivity analysis using SWAT CUP. Modeling Earth Systems and Environment, 4(3), 1111–1124. https://doi.org/10.1007/S40808-018-0474-5/TABLES/8

Krysanova, V., Hattermann, F. and Wechsung, F. 2005, Development of the ecohydrological model SWIM for regional impact studies and vulnerability assessment. Hydrological Processes, 19: 763-783. https://doi-org.ledproxy2.uwindsor.ca/10.1002/hyp.5619

Li, Z., Shao, Q., Xu, Z., and Cai, X. 2010. Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China. Journal of Hydrology, 385(1–4), 76–83. https://doi.org/10.1016/J.JHYDROL.2010.01.025

Li, Z., Xu, Z., Shao, Q., Yang, J., 2009. Parameter estimation and uncertainty analysis of SWAT model in upper reaches of the Heihe river basin. Hydrological Processes. 23, 2744–2753. https://doi.org/10.1002/hyp.7371

Liu, B., Zhang, Q., and Gielen, G. G. E. 2014. A gaussian process surrogate model assisted evolutionary algorithm for medium scale expensive optimization problems. IEEE Transactions on Evolutionary Computation, 18(2), 180–192. https://doi.org/10.1109/TEVC.2013.2248012

Liu, Y., and Gupta, H. V. 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research, 43(7), 7401. https://doi.org/10.1029/2006WR005756

López-Cruz, I.L., Rojano-Aguilar, A., Salazar-Moreno, R., Ruiz-García, A., Goddard, J., 2012. A comparison of local and global sensitivity analyses for greenhouse crop models. Acta Horticulturae. 957, 267–273. https://doi.org/10.17660/ActaHortic.2012.957.30

Maier, H. R. and Dandy, G. C., 1997. Author's reply to comments by Fortin, V., Ouarda, T.B.M.J. and Bobee, B. on "The use of artificial neural networks for the prediction of water quality parameters" by Maier, H.R. and Dandy, G.C. Water Resources Research, 33 (10), 2425-2427.

Matott, L. S., and Rabideau, A. J. 2008, Calibration of complex subsurface reaction models using a surrogate-model approach, Advanced Water Resources,31(12), 1697–1707.

Mike Herrmann, 2008, Indian Creek and Howard's Creek Local Watershed Plan: Preliminary Findings Report

Minns, A. W. and Hall, M. J., 1996. Artificial neural networks as rainfall-runoff models. Hydrological Sciences Journal, 41, 399-417.

Mishra, S. 2009. Uncertainty and sensitivity analysis techniques for hydrologic modeling. Journal of Hydroinformatics. https://doi.org/10.2166/hydro.2009.048

Moges E, Demissie Y, Larsen L, Yassin F. Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. Water 2021; 13(1):28. https://doi-org.ledproxy2.uwindsor.ca/10.3390/w13010028

Montanari, A. 2005. "Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations." Water Resources Research, 41(8).

Montáns, F.J., Chinesta, F., Gómez-Bombarelli, R., Kutz, J.N., 2019. Data-driven modeling and learning in science and engineering. Comptes Rendus - Mec. 347, 845–855. https://doi.org/10.1016/j.crme.2019.11.009

Moore, Robert. (1985). The Probability-Distributed Principle and Runoff Production at Point and Basin Scales. Hydrological Sciences Journal. 30. 10.1080/02626668509490989.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. 2005. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources, 28(2), 135–147. https://doi.org/10.1016/J.ADVWATRES.2004.09.002

Mugunthan, P., Shoemaker C. A., and Regis R. G. 2005, Comparison of function approximation, heuristic, and derivative-based methods for automatic calibration of computationally expensive groundwater bioremediation models, Water Resources Research, 41, W11427, doi:10.1029/2005WR004134

Razavi, S., Tolson B. A., and Burn D. H. 2012a, Review of surrogate modeling in water resources, Water Resources Research.,48, W07401, doi:10.1029/2011WR011527.

Refsgaard, J. C., Storm, B. and Refsgaard, A. 1995. Recent developments of the Système Hydrologique Européen (SHE) towards the MIKE SHE. In: Modelling and Management of Sustainable Basin-Scale Water Resource Systems Boulder Symp., July 1995 (ed. by S. P. Simonovic, Z. Kundzewicz, D. Rosbjerg and K. Takeuchi), 425-434. IAHS Publ.no. 231.

Roach, J., and V. Tidwell. 2009, A compartmental-spatial system dynamics approach to ground water modeling, Groundwater,47(5), 686–698, doi:10.1111/j.1745-6584.2009.00580.x.

Robinson T. D., Eldred M. S., Willcox K.E, and Haimes. R, Surrogate-Based Optimization Using Multifidelity Models with Variable Parameterization and Corrected Space Mapping. American Institute of Aeronautics and Astronautics Journal 2008 46:11, 2814-2822

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. 2004. Sensitivity analysis in practice: a guide to assessing scientific models. Chichester, England.

Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B. J., Ricciuto, D., and Thornton, P. 2014. Dimensionality reduction for complex models via bayesian compressive sensing. International Journal for Uncertainty Quantification, 4(1), 63–93. https://doi.org/10.1615/INT.J.UNCERTAINTYQUANTIFICATION.2013006821

Scire, J.J., Dryer, F.L., Yetter, R.A., 2001. Comparison of global and local sensitivity techniques for rate constants determined using complex reaction mechanisms. International Journal of Chemical Kinetics. 33, 784–802.

Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall-runoff modeling. Journal of Hydrology (Amsterdam), 199 (3/4), 272-294.

Shi, L., Yang, J., Zhang, D., and Li, H. 1996. Probabilistic collocation method for unconfined flow in heterogeneous media. Journal of Hydrology, 365, 4–10. https://doi.org/10.1016/j.jhydrol.2008.11.012

Sikorska, A. E., Montanari, A., and Koutsoyiannis, D. 2014. "Estimating the uncertainty of hydrological predictions through data-driven resampling techniques." Journal of Hydrologic Engineering, 20(1), A4014009.

Singh, V., Bankar, N., Salunkhe, S.S., Bera, A.K., Sharma, J.R., 2013. Hydrological stream flow modelling on Tungabhadra catchment: parameterization and uncertainty analysis using SWAT CUP. Current Science. 104, 1187–1199.

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling : some past experiences and new approaches. Journal of Hydroinformatics 10, 3–22. https://doi.org/10.2166/hydro.2008.015

Stojkovic, M., and Simonovic, S. P. 2020. Understanding the uncertainty of the lim river basin response to changing climate. Journal of Hydrologic Engineering, 25(9), 05020023.

Sudret, Bruno. 2007. Uncertainty propagation and sensitivity analysis in mechanical models -- Contributions to structural reliability and stochastic spectral methods.

Sudret, B. 2008. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering and System Safety, 93(7), 964–979. https://doi.org/10.1016/J.RESS.2007.04.002

Talebizadeh, M., Morid, S., Ayyoubzadeh, S.A., Ghasemzadeh, M., 2010. Uncertainty analysis in sediment load modeling using ANN and SWAT model. Water Resource Management. 24, 1747–1761. https://doi.org/10.1007/s11269-009-9522-2

Taormina, R., Chau, K.W., 2015. Data-driven input variable selection for rainfall runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. Journal of Hydrology. 529, 1617–1632. https://doi.org/10.1016/j.jhydrol.2015.08.022

Tarek, Mostafa and Brissette, François and Arsenault, Richard. (2021). Uncertainty of gridded precipitation and temperature reference datasets in climate change impact studies. Hydrology and Earth System Sciences. 25. 3331-3350. 10.5194/hess-25-3331-2021

Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models. Water Resources Research, 37(10), 2521–2535. https://doi.org/10.1029/2000WR900405

Tolessa, O., Nossent, J., Velez, C., Kumar, N., Griensven, A. Van, Bauwens, W., 2015. Environmental Modelling and Software Assessment of the different sources of uncertainty in a SWAT model of the River Senne (Belgium). Environmental Modeling Software. 68, 129–146. https://doi.org/10.1016/j.envsoft.2015.02.010

Torre, E., Marelli, S., Embrechts, P., and Sudret, B. 2018. Data-driven polynomial chaos expansion for machine learning regression. Journal of Computational Physics, 388, 601–623. https://doi.org/10.1016/j.jcp.2019.03.039

Tran, V. N., and Kim, J. 2019. Quantification of predictive uncertainty with a metamodel: toward more efficient hydrologic simulations. Stochastic Environmental Research and Risk Assessment 2019 33:7, 33(7), 1453–1476. https://doi.org/10.1007/S00477-019-01703-0

Tran, V. N., and Kim, J. 2021. A robust surrogate data assimilation approach to real-time forecasting using polynomial chaos expansion. Journal of Hydrology, 598, 126367. https://doi.org/10.1016/J.JHYDROL.2021.126367

Van Griensven, A. and Bauwens, W. 2001. Integrating modeling of catchments, Water Science and Technology, 43, 7, 321-328.

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., Koch, H., Pechlivanidis, I. G., Plötner, S., Seidou, O., et al. 2017: Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins, Climatic Change, 141, 419–433, 2017.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W. and Verstraten, J. M. 2005 Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. Water Resour. Res. 41, W01017, doi:10.1029/2004WR003059.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research, 39(8), 1201. https://doi.org/10.1029/2002WR001642

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., and Robinson, B. A. 2009. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stochastic Environmental Research and Risk Assessment, 23(7), 1011–1026. https://doi.org/10.1007/S00477-008-0274-Y/TABLES/3

Vuik, C., A. Segal, and J. Meijerink 1999, An efficient preconditioned cg method for the solution of a class of layered problems with extreme contrasts in the coefficients, Journal of Computational Physics,152(1), 385–403.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S. 2001. "A framework for development and application of hydrological models." Hydrology and Earth System Sciences Discussions, 5(1), 13–26.

Wang, H.-M., Chen, J., Xu, C.-Y., Zhang, J., Chen, H. 2020. A framework to quantify the uncertainty contribution of GCMs over multiple sources in hydrological impacts of climate change. Earth'sFuture, 8,e2020EF001602. https://doi-org.ledproxy2.uwindsor.ca/10.1029/2020EF001602

Wang, S., Huang, G. H., Baetz, B. W., and Ancell, B. C. 2017. Towards robust quantification and reduction of uncertainty in hydrologic predictions: Integration of particle Markov chain Monte Carlo and factorial polynomial chaos expansion. Journal of Hydrology, 548, 484–497. https://doi.org/10.1016/j.jhydrol.2017.03.027

Wang, S., Huang, G. H., Baetz, B. W., and Huang, W. 2015. A polynomial chaos ensemble hydrologic prediction system for efficient parameter inference and robust uncertainty assessment. Journal of Hydrology, 530, 716–733. https://doi.org/10.1016/J.JHYDROL.2015.10.021

Wang, Y., S. Guo, L. Xiong, P. Liu, and D. Liu. 2015. "Daily runoff fore- casting model based on ANN and data preprocessing techniques." Water 7 (8): 4144–4160. https://doi.org/10.3390/w7084144.

Watson, T. A., Doherty J.E., and Christensen S. 2013, Parameter and predictive outcomes of model simplification, Water Resources Research,49,3952–3977, doi:10.1002/wrcr.20145.

Weinan, E., and Engquist. B. 2003, Multiscale modeling and computation, Notices of the American Mathematical Society, 50(9), 1062–1070.

White, K. L., and Chaubey, I. (2005). Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model 1. JAWRA Journal of the American Water Resources Association, 41(5), 1077-1089.

White, M. J., Gambone, M., Haney, E., Arnold, J., and Gao, J. 2017. Development of a Station Based Climate Database for SWAT and APEX Assessments in the US. Water 2017, Vol. 9, Page 437, 9(6), 437. https://doi.org/10.3390/W9060437

Wiener, N. 1938. The Homogeneous Chaos. American Journal of Mathematics, 60(4), 897. https://doi.org/10.2307/2371268

Wilby, R.L., Harris, I., 2006. A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. Water Resour. Res. 42, doi:10.1029/2005WR004065. https://doi.org/10.1029/2005WR004065

Willcox, K., and Peraire, J. 2002. Balanced model reduction via the proper orthogonal decomposition, American Institute of Aeronautics and Astronautics.,40(11), 2323–2330.

Wu, C.L., Chau, K.W., 2010. Data-driven models for monthly streamflow time series prediction. Engineering Applications of Artificial Intelligence 23, 1350–1367. https://doi.org/10.1016/j.engappai.2010.04.003

Wu, Y., Liu, S., 2012. Environmental Modelling and Software Automating calibration, sensitivity and uncertainty analysis of complex models using the R package Flexible Modeling Environment (FME): SWAT as an example. Environmental Modelling and Software 31, 99–109. https://doi.org/10.1016/j.envsoft.2011.11.013

Xiao, D., Ferlauto, M., Song, L., and Li, J. 2021. Multi-fidelity sparse polynomial chaos expansion based on Gaussian process regression and least angle regression. Journal of Physics: Conference Series, 1730(1), 012091. https://doi.org/10.1088/1742-6596/1730/1/012091

Xie, K., Liu, P., Zhang, J., Libera, D. A., Wang, G., Li, Z., and Wang, D. 2020. Verification of a New Spatial Distribution Function of Soil Water Storage Capacity Using Conceptual and SWAT Models. Journal of Hydrologic Engineering, 25(3), 04020001. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001887

Xie, X., Schenkendorf, R., and Krewer, U. 2017. Robust Design of Chemical Processes Based on a One-Shot Sparse Polynomial Chaos Expansion Concept. Computer Aided Chemical Engineering, 40, 613–618. https://doi.org/10.1016/B978-0-444-63965-3.50104-5

Xiu, D. 2010. Numerical methods for stochastic computations. In Numerical Methods for Stochastic Computations. Princeton university press.

Xiu, D., and Karniadakis, G. E. 2002. The Wiener-Askey polynomial chaos for stochastic differential equations. Society for Industrial and Applied Mathematics, 24(2), 619–644.

Xu, C. Y. (1999). Climate change and hydrologic models: A review of existing gaps and recent research developments. Water Resources Management, 13(5), 369-382.

Xu, T., A. J. Valocchi, J. Choi, and Amir, E. 2012, Improving groundwater flow model prediction using complementary data-driven models, paper presented at 19th International Conference on Water Resources, Univ. of Ill., Urbana-Champaign.

Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology. 358, 1–23. https://doi.org/10.1016/j.jhydrol.2008.05.012

Young, P. C., and Ratto, M. 2011. Statistical emulation of large linear dynamic models, Technometrics,53(1), 29–43, doi:10.1198/TECH.2010.07151.

Zeroual, A., Meddi, M., Assani, A.A. 2016. Artificial neural network rainfall discharge model assessment under rating curve uncertainty and monthly mischarge volume predictions. Water Resources Management. 3191–3205. https://doi.org/10.1007/s11269-016-1340-8

Zhang, D., Chen, X., and Yao, H. 2016. SWAT-CS (enm): Enhancing SWAT nitrate module for a Canadian Shield catchment. The Science of the Total Environment, 550, 598–610. https://doi.org/10.1016/J.SCITOTENV.2016.01.109

Zhang, D., Chen, X., Yao, H., James, A., 2016. Moving SWAT model calibration and uncertainty analysis to an enterprise Hadoop-based cloud. Environmental Modelling and Software. 84, 140–148. https://doi.org/10.1016/j.envsoft.2016.06.024

Zhang, X., Srinivasan, R., Bosch, D., 2009. Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. Journal of Hydrology. 374, 307–317. https://doi.org/10.1016/j.jhydrol.2009.06.023

VITA AUCTORIS

NAME: Rahul Narula

PLACE OF BIRTH: Sri Muktsar Sahib, India

YEAR OF BIRTH: 1994

EDUCATION: Birla Institute of Technology and Science, B.E.(HONS), Pilani, India, 2018

University of Windsor, M.Sc., Windsor, ON, 2022

EXPERIENCE: Graduate Assistant, Dept. of Civil and Env. Engg., University of Windsor, Jan 2020 to Dec 2020