DOCTORAL THESIS

Estimating areal rainfall time series using input data reduction, model inversion, and data assimilation

Author:
Ashley Wright

Supervisor:
A/Prof. Valentijn Pauwels
Prof. Jeffrey Walker

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Monitoring, prediction and protection theme
Department of Civil Engineering

November 27, 2017

Copyright

Declaration of Authorship

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 2 original papers published in peer reviewed journals and 1 manuscript that has been submitted for publication. The core theme of the thesis is rainfall estimation through input data reduction, model inversion ,and data assimilation techniques. The ideas, development, and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the department of civil engineering under the supervision of Assoc. Professor Valentijn Pauwels and Professor Jeff Walker.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of chapters 2-4 my contribution to the work involved the following:

Thesis chapter	Publication title	Status	Nature and % of student contribution	Co-author name(s), Nature and % of co-authors contribution
2	A comparison of the discrete cosine and wavelet transforms for hydrologic model input data reduction	Published	80 %	Valentijn Pauwels, Theory and editing 10% David Robertson, editing 5% Jeff Walker, editing 5%
3	Estimation of temporal rainfall and model parameter distributions	Published	85 %	Valentijn Pauwels, Theory and editing 10% Jeff Walker, editing 5%
4	A multi-model hydrological analysis of rainfall estimates using ensemble Kalman filter soil moisture innovations	Submitted	85%	Valentijn Pauwels, Theory and editing 10% Jeff Walker, editing 5%

I have provided text to link chapters and identify the relation of each paper to the research questions. I have not renumbered sections of submitted or published papers.

Ashley Wright: Date: 27/11/2017

Valentijn Pauwels: Date: 27/11/2017

Abstract

Floods are devastating natural hazards that can have severe socio-economic impacts and lead to a loss of life. Consequently, the key driver for this research is to provide techniques that will lead to an increase in flood forecast skill. As fluvial floods are a direct result of rainfall, detailed knowledge of uncertainties in rainfall observations provides a fundamental foundation for improving both rainfall and flood forecast skill. As the understanding of uncertainties present in rainfall time series is developed, so too will the confidence in rainfall forecasts, and short- and long-term streamflow forecasts. For a description of rainfall uncertainty to be complete it must take into account uncertainty when rainfall was not observed, thus allowing, model structural errors to be correctly identified, analyzed, and treated. Therefore the focus of this thesis is to develop a robust methodology to estimate rainfall time series and its uncertainty such that it is consistent with both streamflow and soil moisture observations.

To effectively estimate rainfall time series, a method to reduce hydrological input data dimensionality was identified. The effective reduction of hydrological input data dimensionality allows modern parameter estimation algorithms to simultaneously estimate rainfall time series and model parameters. Due to their wide-spread use as model input data reduction techniques in other fields, the discrete cosine transform (DCT) and discrete wavelet transform (DWT) were used, for comparative purposes, to reduce the dimensionality of observed rainfall time series for the 438 catchments in the Model Parameter Estimation Experiment (MOPEX) data set. Once the time series were reduced to a small number of parameters, the rainfall time series were reconstructed for comparison with the observed hyetographs. The rainfall time signals are then reconstructed and compared to the observed hyetographs using standard simulation performance summary metrics and descriptive statistics. Analysis of the results demonstrate that, when compared to the DCT, the DWT is superior at preserving both short- and long-term rainfall patterns.

Second, the DWT was used to reduce the dimensionality of the input rainfall time

series for the catchment of Warwick, Queensland, Australia. The $\mathrm{DREAM}_{(\mathrm{ZS})}$ sampling algorithm, in conjunction with a likelihood function that considers both rainfall and streamflow, was then used to estimate the input rainfall time series. Model parameters and rainfall time series were simultaneously estimated. The inclusion of rainfall in the estimation process improved the root mean square error (RMSE) of streamflow simulations by a factor of up to 1.78. This was achieved while estimating an entire rainfall time series, inclusive of days when none was observed.

Last, rainfall time series for the catchment of Warwick were estimated using three different rainfall-runoff models. Using the rainfall time series and model parameter estimates, remotely sensed soil moisture observations from the Soil Moisture Ocean Salinity (SMOS) and Advanced Microwave Scanning Radiometer - Earth observing system (AMSR-E) satellites were assimilated into each of the models using an ensemble Kalman filter (EnKF). Through analysis of the innovations from the observed and simulated soil moisture it was found that the combination of model choice and remotely sensed soil moisture product had a significant impact on the quality of rainfall estimated. When compared to streamflow simulations obtained via the sole estimation of model parameters, all models that jointly estimated rainfall time series and model parameters produced superior streamflow estimates. Rainfall estimates obtained using the Sacramento Soil Moisture Accounting (SAC-SMA) model were the most realistic. When the SMOS remotely sensed soil moisture product was assimilated into the SAC-SMA, innovations that indicated errors are of a Gaussian nature were obtained. Further, streamflow simulations obtained from the SAC-SMA had the best RMSE.

The research presented in this thesis developed a methodology that can be used to estimate and evaluate rainfall estimates obtained using model input data reduction, model inversion, and data assimilation techniques. These rainfall estimates can be used to condition rainfall forecasts and consequently improve flood forecast skill.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Assoc. Professor Valentijn Pauwels for the continuous support of my Ph.D study and related research, for his dedication, patience, motivation, and immense knowledge. His guidance helped me throughout the research and writing components of this thesis.
Besides my supervisor, I would like to thank my co-supervisor Professor Jeff Walker for his insightful comments and encouragement, but also for the hard questions which encouraged me to widen my research from various perspectives.
My sincere thanks also goes to Dr. David Robertson, Assoc. Professor Jasper Vrugt, Dr Yuan Li and Dr. Stefania Grimaldi, who provided advice and support on my research and it's direction.
I would like to thank Dr. Hamid Bazargan and the anonymous reviewers for their comments and recommendations for my journal articles.
The research I conducted was supported by the Multi-modal Australian Sciences Imaging and Visualisation Environment (MASSIVE) (www.massive.org.au), a Monash University Engineering Research Living Allowance stipend and a top up scholarship from the Bushfire & Natural Hazards Cooperative Research Centre. For that I am most grateful.
I thank my fellow researchers for the stimulating discussions and support throughout this journey.
Last but not the least, I would like to thank my family for their support and encouragement throughout writing this thesis and my life in general.

Contents

List of Figures

List of Tables

List of Abbreviations

AMSL	**A**bove **M**ean **S**ea **L**evel
AMSR-E	**A**dvanced **M**icrowave **S**canning **R**adiometer - **E**arth observing system
API	**A**ntecedent **P**recipitation **I**ndex
ARC	**A**ustralian **R**esearch **C**ouncil
ASCAT	**A**dvanced **SCAT**terometer
AWAP	**A**ustralian **W**ater **A**vailability **P**roject
BATEA	**B**ayesian **T**otal **E**rror **A**nalysis
BNHCRC	**B**ushfire & **N**atural **H**azards **C**o-operative **R**esearch **C**entre
BoM	**B**ureau of **M**eteorology
CATDS	**C**entre **A**val de **T**raitement des **D**onnées
CDF	**C**umulative **D**istribution **F**unction
DCT	**D**iscrete **C**osine **T**ransform
DREAM	**D**ifferential **E**volution **A**daptive **M**etropolis
DSD	**D**rop **S**ize **D**istribution
DWT	**D**iscrete **W**avelet **T**ransform
ECMWF	**E**uropean **C**entre for **M**edium-range **F**orecasts
EnKF	**E**nsemble **K**alman **F**ilter
ERAI	ECMWF **Re**- **A**nalysis **I**nterim
FAR	**F**alse **A**larm **R**atio
GPCC	**G**lobal **P**recipitation **C**limatology **C**entre
GPCP	**G**lobal **P**recipitation **C**limatology **P**roject
GPM	**G**lobal **P**recipitation **M**easurement
HyMod	**Hy**drological **Mod**el
IDW	**I**nverse **D**istance **W**eighting
LPRM	**L**and **P**arameter **R**etrieval textbfModel
MAP	**M**aximum **A** **P**osteriori
MASSIVE	**M**ulti-modal **A**ustralian **S**ciences **I**maging and **V**isualisation **E**nvironment
MCMC	**M**arkov **C**hain **M**onte **C**arlo
MOPEX	**Mo**del **P**arameter **E**stimation **EX**periment
NASA	**N**ational**A**eronautics and **S**pace **A**dministration
NSE	**N**ash-**S**utcliffe **E**fficiency
NWSRFS	**N**ational **W**eather **S**ervice **R**iver **F**orecast **S**ystem
PoD	**P**robability of **D**etection
PDM	**P**robability **D**istributed **M**odel
PE	**P**eak **E**rror
PET	**P**otential **E**vapo**t**ranspiration
POP	**P**ercentage of **O**riginal **P**arameters
QPF	**Q**uantitative **P**recipitation **F**orecast
RMSE	**R**oot **M**ean **S**quare **E**rror
RS	**R**emotely **S**ensed
RSR	**R**MSE to **S**D **R**atio
SAC-SMA	**S**acramento **S**oil **M**oisture **A**ccounting

SCEM-UA	**S**huffled **C**omplex **E**volution **M**etropolis - **U**niversity of **A**rizona
SD	**S**tandard **D**eviation
SM	**S**oil **M**oisture
SMAP	**S**oil **M**oisture **A**ctive **P**assive
SMART	**S**oil **M**oisture **A**nalysis **R**ainfall **T**ool
SMOS	**S**oil **M**oisture **O**cean **S**alinity
SM2RAIN	**S**oil **M**oisture to **R**ain
SODA	**S**imultaneous **O**ptimization and **D**ata **A**ssimilation
STEPS	**S**hort **T**erm **E**nsemble **P**rediction **S**ystem
TRMM	**T**ropical **R**ainfall **M**easurement **M**ission
TS	**T**hreat **S**core
USA	**U**nited **S**tates of **A**merica
WFT	**W**indowed **F**ourier **T**ransform

Chapter 1

Introduction

1.1 Overview

There exists an opportunity to increase the skill of flood forecasts by improving rainfall forecasts and their ability to simulate streamflow events. The research conducted for this thesis makes sequential steps towards improved rainfall forecasts. To be able to improve rainfall forecasts a greater understanding of the uncertainties in rainfall observations and how these uncertainties propagate through hydrological models needs to be developed. The first component of this research evaluates the Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) for hydrologic model input data reduction. The identification of an effective method to reduce hydrologic input data allows the second research task to be undertaken. The second research task simultaneously estimates areal rainfall time series and rainfall-runoff model parameter distributions. These estimates of areal rainfall time series and model parameter distributions are able to simulate streamflow that is superior to streamflow that is simulated by models that are forced with the rainfall observations. The final research task conducts an analysis of Ensemble Kalman Filter (EnKF) innovations to identify the presence of bias in rainfall estimates, model choice and Remotely Sensed (RS) Soil Moisture (SM) observations. Future research can lead to an improvement in rainfall forecasts by conditioning them with unbiased rainfall estimates that simulate superior streamflow.

1.2 Importance of flood forecasting

Natural hazards impinge on the ability of developing countries and countries that depend on climate reliant industries to grow socio-economically. Fourteen of twenty-five mega-cities are located in coastal regions and thus are more exposed to natural hazards (*Chhibber and Laajaj*, 2013). An increase in knowledge and understanding of how these events form and affect society can only lead to more prepared and resilient communities.

Accurate, precise and timely forecasts are the holy grail of flood forecasting. Floods cost the Australian economy on average $377M[1] per year (*Deloitte Access Economics*, 2013). The 2010-2011 Brisbane floods alone resulted in 33 confirmed deaths and $2.38 billion in economic damage (*Queensland Floods Commission of Inquiry*, 2012). The quality and usefulness of flood forecasts are largely constrained by data availability and the implementation of modeling techniques.

An efficient and effective flood warning system is comprised of weather observation systems, weather and flood forecasting models, flood warning dissemination systems, and emergency response procedures (*Sene*, 2008). Due to inherent complexities in model structural formulation and obtaining accurate measurements, significant systematic and random errors are prominent in flood forecasting models and remote sensing data. With advancements in computational power as well as remote sensing capability, considerable improvements in data availability, data quality and flood forecast skill have been made. These improvements can be leveraged to further develop hydrology and flood forecasting (*Cloke and Pappenberger*, 2009).

Flood forecast skill is hinged upon the quality of hydrometeorological observations and predictions (*Ebert and McBride*, 2000).

1.3 Statement of problem

Rainfall time series that are constructed from gauge based observations are the primary input for rainfall-runoff models. The gauges that contribute observations towards the rainfall time series are considered highly accurate at a point. However,

[1] Unless otherwise stated, this thesis uses Australian dollars

they often lack the spatial density required to represent catchment wide rainfall. Conversely, ground and satellite based radar systems are able to capture spatial rainfall patterns and include large inaccuracies.

The rainfall time series that are used to force a flood forecasting rainfall runoff model are called Quantitative Precipitation Forecasts (QPFs) and are generated from Numerical Weather Prediction (NWP) models. To best represent catchment rainfall these QPFs are often conditioned using past gauge based rainfall observations.

Consequently, the primary goal of this research is to develop methodologies to estimate rainfall time series and their associated uncertainty such that they can be used to condition QPFs and ultimately improve flood forecasts.

1.4 Objectives and scope of research

The main research questions and hypotheses for this thesis are

1. Is the DWT or DCT most appropriate for reducing hydrological input data?
   It is hypothesized that the DWT will be most appropriate for reducing hydrological input data due to its superior ability to decompose an input signal into multi-resolution components.

2. Can an increased understanding of hydrologic uncertainty be gained by estimating rainfall from streamflow observations using the DWT and model inversion techniques?
   It is hypothesized that use of model input data reduction and model inversion techniques will provide enhanced estimates of rainfall by estimating rainfall and its associated uncertainty for an entire rainfall time series, including when rainfall was not observed.

3. Using the results from the previous research question, can rainfall estimates be constrained by soil moisture observations?
   It is hypothesized that the ability to constrain rainfall estimates using soil moisture observations will depend on the rainfall-runoff model that is chosen to be

inverted as well as the quality of the soil moisture observations. Models that are most able to characterize a catchments soil moisture characteristics will have a better chance at estimating rainfall time series and model parameters that can be used to simulate soil moisture states which present little bias when compared with soil moisture observations. It is also expected that these combinations of model, rainfall and model parameter estimates, and soil moisture observations will be able to simulate streamflow that is superior to that which would be obtained if the rainfall estimation and data assimilation process was not used.

1.5 Outline of research

The research conducted for this thesis was comprised of three sequential tasks that can be broken down as follows

1. Evaluation of transforms for effective model input data reduction
   (a) Review transforms that have been used in literature to reduce the dimensionality of input data
   (b) Present a theoretical comparison of the DWT and DCT
   (c) Collect data from the MOPEX data set
   (d) Use transforms to reduce the dimensionality of input data
   (e) Compare the ability of each transform to reconstruct the observed rainfall data using a number of simulation performance summary metrics
2. Simultaneously estimate rainfall time series and model parameter distributions
   (a) Select a rainfall-runoff model that is widely accepted by the hydrological community, characterizes key components of the rainfall-runoff process, can be adapted to assimilate RS SM and, has been demonstrated to simulate streamflow well

   (b) Develop a methodology that is built upon Bayesian inference to simultaneously estimate rainfall time series and model parameters
   (c) Select an Australian catchment that has both, been subjected to multiple floods in recent years and has good quality data available
   (d) Develop and conduct a synthetic case study to demonstrate the applicability of the rainfall estimation methodology
   (e) Present the results of the synthetic case study and provide detailed analysis and discussion
   (f) Develop and conduct a real world case study to demonstrate the applicability of the rainfall estimation methodology
   (g) Present the results of the real world case study and provide detailed analysis and discussion

3. Analysis of data assimilation innovations obtained using the rainfall estimates
   (a) Select multiple models with which the data assimilation innovations and rainfall estimates may be compared against each other
   (b) Collect RS SM observation products from at least two missions
   (c) Estimate rainfall time series and model parameter sets for each model using the methodology developed
   (d) Assimilate both RS SM products into each model for all rainfall time series and model parameter sets
   (e) Analyze the data assimilation innovations for Gaussianity
   (f) Present the analysis results and discuss their implications

1.6 Structure of thesis

Including the Introduction chapter, this thesis is comprised of 6 chapters. The remaining 5 chapters are

- Chapter 2: Literature review

  The literature review provides a broad overview of floods and flood forecasting before identifying key areas of hydrological uncertainty. Following this a detailed analysis of rainfall estimation methodologies is given. Existing knowledge gaps are identified and research questions that will fill these gaps are formulated. Lastly, opportunities for utilization of the research are given.

- Chapter 3: Hydrological model input data reduction

  This chapter begins with a statement outlining where the research presented in this chapter sits in relation to the research questions and thesis as a whole. The main body of the chapter is a verbatim reproduction of the published journal article, *A comparison of the discrete cosine and wavelet transforms for hydrologic model input data reduction* and presents a theoretical and numerical evaluation of the DWT and the DCT for model input data reduction using the Model Parameter Estimation Experiment (MOPEX) data set. The introduction of this paper supplements the literature review and presents a more targeted review of the relevant literature.

- Chapter 4: Rainfall estimation

  This chapter begins with a statement outlining where the research presented in this chapter sits in relation to the research questions and thesis as a whole. The main body of the chapter is a verbatim reproduction of the published journal article, *Estimating rainfall time series and model parameter distributions using model data reduction and inversion techniques* and presents a novel methodology to simultaneously estimate entire rainfall time series and model parameter distributions. The introduction of this paper supplements the literature review and presents a more targeted review of the relevant literature.

- Chapter 5: Data assimilation analysis of rainfall estimates

  This chapter begins with a statement outlining where the research presented in this chapter sits in relation to the research questions and thesis as a whole. The main body of the chapter is a verbatim reproduction of the submitted journal article, *A multi-model hydrological analysis of rainfall estimates using Kalman filter soil moisture innovations* and presents a method that can be used to detect bias present in rainfall estimates, model choice or remotely sensed soil moisture observations. The introduction of this paper supplements the literature review presents a more targeted review of the relevant literature.

- Conclusions

  This chapter presents a discussion of the main conclusions that can be drawn from this research and outlines possible outcomes and directions of future work.

Chapter 2

Literature review

In a white paper commissioned by the Australian Business Round table for Disaster Resilience & Safer Communities, *Deloitte Access Economics* (2013) indicate that natural disasters, on average, cost the Australian economy $6.3 billion annually. A major proportion of this expenditure is as a result of flooding (*Gentle*, 2001). It is estimated that this figure could inflate to $23 billion by 2050. It is identified that significant socio-economic savings, of approximately 50%, can be realized by improved planning, preparedness, and resilience measures. Adding skill to the flood forecasting knowledge base is one such measure.

This chapter provides an overview of the flood forecasting knowledge base before identifying specific knowledge gaps. Research questions that address these knowledge gaps will ensure that the research conducted adds skill to the flood forecasting knowledge base. Lastly, opportunities for utilization of the research are identified.

2.1 Floods

Floods are generally distinguished from each other based on their location and the combination of mechanisms that generate them. The focus of this thesis is on fluvial floods that occur in inland catchments. Fluvial floods occur in river valleys or flood plains as a combined result of excess rainfall not being able to infiltrate into the soil, and the channels capacity to convey this water being exceeded (*World Meteorological Organization*, 2011). Heavy rainfall from convective, stratigraphic and orographic events can lead to fluvial flooding. Melted water from areas containing high levels of snow can also lead to fluvial flooding. However this is not a focus of this thesis.

Estuarine flooding occurs around river deltas and is caused by the combined effect of fluvial flows meeting coastal tides. Urban floods occur in developed areas that have large quantities of impervious surfaces. They are largely influenced by rainfall intensity, the connection of impervious areas to the drainage system ,and the capacity of the drainage system. The devastation of many coastal cities by urban flooding can be magnified by the combined effects of estuarine and/or coastal flooding. Whilst still occurring in coastal areas, coastal floods are distinguished from estuarine floods by the mechanism that generates them. The combination of high tides with storm surges such as those that occur in cyclonic regions produce coastal floods.

2.1.1 Value of flood forecasts

Timely, accurate and precise flood forecasts are critical, in the context of a natural disaster the value of information depreciates with time. Consequently, in operational settings progressive forecasts and warnings are issued. The Australian Bureau of Meteorology (BoM) issues less informative flood watches with long lead times when there is a significant chance of a flooding event. More detailed flood warnings are issued closer to the event (*Emergency Management Australia*, 2005).
Difficulties associated with longer term weather forecasts contribute to lower confidence levels prior to an event. In the time leading up to an event, rainfall and streamflow forecasts progressively become more reliable. *Martina et al.* (2006) proposed that flood damage perceived by stakeholders can be modeled by a cost function. No actual value is attributed to the cost function as it is an amalgamation of economic, social and environmental loss. The cost function is represented by

$$U(q) = C_0 + \frac{a}{1 + b \exp -c(q - Q^\star)}, \tag{2.1}$$

where the flood damage U is a function of the flood volume or extent q. Parameters a, b and c dictate the shape of the cost function. $Q^\star$ is the critical threshold flood volume or extent upon which if q is in excess the flood damage grows and C_0 is the minimum damage perceived by the stakeholders for which it is necessary to issue a flood watch or warning. For this cost function it is assumed that the flood volume or extent being forecasted is the same as that which is observed. This function and its relation to the timing of a flood watch or forecast being issued is depicted in Figure 2.1. It can be seen that there is a cost associated with flood watches and warnings being issued for flows below the critical threshold. Where as for flows above the critical threshold it becomes increasingly important to issue flood watches and warnings with larger lead times. If the accuracy and precision of the flood watches and warnings is not sufficient then the damage incurred is likely to increase for all flood volumes and extents.

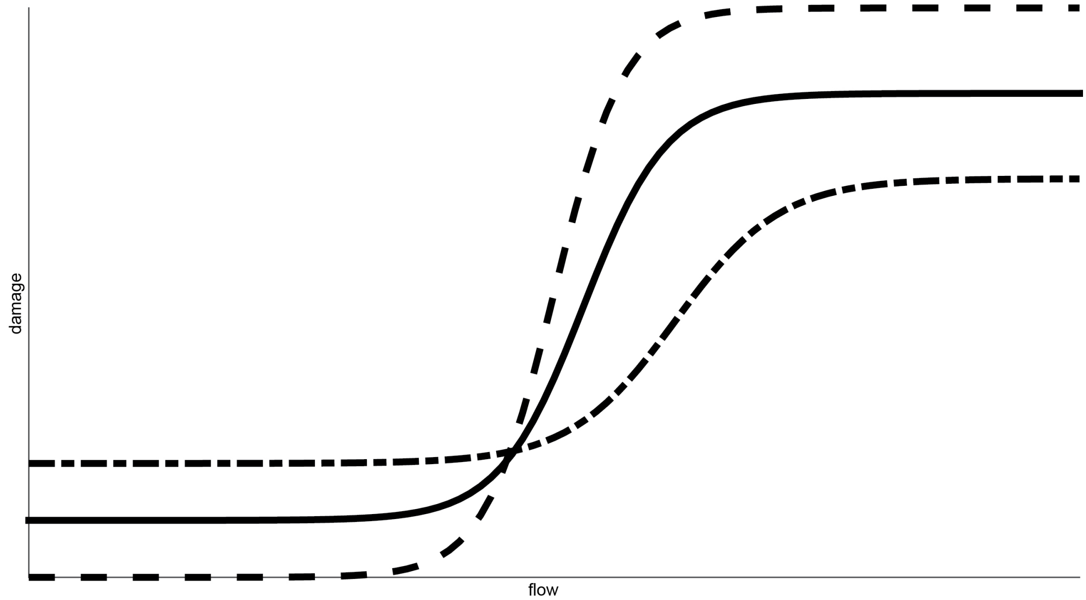

FIGURE 2.1: The flood damage cost function where the solid line depicts damage for floods in which no flood watch or warning is issued, the and −− lines depicts damages incurred for when flood watches or warnings are or are not issued respectively. Adapted from *Martina et al.* (2006)

2.1.2 Flood forecasting models

For many catchments, the antecedent soil moisture conditions play a large role in the partitioning of incident rainfall into flow types. Due to difficulties in the specification of antecedent soil moisture it is often difficult to determine precisely how a catchment will respond to incident rainfall. For operational flood forecasters the correct specification of the initial soil moisture conditions at the beginning of a rainfall event is of critical importance. Since event-based models can easily be calibrated to individual storms as they progress the BoM currently prefer to use an event-based modeling approach in preference over a continuous based modeling approach. However, it has been shown (*Pauwels et al.*, 2001; *Berthet et al.*, 2009) that accounting for antecedent soil moisture conditions in a continuous based modeling approach can lead to increased streamflow forecasting performance. Improving the accuracy and precision of QPFs will allow correctly implemented continuous rainfall-runoff models to better forecast the initial soil moisture state. This will provide increased lead time and add value to flood forecasts, a critical piece of information for emergency services (*Cloke and Pappenberger*, 2009).

2.2 Hydrological uncertainty analysis

When discussing hydrological uncertainty analysis it is important to not only consider all components that contribute towards total uncertainty but also how they relate to each other. It is particularly important to distinguish the random aleatory uncertainty that is present in our observations from the epistemic uncertainty that results from a lack of knowledge of the underlying processes. The dominant sources of uncertainty that are present in hydrological modeling are those arising from, forcing data, model parameters, model structure, state estimation, model output, and the data being calibrated against. Consequently, this section discusses uncertainty in parameter estimation, model structure state estimation, forcing data, and calibration data.

2.2.1 Parameter estimation

Parameter estimation is the field of research that searches for optimal parameter set/s. Equifinality (*Beven*, 2006) dictates that for a given set of forcing and calibration data there may exist multiple models and model parameter sets that each present equally plausible final simulations. Keeping equifinality in mind, parameter estimation algorithms have taken advantage of the advancement of computational power and moved away from deterministic solutions and towards stochastic solutions that describe a posterior parameter distribution.

Deterministic solutions (*Duan et al.*, 1994; *Gan and Biftu*, 1996; *Thyer et al.*, 1999) are idealistic and believe that a unique solution can be obtained. They are focused on finding a global optimum set of parameters through the determination of a global optimum set of parameters for a given objective function. A pitfall of this approach is that the parameter estimation algorithms often estimate parameters that perform well in the model calibration period and poorly in the model evaluation period. *Pappenberger and Beven* (2006) debate the feasibility of determining an optimal parameter set due to the quality and availability of data, model assumptions, current knowledge, computational and time restrains or if it is a problem of equifinality (*Beven*, 2006).
The choosing of an objective function is quite subjective for both deterministic and stochastic parameter estimation problems. Each objective function often leads to a parameter set that is able to only partially describe the hydrologic system, perform well in some catchments or flow situations and poorly in others. Thus deterministic parameter estimation often leads to poor simulation of streamflow in forecasting situations. *McInerney et al.* (2017) evaluate eight common objective functions and make recommendations for hydrologists based on the conclusion that no one function is able to perform well for all performance metrics.

The aim of stochastic parameter estimation is to select all parameter sets that are able to adequately describe the hydrologic system. Sampled parameter sets are ranked based on a likelihood function, the effectiveness of which is dependent upon assumptions made about the errors present in a model and observations (*Vrugt*, 2016). Stochastic parameter estimation methods include: Bayesian recursive parameter estimation (*Thiemann et al.*, 2001), the limits of acceptability approach (*Beven*, 2006; *Blazkova and Beven*, 2009), the BAyesian Total Error Analyis (BATEA) framework (*Kavetski et al.*, 2006a,b; *Kuczera et al.*, 2006; *Thyer et al.*, 2009; *Renard et al.*, 2011), the Simultaneous Optimization and Data Assimilation (SODA) (*Vrugt et al.*, 2005), the Differential Evolution Adaptive Metropolis (DREAM) algorithm and its variations (*Vrugt et al.*, 2005, 2008, 2009a,b; *Vrugt and Ter Braak*, 2011; *Laloy and Vrugt*, 2012; *Sadegh and Vrugt*, 2014), Bayesian model averaging (*Butts et al.*, 2004; *Ajami et al.*, 2007; *Vrugt and Robinson*, 2007), the hypothetico-inductive data based mechanistic modeling framework of *Young* (2013) and Bayesian data assimilation (*Bulygina and Gupta*, 2011). A major advantage of stochastic parameter estimation methods is that, if input is represented as parameters, hydrologists are able to explore input uncertainty. A few studies have focused on elucidating the link between parameter estimation and input error. It is likely that when combined with efforts to constrain state estimates that these techniques will become more valuable.

2.2.2 Model structure

Hydrological models can be used for simulation of rainfall-runoff events, projection of future scenarios, forecasting future events and hind-casting previous events (*Beven and Young*, 2013). A projection may be thought of as a "what if" scenario, e.g. what are water storages expected to look like under severe drought or flood conditions, whereas a forecast is the prediction of behavior that is likely to occur given current conditions. Hind-casting of events is used to evaluate the usefulness of a model for a future forecasting scenario. Modelers use ex-post and ex-ante hind-casting for evaluating respective events where it is assumed that either *"perfect input"* or forecasted data are available (*Young*, 2013), respectively.

No one rainfall-runoff model is able to perfectly deal with every catchment or flow event. It is therefore necessary to understand the variety of models available and their limitations. Models are categorized based on how their structures deal with space, lumped or distributed; how model states are treated, event-based (static) or continuous (dynamic); modeling philosophy, deterministic or stochastic; and how they are formulated, inductively or deductively (*Beven and Young*, 2013). Some models are distributed only at areas of key interest, while some may be partially deductive and utilize a more detailed inductive approach for perceived critical processes. Quite often models are purpose built, e.g. some are more appropriate in a scientific research setting whilst others an operational setting (*Moradkhani and Sorooshian*, 2008).

System identification extends the concept of equifinality to allow for not just more than one viable parameter set for a model but to allow for a range of models each with numerous parameter sets being able to describe a hydrologic system. Thus it is now commonplace for multiple models to be considered in a model averaging scheme. Model averaging schemes assign weights to different models with the aim to exploit the fact that some models will outperform other models in different flow scenarios, lending credibility to the idea that different models will be more or less suitable for streamflow simulation in different catchments and for different events. Thus, the averaging of models allows for more robust streamflow predictions and estimation of uncertainty. A shortcoming of model averaging techniques is the lack of distinction between input errors and model structure.

2.2.3 State estimation

The process of state estimation involves simulated model states being updated with observational data. The underlying assumption being that model states have been incorrectly simulated due to some combination of errors in forcing data, initial conditions and model structure. Consequently simulated model states are updated with observations of those states. Alternatively, states can be updated based on their ability to simulate the desired output data. Commonly used data assimilation techniques include, direct insertion (*Heathman et al.*, 2003), Newtonian nudging (*Houser et al.*, 2001; *Pauwels et al.*, 2001; *Paniconi et al.*, 2003) optimal interpolation (*Seuffert et al.*, 2004) Kalman filtering (*Galantowicz et al.*, 1999), Kalman smoothing (*Dunne and Entekhabi*, 2005; *Li et al.*, 2013) and variational data assimilation (*Castelli et al.*, 1999). The complexity, and consequently applicability of these techniques differ greatly. In some situations directly inserting observations in place of states may suffice, more detailed problems may require an update of states that balances the observations and states and their expected errors. Lastly, the most complex problems may require updating of states that considers errors in both past and present observations and states.

Whilst not all assimilation techniques are derived from ***Bayes Law***, the fundamental notion that the probability associated with *a priori* states can be improved upon, given new information, to form *a posterior* states is a central concept of data assimilation. All data assimilation techniques either filter observations at a given time or smooth observations over a given window of time and are selected for use based on computational cost and the required level of complexity.

2.2.4 Forcing data

Forcing data for hydrological rainfall-runoff models are largely comprised of precipitation and potential evapotranspiration PET data. This study will focus on precipitation or more specifically rainfall data. Studies (*Oudin et al.*, 2005; *Samain and Pauwels*, 2013) have shown that streamflow simulations from rainfall-runoff models are relatively insensitive to more detailed temporal variances in PET data when compared to temporal variances in precipitation data.

Precipitation is the collective term for drizzle, rain, sleet, snow, graupel and hail. Whilst all of these forms of precipitation may make their contribution to runoff, this study focuses on rainfall, the primary forcing of fluvial flooding. Research into the science of rainfall is divided into three main areas; rainfall microphysics, rainfall measurement and estimation, and statistical analysis.

The microphysics concerning rainfall involves understanding the forces that shape a raindrop, its chemical composition, oscillations that raindrops undergo when approaching terminal velocity and the drop size distribution (DSD) of raindrops within a hydrometeor (*Gebremichael and Testik*, 2010). A good understanding of rainfall microphysics and when certain behavior is most likely to occur is particularly pertinent for the accurate estimation of rainfall via microwave remote sensing.

In order to effectively estimate rainfall it is important to gain an understanding of the different mechanisms behind rainfall. **Stratiform** rainfall occurs when different pressure systems meet, consequently stratiform rainfall intensity is quite often uniform over a large spatial area and has the best forecast accuracy. **Convective** rainfall occurs due to localized heating of the Earth's surface via radiation from the sun. Convective events most often occur in summertime and equatorial regions, are often difficult to predict, and are characterized by short intense rainfall events. The last rainfall mechanism is **orographic**. Orographic rainfall is caused by masses of air being pushed up topographical land formations by wind fronts. Hence the direction of wind and the presence of any shielding can have a large impact on orographic rainfall events. Orographic rainfall events generally form around mountainous regions, may be localized to small areas, and have varying levels of prediction accuracy (*Testik and Gebremichael*, 2010).

The most common rainfall measurement techniques in use today are tipping bucket gauges, ground based radar and satellite. Acknowledged systematic errors for tipping bucket gauges include but are not limited to; site selection, light rainfall and splashing, and insufficient spatial density. Rainfall measurement via the use of a tipping bucket is generally considered to be the most accurate (*Habib et al.*, 2013). Without sufficient density of rain gauges, recording of rainfall intensities and their corresponding maxima and minima may be missed. This may potentially dampen or magnify the simulated hydrograph response. Consequences of this are less significant for stratiform rainfall events than orographic and convective rainfall events. Rainfall estimation via microwave remote sensing, using ground radar or satellite, provides additional spatial resolution. However, even after careful bias removal and calibration, ground based microwave remote sensing estimation of rainfall may have large inaccuracies due to bright band reflection, distance, shielding, insufficient sample size and rainfall being measured at the hydrometeor rather than catchment surface (*Seo et al.*, 2013). Further, difficulties may be seen in satellite rainfall estimation due to the temporal repeat at which a satellites passes over a catchment. Depending on catchment characteristics, both spatial and temporal averaging of rainfall can lead to difficulties when simulating or forecasting streamflow. For this reason it has been proposed that rainfall should be given a more rigorous uncertainty treatment (*Pappenberger and Beven*, 2006).

QPFs are most reliable for macroscale events. Increased confidence in a QPF can also obtained by using an ensemble of weather predictions. Originally formulated by *Bowler et al.* (2006) and developed by the UK Met Office and the Australian BoM, the Short Term Ensemble Prediction System (STEPS) has been enhanced by *Seed et al.* (2013) to include multiple QPFs and account for radar observation error. STEPS uses a decomposition framework to extrapolate radar nowcasts out to a QPF. As will be discussed later, skill can be added to QPFs by processing with rain guage measurements.

2.2.5 State data

For a lot of catchments soil moisture content dictates how rainfall becomes runoff. Soil moisture measurements can be used in the calibration of rainfall-runoff models to ensure that the model is able to represent the catchments soil dynamics. Unfortunately, doing so often leads to the rainfall-runoff model not being able to simulate streamflow as well as it does when only streamflow data is used for calibration (*López et al.*, 2017). Alternatively, *Alvarez-Garreton et al.* (2015) have demonstrated that the assimilation of soil moisture measurements into a distributed rainfall-runoff modeling scenario can have positive impacts on the simulation of streamflow. Soil moisture data are obtained via in situ measurements or remotely sensed satellite measurements. In situ measurements have a high level of accuracy, are able to continuously measure soil moisture at depth, yet typically are unable to provide observations that are representative of the catchment. Conversely, at the expense of spatial variability, remotely sensed satellite soil moisture products are able to provide observations that are representative of catchment soil moisture. Consequently, in situ soil moisture observations are used to calibrate and evaluate remotely sensed soil moisture products (*Albergel et al.*, 2012). Unfortunately, the penetration depth of satellite soil moisture observations is restricted by the wavelength of the instrument used. For satellites with sensors that operate in the L-band and C- or X-band, soil moisture can be observed for the near surface layer of 0-5 cm or 0-2 cm, respectively (*Yee et al.*, 2017). The revisit time of the satellite restricts the collection of observations to every few days. A review of the applications of remotely sensed soil moisture to rainfall-runoff models is provided by *Li et al.* (2016).

2.2.6 Calibration data

Streamflow or river discharge data are used in a calibration period to test model effectiveness and calibrate their parameters, with the aim being superior performance to other models and parameter sets in the calibration and independent evaluation periods. It is also possible for streamflow data to be used for state estimation. As is the case in many hydrological studies, this study assumes that uncertainty in streamflow measurements can be considered negligible when compared to other sources of uncertainty. It is however important to acknowledge that possible errors may arise from the changes in roughness arising from vegetation, the interpretation of river stage height, the interpolation and/or extrapolation from a rating curve and/or unsteady flow conditions (*Di Baldassarre and Montanari*, 2009).

2.3 Rainfall estimation

Despite the vast amount of literature on rainfall measurement (*Testik and Gebremichael*, 2010) and quality control procedures (*World Meteorological Organization*, 2008), a shroud of uncertainty still surrounds how rainfall should be used in rainfall-runoff modeling. Following the general review of the inherent sources of uncertainty present in rainfall-runoff modeling this chapter will now narrow the focus towards modeling methods that have been used to estimate rainfall and its uncertainty.

Drawing on the philosophy of the famous physicist Neils Bohr, as quoted by *Petersen* (1963), *"it is wrong to think that the task of physics is to understand how nature is, physics concerns what we can say about nature"*, would suggest that it is folly to think, as hydrologists, that the rainfall-runoff process can be understood. Instead of focusing on a complete understanding of the rainfall-runoff process, hydrological observations should be used to determine what can be said about the rainfall-runoff process.
Further, it was stated by *Einstein et al.* (1935) that *" If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity"*. So how far can the frequency and quality of hydrometeorological

observations be advanced without significantly altering the system? Is there a point of diminishing return? Given current observations, what can be said about rainfall? Is it best to use catchment average rainfall rate or a rainfall rate that is consistent rainfall derived from observed streamflow and soil moisture?

2.3.1 Streamflow

The response of a catchment to rainfall is determined, to varying extents, by Hortonian overland flow, saturation excess overland flow, interflow and groundwater flow. The time differential between peak rainfall and the resulting peak streamflow is referred to as lag. The lag process is often mathematically simulated by unit hydrographs and/or sequential filling and emptying of a series of reservoirs, otherwise known as a Nash cascade (*Li et al.*, 2013). Due to complex interactions involved in each of these processes, discrepancies are quite often noticed between similar rainfall events and the corresponding runoff. Hence the process of retrieving rainfall from streamflow observations is considered an ill-posed problem. Most hydrologists favor deterministic models (*Pappenberger and Beven*, 2006). Consequently attempts to retrieve rainfall from runoff have taken a deterministic approach.

In perhaps the earliest example of rainfall retrieval from streamflow *Hino* (1986) separated time series of daily discharge into their respective runoff components using coefficients obtained by fitting an Auto-regressive Moving Average model to streamflow data. Streamflow components were separated based on abrupt changes of autocorrelation coefficients. The runoff components were then routed to produce rainfall estimates using unit hydrographs. The results found by *Hino* (1986) indicate that this process is able to achieve achieve results that agree relatively well with gauged rainfall data.

In an attempt to produce robust streamflow simulations, deductive hydrologic models have often been developed with Occam's Razor in mind; also known as the law of parsimony, i.e. *"Entities should not be multiplied unnecessarily"*. Perhaps the best test for parsimony is the potential for a model to be analytically inverted to obtain rainfall rates from streamflow. *Kirchner* (2009) used a first order approximation to solve the water balance equation.

$$\frac{dS}{dt} = P_t - E_t - Q_t, \tag{2.2}$$

where dS/dt $[l^3/s]$ is the change in the total volume of water stored in the catchment and P_t $[L/s]$, E_t $[L/s]$ and Q_t $[l^3/s]$ are the rates of precipitation, evapotranspiration and discharge respectively. The success of this approach relies on the assumption that streamflow is largely dependent upon change in storage. Using this assumption, a sensitivity function which describes the change of flow with respect to a change in storage can be formulated as

$$g(Q) = \frac{dQ}{dS} = \frac{dQ/dt}{P_t - E_t - Q_t}. \tag{2.3}$$

Consequently, catchments with a significant proportion of impermeable surface or are highly saturated are unlikely to yield meaningful results. The sensitivity function, $g(Q)$, can thus be best estimated when $P_t - E_t \ll Q_t$, i.e. when the only contributions to streamflow are interflow and baseflow. An example of when this may occur is at night when it is not raining and evapotranspiration rates are low or zero.

Using this sensitivity function both evapotranspiration and precipitation are estimated via

$$P_t - E_t = \frac{Q_{t+\ell+1} - Q_{t+\ell-1}/2}{g(Q_{t+\ell+1}) - g(Q_{t+\ell-1})/2} + Q_{t+\ell+1} - Q_{t+\ell-1}/2. \tag{2.4}$$

where ℓ $[s]$ represents the time lag between precipitation and the observation of streamflow. One shortcoming of the approach taken by *Kirchner* (2009) is that the lag is given a fixed approximate value. Depending on the catchment, lag can be a highly dynamic parameter. Smoothing (*Li et al.*, 2013) and some data based mechanistic models pay significant attention to lag. Often overlooked by hydrologists, *Kirchner* (2009) draws attention to the fact that, of the components of the water balance, only streamflow can be considered a catchment scale observation.

Using a different framework *Kavetski et al.* (2006a) and *Vrugt et al.* (2008) used an alternative approach to simplifying the water balance and chose to represent true catchment rainfall and its associated uncertainty using parameters. One parameter or rainfall multiplier was assigned to each storm event. Taking advantage of an effective stochastic sampling algorithm and a likelihood function that considers streamflow and rainfall, the reduction of storm events to rainfall multipliers allows hydrologists to jointly estimate hydrologic model parameter distributions as well as input rainfall and its uncertainty. *Kavetski et al.* (2006a) and *Vrugt et al.* (2008) determined that there was sufficient data to estimate both hydrological model parameters and rainfall input. The reduction of input data using mathematical transforms offers an alternative to the storm multiplier method that provides the potential to reduce the dimensionality of the parameter estimation problem, thus enabling a more robust inference. Signal transforms, such as Fourier and wavelet transforms, are examples of data reduction transformations that have been applied in hydrology. Yet, not for data reduction purposes, their suitability to reduce hydrological input data has not been assessed to date.

Rainfall estimation methods that use streamflow measurements as the main input towards inverting the water balance are able to maintain good resolution in the time domain. They yield good results for catchments with simple dynamical systems that exhibit linear behavior. Conversely they perform poorly for complex catchments that exhibit highly nonlinear rainfall-runoff behavior. *Renard et al.* (2010, 2011) built on the idea of using storm multipliers by characterizing the storm multipliers with a hyper-distribution. The use of informative priors for the storm multipliers ensures that the sampled multipliers remain realistic and convergence is able to be achieved. The storm multiplier method assumes that a multiplicative error structure is appropriate for rainfall. A pitfall of this assumption is that poorly gauged catchments often overestimate, underestimate or completely miss localized rainfall events. Thus it is imperative that the development of any rainfall estimation method allows for rainfall to be estimated even when none was observed. When compared to storm multipliers, the reduction of rainfall by transfer functions allows for rainfall to be estimated when no rainfall was recorded at the gauge.

2.3.2 Soil moisture

The root zone soil moisture state has a large impact on a catchment's rainfall-runoff characteristics (*Grayson et al.*, 2006) and typically governs the proportion of rainfall that is available to contribute to surface and subsurface flows (*Tebbs et al.*, 2016). Consequently there is much that can be learned from analysis of the interactions between rainfall and soil moisture (*Kucera et al.*, 2013). As coined by *Brocca et al.* (2014), the notion of using '*Soil as a Natural Rain Gauge*' has gained traction. Methodologies to extract informative rainfall data from soil moisture observations have predominantly taken one of two approaches. Either updating satellite based rainfall products using changes in relative soil moisture or extracting rainfall data from soil moisture using approximations to the water balance. With the continued support and improvement of satellite rainfall and soil moisture measurement missions such as the Global Precipitation Measurement (GPM) mission and the Soil Moisture Active Passive mission (SMAP), it is expected that the value of methods outlined by *Crow et al.* (2011) and *Brocca et al.* (2014) will become more valuable.

Taking the first approach *Crow* (2007) used a Kalman filter based updating scheme to develop a Soil Moisture Analysis Rainfall Tool (SMART) (*Crow et al.*, 2009, 2011) which uses remotely sensed soil moisture observations to update an Antecedent Precipitation Index (API) model forced by satellite rainfall. The innovations from the updating scheme showed correlation with the errors between the observed satellite based rainfall and a high quality gauge based rainfall product. This updating scheme relies on the assumption that the observed soil layer is able to remember past rainfall and will be most accurate when rainfall events occur immediately prior to the acquisition of a remotely sensed soil moisture image. As the time difference between the rainfall event and soil moisture observation increase the assumption that losses due to percolation and PET are negligible becomes less valid. Further, this methodology is most appropriate for catchments or rainfall events in which minimal surface flow occurs. Enhancements were made to earlier versions of the SMART by addressing seasonal fluctuations, conditioning of rainfall forecasts, filter calibration, accounting for non Gaussian errors, PET and temperature influences, as well as covariance with adjacent grid cells, and drainage rate. Also, using remotely sensed soil moisture observations to correct satellite based rainfall estimates *Pellarin et al.* (2008) demonstrated that soil moisture observations can be used to suppress satellite based rainfall observations that were not observed by ground based rainfall gauges. *Pellarin et al.* (2009, 2013) also demonstrated that the satellite based rainfall estimates can be used to add value to the remotely sensed soil moisture observations.

The second methodology builds upon the seminal work conducted by *Kirchner* (2009) on catchments as simple dynamical systems. Instead of correcting satellite rainfall observations, the Soil Moisture to Rain (SM2RAIN) algorithm (*Brocca et al.*, 2014) makes simplifications to the soil water balance equation to allow for the direct estimation of rainfall from the knowledge of relative soil moisture. A major assumption is that all rainfall infiltrates. *Ciabatta et al.* (2015) uses the SM2RAIN algorithm to nudge satellite precipitation estimates in order to estimate daily rainfall; *Abera et al.* (2016) evaluated this product in a comparative study. The inversion of soil moisture data to estimate rainfall depends on knowledge of the relative change of soil moisture and the runoff flow rate

$$Z(L)\frac{ds(t)}{dt} = p(t) - r(t) - e(t) - g(t), \tag{2.5}$$

where $Z(L)$ $[m]$ is the soil layer depth, $ds(t)/dt$ $[t^{-1}]$ is the change of relative saturation of soil with respect to time, $p(t)$ $[L \cdot s^{-1}]$ and $e(t)$ $[L \cdot s^{-1}]$ are the intensity of precipitation and evapotranspiration, whilst $r(t)$ $[L \cdot s^{-1}]$ and $g(t)$ $[L \cdot s^{-1}]$ are the runoff and drainage flow rates. It is assumed that all precipitation infiltrates into the soil and consequently the runoff rate is zero. When precipitation occurs evapotranspiration is assumed negligible. Drainage to groundwater stores is expressed as

$$g(t) = as(t)^b, \tag{2.6}$$

where a $[L \cdot s^{-1}]$ and b $[-]$ are the only two parameters that need calibration. Consequently, the formula that is used to describe the retrieval of rainfall from soil moisture is

$$p(t) \simeq Z(L)\frac{ds(t)}{dt} + as(t)^b. \tag{2.7}$$

In contrast to the SMART algorithm, each individual grid cell is able to be calibrated according to the dominant soil characteristics. Similar to SMART, the performance metrics are based on the False Alarm Ratio (FAR), Probability of Detection (PoD), Threat Score (TS), and Root Mean Square Error (RMSE).

The SM2RAIN tool has been evaluated using the Advanced Scatterometer (ASCAT), Advanced Microwave Scanning Radiometer - Earth observing system (AMSR-E) and Soil Moisture and Ocean Salinity (SMOS) products against the Tropical Rainfall Measurement Mission (TRMM)-3B42RT product using the Global Precipitation Climatology Centre (GPCC), Global Precipitation Climatology Project (GPCP) and European Centre for Medium-range Forecasts (ECMWF) Re-Analysis Interim (ERAI) products as benchmark data sets. The AMSR-E and TRMM-3B42RT rainfall estimates showed similar results, whilst ASCAT and SMOS products produced inferior rainfall estimates.

If rainfall estimates that are based on a satellite soil moisture product are to be used in a flood forecasting situation, it is imperative that the rainfall estimate is up to date and that the satellite soil moisture images are obtained immediately prior to the flood. With the continued improvement of satellite rainfall and soil moisture measurement missions, such as the GPM mission (*Hou et al.*, 2014) and the SMAP (*Entekhabi et al.*, 2010) mission, it is expected that the methods outlined by *Crow et al.* (2011); *Brocca et al.* (2014) and *Ciabatta et al.* (2015) will become more valuable for estimating rainfall time series in the future. Yet, there are currently no methods that use both streamflow and soil moisture to estimate rainfall.

2.3.3 Processing of quantitative precipitation forecasts

As a precursor for flood forecasting QPFs are developed and used as a forcing for hydrological models. If an ensemble QPF is used the modeler runs the risk of forecasting median flows in adjacent catchments rather than a flood in one catchment and low flow in the other catchment. Thus, the *Schaake Shuffle* (*Clark et al.*, 2004) was developed to preserve historical spatial and temporal patterns between a given QPF and observed rainfall at the gauge.

It is common for QPFs to be processed using a two part probability model to correct bias and variance. Logistic regression is used to determine the probability of rainfall occurrence, whilst a transformation function is used to normalize and model the relationship between QPF and observed rainfall intensities (*Robertson et al.*, 2013). A deterministic QPF can be transformed into ensembles by stochastically adding Gaussian noise to a least squares fit between observed and predicted rainfall. *Robertson et al.* (2013) used a log sinh transformation within a Bayesian Joint Probability (BJP) framework to process a deterministic QPF before using the *Schaake Shuffle* to develop a spatially and temporally correlated QPF that is more skillful than the QPF that it was made from.

2.4 Knowledge gap

This literature review has explored key studies pertinent to flood forecasting, hydrological uncertainty analysis, and rainfall estimation methods. The three gaps in literature to be explored throughout this thesis are;

1. the lack of methods to effectively reduce rainfall data to a small number of parameters for estimation,
2. the estimation of rainfall data using transforms that account for an entire rainfall series through model inversion techniques using streamflow data, and
3. the ability to constrain rainfall estimates obtained through model inversion with soil moisture data.

There have been a number of studies dedicated to estimating rainfall from streamflow observations. An extension of these studies was to use streamflow observations and model inversion techniques to infer rainfall uncertainty. Doing so required input rainfall data to be characterized by a set of parameters. Early approaches applied a storm multiplier to adjust the volume of storms. In consequent studies a rainfall multiplier was applied to each observation and the distribution of rainfall multipliers was estimated using hyper-parameters. Throughout these studies none of them sought to determine the most efficient way to represent an entire rainfall time series for estimation purposes. The first research question of this thesis will address this knowledge gap.

Research conducted in response to the first knowledge gap identifies a transform that is effective at reducing hydrological input data to parameters. The application of the identified transform as a model input data reduction technique for the estimation of rainfall and its uncertainty forms the basis of the second knowledge gap. The second research question for this thesis explored how the identified transform can be best used to estimate an entire rainfall time series and its associated uncertainty. Further, a common problem with methods that estimate rainfall and its associated uncertainty using model inversion techniques is that the use of storm multipliers assumes that there is no uncertainty in rainfall when none is observed. The exploration of methods that estimate rainfall and its associated uncertainty using model inversion techniques that do not assume there is zero uncertainty in rainfall when none is observed is another knowledge gap that was addressed by the second research question.

The literature review revealed that there have been a number of studies that have attempted to estimate rainfall and its uncertainty through various model inversion techniques. Further, soil moisture observations have been used to both constrain satellite based rainfall estimates as well as estimate rainfall itself. The main obstacles that are faced when attempting to estimate rainfall from streamflow data are the complexities and non-linearities that soil moisture stores introduce. Conversely, when estimating rainfall from remotely sensed soil moisture observations, insufficient resolution in the time domain is able to achieved. There are no studies in the literature that have attempted to estimate rainfall and its associated uncertainty by inverting streamflow observations and constrain those rainfall estimates by assimilating remotely sensed soil moisture observations. Thus, the last knowledge gap that is to be addressed in this thesis has been identified.

2.5 Outline of approach

The main research objective of developing techniques to aid in the development of a robust methodology to improve flood forecasting skill was realized through three sequentially linked research tasks. Prior to estimating rainfall and its associated uncertainty an efficient technique to reduce hydrological input data was identified. The search for this formed the basis of the first research question. The second research question focused on using the DWT as a model input data reduction technique for the estimation of rainfall and its associated uncertainty in a complex catchment that exhibits non-linear characteristics. The constraint of the rainfall estimates obtained from inverting streamflow observations with remotely sensed soil moisture observations was the focus of the last research question.

2.5.1 Efficient techniques to reduce hydrological input data

Before input rainfall can be estimated using parameter estimation algorithms it is necessary to determine an effective method to reduce hydrological input data into parameters. In other scientific fields the DWT and DCT are two common transform that are used to reduce input signals into parameters. Consequently it is asked;

1. Is the DWT or DCT most appropriate for reducing hydrological input data?
    (a) Does either transform represent transient events better than the other?
    (b) Does either transform represent seasonal patterns better than the other?

To effectively estimate rainfall using parameter estimation algorithms it is essential that rainfall is represented using the most effective parameters. As the DWT and DCT are two transforms used for model input data reduction in other scientific fields their ability to represent hydrological input data using as few parameters as possible is evaluated. As the DCT is a Fourier based transform it is expected that it is most applicable for rainfall time series that do not exhibit time specific information. Alternatively, the DWT is expected to show promising results for rainfall time series with irregular rainfall patterns. Due to the sinusoidal nature of the DCT it is plausible that the DCT is able to represent seasonal patterns better than the DWT.

The transform that was able to best preserve hydrological input data was used in the rainfall estimation process.

2.5.2 Estimation of an entire rainfall time series using hydrological model input data reduction

Once a suitable methodology to reduce hydrological input data to a small number of parameters has been determined, its ability to be used in a rainfall estimation scenario needs to be tested. Consequently, the basis of the second research question is formed around the estimation of representative rainfall incident upon a catchment using the DWT and model inversion techniques.

2. Can an increased understanding of hydrologic uncertainty be gained by estimating rainfall from streamflow observations using the DWT and model inversion techniques?

    (a) Can uncertainty in rainfall events be estimated when no rainfall was observed?

    (b) Is there an upper limit to which incident rainfall can be estimated?

Regularization and linear inversion are the most commonly used techniques in literature to estimate rainfall from streamflow observations. Yet, due to the strong non-linear behavior and the ill-posed nature of the problem it is expected that estimating incident rainfall via the reduction of input rainfall to DWT parameters, to be solved for in the estimation routine, and constraining rainfall estimations with a likelihood function that considers both rainfall and streamflow will provide superior results. It is expected that estimating the distribution of an entire rainfall time series will capture subtle yet complex and indeterminate variations, both within a catchment and between events. These variations are unable to be captured by methods that only consider rainfall uncertainty when rainfall was observed. Thus, value will be added to the understanding of rainfall and its temporal distribution. By including the DWT parameters in the parameter estimation routine it is expected that uncertainty in rainfall events will be able to be estimated even if rainfall was not observed. It is expected that as more DWT parameters are estimated in the rainfall estimation process that a greater resolution and understanding of rainfall estimates and their associated uncertainty will be able to be achieved. The use of an increased number of DWT parameters in the estimation algorithm does however come at a cost. As more parameters are used the computational requirements increase greatly. Consequently the trade off between computational cost and increased resolution in rainfall estimates will need to be optimized. In order to retrieve superior rainfall estimates it is expected that more frequent streamflow measurements from the headwaters of small flashy catchments and catchments that show minimal signs of dampening will produce the best results. The estimation of input rainfall along with model parameters is expected to be comprised of some model structural errors. As such it is expected that answers to the third research question will provide a step towards untangling uncertainty in rainfall estimates and model structural errors.

2.5.3 Constraining rainfall estimates from model inversion using soil moisture observations

Due to clear limitations involved with the determination of state variables, both rainfall-runoff models and the inverse process of retrieving rainfall inputs from streamflow outputs can be thought of as ill-posed problems. Since good results have been found when rainfall-runoff models have been constrained with soil moisture, it is only natural to ask;

3. Using the results from 2.5.2, can rainfall estimates be constrained by soil moisture observations?
   (a) Using an EnKF, do innovations exhibit white noise?
   (b) Does the assimilation of soil moisture observations provide an element of physical realism to the rainfall retrieval process?
   (c) Does the assimilation of soil moisture observations restrict the efficacy of the rainfall retrieval process due to model structural inadequacy?

Rainfall estimates from 2.5.2 were constrained by soil moisture observations using an EnKF and the noise of the innovations were tested for whiteness. Due to the addition of good quality data it is expected that rainfall estimates will be meaningfully constrained by soil moisture observations. Further, it is expected that only combinations of rainfall estimates, model structure, model parameterization and soil moisture observations that are in agreement with the physical *'truth'* will show similar or improved capacity to model streamflow. Rainfall realizations that do not produce white noise when soil moisture observations are assimilated and do not produce streamflow within the bounds of uncertainty will be rejected.

2.6 Opportunities for utilization of research

Throughout this literature review areas have been highlighted that can contribute to the scientific understanding and development of flood forecasting using remote sensing data. Of particular importance to this study is the development and progression of flood forecasting skill. The optimal way to do this is to provide an increased understanding of rainfall uncertainty and eventually rainfall forecasting skill. An increase in rainfall forecasting skill will translate to an increase in flood forecasting skill. Rainfall estimates will be gathered using a methodology that involves model input data reduction, model inversion and data assimilation. The rainfall product can then be used to post process QPFs before being used in an operational flood forecasting scenario.

Chapter 3

A comparison of the discrete cosine and wavelet transforms for hydrologic model input data reduction

Overview

This chapter addresses the first research question by reducing rainfall data from the MOPEX data set to DWT and DCT parameters. The parameters are compressed before the rainfall observations are reconstructed. The reconstructed rainfall time series are evaluated to determine which transform is most appropriate for reducing hydrological input data. The findings of this chapter form the foundations for the research presented in subsequent chapters.

Citation: Wright, A., Walker, J. P., Robertson, D. E., and Pauwels, V. R. N.: A comparison of the discrete cosine and wavelet transforms for hydrologic model input data reduction, Hydrol. Earth Syst. Sci., 21, 3827-3838, https://doi.org/10.5194/hess-21-3827-2017, 2017.

3.1 Abstract

The treatment of input data uncertainty in hydrologic models is of crucial importance in the analysis, diagnosis and detection of model structural errors. Data reduction techniques decrease the dimensionality of input data, thus allowing modern parameter estimation algorithms to more efficiently estimate errors associated with input uncertainty and model structure. The discrete cosine transform (DCT) and discrete wavelet transform (DWT) are used to reduce the dimensionality of observed rainfall time series for the 438 catchments in the Model Parameter Estimation Experiment (MOPEX) data set. The rainfall time signals are then reconstructed and compared to the observed hyetographs using standard simulation performance summary metrics and descriptive statistics. The results convincingly demonstrate that the DWT is superior to the DCT in preserving and characterizing the observed rainfall data records. It is recommended that the DWT be used for model input data reduction in hydrology in preference over the DCT.

3.2 Introduction

Rainfall uncertainty is the biggest obstacle hydrologists face in their pursuit of accurate, precise and timely streamflow forecasts (*McMillan et al.*, 2011). Unfortunately, errors in rainfall time series data may lead to hydrological model parameter estimates that produce adequate streamflow simulations only during the calibration period (*Beven*, 2006). This can lead to poor-quality streamflow predictions for independent periods, and low confidence in the ability of streamflow forecasts. Consequently, a precise and accurate representation of rainfall uncertainty is paramount for robust hydrological model parameter estimation, streamflow forecasting and quantitative precipitation forecasts (QPFs). *Robertson et al.* (2013) and *Shrestha et al.* (2015) have demonstrated that skill can be added to QPFs by postprocessing with past observations. As such, skill can be added to QPFs and consequently flood forecasts, through developing a greater understanding of rainfall uncertainty.

The propagation of input errors in rainfall runoff modeling impedes the hydrologic community's ability to validate model structural error. Despite the vast amount of literature on rainfall measurement, estimation, statistical analysis (*Testik and Gebremichael*, 2010) and quality control procedures (*World Meteorological Organization*, 2008), a shroud of uncertainty still surrounds how rainfall and its associated uncertainty should be addressed in rainfall runoff modeling. The implementation of uncertainty analysis in many hydrological applications is also often limited by computational power.

Recent advancements in computational power as well as remote sensing have led to considerable improvements in availability and quality of hydrological observations (*Cloke and Pappenberger*, 2009). These improvements can be leveraged to increase the hydrological and flood forecasting knowledge base and consequently provide water policy decision makers and emergency management services with higher-quality information.
The advancement of computational power has also aided the search for hydrological model parameters that optimally simulate hydrological observations. These approaches initially focused on finding only the global optimum values of the parameters for a given objective function (*Duan et al.*, 1994; *Gan and Biftu*, 1996; *Thyer et al.*, 1999). However, in the past two decades it has been recognized that the uncertainties in model parameters and predictions need to be estimated. Methods that seek to estimate parameter and prediction uncertainty include: Bayesian recursive parameter estimation (*Thiemann et al.*, 2001), the limits of acceptability approach (*Beven*, 2006; *Blazkova and Beven*, 2009), the Bayesian total error analysis (BATEA) framework (*Kavetski et al.*, 2006b,a; *Kuczera et al.*, 2006; *Thyer et al.*, 2009; *Renard et al.*, 2011), the simultaneous optimization and data assimilation (SODA) (*Vrugt et al.*, 2005), the DREAM algorithm and its variations (*Vrugt et al.*, 2005, 2008, 2009a,b; *Vrugt and Ter Braak*, 2011; *Laloy and Vrugt*, 2012; *Sadegh and Vrugt*, 2014), Bayesian model averaging (*Butts et al.*, 2004; *Ajami et al.*, 2007; *Vrugt and Robinson*, 2007), the hypothetico-inductive data-based mechanistic modeling framework of *Young* (2013) and Bayesian data assimilation (*Bulygina and Gupta*, 2011). It is through the development of these parameter estimation algorithms that hydrologists are able to explore input uncertainty.

Kavetski et al. (2006a) and *Vrugt et al.* (2008) identified the need to represent true catchment rainfall and its associated uncertainty using parameters, both applied a parametric approach to estimating true catchment rainfall and its associated uncertainty using a rainfall multiplier to storm events. The use of a parametric representation of rainfall with an effective sampling algorithm provides the ability to jointly estimate hydrologic model parameter distributions as well as input uncertainty. As in most hydrological problems there is a lack of sufficient data to obtain a unique solution. However, *Kavetski et al.* (2006a) and *Vrugt et al.* (2008) found there were sufficient data to estimate both hydrological model parameters and rainfall input. Data reduction transformations offer the potential to reduce the dimensionality of the parameter estimation problem and thus enable a more robust inference. Signal transforms, such as Fourier and wavelet transforms, are examples of data reduction transformations that have been applied in hydrology; however they have not previously been used to reduce the dimensionality of input data.

Fourier transforms use sinusoidal functions to represent the spectral component of an input signal; thus, a periodic signal could be represented using a smaller number of Fourier coefficients than the number of input data points. A pitfall of the Fourier transform is that it represents the spectral components of a signal, without any indication of the time localization of those specific spectral components. In order to account for this, the windowed Fourier transform (WFT), sometimes referred to as the short-time Fourier transform, segments the signal into discrete time windows before performing the Fourier analysis. A major drawback to this approach is that the uncertainty principle of signal processing imposes a limitation on the time and frequency resolutions that can be obtained for a given signal. As a response to this *Daubechies* (1990) produced discrete basis functions with good time and frequency localization. In conjunction with the pyramid algorithm, as described by *Mallat* (1989), this work formed the basis for multi-resolution analysis with the discrete wavelet transform (DWT) (*Polikar*, 1999). The DWT decomposes an input signal into high- and low-frequency components.

Wavelet analysis was first introduced to the geophysical sciences by *Kumar and Foufoula-Georgiou* (1997) and has been adopted for several different applications. Wavelet analysis has been used to assess the performance of hydrological models for parameter estimation (*Schaefli and Zehe*, 2009) to analyze changes over different time periods for both streamflow and preciptation data (*Nalley et al.*, 2012). Various spectral methods have also been applied in hydrology, including the application of discrete Fourier transforms to calibrate water and energy balance models (*Pauwels and De Lannoy*, 2011) and for the calibration of the conceptual rainfall runoff model known as the probability distributed model (PDM) (*De Vleeschouwer and Pauwels*, 2013). While wavelet and spectral methods have been applied in the hydrological sciences, to date there have been no instances in which the suitability of different transforms has been compared for hydrological data reduction applications. *Labat* (2005) has pointed out that Fourier transforms and their derivatives are not well suited to reconstruct hydrologic data, which are generated by transient mechanisms. This is due to the Fourier transforms poor capability to represent sporadic high-frequency events when dimensionally reduced. If model input data reduction techniques are to be accepted by the hydrologic community it is of critical important that the transform used is able to reconstruct transient events. Through a comparative study it will be shown that DWTs are a good multi-resolution alternative to the discrete cosine transform DCT.

Traditionally, transform coefficients are the result of a convolution operation on an input signal. However, the aim of model input data reduction is to estimate these transform coefficients. Hence, they shall be referred to as transform parameters from herein. This paper provides novel theoretical and numerical comparisons of the DCT and DWT in a hydrological context. The ability of both transforms to reproduce key components of hydrological data sets is investigated. The extent to which each transform can reproduce hydrologic data using a decreasing number of parameters will serve as a metric upon which their ability to be used as a tool for model input data reduction for hydrological data will be evaluated. To address the requirements for hydrologic model input data reduction, this paper details (i) theoretical differences between the DCT and DWT, (ii) methodologies to reduce input rainfall to parameters, and (iii) an evaluation of the proposed methodologies using several simulation performance summary metrics.

3.3 Model input data reduction theory

For this study, model input data reduction theory is introduced using a lumped conceptual watershed model. Consider a nonlinear model $\mathcal{F}(\cdot)$, which simulates n discharge values, $\widehat{\vec{Y}} = \{\widehat{y}_1, \ldots, \widehat{y}_n\}$ in mm day^{-1} according to

$$\widehat{\mathbf{Y}} = \mathcal{F}(\vec{\theta}, \widetilde{\mathbf{x}_0}, \widehat{\vec{E}}, \widehat{\vec{R}}), \tag{3.1}$$

where the model input arguments are the $1 \times d$ vector $\vec{\theta}$, with arbitrary model parameter values, the $1 \times m$ vector $\widetilde{\vec{x}_0}$ with values of the initial states in millimeters, and the $1 \times n$ vectors $\widehat{\vec{E}} = \{\widehat{e}_1, \ldots, \widehat{e}_n\}$ and $\widehat{\vec{R}} = \{\widehat{r}_1, \ldots, \widehat{r}_n\}$ which store the observed values of the potential evapotranspiration (PET) and rainfall in mm/day, respectively. Note that $\widehat{\vec{R}}$ is used to represent rainfall and not precipitation, as snow, hail and other forms of precipitation are not considered. The $\widehat{}$ (hat) symbol is used to denote measured quantities and the $\widetilde{}$ (tilde) symbol reflects variables that are either reconstructed or could, in theory, be observed in the field but due to their conceptual nature are difficult to determine accurately. If the traditional hydrological perspective in which the inputs $\vec{E}$ and $\vec{R}$ are considered to be fixed and known quantities is relaxed, and rainfall is now considered unknown, then a new inference problem arises in which the input rainfall is estimated via the treatment of the input rainfall as a series of parameters. Inference problems in which the input is considered unknown can be dealt with using a Bayesian framework. Such inference problems have been considered by *Kavetski et al.*

(2006b) and *Vrugt et al.* (2008) but are outside the scope of this paper. Consequently, for rainfall to be inferred a suitable parametric representation of rainfall must be determined. Given a daily rainfall data record with n observations in millimeters, n rainfall parameters could be used to represent the input hyetograph. This approach would be particularly elegant and parsimonious. Yet, for a 10-year record of daily discharge data, the inference problem would grow from d model parameters to roughly $10 \times 365 + d = 3650 + d$ parameters. These values would need to be estimated from the observed rainfall and discharge data record, respectively. As many hydrological models are already underdetermined the introduction of additional parameters would make the model even less determinable. Additionally an excessive amount of CPU time is required to solve for a 3600+ dimensional posterior parameter distribution. An alternative approach is therefore necessary. Sparse transforms convey large amounts of data using fewer parameters than data points in the observed signal. An input rainfall signal can be reduced to sparse transform parameters. Doing so allows multiple rainfall observations to be modified using a single parameter. Some or all of these transform parameters can be altered before the transform is inverted to produce a new input signal for streamflow simulation and posterior analysis. The use of sparse transforms to represent input time series enables input uncertainty to be explored in great detail. The ability of discrete wavelet and Fourier transformations to reduce hydrological input data to a set of parameters for uncertainty estimation is compared using theoretical and analytical methods.

3.3.1 Overview of the DCT and DWT

Wavelet and Fourier transforms are invertible transforms in which a forward convolution operation can be used to decompose a signal into various components. Similarly, a backwards deconvolution operation can be applied to retrieve the original signal. Fourier-based transforms decompose signals into frequency components and are best used for regular time-invariant signals that do not exhibit time-specific information. Alternatively, wavelet-based transforms decompose signals into frequency and time components. The advantage of using wavelet functions to transform data is that time-specific information about when higher frequency components occur can be preserved. To obtain time-specific information, Fourier-based transforms can be applied over pre-specified temporal windows. Yet, this approach is limited by the uncertainty principle of signal processing. The uncertainty principle of signal processing imposes a lower limit on obtainable resolutions in the time-frequency domain such that

$$\sigma_{\mathrm{t}}\sigma_{\omega} \geq \frac{1}{2}, \tag{3.2}$$

where σ_{t} (s) and σ_{ω} (s^{-1}) are the respective temporal and frequency widths used in the sparse transform.

Applying the uncertainty principle of signal processing (Equation 3.2) it is clear that any attempt to narrow the temporal period analyzed to gain increased resolution in the time domain would be met by a widening of the frequency spectrum, and consequently a loss of resolution in the frequency domain.

Considering that there is no time-frequency window that is able to obtain limitless resolution in both the time and frequency domains, it is clear that an alternative solution must be found. Wavelet transforms can be used to decompose a signal into different levels that consist of different time and frequency resolution windows. Thus, the wavelet transform is able to be configured to simultaneously obtain high levels of resolution in both the time and frequency domains. For a more detailed discussion on wavelets and sparse transforms the reader is referred to *Mallat* (2009).

3.3.2 Discrete cosine transform

The DCT (*Ahmed et al.*, 1974) is a version of the WFT that has advantageous properties for the field of data compression. Due to the boundary conditions of the cosine function, the DCT is well suited to represent an observed input signal with a minimal number of parameters; in this case rainfall, $\widehat{\vec{R}}(t)$. The DCT parameters $\vec{p}(i)$ are calculated as

$$\vec{p}(i) = \mathbf{w}(i) \sum_{t=1}^{\mathrm{n}} \widehat{\vec{R}}(t) \cos\left[\frac{\pi}{2\mathrm{n}}(2t-1)(i-1)\right], \tag{3.3}$$

where $i = 1, 2, \ldots, \mathrm{n}$ and

$$\mathbf{w}(i) = \begin{cases} \frac{1}{\sqrt{\mathrm{n}}}, & i = 1 \\ \sqrt{\frac{2}{\mathrm{n}}}, & 2 \leqslant i \leqslant \mathrm{n}. \end{cases} \tag{3.4}$$

The convolution process can be reversed to reconstruct the observed signal using the inverse transform

$$\widetilde{\vec{R}}(t) = \sum_{i=1}^{\mathrm{n}} \mathbf{w}(i)\vec{p}(i)\cos\left[\frac{\pi(2t-1)(i-1)}{2\mathrm{n}}\right], \tag{3.5}$$

where $t = 1, 2, \ldots, \mathrm{n}$.

3.3.3 Discrete wavelet transform

Using the pyramid algorithm, depicted in Fig 3.1, *Mallat* (1989) first described the decomposition of an input signal into multi-resolution components using high- and low-pass filters. Each stage of decomposition is referred to as a level. An advantage of using wavelets is that decomposition can be performed using a variety of different wavelet families. This allows for signals with differing properties to be analyzed using the same methodology. The most commonly used wavelet family is the Daubechies wavelets (*Daubechies*, 1990). Each wavelet within each family consists of a scaling $\mathrm{h}(m)$ and wavelet $\mathrm{w}(m)$ function, where m denotes the length along the scaling and wavelet function. The scaling and wavelet functions are used in the low- and high-pass filtering sequences, respectively. Whilst there are numerous wavelet families that can be chosen for analysis, this study applies the most commonly used Daubechies wavelets. Depending on the choice of wavelet stepwise convolutions of the input signal are performed over the filter length L. $j_{\max}$ imposes an upper limit on the level of decomposition j that a signal can be decomposed into, where

$$j_{\max} = \left\lfloor \log_2\left(\frac{\mathrm{n}+\mathrm{L}-1}{2}\right)\right\rfloor, \tag{3.6}$$

in which $\lfloor . \rfloor$ is the floor operator. The input signal is then convoluted by being passed through high- and low-pass filters, where

$$\vec{p}_j^{\,\mathrm{L}}(i) = \begin{cases} \sum_{m=1}^{\mathrm{L}} \widetilde{\vec{R}}(2i-m-1)\mathrm{w}(m), & j = 1 \\ \sum_{m=1}^{\mathrm{L}} \vec{p}_{j-1}^{\,L}(2i-m-1)w(m), & j > 1. \end{cases} \tag{3.7}$$

is the low pass and

$$\vec{p}_j^{\,\mathrm{H}}(i) = \begin{cases} \sum_{m=1}^{\mathrm{L}} \widetilde{\vec{R}}(2i-m-1)\mathrm{h}(m), & j = 1 \\ \sum_{m=1}^{\mathrm{L}} \vec{p}_{j-1}^{\,L}(2i-m-1)\mathrm{h}(m), & j > 1. \end{cases} \tag{3.8}$$

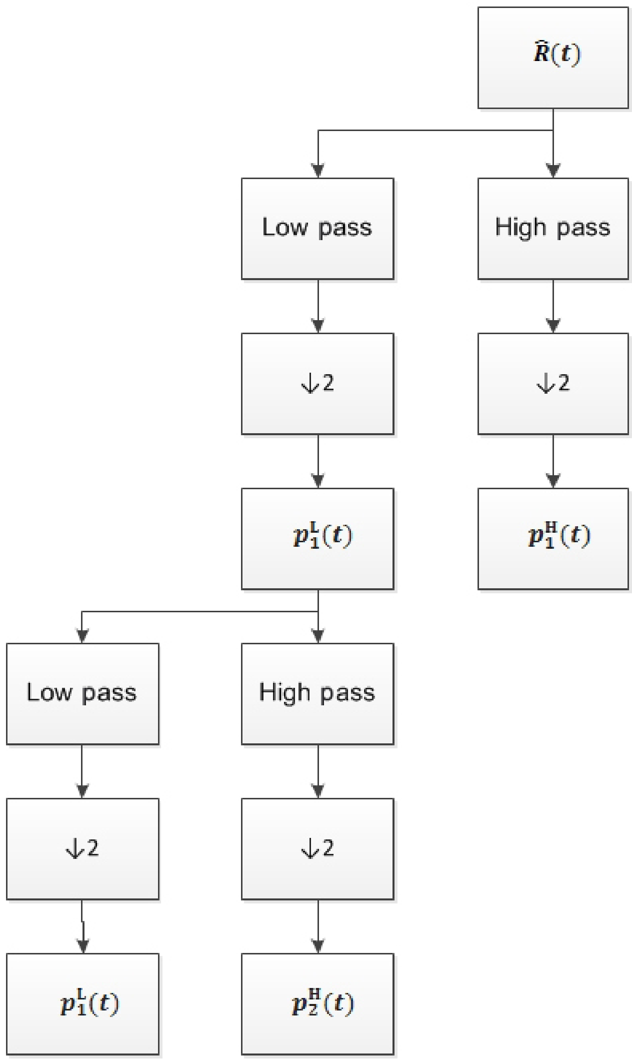

FIGURE 3.1: A schematic showing the pyramid algorithm used to decompose and downsample ($\downarrow$ 2) an input signal ($\widehat{\vec{R}}$) into high- and low-frequency components. The input signal is filtered using the high- and low-pass filters described in Equations 4.6 & 4.7 before being downsampled to produce the level one high- and low-pass parameters. The low pass parameters are now used as input for the high- and low-pass filters. This process of filtering and downsampling is repeated until the desired level of decomposition is met.

is the high pass, $i = 1, \ldots, \mathrm{n}_{j-1} + \mathrm{L} - 1$ and refers to the ith parameter, $j = 1, \ldots, j_{\max}$ and refers to the jth level, m refers to the mth filter coefficient. The resultant low-pass $\vec{p}_j^{\,L}(i)$ and high-pass $\vec{p}_j^{\,\mathrm{H}}(i)$ parameters are commonly referred to as approximation and detail parameters, respectively. After the input signal is passed through the high- and low-pass filters there is an issue of redundancy that needs to be dealt with. The filters split the input signal into high- and low-frequency components that each contain roughly half the information of the input signal. As the length of each of the resultant approximation and detail parameter series is equivalent to the length of the input signal, each of the parameter series must be downsampled. The process of downsampling removes every other parameter. It is the process of high- and low- pass filtering followed by downsampling that enables the DWT to analyze multi-resolution components of a signal. After downsampling the length of the resultant approximation and detail parameter series is

$$\mathrm{n}_j = \begin{cases} \left\lfloor \frac{\mathrm{n}+\mathrm{L}-1}{2} \right\rfloor, & j = 1 \\ \left\lfloor \frac{\mathrm{n}_{j-1}+\mathbf{L}-1}{2} \right\rfloor, & j > 1. \end{cases} \tag{3.9}$$

where n_j refers to the length of the series at the $j\mathrm{th}$ level. If further decomposition is required the downsampled low-pass may be fed back into the filters until the resultant parameters can no longer be split any further. An iteration of this process is shown in Fig 3.1. To reverse the decomposition process and reconstruct a signal, upsampling is performed on the parameter series before the lower level parameters are obtained through

$$\vec{p}_{j-1}(i) = \sum_{m=\lceil i/2 \rceil}^{\lfloor (\mathrm{L}-1+i)/2 \rfloor} \left(\vec{p}_j^{\,\mathrm{H}}(i)\mathrm{h}(2m-i) \right) \left(\vec{p}_j(i)\mathrm{w}(2m-i) \right), \quad j > 1, \tag{3.10}$$

where $\lceil . \rceil$ is the ceiling operator and the input signal is reconstructed using

$$\widetilde{\vec{R}}(i) = \sum_{m=\lceil i/2 \rceil}^{\lfloor (\mathrm{L}-1+i)/2 \rfloor} \left(\vec{p}_j^{\,\mathrm{H}}(i)\mathbf{h}(2m-i) \right) \left(\vec{p}_j(i)w(2m-i) \right), \quad j = 1. \tag{3.11}$$

3.4 Data

This study utilizes data from the Model Parameter Estimation Eperiment (MOPEX) data set. The 10 years of rainfall data spanning the 1990s for 438 catchments in the United States of America (USA) are used to compare the suitability of the DWT and DCT to represent rainfall time series. The catchments used in this study were chosen to ensure they had sufficient rain gauge density and represented a range of catchment sizes and climates. Rainfall for the Leaf River catchment (Collins, Mississippi), a catchment that is frequently used for hydrological studies (*Sivakumar*, 2001; *Tang et al.*, 2006; *Bulygina and Gupta*, 2011), is used to compare the DWTs' and DCT's ability to reconstruct high-magnitude rainfall events. A single rainfall product for each catchment is used for analysis at a daily time step. A complete description of the selection process and MOPEX data set is given by *Schaake et al.* (2006). No streamflow data are used in the experiment.

3.5 Experiment design

This experiment does not involve the use of any hydrological models. Due to this and the nature of the transforms there are no calibration and evaluation periods. A major use of both the DWT and DCTs has been in image compression consequently, the observed input signals were compressed and decompressed using a methodology similar to that used in image compression. In order to determine which transform's parameters are able to effectively store the most hydrological input data, both DWT and DCT parameters will be compressed to varying extents for the MOPEX rainfall time series.
The process undertaken involves a number of steps. Firstly, before any compression is applied, the original rainfall signal for a given catchment is transformed into DCT and DWT parameters using equations 3.3 and 3.4 and equations 3.7 to 3.9 for the DCT and DWT respectively. Secondly, each transform is compressed by iteratively zeroing out parameters that provide a low degree of information, these parameters are those closest to zero. A threshold value T (mm) applies a lower limit, for which transform parameters above the threshold are retained. This threshold is iteratively increased until the compressed transform is composed of the desired number of remaining parameters k and percent of original parameters POP is met.

$$\mathrm{POP}(T) = 100 \cdot \left(\frac{k}{\mathrm{n}}\right), \tag{3.12}$$

where k becomes smaller as the threshold T increases and $\lim_{\mathrm{T}\to\infty} \mathrm{POP} = 0$. The next step is to reconstruct the observed signal from the compressed transform parameters using Equations 3.5 and 4.9 for the DCT and DWT respectively. After the reconstruction has been performed, a comparison between the reconstructed and observed rainfall can be made. Lastly, this process is iterated for different POPs as well as for each catchment within the data set.

To provide a meaningful comparison between the DCT's and DWT's ability to reproduce different rainfall time series with an increasing POP, a number of simulation performance summary metrics are used. Following *Moriasi et al.* (2007), a combination of graphical techniques and dimensionless and error index statistics that are widely accepted by the hydrological community were adopted for model evaluation. The Nash-Sutcliffe efficiency (NSE) and the root mean square error (RMSE) to standard deviation ratio (RSR) of the observed input signal $(\mathrm{RSR} = \mathrm{RMSE}/\sigma_{obs})$ are used to compare the performance of the reconstructed rainfall signal with the observed rainfall signal. Once the reconstructed signals are obtained, further comparison with the observed rainfall will be made using the bias summary metric. The variance, kurtosis and skewness of the reconstructed signals will be compared with those of the observed signal. The bias is calculated as $\sum_{t=1}^{n}[\widehat{\vec{R}}(t) - \widetilde{\vec{R}}(t)]/n$, where $\widehat{\vec{R}}(t)$ and $\widetilde{\vec{R}}(t)$ are the observed and reconstructed rainfall signals, respectively. The reconstructed variance, kurtosis and skewness are all normalized by the observed input signals variance, kurtosis and skewness respectively. The peak error (PE) is the peak rainfall error over the 10-year period. It is used to compare the reconstructed and observed signals for seasonal and flood forecasting situations. The PE is normalized by the peak height of the observed input signal. Further, the number of rain events missed is computed for each reconstruction by; flagging original or reconstructed observations that exhibit no rainfall. Either the absolute difference between the reconstructed and original observation is less than 0.01 or the ratio of the reconstructed and original observation is equal to 0 or larger than 10. Lastly, reconstructed rainfall using the DCT and DWT will be presented for the Leaf River catchment to compare each transform's ability to reconstruct high-magnitude rainfall events.

3.6 Results

Fig 3.2 shows the relationships between RSR and the number of transform parameters using the DCT and DWT for three different catchments, Arroyo Chico, Skykomish River and Ohio Brush Creek. These catchments represent the smallest, largest and mean rainfall volumes for the MOPEX data set, respectively. It is clear that for all but the highest POP the DWT is able to reconstruct the observed signal with lower RSR than the DCT and that as the rainfall volume increases the RSR decreases. For intermediate POPs the DWT is able to reconstruct the observed signal with significantly better RSR than the DCT. As the POP approaches both 100% and 0% there is little discernible difference between the DCT and DWT reconstructions.

By comparing the reconstructed DWT and DCT signals, using 20 POP and the observed rainfall signal as a reference, a histogram for the NSE is shown for all catchments in Fig 3.3. Each frequency count in the histogram represents a catchment from the MOPEX data set. The reconstructed DWT signals are clearly able to better simulate the observed rainfall signal. All DWT reconstructed rainfall signals obtained a higher NSE than the DCT reconstructed rainfall signals. Table 3.1 shows that as the transforms are compressed and fewer parameters are used in the reconstruction, the mean NSE for the DWT stays much closer

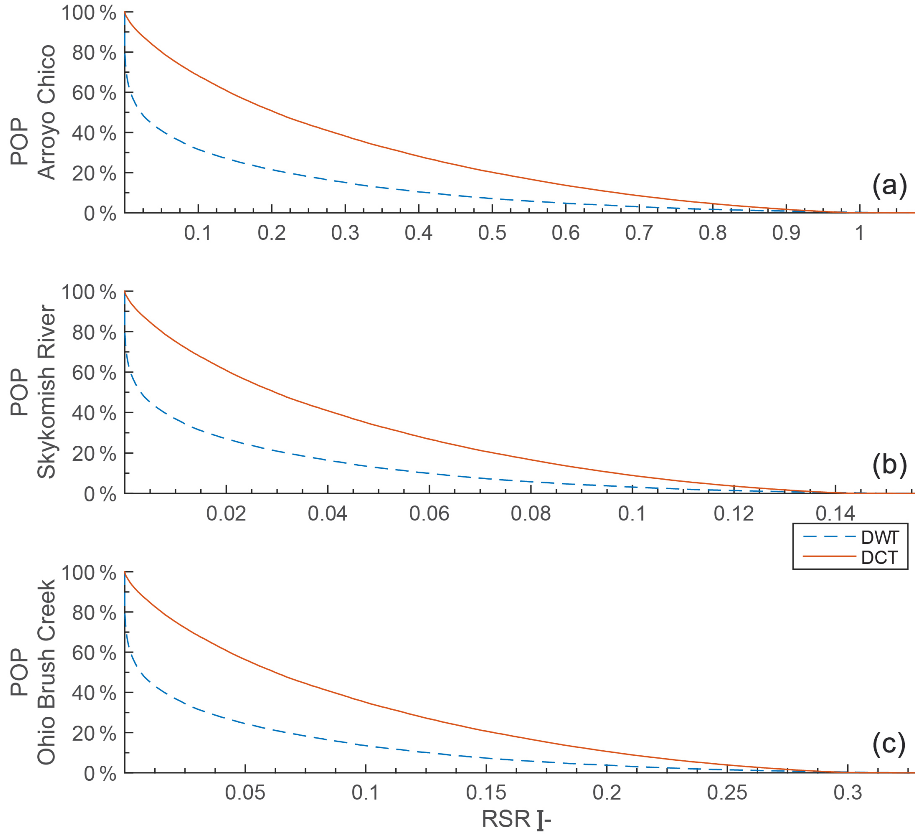

FIGURE 3.2: Empirical plots showing the relationship between RSR and the POP used for reconstructing an input rainfall signal using the DWT and DCT. The three catchments, from the top to the bottom of the figure, represent the smallest, largest and mean rainfall volumes throughout the 1990s for the MOPEX data set.

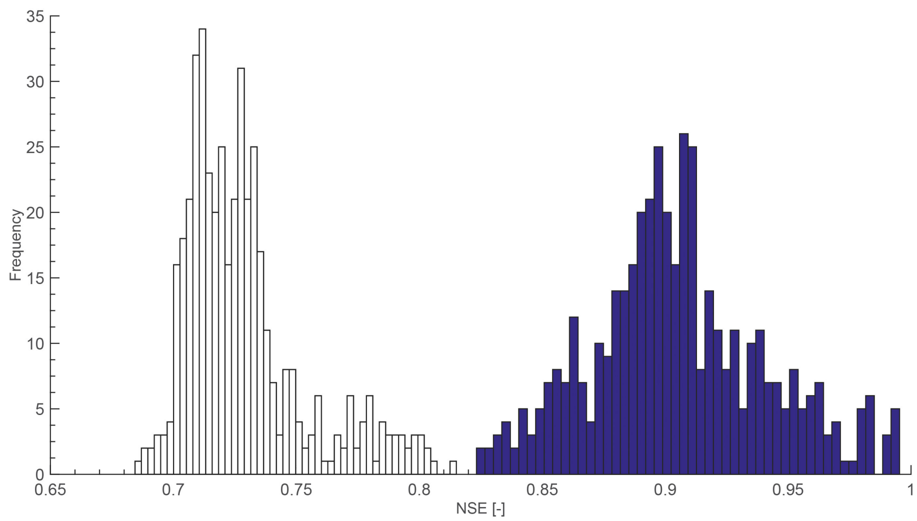

FIGURE 3.3: Histogram representing the reconstructed DWT (dark bins) and DCT (clear bins) NSE when compared to the observed rainfall signal. Rainfall is reconstructed after the input signal is compressed to 20 POP. Each frequency count represents a catchment from the MOPEX data set.

to the ideal value of 1 than the DCT. Further the standard deviation of NSE becomes much

TABLE 3.1: The mean and standard deviation (SD) of NSE for the DWT and DCT using a different POP.

	NSE DWT		NSE DCT	
POP	Mean	SD	Mean	SD
40%	0.988	0.007	0.918	0.010
30%	0.965	0.017	0.844	0.016
20%	0.905	0.036	0.729	0.025
10%	0.746	0.070	0.522	0.037

larger for the DWT.

Fig 3.4 compares the RSR for the DCT and DWT using four different POPs. A 1:1 line is included in all subplots and each point represents a catchment from the data set. If the data

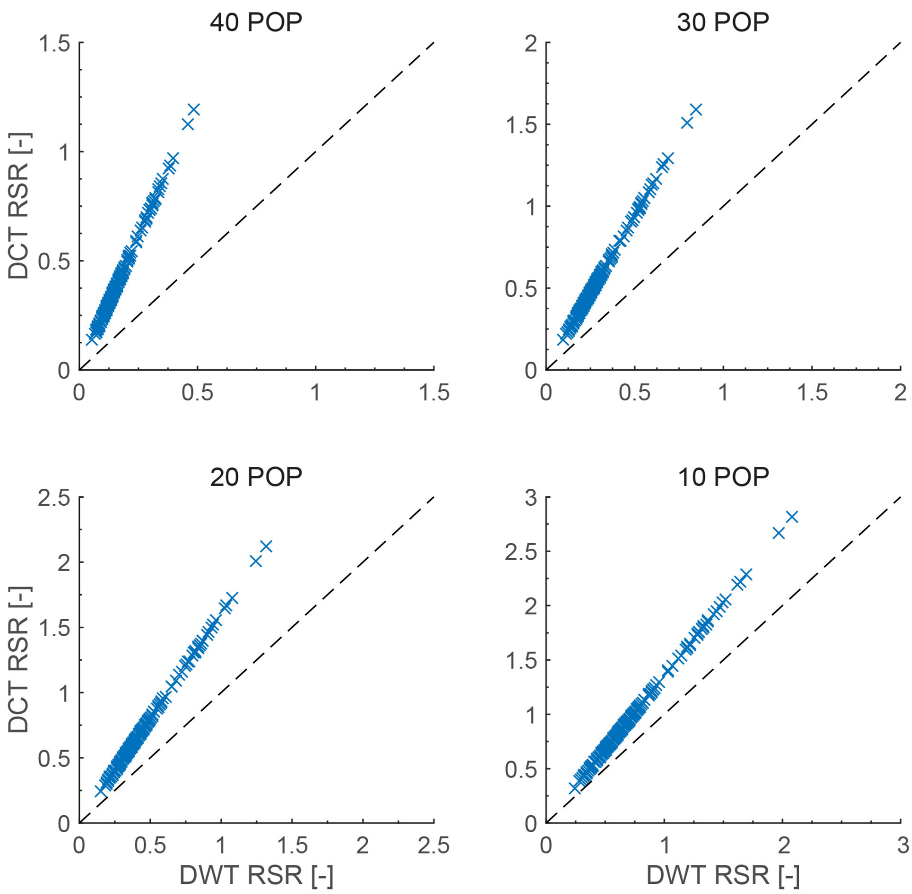

FIGURE 3.4: Comparative plots of RSR for the DCT and DWT using a different POP. Each data point represents a catchment.

points fall above the 1:1 line, then for that catchment and POP the DWT is able to reconstruct the input rainfall signal with lower RSR. Again it is found that DWT is always able to reconstruct the original signal with lower RSR than the DCT reconstructions for all POPs. In a similar fashion to that discussed regarding Fig 3.2 it is observed that as the POP approaches 0% the difference between the DWT and DCT reconstructions becomes smaller.

The bias, variance and skewness observed in the reconstructed signals for each catchment are shown in Fig 3.5 for different POPs. The DWT reconstructions are able to maintain a

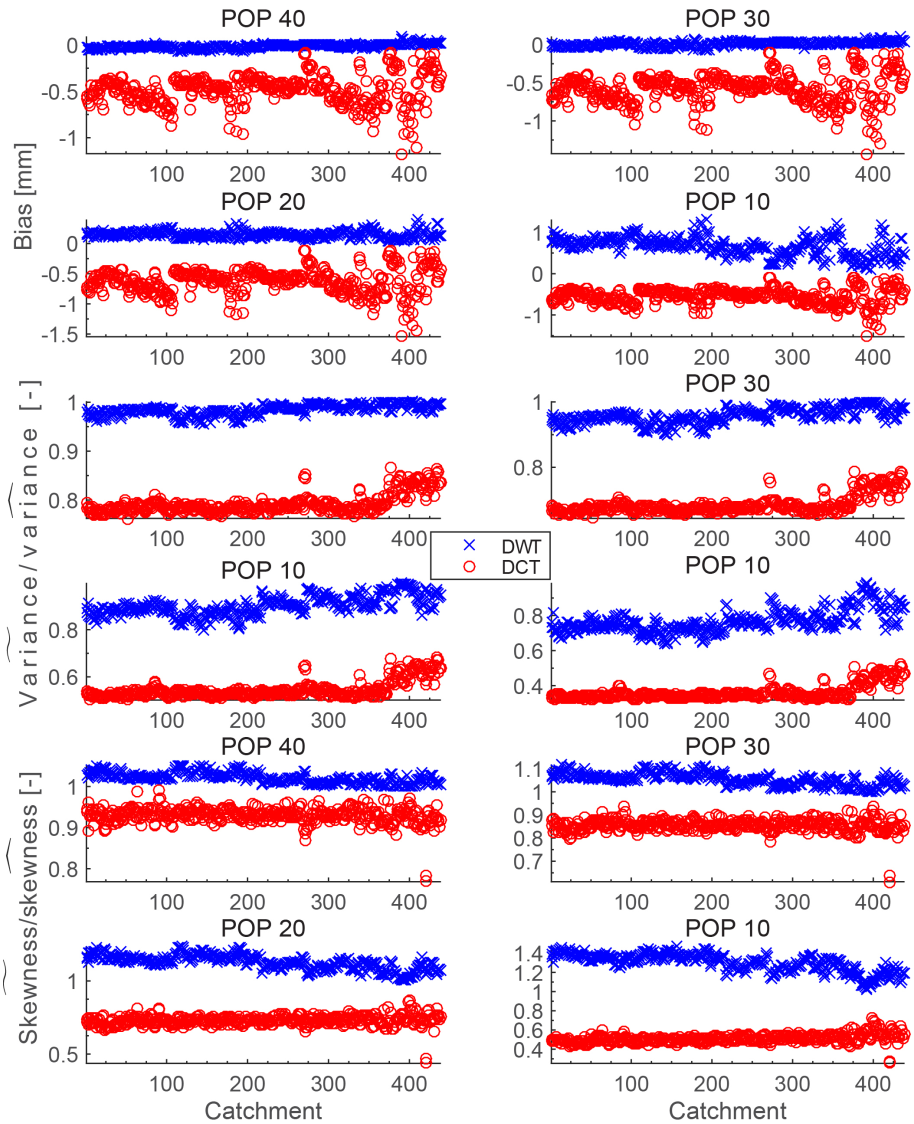

FIGURE 3.5: Bias and normalized variance and skewness of the reconstructed DWT and DCT signals for each catchment using a different POP.

smaller bias than the DCT reconstructions at different POPs for all of the catchments. As the POP decreases the bias becomes increasingly positive and negative for the DWT and DCT, respectively. The distribution of the bias becomes more dispersed for both the DCT and DWT as the POP decreases. The bias can be seen to be dependent on the transform and POP used as well as the catchment being analyzed. Both the DWT and DCT never reconstruct the observed signal with greater variance than that of the observed rainfall signal.

As the POP decreases the normalized variance for the DCT moves further away from unity than the normalized variance for the DWT. The reduction in normalized variance means that, as the POP decreases, both the DWT and especially the DCT reconstructions will have fewer extreme values, when compared to the observed rainfall. The normalized skewness is a measure of symmetry that describes whether or not the reconstructed signal is more positively skewed (more than 1) or less positively skewed (less than 1) than the observed input signal. All of the reconstructed and observed signals had a positive normalized skewness. When compared to the observed signal the DWT becomes increasingly skewed as the POP is reduced. The opposite of this is observed for the DCT. This indicates that, when compressed the DWT and DCT will reconstruct the observed rainfall signal with a greater and lower number of values close to zero when compared to the observed signal, respectively. This does not mean that the total volume will be any lower than the total volume of rainfall observed. This is made evident by the low bias observed in Fig 3.5.

The normalized kurtosis and PE for all catchments using different POPs are shown in Fig 3.6. The measure of kurtosis describes how much the fraction of the distributions variance is explained by extreme deviations. Consequently, a normalized kurtosis value larger than 1 indicates that the reconstructed signals variance is explained more by extreme deviations than the observed input signal. This is likely to be the result of more rainfall values being reconstructed at the extremities than those of the observed rainfall series. A value smaller than one indicates that the variance is described less by extreme deviations than the observed input signal. Similarly, this is likely to be the result of fewer rainfall values being reconstructed at the extremities than those of the observed rainfall series. It is worth noting that a reconstructed time series can have the same variance yet different kurtosis than the observed rainfall time series. As the POP decreases, the dispersion of normalized kurtosis and skewness increases, and the normalized kurtosis and skewness for the DWT and DCT reconstructions become larger and smaller than unity, respectively. With decreasing POP the normalized PE for the reconstructed DWT signal remains small and relatively consistent when compared to the normalized PE for the reconstructed DCT signal.

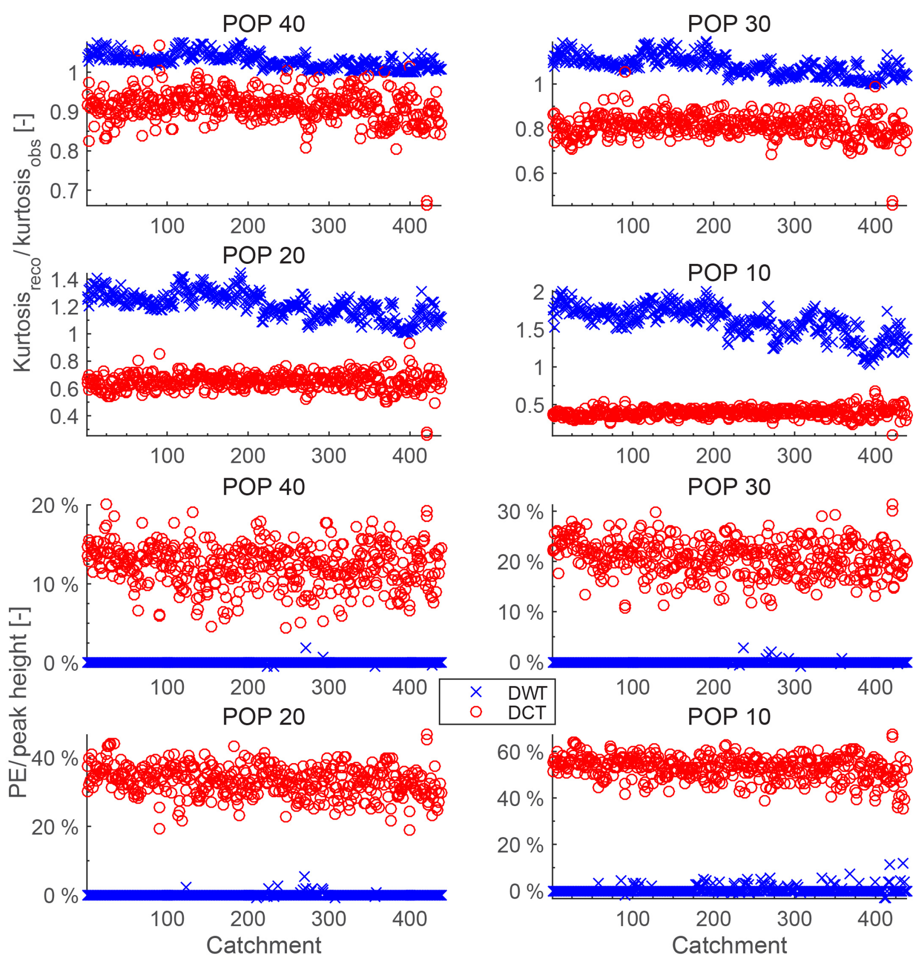

FIGURE 3.6: Normalized kurtosis of the reconstructed DWT and DCT signals and percentage PE for the reconstructed DWT and DCT signals for each catchment using a different POP.

3.7 Discussion

Fig 3.3 shows that the DWT and DCT are able to reconstruct the observed input signals with good efficiency using 20 POP. However, the DWT consistently outperforms the DCT. Fig 3.2 shows that as the POP decreases from 100% the DWT is able to reconstruct the input signal with increasingly lower RSR than the DCT, the gap in performance is largest for 40 POP. As the POP continues to decrease towards 0%, the gap in RSR reduces to zero. It is interesting to note that the DWT perfectly reconstructs the observed rainfall signal with as many parameters as there are rainy days whereas the DCT does not.

As the bias for the DWT is consistently close to zero, the use of the DWT for rainfall input data reduction is likely to be beneficial for hydrologic studies that have short time steps and involve rainfall as an input. Whilst modification of the DWT parameters may slightly overestimate input rainfall, it is not as significant as the consistent underestimation of input rainfall by the DCT. The diminishing ability of both the DWT and DCT to match the input rainfall signal variance indicates that both transforms smooth out input data towards the mean. This behavior is more significant for the DCT than the DWT. Consequently, when used as a technique for input data reduction, the DWT will reconstruct temporal variances better than the DCT. The increased skewness for the reconstructed DWT signals compared to the observed input signals indicates that there is an increased reconstruction of low-magnitude rainfall events. On the contrary, the decreased normalized skewness for the reconstructed DCT signals indicates that a number of the low-magnitude rainfall events are tending to be reconstructed towards the mean. The kurtosis results shown in Figure 3.6 demonstrate that, when compared to the observed input signal, events of extreme deviation explain more of the variance for the reconstructed DWT and less of the variance for the reconstructed DCT. Consequently, as the nature of the extreme deviations is a critical piece of information, the use of the DCT for model input data reduction for hydrologic studies that have short time steps and involving rainfall as an input is not recommended. It is also seen in Figure 3.6 that the DCT is more likely to miss peak rainfall height information. Consequently, care needs to be taken when choosing a transform when peak height is critical. Further, the DCT should not be used for studies involving flood forecasting situations where the accuracy of peak height is critical.

Whilst it is important that rain gauges measure high-magnitude rainfall events with accuracy and precision, it is also important that low-magnitude rainfall events are recorded. Consequently, when evaluating the merits of the DCT and DWT to reconstruct rainfall it would be prudent to analyze the frequency in which each transform is either unable to reconstruct a rainfall event or erroneously constructs a rainfall event. Table 3.2 illustrates that, at times, both transforms will either fail to reconstruct a low-magnitude rainfall event or will erroneously construct a rainfall event when there was none observed in the original rainfall time series.

TABLE 3.2: The mean and standard deviation (SD) for the number of missed rainfall events for the DWT and DCT using a different number of parameters.

	Number of missed rainfall events			
	DWT		DCT	
POP	Mean	SD	Mean	SD
40%	239.004	117.317	587.934	155.375
30%	398.005	138.793	645.495	159.378
20%	581.591	145.769	696.288	168.524
10%	852.340	168.590	748.075	184.910

In general, the DWT outperforms the DCT. The exception to this is at 10 POP. This is a result of the discrete nature of the DWT analysis function as opposed to the continuous analysis function used in the DCT. As the POP decreases towards zero both transforms miss more rainfall events.

Due to rapid increases in rainfall intensity, high-magnitude rainfall events tend to have high-frequency components. In Fig 3.7 the smoothing of high-frequency, high-magnitude rainfall events by the DCT is made evident by the lower slope of the linear least squares fit for the DCT reconstruction of Leaf River observed rainfall data when compared to the DWT. This

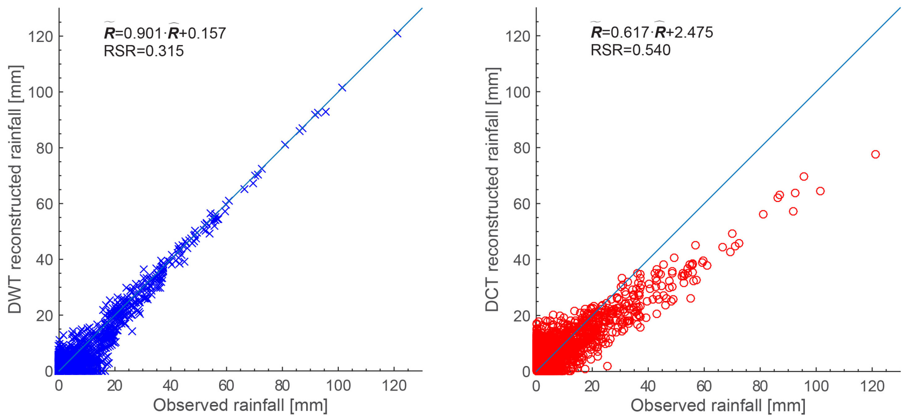

FIGURE 3.7: Comparison of the reconstructed DCT and DWT signal for the Leaf River (Collins) catchment using 20 POP.

shows that the compressed DWT is able to retain more detail for high-magnitude rainfall events than the DCT. Using 20 POP, 730 DWT parameters are able to reconstruct observed rainfall with an $\text{RSR} = 0.315$, whereas 730 DCT parameters are able to reconstruct observed rainfall with an $\text{RSR} = 0.540$. Fig 3.7 and Fig 3.8 shows that the DWT often misses and sometimes smooths out low-magnitude rainfall events; the DCT however does, reconstruct inaccurate rainfall at these times. Figure 3.8 also demonstrates, that at lower POPs, the DCT will smooth out and underestimate high-magnitudes events whilst the DWT will maintain accuracy and precision.

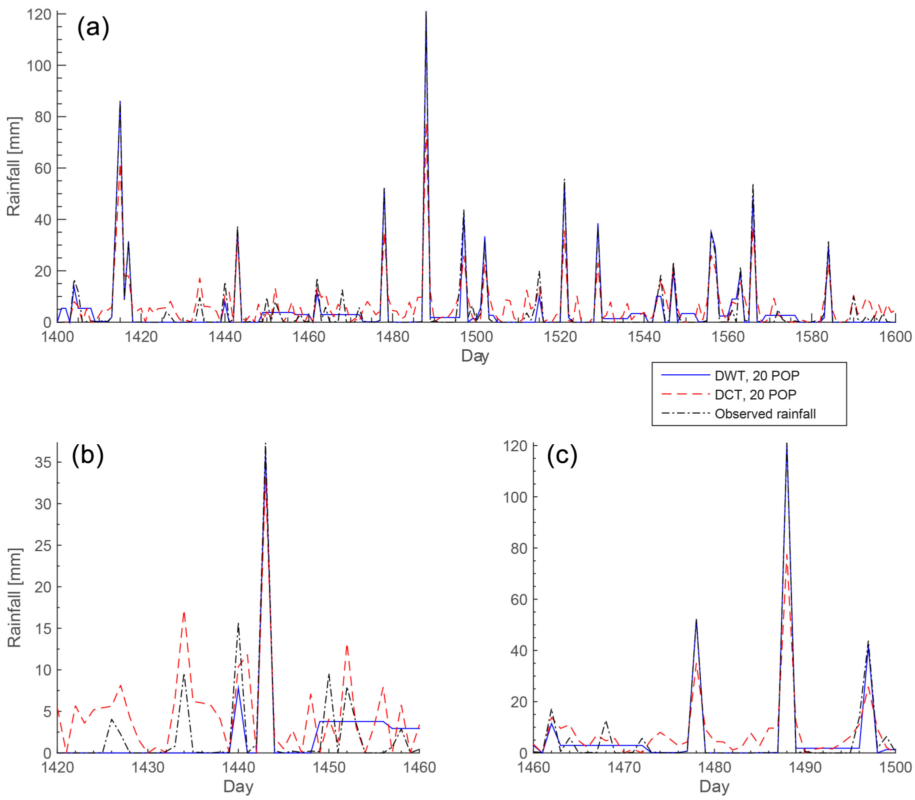

FIGURE 3.8: Panel (**a**) shows a time series comparison of the reconstructed DCT and DWT signals for the Leaf River (Collins) catchment using 20 POP for a period of 200 days. Panels (**b** and(**c**) are smaller windows of the same time series during both low- and high-rainfall periods.

3.8 Conclusions

Succinct descriptions of the DCT and DWT were provided to determine the suitability of each transform to be used as a tool for hydrologic model input data reduction. Due to their different construction, each transform provides different possibilities for use in model input data reduction. Since it is infeasible to estimate all transform parameters, the modeller could choose to estimate high- or low frequency parameters of the DCT. This would result in minimal control of the temporal component being modified. Due to the multi-level decomposition of an input signal into high- and low-frequency parameters by the DWT, the modeller is able to specify the estimation of both time and frequency components. Hence, portions of the input data record can be targeted for estimation. The use of the DWT as a hydrologic model input data reduction technique allows the modeller more flexible options. A comparison of the DWTs' and DCTs' ability to reconstruct MOPEX rainfall data using standard simulation performance summary metrics, descriptive statistics and peak errors was then made and it was found that the DWT is most efficient at preserving high-magnitude and transient rainfall events. Thus, it is recommended that the DWT be used as a model input data reduction technique for hydrologic studies that have short time steps and involve rainfall as an input. Considering that the bias for the reconstructed DWT rainfall signal is consistently lower than that of the reconstructed DCT signal and that the skewness, kurtosis and variance are also closest to the input rainfall signal, it is recommended that the DWT also be used as a model input data reduction technique for hydrologic studies that have long time steps with rainfall as an input.

Acknowledgements

The authors would like to extend their gratitude to Jasper Vrugt, Hamid Bazargan and the anonymous reviewers for their comments and recommendations. This work was supported by the Multi-modal Australian Sciences Imaging and Visualisation Environment (MASSIVE) (www.massive.org.au), a Monash University Engineering Research Living Allowance stipend and a top up scholarship from the Bushfire & Natural Hazards Cooperative Research Centre. Valentijn Pauwels is funded by ARC grant FT130100545.

Chapter 4

Estimating rainfall time series and model parameter distributions using model data reduction and inversion techniques

Overview

This chapter builds upon findings from Chapter 3 to answer the second research question. The DWT is used to reduce rainfall data for the catchment of Warwick. In conjunction with a likelihood function that considers rainfall and streamflow, the reduction of rainfall data allows for rainfall time series to be estimated along with model parameters. The estimated rainfall time series provide an increased understanding of hydrologic uncertainty. The findings of this chapter form the foundations for the research presented in subsequent chapters.

Citation: Wright, A., Walker, J. P., Robertson, D. E., and Pauwels, V. R. N.: A comparison of the discrete cosine and wavelet transforms for hydrologic model input data reduction, Hydrol. Earth Syst. Sci., 21, 3827-3838, https://doi.org/10.5194/hess-21-3827-2017, 2017.

4.1 Abstract

Floods are devastating natural hazards. To provide accurate, precise and timely flood forecasts there is a need to understand the uncertainties associated within an entire rainfall time series, even when rainfall was not observed. The estimation of an entire rainfall time series, and model parameter distributions from streamflow observations in complex dynamic catchments adds skill to current areal rainfall estimation methods, allows for the uncertainty of entire rainfall input time series to be considered when estimating model parameters, and provides the ability to improve rainfall estimates from poorly gauged catchments. Current methods to estimate entire rainfall time series from streamflow records are unable to adequately invert complex non-linear hydrologic systems. This study aims to explore the use of wavelets in the estimation of rainfall time series from streamflow records. Using the Discrete Wavelet Transform (DWT) to reduce rainfall dimensionality for the catchment of Warwick, Queensland, Australia, it is shown that model parameter distributions and an entire rainfall time series can be estimated. Including rainfall in the estimation process improves streamflow simulations by a factor of up to 1.78. This is achieved while estimating an entire rainfall time series, inclusive of days when none was observed. It is shown that the choice of wavelet can have a considerable impact on the robustness of the inversion. Combining the use of a likelihood function that considers rainfall and streamflow errors with the use of the DWT as a model data reduction technique allows the joint inference of hydrologic model parameters along with rainfall.

4.2 Introduction

Floods can have significant economic, social and environmental impacts (*Brouwer and Van Ek*, 2004). Cost benefit analyses and environmental and social impact assessments are common evaluation methods available to water policy decision makers (*Hajkowicz and Collins*, 2007). Flood forecast skill greatly influences societal resilience to floods. However, without accurate, precise and timely rainfall information, the value of such analytical tools is rendered subjective.

Currently, rainfall uncertainty is the biggest obstacle hydrologists face in their pursuit toward obtaining accurate, precise and timely flood forecasts (*McMillan et al.*, 2011). Operational flood forecasters tend to adhere to familiar flood forecasting procedures, including semi-distributed event-based hydrological models (*Pagano*, 2009). Consequently, it is often not possible for reliable flood forecasts to be issued until the catchment's response to rainfall has been observed (*Elliott*, 1997). Hydrologists look to overcome this by using continuous hydrological models, but the lack of reliable rainfall inputs from quantitative precipitation forecasts (QPFs) impedes the development of robust flood forecasts (*Hapuarachchi et al.*, 2011). *Robertson et al.* (2013) and *Shrestha et al.* (2015) have demonstrated that skill can be added to raw QPFs by postprocessing the raw QPFs using past observations as input into a methodology that combines a simplified version of the Bayesian joint probability with the Schaake shuffle (*Clark et al.*, 2004). The Schaake Shuffle is a methodology to reconstruct space-time variability in forecasted precipitation and temperature fields. The combination of the use of model input data reduction techniques with parameter estimation algorithms allows links to be explored between rainfall input error, QPF postprocessing algorithms and errors associated with model structure, parameter estimation, and systematic and random errors associated with observations.

Due to complex interactions between Hortonian overland flow, saturation excess overland flow, interflow and groundwater flow, discrepancies are quite often noticed between similar rainfall events and the corresponding runoff, and vice versa. Hence the process of estimating rainfall from streamflow observations is an ill-posed problem. As a large proportion of hydrologists favor deterministic models (*Pappenberger and Beven*, 2006), it is not surprising that some attempts to estimate rainfall from runoff have taken a deterministic rather than probabilistic approach (*Hino*, 1986; *Kirchner*, 2009). While using analytical inversion to estimate rainfall from streamflow, *Kirchner* (2009) draws attention to the fact that, of the components of the water balance, only streamflow can be considered a catchment-scale observation. Hence the authors ask, can streamflow and / or soil moisture be used to estimate catchment scale rainfall time series?

The root zone soil moisture state can have a large impact on a catchment's rainfall-runoff characteristics (*Grayson et al.*, 2006). Recent studies e.g. (*Crow*, 2007; *Pellarin et al.*, 2008; *Crow et al.*, 2009, 2011; *Kucera et al.*, 2013; *Pellarin et al.*, 2013; *Brocca et al.*, 2014) focus on using soil moisture to correct and estimate rainfall accumulations. *Brocca et al.* (2014) coined the phrase '*Soil as a Natural Rain Gauge*'. In the Soil Moisture to Rain (SM2RAIN) algorithm (*Brocca et al.*, 2014), rainfall estimates are retrieved from the inversion of the soil water balance equation, assuming that all rainfall infiltrates. *Ciabatta et al.* (2015) uses the SM2RAIN algorithm to nudge satellite precipitation estimates in order to estimate daily rainfall; *Abera et al.* (2016) validates this product in a comparative study. If rainfall estimates that are based on a satellite soil moisture product are to be used in a flood forecasting situation it is imperative that the rainfall estimate is up to date and that the satellite soil moisture images are obtained immediately prior to the flood. With the continued improvement of satellite rainfall and soil moisture measurement missions, such as the Global Precipitation Measurement (GPM) mission (*Hou et al.*, 2014) and the Soil Moisture Active Passive (SMAP) (*Entekhabi et al.*, 2010) mission, it is expected that the methods outlined by *Crow et al.* (2011); *Brocca et al.* (2014) and *Ciabatta et al.* (2015) will become more valuable for estimating rainfall time series in the future. Yet there are currently no methods that use both streamflow and soil moisture to estimate rainfall. (*Kavetski et al.*, 2006b,a; *Vrugt et al.*, 2008; *Renard et al.*, 2010, 2011) Rainfall estimation methods that solely rely on streamflow measurements maintain good temporal resolution, yet have been subject to poor performance in catchments that have complex rainfall-runoff characteristics and exhibit highly nonlinear rainfall-runoff behavior. In an early attempt to estimate rainfall from streamflow, *Hino* (1986) separated time series of daily discharge into their respective runoff components using coefficients obtained from fitting an Auto-Regressive Moving Average (ARMA) model. Using the law of parsimony, otherwise known as Occam's razor *"Entities should not be multiplied unnecessarily"* *Kirchner* (2009); *Teuling et al.* (2010); *Adamovic et al.* (2015), and *Rusjan and Mikoš̃* (2015) used first order approximations to analytically invert the water balance equation. These approximations also assume that all rainfall infiltrates and are not able to estimate rainfall when streamflow is generated by other mechanisms. Using the Bayesian Total Error Analysis (BATEA) framework (*Kavetski et al.*, 2006b), *Kavetski et al.* (2006a) are able to estimate rainfall time series by identifying storms within a rainfall time series and estimating a storm multiplier that acts to modify each of the observations within that storm. Using a Markov Chain Monte Carlo (MCMC) sampler known as the Differential Evolution Adaptive Metropolis (DREAM) (*Vrugt et al.*, 2009b), *Vrugt et al.* (2008) also estimated rainfall time series using storm multipliers. This methodology results in prediction uncertainty bounds for storm events as well as significantly altering the posterior parameter distributions for the hydrological model parameters. Work by *Renard et al.* (2010, 2011) have built on the idea of using storm multipliers by using rainfall multipliers characterized by a hyperdistribution. The importance of specifying informative prior distributions on rainfall errors was demonstrated. Additionally conditional simulation was proposed as an effective method to build such priors for daily rainfall estimation. Using multiplicative error structures for rainfall has shown promise, yet is unable to ascertain uncertainty when no rainfall is recorded. This is a critical gap that has not been addressed in literature, particularly for poorly gauged catchments. Depending on the location of a rainfall gauge, poorly gauged catchments are particularly prone to overestimate, underestimate or completely miss localised rainfall. Thus it is imperative that a characterization of rainfall error allows for uncertainty that is independent of the observed rainfall magnitude to be developed. Further, rainfall observations of the same magnitude may have different uncertainty. The use of transfer functions to reduce input model data into parameters allows for a window of input data to be adjusted for each parameter. In contrast to the use of storm multipliers, the use of transfer functions allows for uncertainty in rainfall events to be accounted for when no rainfall is recorded at the gauge.

This paper explores the use of wavelets to estimate rainfall time series in the context of a lumped catchment scale rainfall runoff model. To address the need for estimating rainfall time series, this paper will address (i) the use of the DWT to reduce model input data to parameters for estimation of input uncertainty, (ii) possible methodologies to estimate input uncertainty using DWTs and $\mathrm{DREAM}_{(\mathrm{ZS})}$ and (iii) estimation of rainfall input series and the validation of results against rainfall and streamflow observations.

4.3 Hydrologic model description

For this study, the Sacramento Soil Moisture Accounting (SAC-SMA) model was used with a fixed integration time step of 1 day. This lumped conceptual watershed model is used by the National Weather Service River Forecast System (NWSRFS) for flood forecasting throughout the United States. This model has been used as well to model the rainfall-runoff transformation throughout Australia (*Herron et al.*, 2002) and has shown promising results for soil moisture data assimilation (*Crow and Ryu*, 2009).

The SAC-SMA model can be described as a nonlinear regression model, $\mathcal{F}(\cdot)$, which simulates a n-record of discharge values, $\vec{Y} = \{y_1, \ldots, y_n\}$ in mm/d:

$$\vec{Y} = \mathcal{F}(\boldsymbol{\theta}, \tilde{\mathrm{x}}_0, \hat{\vec{E}}, \hat{\vec{R}}). \tag{4.1}$$

The model input arguments are the $1 \times d$ vector, $\boldsymbol{\theta}$ with SAC-SMA parameter values, the $1 \times m$ vector $\widetilde{\vec{x}_0}$ with values of the initial states (at $t = 0$) in mm, and $1 \times n$ vectors $\widehat{\vec{E}} = \{\widehat{e}_1, \ldots, \widehat{e}_n\}$ and $\widehat{\vec{R}} = \{\widehat{r}_1, \ldots, \widehat{r}_n\}$ that store the observed values of the potential evapotranspiration (PET) and rainfall in mm/d, respectively. Note, the $\widehat{\ }$(hat) symbol is used to denote measured quantities, and a ~(tilde) symbol reflects variables that could, in theory, be observed in the field but due to their conceptual nature are difficult to determine accurately.

The SAC-SMA model is comprised of three layers: surface, upper and lower soil moisture layers. A variable impervious area alters the percentage of precipitation that contributes to direct runoff and infiltration into the upper soil layer. Evapotranspiration is able to occur from surface water as well as the lower and upper zone tension water stores. The upper soil layer is comprised of tension and free water. For free water to be able to contribute to the lower zone via percolation, total channel flow via interflow or surface runoff, the tension water store must first be full. Losses in tension water through evapotranspiration can be replenished by free water in both the lower and upper soil layers. The lower soil layer is comprised of tension water as well as primary and supplementary free water stores. Unlike the upper layer, the lower layer has a reserve on the percentage of free water that can supplement tension water losses due to evapotranspiration. Both the primary and supplementary free water stores contribute to a primary and supplementary base flow. A portion of base flow contributes to the total channel flow, while another portion contributes to subsurface discharge. For a more detailed description of the SAC-SMA model the reader is referred to *NWSRFS* (2002).

Based on the recommendations of *Peck* (1976), the 16 parameter SAC-SMA model has been reduced to 13 parameters by fixing SIDE, RIVA and RSERV, the parameters that control the ratio of deep recharge to channel base flow, Riparian vegetation area and fraction of lower zone free water not transferable to lower zone tension water, respectively. Consequently the SAC-SMA model used has 13 parameters and 6 state variables, hence, $d = 13$ and $m = 6$. The parameter distribution for the remaining 13 parameters is obtained using the $\mathrm{DREAM}_{(\mathrm{ZS})}$ algorithm (*Laloy and Vrugt*, 2012; *Vrugt*, 2016). The initial parameter space was selected based on recommendations by *Boyle et al.* (2000). Since the Maximum A Posteriori Probability (MAP) solution involved a large number of parameters that were hitting their respective upper or lower bound, adjustments of the parameter space were made based on recommendations of *Anderson et al.* (2006). Even with these more relaxed ranges, some parameters continued to find their MAP value at the edge of their respective search domains. For those parameters, the search ranges were further increased-making sure that values remain physically plausible. Table 4.1 summarizes the parameters of the SAC-SMA model including their prior uncertainty ranges. Note, these enlarged ranges of the parameters are justified given the rather contrasting characteristics of the Warwick catchment as compared to the watersheds studied by *Boyle et al.* (2000) and *Anderson et al.* (2006).

TABLE 4.1: Parameters of the SAC-SMA model and the range used for the estimation process.

Parameter	Description	Units	Initial range
Capacity thresholds			
UZTWM	Upper zone tension water capacity	mm	$1.00 - 150.00$
UZFWM	Upper zone free water capacity	mm	$1.00 - 150.00$
LZTWM	Lower zone tension water capacity	mm	$10.00 - 500.00$
LZFPM	Lower zone free water primary capacity	mm	$10.00 - 1.00 \times 10^{4}$
LZFSM	Lower zone free water supplemental capacity	mm	$5.00 - 400.00$
Recession parameters			
UZK	Upper zone free water withdrawal rate	day^{-1}	$1.00 \times 10^{-1} - 7.50 \times 10^{-1}$
LZPK	Lower zone primary free water withdrawal rate	day^{-1}	$1.00 \times 10^{-4} - 2.50 \times 10^{-2}$
LZSK	Lower zone supplemental free water withdrawal	day^{-1}	$1.00 \times 10^{-2} - 8.00 \times 10^{-1}$
Percolation			
ZPERC	Maximum percolation rate	-	$1.00 - 500.00$
REXP	Exponent of the percolation equation	-	$1.00 - 5.00$
PFREE	Fraction percolation from upper to lower zone free water storage	-	$0.00 - 8.00 \times 10^{-2}$
Impervious area			
PCTIM	Minimum impervious fraction of the watershed area	-	$0.00 - 1.00 \times 10^{-1}$
ADIMP	Additional impervious area	-	$0.00 - 4.00 \times 10^{-1}$
Fixed parameters			
RIVA	Riparian vegetation area	-	0.00
SIDE	Ratio of deep recharge to channel base flow	-	0.00
RSERV	Fraction of lower zone free water not transferable to tension water	-	3.00×10^{-1}

4.4 Bayesian inference of SAC-SMA model parameters and rainfall time series

The rainfall-runoff parameter estimation problem has been studied extensively in the literature. Many different approaches have been developed to find the optimal parameter estimates. These approaches initially focused on finding only the global optimal values of the parameters for some given objective function (*Duan et al.*, 1994; *Gan and Biftu*, 1996; *Thyer et al.*, 1999). In the past two decades the interest has switched to assessment of parameter and prediction uncertainty. Examples of such methods include Bayesian recursive parameter estimation (*Thiemann et al.*, 2001), the limits of acceptability approach (*Beven*, 2006; *Blazkova and Beven*, 2009), the Bayesian Total Error Analyis (BATEA) framework (*Kavetski et al.*, 2006b,a; *Kuczera et al.*, 2006; *Thyer et al.*, 2009; *Renard et al.*, 2011), the Simultaneous Optimization and Data Assimilation (SODA) (*Vrugt et al.*, 2005), the DREAM algorithm and its variations (*Vrugt et al.*, 2005, 2008, 2009a,b; *Vrugt and Ter Braak*, 2011; *Laloy and Vrugt*, 2012; *Sadegh and Vrugt*, 2014), Bayesian model averaging (*Butts et al.*, 2004; *Ajami et al.*, 2007; *Vrugt and Robinson*, 2007), the hypothetico-inductive data based mechanistic modeling framework of *Young* (2013) and Bayesian data assimilation (*Bulygina and Gupta*, 2011). This paper adopts a Bayesian viewpoint to quantify model parameter and predictive uncertainty. If the SAC-SMA parameters, initial states, PET and rainfall are considered to be unknown, then their posterior probability distribution, $p(\boldsymbol{\theta}, \widetilde{\mathbf{x}}_0, \widetilde{\vec{E}}, \widetilde{\vec{R}} | \widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}})$ can be estimated from the observed discharge, PET, and rainfall time series using Bayes Law

$$p(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}} | \widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}}) = \frac{p(\boldsymbol{\theta}, \tilde{\mathbf{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}}) L(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}} | \widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}})}{p(\widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}})}, \tag{4.2}$$

where the $p(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}})$ signifies the joint prior distribution of the parameters, initial states, potential evapotranspiration and rainfall, respectively, $L(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}} | \widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}})$ denotes the likelihood function, and the denominator $p(\widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}})$ represents the evidence or the marginal likelihood. This formulation of Bayes law takes into consideration explicitly the major sources of uncertainty involved in the modeling of the rainfall-runoff transformation. Indeed, rainfall and PET observations are subject to considerable uncertainty, and if their errors are not properly treated then the SAC-SMA parameters will compensate, in part, for their misspecification.

The prior distribution, $p(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}})$ summarizes all the information about the SAC-SMA parameters, initial states, potential evapotranspiration and rainfall data records and their multivariate dependencies before the primary data (hydrologic measurements) and/or secondary data (watershed characteristics) are collected. The likelihood function quantifies in probabilistic terms the distance between the observed and simulated data. Finally, the evidence, $p(\widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}})$ normalizes the posterior distribution so that it integrates to unity, and represents a proper statistical distribution. This constant is independent of the parameter values; hence, the marginal likelihood can be removed from equation (4.2) and a proportionality sign used instead

$$p(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}}|\widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}}) \propto p(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}}) L(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}}|\widehat{\vec{Y}}, \widehat{\vec{E}}, \widehat{\vec{R}}) \tag{4.3}$$

Equation (4.3) considers joint inference of the parameters of the SAC-SMA model, its initial states, and rainfall and potential evapotranspiration data records. This would involve the estimation of a very large number of unknowns, and result in issues such as overfitting.

To proceed the following is taken advantage of

1 Watershed-scale hydrologic processes exhibit generative, negative feedbacks that gravitate the moisture status to a stable state, also called attractor. Numerical results of watershed models indeed demonstrate that the effect of the initial states on the model results rapidly diminishes with increasing "distance" from the start of simulation. Therefore advantage can be taken of a spin-up period of Q days to remove sensitivity of the modeling results and error residuals to state value initialization.

2 The inherent low-pass filter properties of watershed models and buffer capacity of soil moisture stores causes the governing state dynamics and output fluxes to be relatively insensitive to random and systematic errors in the PET data (*Oudin et al.*, 2006; *Samain and Pauwels*, 2013), and it can be conveniently assumed that $\delta_t(\widetilde{\tilde{\mathrm{E}}}_{(t-\Delta t:t)}, \tilde{\mathrm{E}}_{(t-\Delta t:t)}) \approx 0$. Yet the framework presented herein can be easily extended to explicitly treat errors in PET observations as well.

If these two assumptions are adopted, then equation (4.3) simplifies to

$$p(\boldsymbol{\theta}, \widetilde{\vec{R}}|\widehat{\vec{Y}}, \widehat{\vec{R}}) \propto p(\boldsymbol{\theta}, \widetilde{\vec{R}}) L(\boldsymbol{\theta}, \widetilde{\vec{R}}|\widehat{\vec{Y}}, \widehat{\vec{R}}). \tag{4.4}$$

It can further safely be assumed that the prior information of the parameters and rainfall record are independent. Thus, the multivariate joint prior distribution, $p(\boldsymbol{\theta}, \widehat{\vec{R}})$, can be replaced with two individual prior distributions for the parameters and the hyetograph. If it is further assumed that the prior parameter distribution, $p(\boldsymbol{\theta})$ is uniform, flat and non informative, then this leaves the following definition of the posterior distribution, $p(\boldsymbol{\theta}, \widetilde{\vec{R}} | \widehat{\vec{Y}}, \widehat{\vec{R}})$, of the parameters and rainfall record given the observed discharge and rainfall record. Equation (4.4) can be further simplified by decomposing the likelihood function into two separate likelihood functions for the discharge data and rainfall record as follows: $L(\boldsymbol{\theta}, \widetilde{\vec{R}} | \widehat{\vec{Y}}, \widehat{\vec{R}}) = L(\boldsymbol{\theta} | \widehat{\vec{Y}}) L(\boldsymbol{\theta} | \widehat{\vec{R}})$. This decomposition is appropriate as it is highly plausible that the rainfall and discharge measurement data errors are indepedent. Thus, the following equation remains

$$p(\boldsymbol{\theta}, \widetilde{\vec{R}} | \widehat{\vec{Y}}, \widehat{\vec{R}}) \propto p(\widetilde{\vec{R}}) L(\boldsymbol{\theta} | \widehat{\vec{Y}}) L(\boldsymbol{\theta} | \widehat{\vec{R}}) \tag{4.5}$$

and requires the user to define the rainfall data prior, $p(\widetilde{\vec{R}})$, and the pair of likelihood functions, $L(\boldsymbol{\theta} | \widehat{\vec{Y}})$, and $L(\boldsymbol{\theta} | \widehat{\vec{R}})$, respectively. Before the mathematical definition of these three distributions is further discussed, the parameterization of the rainfall record is presented. This is of crucial importance and prerequisite to the numerical implementation of Equation (4.5).

4.5 Model input data reduction using the DWT

Wright et al. (2017a) provided a comparison of the discrete wavelet and discrete cosine transforms for hydrologic model input data reduction and recommended that the Discrete Wavelet Transform be used for hydrologic studies that have both short and long temporal durations that also involve rainfall as an input. Using the pyramid algorithm developed by *Mallat* (1989) along with the Daubechies wavelets (*Daubechies*, 1990) an input rainfall signal can be transformed into a set of rainfall parameters. This algorithm can be summarized as follows. The input rainfall $\widetilde{\vec{R}}$ is passed through high- and low-pass filters where

$$\vec{p}_j^{\,L}(i) = \begin{cases} \sum_{m=1}^{\mathrm{L}} \widetilde{\vec{R}}(2i - m - 1)\mathrm{w}(m), & j = 1 \\ \sum_{m=1}^{\mathrm{L}} \vec{p}_{j-1}^{\,L}(2i - m - 1)\mathbf{w}(m), & j > 1. \end{cases} \tag{4.6}$$

is the low-pass and

$$\vec{p}_j^{\mathrm{H}}(i) = \begin{cases} \sum_{m=1}^{\mathrm{L}} \tilde{\vec{R}}(2i-m-1)\mathrm{h}(m), & j=1 \\ \sum_{m=1}^{\mathrm{L}} \vec{p}_{j-1}^{L}(2i-m-1)\mathrm{h}(m), & j>1. \end{cases} \tag{4.7}$$

is the high-pass, where $\mathrm{h}(m)$ and $\mathrm{w}(m)$ are the scaling and wavelet functions used in the high and low-pass filters, respectively. $\vec{p}_j^{\mathrm{L}}(i)$ and $\vec{p}_j^{\mathrm{H}}(i)$ refer to the low and high-pass parameters at the j^{th} level respectively. This process decomposes the original signal into levels of parameters that preserve resolution in both the temporal and frequency domains. Due to the length of each resultant parameter series being equivalent to the length of the input series, every other parameter is removed to avoid redundancy. This process is referred to as down sampling. At this stage, further decomposition can be achieved by iteratively passing the low-pass parameters through the filtering equations. At each level the low- and high-pass parameters can be referred to as approximation or detail parameters. A number of different combinations of these DWT approximation and detail parameters can be sampled to alter different components of the rainfall time series. After the parameters are sampled the DWT decomposition process is reversed by iterating through

$$\vec{p}_{j-1}(i) = \sum_{m=\lceil i/2 \rceil}^{\lfloor (\mathrm{L}-1+i)/2 \rfloor} \left(\vec{p}_j^{\mathrm{H}}(i)\mathrm{h}(2m-i)\right)\left(\vec{p}_j^{\mathrm{L}}(i)\mathrm{w}(2m-i)\right), \quad j>1, \tag{4.8}$$

where $\lceil . \rceil$ is the ceiling operator. Finally, the input signal is reconstructed using

$$\tilde{\vec{R}}(i) = \sum_{m=\lceil i/2 \rceil}^{\lfloor (\mathrm{L}-1+i)/2 \rfloor} \left(\vec{p}_j^{\mathrm{H}}(i)\mathbf{h}(2m-i)\right)\left(\vec{p}_j^{\mathrm{L}}(i)w(2m-i)\right), \quad j=1, \tag{4.9}$$

before the resulting rainfall time series is able to passed into equation (4.5) for evaluation. A major advantage of using discrete wavelet decomposition is that the user is able to alter the number of parameters used to sample the posterior rainfall time series. As more levels of decomposition are used, a lower number of approximation parameters describe the low-pass component of the rainfall time series. One drawback of estimating the approximation parameters with more levels of decomposition is that lower resolution can be achieved. For a more detailed discussion on the DWT, the reader is referred to *Mallat* (2009).

4.6 Formulation of posterior distribution

Now that a sparse parameterization for the rainfall record has been defined, there remains the definition for the prior distribution, $p(\widetilde{\boldsymbol{R}})$ and two likelihood functions, $L(\boldsymbol{\theta}|\widehat{\vec{Y}})$ and $L(\boldsymbol{\theta}|\widehat{\vec{R}})$, in equation (4.5), respectively. In this paper the inference results for a formulation of the posterior distribution of equation (4.5) is presented to evaluate the sensitivity of the posterior distribution to the underlying assumptions regarding the information content of the discharge and rainfall data.

The formulation of $p(\boldsymbol{\theta}, \widetilde{\vec{R}}|\widehat{\vec{Y}}, \widehat{\vec{R}})$ in equation ((4.5) is derived from *Kavetski et al.* (2006b) and assumes a Gaussian likelihood for $L(\boldsymbol{\theta}|\widehat{\vec{Y}})$ and $L(\boldsymbol{\theta}|\widehat{\vec{R}})$, respectively,

$$L(a|b) = -\frac{1}{2}n\log\left(\sum_{t=1}^{n}(a_t - b_t)^2\right) \tag{4.10}$$

with n-input vectors, a and b. Using $\boldsymbol{\beta}$ to describe the ratio of the n rainfall depths (in mm/d) predicted by the k wavelet parameters and their corresponding measured values where,

$$\boldsymbol{\beta} = \left\{\frac{\widetilde{r}_1}{\widehat{r}_1}, \ldots, \frac{\widetilde{r}_n}{\widehat{r}_n}\right\}, \tag{4.11}$$

a vague inverse gamma prior for $p(\widetilde{\boldsymbol{R}})$ is used

$$p(\sigma_\beta^2|\nu_0, s_0) \propto \frac{1}{\sigma_{\boldsymbol{\beta}}^{\nu_0+1}} \exp\left(-\frac{\nu_0 s_0^2}{2\sigma_{\boldsymbol{\beta}}^2}\right), \tag{4.12}$$

where $\sigma_{\boldsymbol{\beta}}$ (mm/d) signifies the rainfall measurement data error, and $\nu_0 > 0$ (-) and $s_0 > 0$ (mm/d) are the scale and shape parameter of the inverse gamma prior, respectively.

If the prior of equation (4.12) is combined with the Gaussian likelihoods of the rainfall and discharge data record then the following formulation of the posterior distribution in equation (4.5) is derived:

$$p(\boldsymbol{\theta}, \widetilde{\vec{R}}|\widehat{\vec{Y}}, \widehat{\vec{R}}) \propto \left[\mathrm{SSE}(\boldsymbol{\beta}, 1) + \nu_0 s_0^2\right]^{-\frac{k+\nu_0-1}{2}} \mathrm{SSE}(\vec{Y}(\boldsymbol{\theta}), \widehat{\vec{Y}})^{-\frac{n}{2}} \tag{4.13}$$

where $\mathrm{SSE}(a,b) = \sum\limits_{t=1}^{n}(a-b)^2$ and $\vec{Y}(\boldsymbol{\theta}) = \mathcal{F}(\boldsymbol{\theta}, \widetilde{\vec{x}_0}, \widetilde{\vec{E}}, \widetilde{\vec{R}})$.

If any of the n rainfall multipliers deviate from unity then the first term on the right hand-side (likelihood of rainfall data) decreases. This is only acceptable if the value of the discharge likelihood (second term, right-hand-side) increases sufficiently such that posterior density increases as a whole. Thus, the formulation of equation (4.13) constrains the rainfall adjustments as large changes to the measured rainfall record are discouraged, unless the fit to the discharge data increases so much so that the product of the two likelihoods increases.

In practice, it is much more convenient to work with the log-formulation of equation (4.13) as this avoids numerical problems with a zero density if n becomes large. *Kavetski et al.* (2006a) is followed and it is assumed that $\nu_0 = 5$ and that the value of s_0 is estimated along with the d model parameters and k wavelet parameters. This thus involves the inference of $k + d + 1$ unknowns.

4.7 Posterior sampling

A key task in Bayesian inference is now to summarize the posterior distribution of the individual SAC-SMA parameters, and rainfall estimates at times $t = \{1, \ldots, n\}$. Unfortunately, for equation (4.13) this task cannot be carried out analytically, and thus Markov chain Monte Carlo (MCMC) simulation with the DREAM$_{\text{(ZS)}}$ algorithm to generate samples of the posterior distribution (*Vrugt et al.*, 2008, 2009b; *Vrugt*, 2016) is used. This method runs $N \geq 3$ different Markov chains in parallel and proposals in each chain are created using parallel direction and snooker updates from an archive of past states of the chains. Snooker updates involve sampling along an axis that is developed from past states in preference to sampling along the coordinate axis. This approach solves a practical problem in Monte Carlo Markov chain (MCMC) simulation, that is choosing a correct orientation and scale of the proposal distribution. To maximize speed up convergence to the target distribution, the DREAM$_{\text{(ZS)}}$ algorithm uses adaptive randomized subspace sampling to only update a random selection of parameters. A detailed description of the DREAM$_{\text{(ZS)}}$ algorithm appears in *Laloy and Vrugt* (2012) *Vrugt* (2016) and related cited publications.

For all numerical studies presented herein default values for the algorithmic parameters (*Vrugt*, 2016) and $N = 3$ Markov chains are used. Convergence of the sampled chain trajectories using the $\hat{R}$ convergence diagnostic (*Gelman and Rubin*, 1992) is used. This statistic compares, for each dimension of the target distribution, the variance of each parameter within each chain to the variance of that same parameter between the N different chains.

All trials were executed until the $\hat{R}$-diagnostic convergence criterion was smaller than the stipulated threshold of 1.2, $\hat{R}_j \leq 1.2 \,\forall\, j = \{1, \ldots, k + d + 1\}$.

4.8 Site and data description

The data set used for the experiment comprises daily rainfall from 14 operational real time rain gauges, Potential Evapotranspiration (PET) and observed streamflow data for the Warwick catchment. Warwick is a small sub-catchment of the Condamine-Culgoa catchment, Figure 4.1. Located in South-East Queensland, Australia, the total drainage area for the Warwick catchment is 1360 km^2. The Warwick basin has been subjected to multiple flood events of significant magnitude in the past decade. The total length of the perennial channels is 78 km. Cease to flow conditions have been observed during times of extended drought. The maximum elevation difference along the channel is 308 m. The highest, lowest and mean elevations in the catchment are 1361, 446 and 650 m Above Mean Sea Level (AMSL), respectively. In the period beginning at the start of November 2000 and finishing at the end of June 2015, the mean, median, 10th percentile and 90th percentile annual rainfall amounts for the Warwick Alert rainfall gauge are 564, 513, 408 and 748 mm/y, respectively. Due to the severe droughts that affected Australia for most of the first decade of the millennium, it is likely that these rainfall statistics are negatively biased and that, over a longer time period the average rainfall at these gauges would be larger than those observed. The analysis period used the data with highest quality and begins 1st of January 2007 and ends 31st of March 2013. Areal rainfall is constructed using the Inverse Distance Weighting (IDW) method, which is current operational practice at the Australian Bureau of Meteorology (BoM). Distance is calculated from the catchment centroid to the rain gauge. Monthly PET data from the Australian Water Availability Project (AWAP) were used. A crump weir was used to record continuous height measurements. These height measurements have been converted to streamflow using periodically updated rating curves.

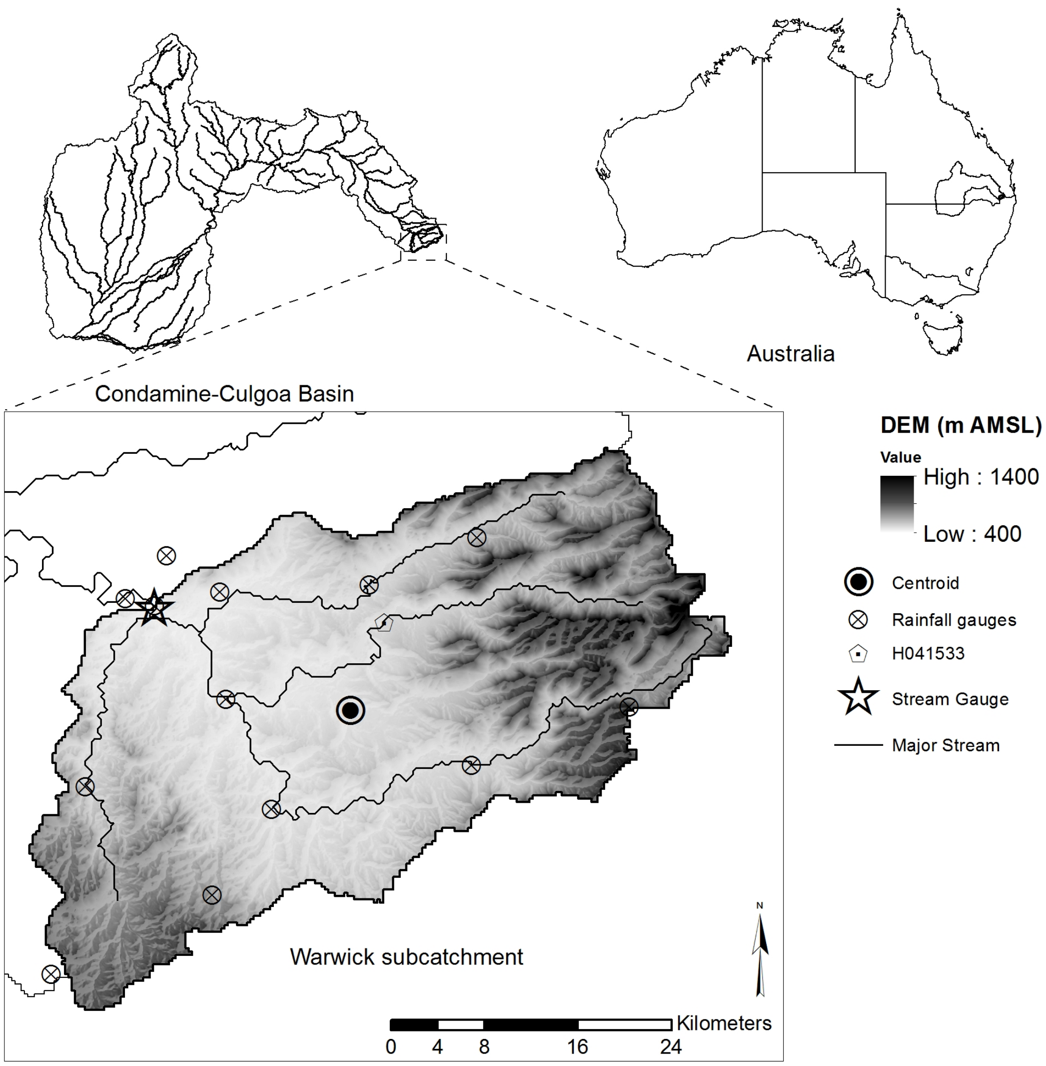

FIGURE 4.1: The location of the Condamine-Culgoa basin in Australia and digital elevation map of the Warwick subcatchment within this region. The notation "m AMSL" in the legend denotes "meters above mean sea level"

4.9 Synthetic case study

4.9.1 Aims

The synthetic case study was designed as a preliminary study to explore some of the different model input data reduction techniques that wavelets make available. A major aim of the synthetic case study was to assess the suitability of the db1 and db2 wavelets to account for a known random multiplicative heteroscedastic Gaussian error. Another aim was to assess the value of rainfall, model parameter and streamflow estimates using a model parameter estimation approach, a segmented rainfall and model parameter estimation approach and via the simultaneous estimation of rainfall and model parameters.

4.9.2 Description of experiments

As there is no definite true rainfall time series and model parameter set for the Warwick catchment, a synthetic case study was conducted. The DWT was used to reduce model input data, to estimate rainfall time series that are representative of the 'synthetic true' rainfall from an imperfect rainfall product. Synthetic streamflow data and model parameters were created by estimating the SAC-SMA model parameters using areal rainfall and observed streamflow. The simulated streamflow, estimated parameters and areal rainfall are considered to be the synthetic truth. A random multiplicative heteroscedastic Gaussian error with standard deviation equivalent to 10% of the observation was added to the synthetic rainfall truth in order to simulate the errors that can be expected in an areal rainfall product. Evaluating different assumptions of error types and distributions is outside the scope of this study and is a possible direction for future studies. The choice of error does not detract from the ability of the synthetic case study to demonstrate that the DWT is a powerful transform that can be used to estimate rainfall time series. Throughout the synthetic case study, the perturbed rainfall and PET are used as the a-priori input data, the synthetic streamflow truth and perturbed rainfall truth are used as the evidence for the posterior estimation of model parameters, rainfall and streamflow.

Throughout the synthetic case study rainfall time series were estimated by inverting the estimated level 4 wavelet approximation parameters. The selection of the parameter level to be estimated and analysis wavelet to be used, determines the number of rainfall observations that each parameter will impact. The level 4 approximation parameters were chosen to maximize the trade-off between the benefit gained by representing the rainfall data set using DWT parameters and creating a highly dimensionalized problem that cannot feasibly be solved. The DWT transform is used with the Daubechies db1 and db2 wavelets. It is necessary to test different wavelets to assess the impact they have on the assumed error structure. The model was allowed a spin up period of 100 days, 5 years of data were used in the calibration period and 357 days of data was used for the validation period. DWT parameters are only estimated in the calibration period, consequently rainfall is only estimated in the calibration period.

The synthetic case study was comprised of four tests and a benchmark. The benchmark synthetic test E simulated streamflow using the perturbed rainfall input and the synthetic truth parameters. Synth 1 used the synthetic true model parameters and the DWT parameters were then constructed and estimated based on the perturbed input rainfall. The second test, Synth 2, estimated only the SAC-SMA parameters using the perturbed rainfall data. The third and last tests, Synth 3 and Synth 4, simultaneously estimated both the SAC-SMA parameters and the DWT parameters for the perturbed input rainfall using the db1 and db2 wavelets, respectively. During the synthetic case study only the level 4 approximation parameters were modified. Of which there were 115 and 116 for the db1 and db2 wavelets respectively.

4.9.3 Results and discussion

The performance of each synthetic test in the calibration and validation period as well as the rainfall and streamflow volume for the calibration and validation periods are shown in Table 4.2. All synthetic experiments apart from Synth 3 were able to simulate streamflow at least as well as the benchmark synthetic test E.

The implications of this result will be elaborated on. As there is no observation error in the synthetic truth streamflow and the synthetic true model parameters are known in Synth 1, Synth 1 therefore tests the ability of the proposed methodology to deal with a random heteroscedastic multiplicative Gaussian error in isolation. If subsequent tests are not susceptible to over fitting then, for the given DWT setup, Synth 1 should place a lower bound on the streamflow RMSE able to be achieved in the calibration period. However, as the calculated RMSE in the calibration period for Synth 2 is much lower than the calculated RMSE for Synth 1 and the benchmark test E, it is clear that the model parameters are being modified in order to satisfy the likelihood function.

Surprisingly Synth 3 both simulates streamflow and estimates rainfall the poorest. As the RMSE of streamflow in the calibration period is closely related to the second term of equation (4.13), the only way that this solution can be returned as the MAP solution is that the first term of equation (4.13) was increased sufficiently so that the resulting posterior density increased. The reason for this is the inability of the db1 wavelet to account for random multiplicative heteroscedastic Gaussian errors. Modification of the DWT level 4 approximation parameters in Synth 3 results in a homogeneous adjustment in rainfall for the window in which the DWT parameter adjusts. Consequently, it is postulated that the db1 wavelet is

TABLE 4.2: Description of the experimental setup of the synthetic case study used herein along with the results for the calibration (Cal) and evaluation (Val) periods.The rainfall, model, wavelet and coef columns indicate which parameters are being estimated, the analysis wavelet being used and which, if any, any wavelet parameters are being estimated. The number of wavelet parameters estimated are shown in brackets. E indicates the simulation in which the perturbed rainfall is used with the synthetic true model parameters; S1, S2, ... etc are used to label the synth studies 1 and 2 etc.The results in the streamflow and rainfall columns are estimated volumes of that designated period. Rainfall is not modified in the validation period.

Experimental Setup					RMSE streamflow $(m^3 s^{-1})$		Streamflow		Rainfall	
Experiment	Rainfall	Model	Wavelet	Coef	Cal (101:1925)	Val (1926:2282)	Cal (GL)	Val (GL)	Cal (mm)	Val (mm)
Truth							688.26	75.82	3205.12	544.48
E		N	-	-	3.54	1.18	699.73	74.05	3208.07	545.18
S1	Y	N	db1	appx4 (115)	3.12	1.20	657.52	72.35	2688.52	545.18
S2	N	Y	-	-	2.54	1.09	624.63	68.92	3208.07	545.18
S3	Y	Y	db1	appx4 (115)	4.67	11.06	639.00	115.66	1941.30	545.18
S4	Y	Y	db2	appx4 (116)	2.39	2.83	644.47	59.07	2902.05	545.18

more suited to correct homoscedastic errors. The validity of this postulation may vary with model structure and distribution of parameters. The good simulation of streamflow and estimation of a realistic rainfall time series in Synth 4, in which the level 4 wavelet parameters for the db2 wavelet are estimated, further validates this observation. The non-linear nature of the db2 wavelet allows the estimation of wavelet parameters to account for random multiplicative heteroscedastic Gaussian errors.

The 13 SAC-SMA synthetic truth model parameters are compared to the MAP solutions obtained using the perturbed rainfall product and the simultaneous estimation of model parameters as well as rainfall in Table 4.3. The parameter distributions obtained in Synth 2 and Synth 4 are rarely able to estimate parameter distributions that describe the synthetic truth model parameters. Given the prevalent nature of equifinality in hydrological systems, this result is not all together unexpected. The only difference between the generation of the synthetic truth model parameters and the estimation of model parameters in Synth 2 is that the input rainfall is perturbed with a random heteroscedastic multiplicative Gaussian error. Since the synthetic truth parameters are not able to be estimated, it is expected that the model parameters were erroneously modified in order to account for input error and consequently produce superior streamflow. Consequently, unless either the input error is removed before simulation or additional constraints, such as using informative priors in a similar fashion to *Renard et al.* (2010), are placed on the system, it is likely that estimations of both rainfall and model parameters will include some erroneous modifications in order to satisfy the likelihood function. It is also seen in Synth 4 that even small modifications to input rainfall are able to vastly change the model parameters estimated. This result does not mean that realistic rainfall time series cannot be estimated, but rather that the rainfall time series estimated may include some errors.

Comparisons of the perturbed rainfall, the MAP rainfall estimations using the synthetic true model parameters and the simultaneous estimation of DWT rainfall parameters and model parameters using the db1 and db2 analysis wavelets are made against the synthetic true rainfall in Figure 4.2. While the perturbed rainfall was best able to represent the synthetic true rainfall, the rainfall estimations from Synth 1, Synth 3 and Synth 4 were still quite reasonable. Due to the slope of the estimations made by the db2 wavelet being closer to unity than those obtained using the db1 wavelet, the larger coefficient of determination and the lower RMSE, it is evident that the db2 wavelet is more suited to representing the random multiplicative heteroscedastic Gaussian errors than the db1 wavelet. Neither Synth 3 or Synth 4 are able to represent rainfall as well as Synth 1 in which the true model parameters are

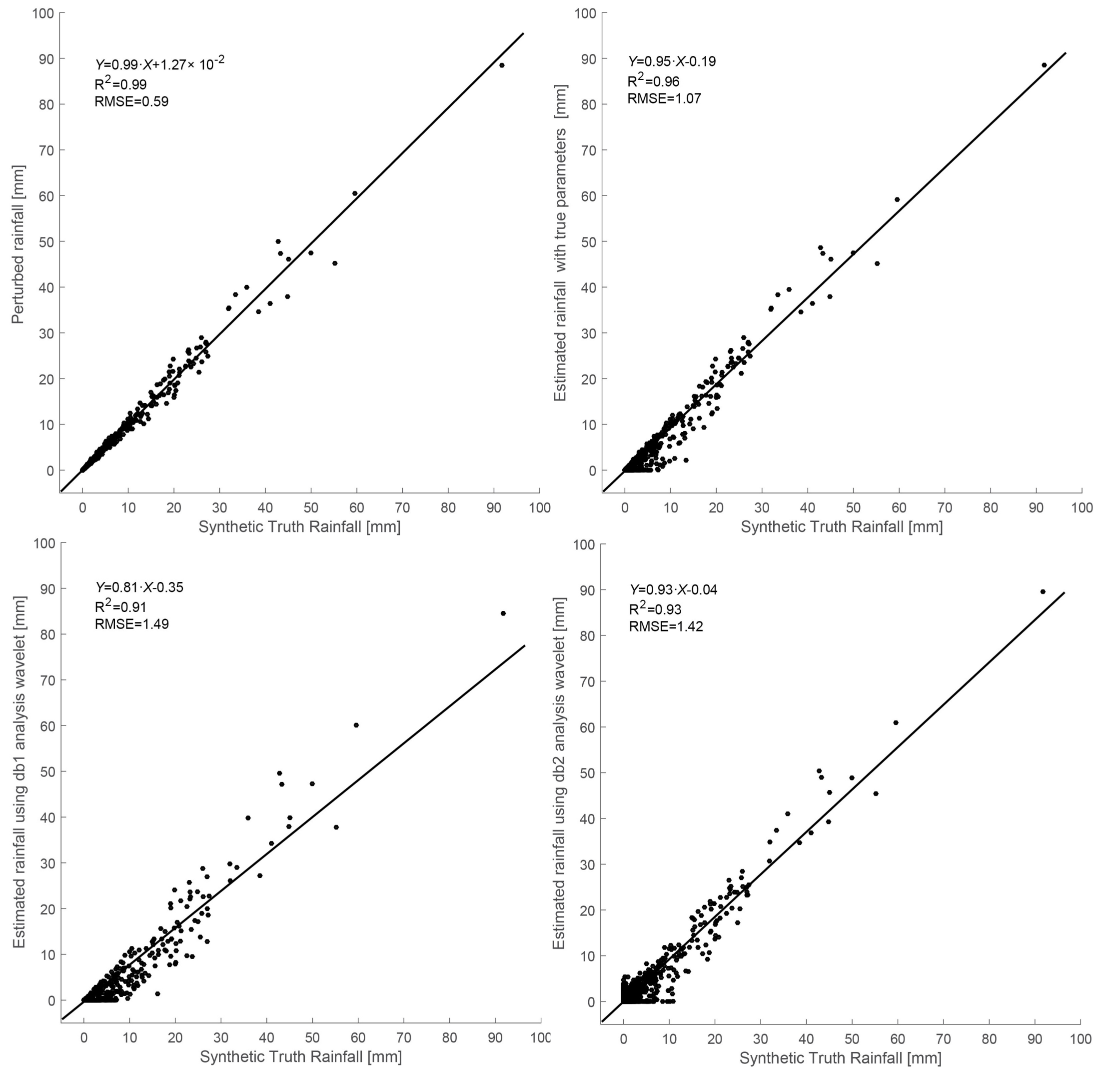

FIGURE 4.2: Comparison of the (top left) perturbed synthetic truth rainfall, and (top right) synthetic and estimated rainfall using inference of only the wavelet parameters (top right plot), joint inference of the SAC-SMA model parameters and wavelet parameters using (bottom left) the db1 wavelet and (bottom right) db2 wavelet. The linear least squares fit and corresponding quality of fit metrics are separately indicated in each plot.

assumed known. This is further proof that the rainfall and model parameter estimations can be erroneously modified in order to produce superior streamflow.

Figure 4.3 supports the hypothesis that rainfall time series can be estimated through the simultaneous modification of DWT rainfall parameters and model parameters. Figure 4.3 (top) shows the total volume of rainfall estimations over the calibration period for the synthetic experiments 2 and 4 next to benchmark test E. Also shown are the maximum (3945 mm) and minimum (2660 mm) rainfall volumes observed throughout the catchment. Using the true synthetic parameters, S2 is able to estimate a rainfall time series that simulates

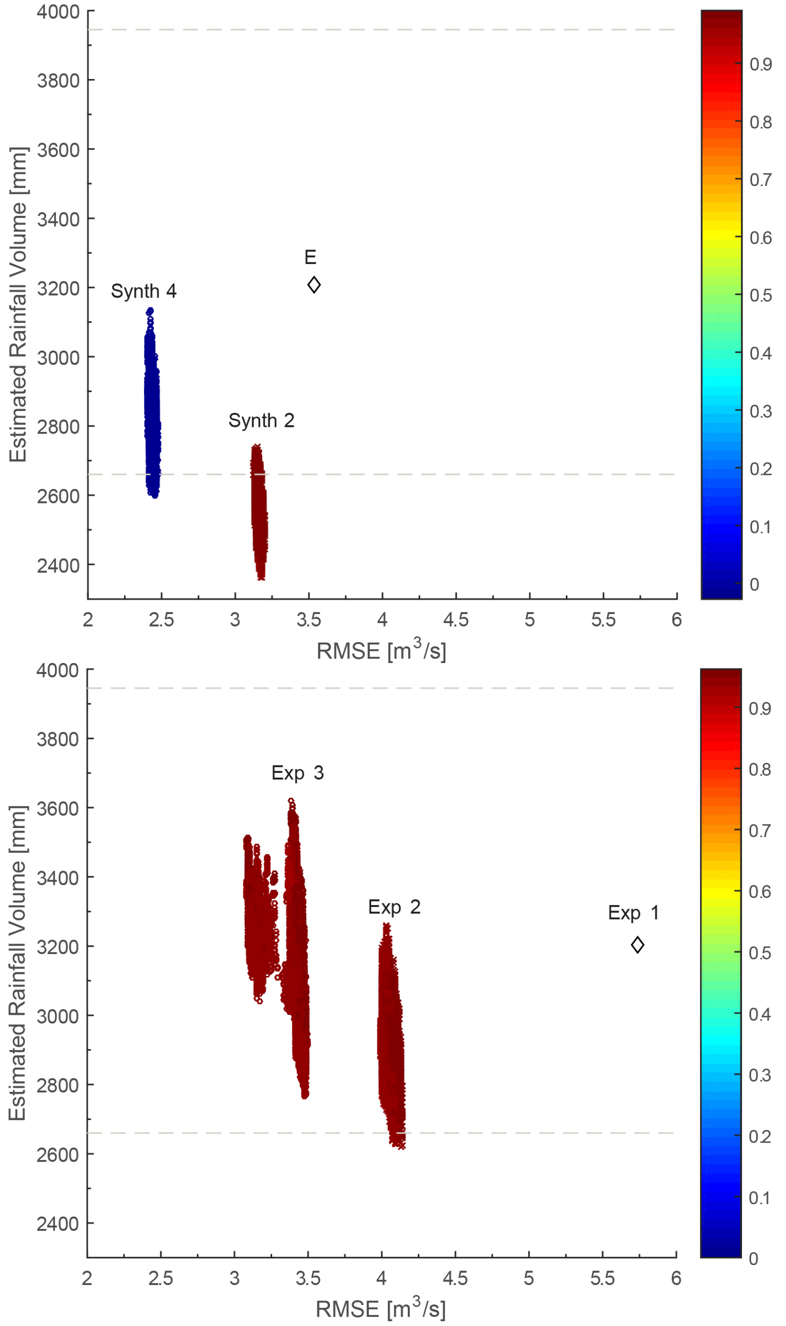

FIGURE 4.3: The estimated rainfall volume for the calibration period plotted against the RMSE of the estimated rainfall's streamflow simulation when compared to observed streamflow. The color bar indicates the correlation coefficient between the estimated rainfall and observed rainfall and the dotted lines in both plots show the maximum and minimum rainfall volumes observed at the gauges within the catchment. (top) The estimated rainfall time series after the convergence criteria has been met for Synth 2 and Synth 4. "E" represents the streamflow simulation in which the perturbed rainfall is used as input to the hydrological model with the synthetic truth parameters. (bottom) The same as the top panel but for experiments 2 and 3. Experiment 1 results are shown to demonstrate the results of a traditional calibration approach.

streamflow better than the benchmark simulation and is in partial agreement with the volumes obsserved at the rainfall gauges. The rainfall time series estimated in S4 is able to estimate rainfall that has a total volumetric range that is generally in agreement with observations while simulating streamflow with a lower RMSE than the perturbed input product.

The synthetic case study has shown that the db1 wavelet does not adequately account for random multiplicative heteroscedastic Gaussian errors for the described model data reduction and model inversion methodology to estimate rainfall time series. The db2 is a more suitable wavelet choice. A model parameter distribution that describes the synthetic true model parameters could not be retrieved. However, this was not entirely unexpected. A rainfall time series that is generally in agreement with observations could be estimated via the simultaneous estimation of rainfall and model parameters. All rainfall time series estimated using the db2 wavelet led to better simulations than the benchmark test experiment, a traditional calibration approach in which only model parameters are estimated, and the estimation of rainfall time series using the synthetic truth model parameters.

Since the synthetic experiments were applied at a relatively coarse temporal resolution, it is expected that using a finer resolution would enable rainfall estimations to meet or exceed the ability of the perturbed rainfall input to model the synthetic true rainfall. Even if the rainfall estimations are not able to reproduce the synthetic true rainfall, the methodology still has a few advantages in that rainfall time series that are similar, yet in this instance have a drier tendency, to those observed at the gauges are produced. These series are able to simulate streamflow which is closer to the synthetic streamflow than that produced by the synthetic rainfall. Thus, it is likely that streamflow forecasts would benefit from rainfall forecasts that are conditioned on rainfall time series that are known to give good results.

TABLE 4.3: Comparison of the synthetic true parameters to the MAP solution alongside the minimum and maximum limits for the marginal density for each parameter for two synthetic studies.

Parameter	Synthetic Truth	Synth 2	min	max	Synth 4	min	max
Capacity thresholds							
UZTWM	24.82	30.06	25.75	31.85	79.35	74.19	82.16
UZFWM	11.87	8.67	8.29	9.30	10.18	10.17	10.50
LZTWM	55.63	84.31	79.46	91.30	10.47	10.00	11.34
LZFPM	42.84	23.81	19.86	27.84	139.20	110.66	139.20
LZFSM	8.01	10.94	9.54	12.35	18.57	15.32	20.93
Recession parameters							
UZK	6.26×10^{-1}	7.50×10^{-1}	7.15×10^{-1}	7.50×10^{-1}	5.32×10^{-1}	5.01×10^{-1}	5.89×10^{-1}
LZPK	6.22×10^{-3}	1.53×10^{-2}	1.17×10^{-2}	1.79×10^{-2}	1.37×10^{-3}	1.01×10^{-3}	1.72×10^{-3}
LZSK	5.26×10^{-1}	5.79×10^{-1}	5.30×10^{-1}	6.24×10^{-1}	5.01×10^{-1}	4.51×10^{-1}	5.66×10^{-1}
Percolation							
ZPERC	497.23	93.12	82.02	127.92	2.07	1.35	4.15
REXP	4.37	5.00	4.95	5.00	4.99	4.98	5.00
PFREE	4.29×10^{-1}	3.86×10^{-1}	3.44×10^{-1}	4.12×10^{-1}	7.74×10^{-1}	7.27×10^{-1}	5.80×10^{-1}
Impervious area							
PCTIM	1.35×10^{-2}	1.56×10^{-2}	1.11×10^{-2}	1.79×10^{-2}	1.91×10^{-2}	1.67×10^{-2}	2.35×10^{-2}
ADIMP	1.85×10^{-2}	3.40×10^{-5}	2.81×10^{-9}	1.04×10^{-3}	8.29×10^{-4}	6.55×10^{-8}	7.30×10^{-3}

4.10 Observation case study

4.10.1 Aims

The observation case study was designed to further explore some of the different model input data reduction techniques that wavelets make available. A major aim of the study was to assess the suitability of the db1 and db2 wavelets to account for unknown errors in input data. Other aims were to determine the impacts of estimating approximation and detail parameters at different levels as well as assessing the value of rainfall and streamflow estimates using a traditional calibration approach, a segmented rainfall and model parameter estimation approach and via the simultaneous estimation of rainfall and model parameters. This study does not aim to nor is able to improve streamflow simulations in the validation period but rather aims to gain understanding of realistic representations of rainfall that can lead to superior streamflow simulations.

4.10.2 Description of experiments

The observation case study begins by estimating an initial parameter distribution set for the 13 SAC-SMA parameters in experiment 1. Experiment 2 then used the estimated MAP parameter set to estimate a rainfall time series. The main difference between experiments for experiment 2 to 8 is that the DWT is either constructed differently or different parameters of the DWT are being estimated. In experiment 3 simultaneous estimation of model parameter distributions and rainfall time series was then performed by estimating the wavelet approximation parameters, experiments 4 and 5 involved estimating the detail parameters of different levels. To this point the level of decomposition was held constant throughout. Next in experiments 6 and 7, the simultaneous estimation of model parameter distributions and rainfall time series was performed by estimating the wavelet approximation parameters under different levels of wavelet decomposition. The number of DWT parameters estimated for each experiment are given in Table 4.4. After this the simultaneous estimation of model parameter distributions and rainfall time series was conducted in experiment 8 using the "db2" wavelet. This was done to assess the ability of the db1 and db2 wavelet to model the errors for the experiments.

Similar to the synthetic case study the model was allowed a spin up period of 100 days, 5 years of data was used in the calibration period and 357 days of data was used for the validation period. Throughout the observation case study observed rainfall and PET are used as the a-priori input data, the observed streamflow and rainfall are used as the evidence for the posterior estimation of model parameters, rainfall and streamflow.

4.10.3 Results and discussion

The performance of the inference approaches in the observation case study for the calibration and validation period (where applicable) as well as the rainfall and streamflow volume for the calibration and validation periods are shown in Table 4.4. All of the experiments were able to estimate model parameter and temporal rainfall distributions or combinations thereof that yield superior streamflow simulations in the calibration period. As expected there was no discernible difference in the validation period. This is because rainfall was not able to be modified in this period.

While the streamflow simulations are consistently improved, the resulting estimated model parameter distributions and rainfall time series or combinations thereof are not all desirable. In experiment 2 a rainfall time series was estimated using the MAP model parameter set found in experiment 1. Thus a set of rainfall time series that agrees with the observed gauged rainfall was estimated. In contrast to the synthetic case study (Synth 3 and Synth 2) the simultaneous estimation of both rainfall time series and model parameter distributions in experiment 3 is able to both simulate streamflow better than experiment 2 and produce rainfall time series that are closer to the rainfall observations at the gauges. While the estimated rainfall series from experiments 4, 5 and 6 are able to simulate superior streamflow, the resultant estimated rainfall time series appears to be unrealistic when compared to the volumes at the rainfall gauges.

In general, the results from experiments 3-6 suggest that the use of the wavelet approximation parameters yield superior results when compared to use of the wavelet detail parameters. When compared to solely estimating model parameters, the streamflow simulated in experiments 3 and 7 showed that RMSE improved by a factor of 1.67 and 1.78, respectively. As expected, using a higher level of decomposition and consequently less parameters in the rainfall reduction in experiment 3 did not produce superior streamflow simulations, or rainfall time series when compared to the use of a lower level of decomposition and estimation of more wavelet parameters in experiment 7. The use of the “db2” wavelet in experiment 8 produced similar streamflow simulations and rainfall time series as was found in experiment 3. This finding suggests that, unlike the introduced error in the synthetic case study, errors in rainfall observations may not be of a random heteroscedastic multiplicative Gaussian nature. Studies conducted by *Renard et al.* (2010) and *McMillan et al.* (2011) have attempted to evaluate multiplicative error models and account for input and structural errors in hydrological modeling, respectively. Their findings indicate that rainfall errors, especially in larger storms, appear to be heteroscedastic. A shortcoming of the studies was that errors in rainfall when no rainfall was observed were not taken into account. Consequently more work is required to determine error models that account for errors when no rainfall is observed. The results of this study indicate that the DWT transform is a tool that can be utilized to further understand rainfall errors. Further work would look at identifying a superior analysis wavelet for the categorization of rainfall errors and rainfall reduction. The unrealistic estimation of rainfall time series in experiments 4-6 further suggests that using informative priors for rainfall measurement error (*Renard et al.*, 2010) may produce fruitful results.

A depiction of the estimated rainfall for a 120 day duration is provided in Figure 4.4 for comparison to other rainfall estimation methods. Unlike the methodology proposed by *Hino*

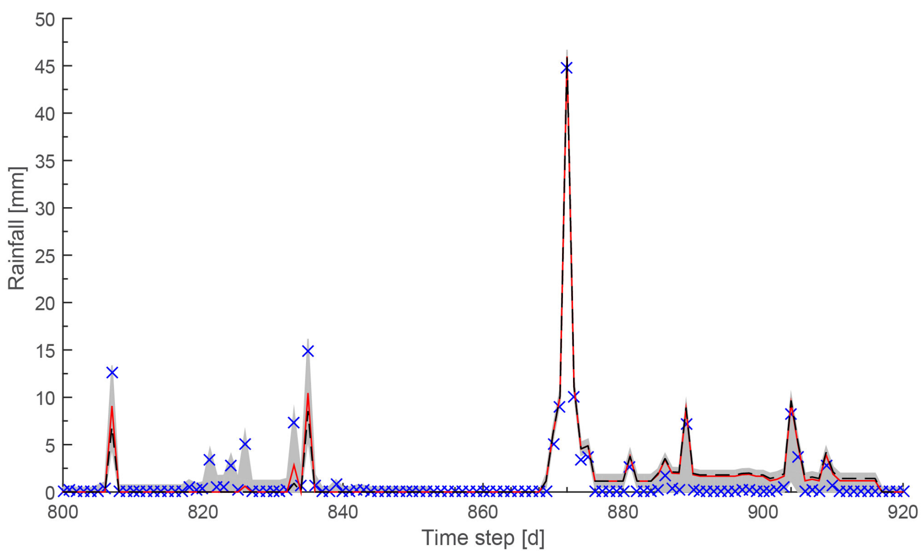

FIGURE 4.4: Rainfall estimates for experiment 3 over a 120 day period. The blue crosses and red line represent the observed rainfall and mean rainfall estimates, respectively. The dashed red line is the MAP rainfall estimate while the grey shading indicates the 5th and 95th percentile rainfall estimates.

(1986), this method does not attempt to separate streamflow into respective runoff components. Further, as was the case in work conducted by *Kirchner* (2009); *Teuling et al.* (2010); *Adamovic et al.* (2015) and *Rusjan and Mikoš̃* (2015), no first-order approximations, to ensure the water balance can be analytically inverted, are made. Figure 4.4 shows that this methodology allows for uncertainty in rainfall to be estimated when no rainfall was observed at the gauges. This is a shortcoming of studies that use the rainfall multiplier method (*Kavetski et al.*, 2006b,a; *Vrugt et al.*, 2008; *Renard et al.*, 2010, 2011). By using the DWT to describe rainfall, this study attempts to move away from the rainfall multiplier methods. The effectiveness of the study is somewhat limited by the use of multipliers in formulation of the likelihood function. Developing a new likelihood function was outside the scope of the paper. Doing so in future studies could enhance the value of the techniques described within this study. Since the resultant streamflow from experiment 3 is superior to that obtained from a traditional calibration approach, the median, MAP 5th and 95th percent rainfall estimates indicate times when streamflow is improved by providing increased or decreased estimates of rainfall as well as the degree of uncertainty associated with the rainfall estimates. It is observed that both the median and MAP rainfall estimates are close to zero when rainfall was observed at the gauge for the time period spanning the 820th to the 830 days. Conversely, rainfall estimates are higher than that observed at the gauge for the time period spanning the 875th to the 885th day. Further, during the time period spanning the 855th to the 865th day, the rainfall estimates completely agree with the observations of zero rainfall. This finding indicates that the rainfall estimation methodology, when applied to the SAC-SMA model, is able to account for rainfall events that are accurately observed as well as under and over observed. For all but a very few time steps the uncertainty bounds estimated by this methodology cover the observed rainfall volumes. Considering that all of these rainfall estimates simulate streamflow that is closer to the observed streamflow, this result is quite significant. This indicates that a significant improvement in streamflow simulation can be made with an improved understanding of rainfall uncertainty. A limitation of this methodology is made evident by examining the constant uncertainty bounds for consecutive days. This is an artefact generated by estimating DWT parameters that apply to a number of consecutive days. The impact on the results can be minimized by increasing the number of estimated parameters or by choosing a more suitable analysis wavelet. The study of this issue is outside the scope of this work.

Figure 4.3 (bottom) shows the converged rainfall volumes for experiments 2 and 3 and the results of a traditional calibration approach in experiment 1. When compared to the synthetic case study in Figure 4.3 (top), the observation case study in the bottom panel of Figure 4.3 (bottom) shows that both the independent estimation of rainfall time series and model parameters (experiment 2) and the simultaneous estimation of rainfall time series and model parameter distributions are able to yield rainfall time series that are generally in agreement with the gauges and consistently produce superior streamflow estimates than their respective benchmarks. Further, as seen in Figure 4.3, the total volumes of the rainfall time series that are estimated in experiments 2 and 3 cover a broad volumetric range. This range is much closer to the range observed at the gauges than their synthetic case study equivalents. The results of experiment 3 indicate that the proposed likelihood function is able to both realistically constrain rainfall estimations and simulate streamflow with RMSE 1.67 times lower than that obtained from only estimating model parameters. Thus, the use of the proposed likelihood function is advantageous when compared to likelihood functions that do not consider input uncertainty.

The observation case study further explored some of the different model input data reduction techniques that wavelets make available. In contrast to the synthetic case study, neither the db1 nor db2 wavelets were able to better account for errors in the input data. The formulation of a complete description of rainfall errors was outside the scope of this study. However, this result suggests that the input error in the observation case study contains both homoscedastic and heteroscedastic errors and that further exploration is warranted. It was found that estimating the approximation parameters of lower level DWT decompositions were able to provide the most realistic rainfall time series with streamflow simulations that are superior to the traditional calibration approach. Both a segmented rainfall and model parameter estimation approach and the simultaneous estimation of rainfall and model parameters were able to estimate realistic rainfall time series that simulated streamflow better than the benchmark. The findings detailed in this discussion indicate that using the proposed likelihood function, realistic rainfall time series and streamflow simulations can be obtained.

TABLE 4.4: Description of the experimental setup of the observation case study used herein along with the results for the calibration (Cal) and evaluation (Val) periods. The rainfall, model, wavelet and coef columns indicate which parameters are being estimated, the analysis wavelet being used and which, if any, wavelet parameters are being estimated. The number of wavelet parameters estimated are shown in brackets. The results in the streamflow and rainfall columns are estimated volumes of that designated period. Rainfall is not modified in the validation period.

	Experimental Setup				RMSE streamflow (m^3/s)		Streamflow		Rainfall	
Experiment	Rainfall	Model	Wavelet	Coef	Cal (101:1925)	Val (1926:2282)	Cal (GL)	Val (GL)	Cal (mm)	Val (mm)
T							685.17	195.29	3205.12	544.48
1	N	Y	-	-	5.74	37.40	688.26	75.82	3205.12	544.48
2	Y	N	db1	appx4 (115)	3.99	37.29	730.27	92.20	2994.80	544.48
3	Y	Y	db1	appx4 (115)	3.43	37.39	697.51	77.32	2957.16	544.48
4	Y	Y	db1	lev4 (115)	4.23	40.28	704.58	54.84	4470.98	544.48
5	Y	Y	db1	lev3 (115)	3.33	41.86	689.15	19.38	4743.98	544.48
6	Y	Y	db1	appx5 (229)	3.91	41.69	658.22	16.56	4523.15	544.48
7	Y	Y	db1	appx3 (58)	3.21	39.79	721.51	56.60	3149.52	544.48
8	Y	Y	db2	lev4 (116)	3.62	37.60	702.91	75.81	3113.49	544.48

4.11 Conclusions

The DWT was used to reduce model input data for the estimation of input uncertainty. Along with $\mathrm{DREAM}_{(\mathrm{ZS})}$, different aspects and configurations of the DWT were explored to outline possible methodologies that may be used to estimate input uncertainty. In this study the methodologies are applied to a gauge-based rainfall estimate yet the methodologies are not limited to gauge-based rainfall estimates. These methodologies could be applied to hydrologic input data such as high-resolution remote sensing of rainfall or even evapotranspiration. It was found that in conjunction with the estimation of DWT rainfall parameters the use of a likelihood function that considers both input rainfall and streamflow error is able to estimate model parameter distributions and entire rainfall time series. The joint estimation of model and wavelet approximation parameters yielded estimates of the most realistic rainfall time series. At the same time streamflow simulations were shown to have improved RMSE by a factor of up to 1.78 when being compared to benchmark simulations in which only model parameters were estimated. The choice of analysis wavelet used for estimation purposes can have a considerable impact on the errors that are corrected for. In most cases, but not all, the proposed likelihood function was able to effectively constrain rainfall estimations while simultaneously producing streamflow simulations that were superior to a traditional calibration approach. Finally, a methodology to create a set of realistic rainfall time series was presented. This methodology will be used in a future study to compare rainfall time series and their respective model parameters with their ability to simulate streamflow and soil moisture observations.

Acknowledgements

The authors would like to extend their gratitude to Jasper Vrugt and the anonymous reviewers for their comments and recommendations. The authors would also like to thank the Bureau of Meteorology (www.bom.gov.au/climate/data) for the provision of data. This work was supported by the Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE) (www.massive.org.au), a Monash University Engineering Research Living Allowance stipend, and a top up scholarship from the Bushfire & Natural Hazards Cooperative Research Centre. Valentijn Pauwels is funded by ARC grant FT130100545.

Chapter 5

A multi model hydrological analysis of rainfall estimates using ensemble Kalman filter innovations

Overview

This chapter addresses the third research question by constraining rainfall estimates with remotely sense soil moisture observations. Model input data reduction and inversion techniques developed in chapters 3 and 4 are used to estimate rainfall time series in the catchment of Warwick for 3 rainfall-runoff models. Using the estimated rainfall time series RS SM observations are assimilated into 3 hydrological models. Analysis of the innovations demonstrate that the choice of model and RS SM product can have a significant impact on the quality of rainfall estimates. This chapter presents a methodology to estimate and evaluate rainfall estimates. The resulting rainfall estimates can be used in future research to condition rainfall forecasts and improve flood forecast skill.

This chapter is reproduced from an article submitted to the Journal of Hydrometeorology, American Meteorological Society.

5.1 Abstract

An increased understanding of the uncertainties present in rainfall time series can lead to improved confidence in both short- and long-term streamflow forecasts. This study presents an analysis that considers errors arising from model input data, model structure, model parameters and model states. Areal rainfall time series were estimated for the study catchment of Warwick, Australia using multiple rainfall-runoff models that take advantage of model input data reduction and model inversion techniques. Remotely sensed soil moisture observations from the Soil Moisture Ocean Salinity (SMOS) and Advanced Microwave Scanning Radiometer-Earth observing system (AMSR-E) satellites were assimilated into three different rainfall-runoff models using an ensemble Kalman filter (EnKF). Innovations resulting from the observed and predicted soil moisture were analyzed for Gaussianity. The findings demonstrate that the combination of remotely sensed soil moisture product and rainfall-runoff model chosen have a significant impact on the quality of rainfall estimates. All models simulated superior streamflow and estimated rainfall to be less than the observed rainfall. Rainfall estimates obtained using the Sacramento Soil Moisture Accounting (SAC-SMA) model were the most realistic. Further, the SAC-SMA model was best able to simulate streamflow. Superior EnKF innovations were obtained when SMOS remotely sensed soil moisture observations were assimilated into the SAC-SMA model and its rainfall estimates.

5.2 Introduction

The analysis and understanding of the uncertainty associated with streamflow observations and simulations can aid in the reduction of socioeconomic and environmental costs of floods and promote robust decision making in water management applications (*McMillan et al.*, 2017). An improved understanding of the uncertainty in streamflow simulations will allow water authorities to make informed and reliable decisions that affect drought management, water allocations, flood resilience and agricultural demand. The major sources of uncertainty in streamflow simulation and forecasting were errors in model input data, model structure, model parameters and model states (*Vrugt*, 2016). A knowledge gap currently exists in the combined analysis of errors arising from these sources.

In rainfall-runoff models, soil moisture governs the proportion of rainfall that contributes to surface and subsurface flows (*Tebbs et al.*, 2016). Consequently, recent studies have focused on skillfully updating rainfall observations using remotely sensed (RS) soil moisture (SM) observations (*Brocca et al.*, 2015; *Ciabatta et al.*, 2015). As such it is expected that rainfall estimates obtained via inverting streamflow observations will benefit from the intermediate soil moisture states being constrained by RS SM observations.

Two dominant techniques to estimate rainfall from soil moisture have emerged. First, RS SM observations have been used to update an Antecedent Precipitation Index (API) forced by satellite rainfall (*Crow et al.*, 2009), with the API innovations assumed to be correlated with the errors between the satellite rainfall and actual rainfall. This assumption implies that the observed soil moisture is influenced by past rainfall, and that losses due to percolation and Potential Evapotranspiration (PET) were negligible. It is therefore expected to work best in catchments and for events in which minimal surface runoff occurs.

Second, is the direct estimate rainfall from the knowledge of relative soil moisture. *Kirchner* (2009) used first order approximations to the water balance equation to describe catchments as simple dynamical systems, thus enabling rainfall to be estimated from streamflow or soil moisture observations. *Brocca et al.* (2013, 2014) made simplifications to the soil water balance equation to enable the direct estimation of rainfall from the knowledge of relative soil moisture. These simplifications assume that all rainfall infiltrates and that PET is zero when rainfall occurs. The technique has successfully been applied at several sites throughout Europe (*Brocca et al.*, 2015; *Ciabatta et al.*, 2015), and has also been demonstrated to improve flood modeling (*Massari et al.*, 2014). Whilst these techniques have shown encouraging results, restricting the analysis to events and catchments in which all rainfall infiltrates places a limitation on the applicability of the techniques. A knowledge gap therefore exists in the utilization of soil moisture observations to estimate or correct rainfall for complex catchments or events that exhibit both surface and subsurface flows.

To effectively utilize soil moisture observations to estimate or correct rainfall for complex catchments or events that exhibit both surface and subsurface flows, it is imperative that the main sources of error, and the methods to account for them, be considered. Errors in rainfall-runoff modeling can arise from model input data, model structure, model parameters and model states (*Vrugt*, 2016). The objective of data assimilation is to incorporate observations of the system to minimize errors. Prior to data assimilation techniques being used to estimate or correct input data, the hydrologic community largely considered the three main types of data assimilation to be; system identification, parameter estimation and state estimation (*Liu and Gupta*, 2007).

System identification suggests that, in addition to the concept of equifinality in which multiple parameter sets tend to arrive at equally acceptable solutions, there were a range of models that have multiple parameter sets that can adequately describe a hydrologic system. Model averaging schemes attempt to extract useful information from each model by assigning a weight to each (*Duan et al.*, 2007). The aim is to take advantage of the fact that some models will outperform other models in different flow scenarios. This suggests that different models will be more or less suitable for different catchments and/or flow events. *Renard et al.* (2010) identified that a shortcoming of model averaging techniques is the lack of distinction between input errors and model structure (*Vrugt and Robinson*, 2007).

The focus of parameter estimation has shifted from deterministic parameter estimation towards stochastic parameter estimation (*Vrugt*, 2016). This shift is largely due to the advancement of computational power and acceptance of equifinality within the modeling community. Deterministic parameter estimation techniques were focused on finding a unique parameter set that best describes a hydrologic system via the minimization of an objective function. However, the choosing of an objective function is subjective (*Vrugt*, 2016) and often leads to finding a parameter set that is able to only partially describe the hydrologic system. Consequently, each objective function may perform well in some catchments or flow situations and poorly in others. Thus, deterministic parameter estimation often leads to unreliable simulation of streamflow in forecasting situations. The aim of stochastic parameter estimation is to select all parameter sets that were able to adequately describe the hydrologic system. Sampled parameter sets were ranked based on an objective function, the effectiveness of which is dependent on assumptions made about model and measurement error (*Vrugt*, 2016). Few studies have focused on elucidating the link between parameter estimation and input error (*Vrugt et al.*, 2008; *Kavetski et al.*, 2006b; *Renard et al.*, 2011). However, it is likely that when combined with efforts to constrain state estimates these techniques will become more valuable.

Pauwels (2008) describes an alternative to traditional calibration schemes in which Monte Carlo simulations, in conjunction with the EKF were used to estimate model parameters instead of the traditional model states. *Moradkhani et al.* (2005) have demonstrated that the EnKF and particle filter can be used to simultaneously estimate model parameters and states. *Vrugt et al.* (2005) demonstrated that data assimilation via the EnKF can be used in conjunction with the Shuffled Complex Evolution Metropolis - Univeristy of Arizona (SCEM-UA) Algorithm (*Vrugt et al.*, 2003). These studies provide techniques that were able to explore links between parameter estimation and the simulation of observed states. The authors expect that combining the strengths of these studies with studies that include the analysis of model structure and input uncertainty will be a step towards further unraveling the links between errors from input data, model structure, model parameters and model states.

This paper is the first to simultaneously explore the links between model estimated rainfall time series, model structure, model parameter estimates and modeled states. Rainfall time series and model parameters were estimated from multiple models by taking advantage of model input data reduction techniques, an objective function that balances rainfall and streamflow estimates and the $\mathrm{DREAM_{ZS}}$ (*Vrugt and Ter Braak*, 2011) algorithm. RS SM observations were assimilated to provide a link between the multiple models, the model estimated rainfall time series, the model parameters and the modeled states. The overarching objective of this paper is to provide a methodology to estimate rainfall through model input data reduction, model inversion and data assimilation techniques, thereby providing a tool to evaluate rainfall estimates, model performance and RS SM observations.

5.3 Model description

5.3.1 General overview

Three models were selected based on their wide-spread acceptance by the hydrologic community as well as their demonstrated ability to assimilate remotely sensed soil moisture (*Li et al.*, 2016). The Sacramento soil moisture accounting (SAC-SMA) model simulates the dominant soil moisture characteristics whilst the probability distributed model (PDM) simulates variable catchment soil moisture using a chosen probability density function. The Hydrological model (HyMod) represents a simplified version of the PDM. Illustrations depicting the models key characteristics were given in Figure 5.1, with only brief descriptions of the models provided here. Table 5.1 describes the parameters and the parameter limits used in the calibration of the models. For more complete descriptions the interested reader is referred to the cited papers.

5.3.2 SAC-SMA

A comprehensive description of the SAC-SMA model is given by *NWSRFS* (2002). The model is applied using the 13 parameters recommended by *Peck* (1976). The model consists of one surface layer and an upper and lower soil moisture layer. The proportion of rainfall that contributes to direct runoff and infiltration is governed by a variable impervious area. The upper soil layer is comprised of tension and free water stores whilst the lower soil layer is comprised of tension and primary and supplementary free water stores. Evapotranspiration is able to occur from both tension water stores, as well as the surface water store. The extent to which free water can supplement tension water due to losses by evapotranspiration is only restricted in the lower layer. The lower layers primary and supplementary free water stores contribute to base flow. The model consists of 6 states and 13 parameters.

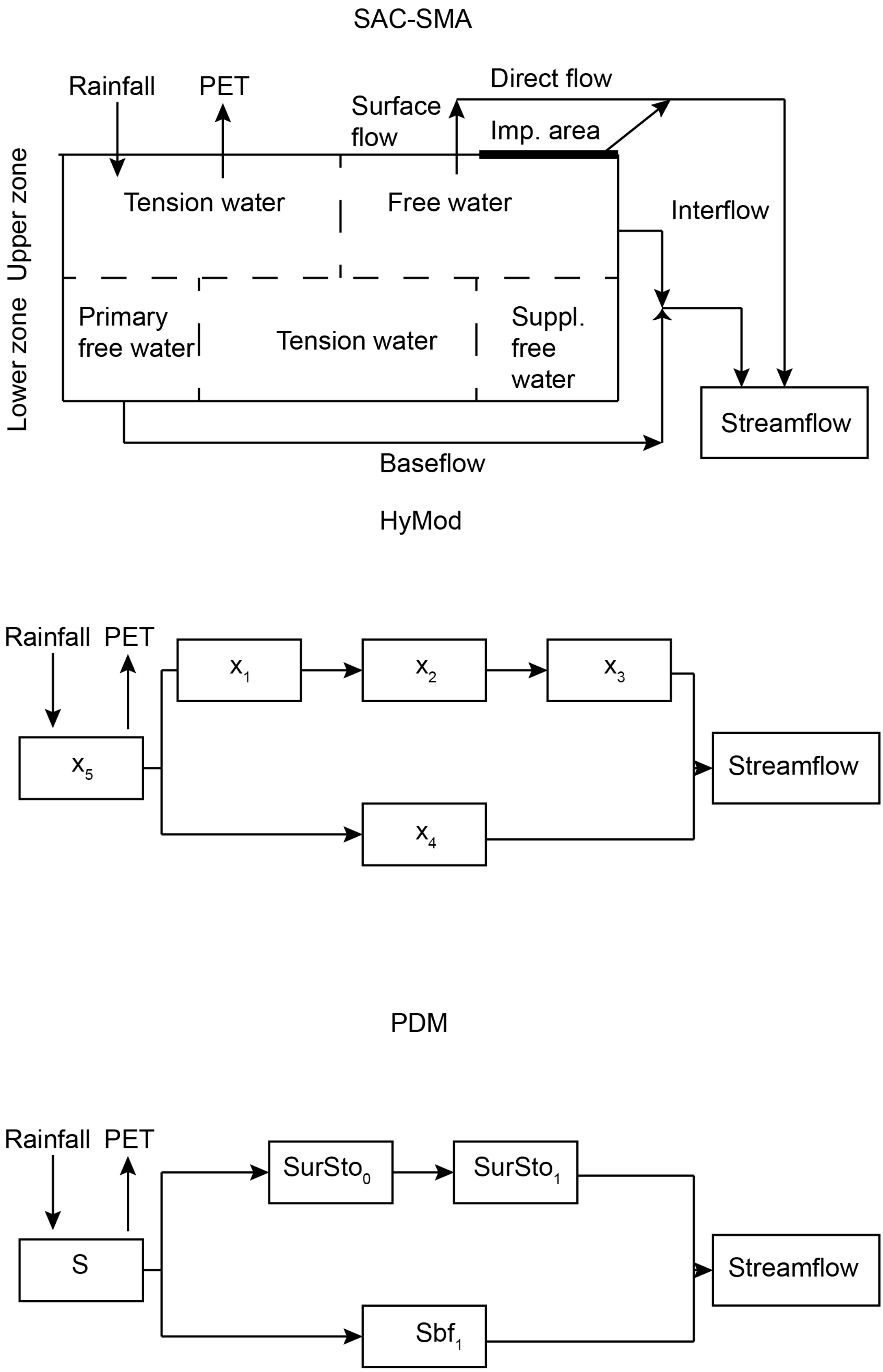

FIGURE 5.1: Diagrams representing the main characteristics of the hydrological models used in the experiment.

5.3.3 PDM

The PDM (*Moore*, 2007) assumes the soil moisture stores within a catchment to have variable capacities that can be represented by a Pareto distribution. Upon incident rainfall, parts of the catchment that have shallow soil moisture stores may be generating runoff whilst other parts were retaining water. The stores were also subject to losing water via groundwater recharge and evapotranspiration. Surface runoff is routed through a cascade of two linear stores, whilst subsurface flow is routed through one linear store. Outflow from both stores were combined as streamflow. The model consists of 4 states, one for each store, as well as 9 parameters.

5.3.4 HyMod

The hydrologic model (HyMod), is a derivative of the PDM (*Moore*, 2007). The model itself consists of a nonlinear soil moisture store succeeded by a series of three linear quick flow stores in parallel with a linear slow flow store. The model consists of 5 states, one for each store, as well as 5 parameters. The parameters govern the maximum storage capacity of the watershed, the spatial variability of the soil moisture store, the separation of flow from the soil moisture store to the quick flow and slow flow stores, and the residence time for the quick flow and slow flow stores, respectively.

TABLE 5.1: Parameters and ranges for hydrological models used in the estimation process.

Parameter	Description	Units	Range
	SAC-SMA		
Capacity thresholds			
$UZTWM$	Upper zone tension water capacity	mm	$1.00 - 150$
$UZFWM$	Upper zone free water capacity	mm	$1.00 - 150$
$LZTWM$	Lower zone tension water capacity	mm	$10.00 - 500$
$LZFPM$	Lower zone free water primary capacity	mm	$10.00-$ 1.00×10^4
$LZFSM$	Lower zone free water supplemental capacity	mm	$5.00 - 400$
UZK	Upper zone free water withdrawal rate	day^{-1}	$1.00 \times 10^{-1}-$ 7.50×10^{-1}
$LZPK$	Lower zone primary free water withdrawal rate	day^{-1}	$1.00 \times 10^{-4}-$ 2.50×10^{-2}
$LZSK$	Lower zone supplemental free water withdrawal	day^{-1}	$1.00 \times 10^{-2}-$ 8.00×10^{-1}
$ZPERC$	Maximum percolation rate	-	$1.00 - 500$
$REXP$	Exponent of the percolation equation	-	$1.00 - 5.00$
$PFREE$	Fraction percolation from upper to lower zone free water storage	-	$0.00-$ 8.00×10^{-2}
$PCTIM$	Minimum impervious fraction of the watershed area	-	$0.00-$ 1.00×10^{-1}
$ADIMP$	Additional impervious area	-	$0.00-$ 4.00×10^{-1}
	PDM		
C_{max}	Maximum store capacity	mm	$1.00 - 500$
b	Pareto distribution exponent that controls spatial variability of C_{max}	-	$1.00 \times 10^{-4}-$ 1.80
b_e	Actual evaporation exponent	-	$0.10 - 5.00$
b_g	recharge function exponent	-	$0.20 - 6.70$
$\mathbf{k_b}$	baseflow constant	hour mm^2	$1.00 - 2000$
C_{min}	minimum store capacity	mm	$0.00 - 500$
S_t	soil tension storage capacity	mm	$0.00 - 500$
k_1	time constant for linear reservoir	hour	$1.00 - 300$
k_2	time constant for linear reservoir	hour	$1.00 \times 10^{-7}-$ 30000
	HyMod		
C_{max}	Maximum store capacity	mm	$1.00 - 500$
b	Pareto distribution exponent that controls spatial variability of C_{max}	-	$0.10 - 2.00$
α	Factor that distributes flow between R_s & R_q	-	0.010 - 0.99
R_s	Residence time of slow flow store	days	0.00 - 0.10
R_q	Residence time of slow flow store	days	0.10 - 0.99

5.4 Data set

5.4.1 General overview

This study used daily rainfall, PET and streamflow data from the study catchment as input to the hydrological models. The $\mathrm{DREAM_{ZS}}$ algorithm in conjunction with model input data reduction methods and a dual objective function were used to estimate $125,000$ unique rainfall time series and model parameter distributions. $125,000$ unique rainfall timer series and model parameter sets were used in order to sufficiently sample the posterior distributions. Remotely sensed soil moisture observations were then assimilated into the models for each of these rainfall time series and model parameter sets.

5.4.2 The catchment of Warwick

Figure 5.2 presents the location of the Warwick catchment within the south-east corner of Queensland, Australia, and the Condamine-Culgoa basin. The experiment presented in this manuscript was conducted on a small subcatchment of the Condamine-Culgoa basin; more specifically the Warwick catchment. The $1360\ \mathrm{km}^2$ catchment fosters a strong agricultural community that has been subjected to several significant flood events. At times of prolonged drought, reaches of the river have ceased to flow. The length of the perennial channels is 78 km whilst the maximum elevation difference along the channel is 308 m. The analysis period is from the 1$^{\mathrm{st}}$ of January 2007 through to the 31$^{\mathrm{st}}$ of March 2013.

5.4.3 Rainfall, PET and streamflow

Daily rainfall data obtained from 8 gauges were aggregated to obtain a catchment areal rainfall estimate using the Inverse Distance Weighting (IDW) method, whilst monthly PET data from the Australian Water Availability Project (AWAP) (*Raupach et al.*, 2012) were used. In applying the IDW method the 5 gauges closest to the catchment centroid were used; if for any given time-step there was no recorded observation at one or more of these gauges then observations from the next nearest gauge/s

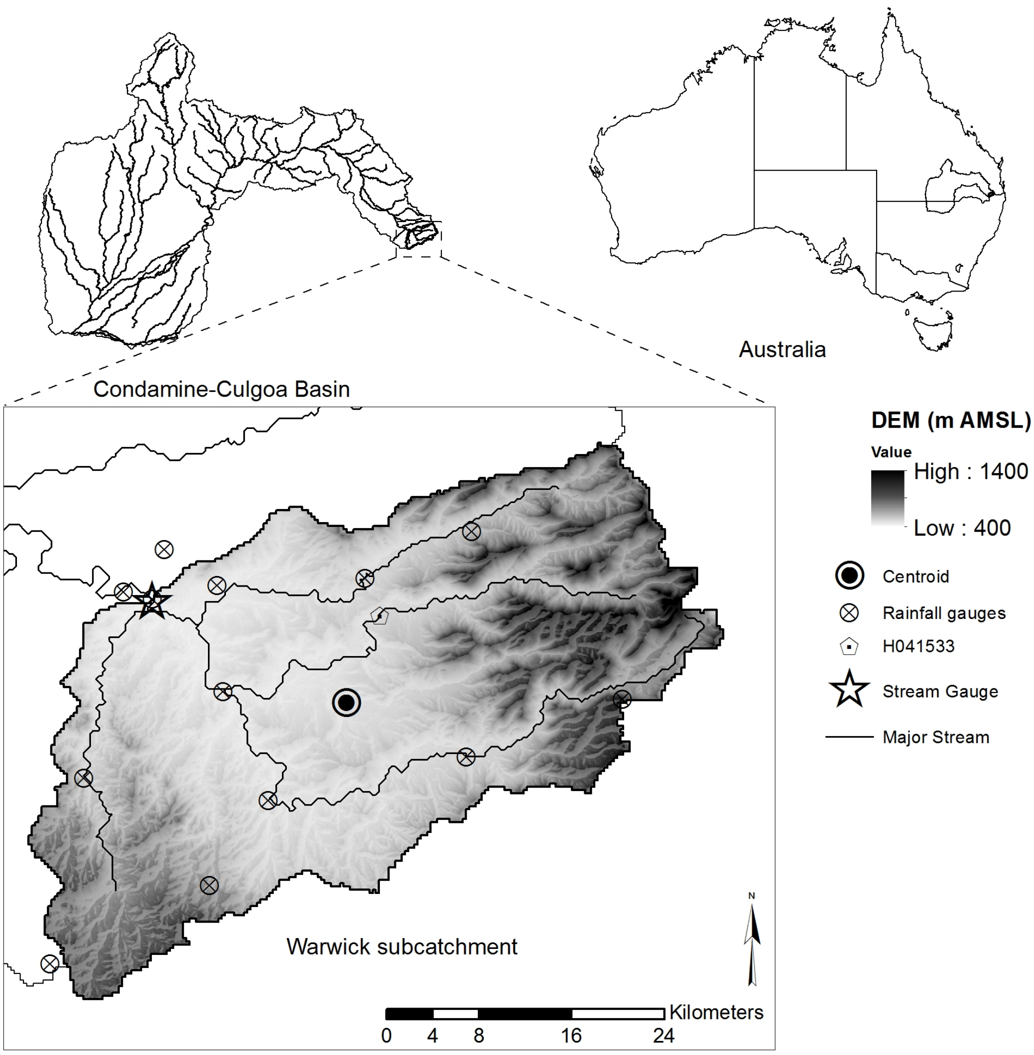

FIGURE 5.2: A locality map showing the study catchment of Warwick, it's location within the Condamine-Culgoa Basin and Australia. The notation "m AMSL" in the legend denotes "meters above mean sea level".

were used. Continuous height measurements from a crump weir were converted to streamflow using periodically updated rating curves (*Queensland Government*, 2017). Daily streamflow and rainfall observations from the 1st of January 2007 to the 31st of March 2013 for the Warwick catchment can be seen in Figure 5.3.

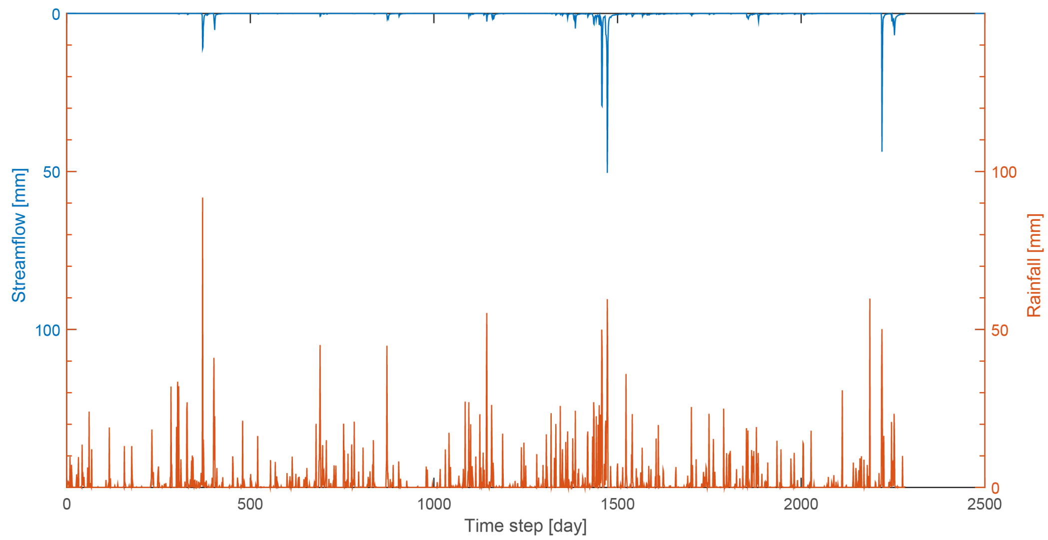

FIGURE 5.3: SMOS (x) and AMSR-E (○) remotely sensed soil moisture observations for the Warwick catchment.

5.4.4 Remotely sensed soil moisture

Remotely sensed soil moisture data from the Soil Moisture Ocean Salinity (SMOS) and Advanced Microwave Scanning Radiometer-Earth observing system (AMSR-E) satellites were separately assimilated into the lumped hydrological models. The AMSR-E soil moisture data set consists of 1078 observations, and were part of the Land Parameter Retrieval Model (LPRM) National Aeronautics and Space Administration (NASA) Level 3 descending product for the time period beginning the 2$^{\text{nd}}$ of January 2007 and ending the 29$^{\text{th}}$ of September 2011 (*Amsterdam and GSFC*, 2012). The SMOS soil moisture data set consists of 581 observations, and were part of the Level 3 descending product obtained from the Centre Aval de Traitement des Données SMOS (CATDS) for the time period beginning the 15$^{\text{th}}$ of January 2010 and ending the 30$^{\text{th}}$ of March 2013. Since the night-time is the time of day when the surface temperature is vertically and horizontally most homogeneous, the descending pass observations were more likely to be representative of spatial soil moisture than the ascending pass observations. Studies have shown the descending pass observations to be more representative of the surface soil moisture (*Draper et al.*, 2009). Consequently, they were selected for this study. Soil moisture for the Warwick catchment was obtained by averaging the 7 SMOS and 4 AMSR-E pixels that cover the Warwick catchment. Even though both products have a similar footprint size, a different number of pixels were used as their centers have different locations. In the time period in which both satellites were active and observing soil moisture there were 387 AMSR-E observations and 307 SMOS observations. There were 194 days in which there were both AMSR-E and SMOS observations. Figure 5.4 shows the AMSR-E and SMOS volumetric soil moisture data used in this study. It can be observed that the AMSR-E soil moisture observations were continuously wetter than the SMOS soil moisture observations.

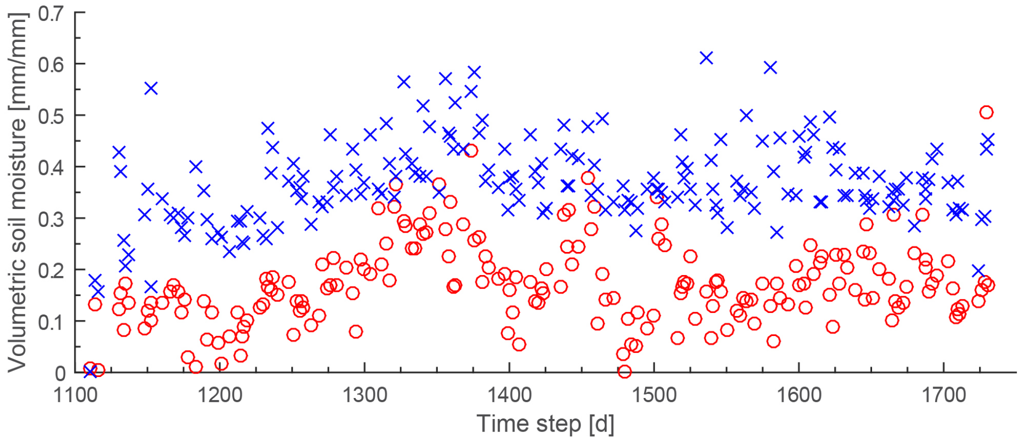

FIGURE 5.4: SMOS (x) and AMSR-E (○) remotely sensed soil moisture observations for the Warwick catchment.

5.5 Experiment design

5.5.1 Rainfall estimation

The rainfall estimation process is described using Figure 5.5. The estimation of rainfall time series along with model parameter distributions began by reducing input data to a dimensionality that is computationally feasible for modern parameter estimation algorithms. As recommended by *Wright et al.* (2017a) the DWT is used to reduce the observed rainfall time series. The rainfall time series for the estimation period was represented by 115 DWT parameters. The Daubechies 1, db1 wavelet and 4 levels of decomposition were chosen to allow for reasonable computational speed. Only the approximation parameters were modified. The hydrological models used a 100 day spin up period before a 1825 day estimation and 357 day evaluation period. Using the $\mathrm{DREAM_{ZS}}$ algorithm a sample of $115\ \mathrm{rainfall} + d$ (model parameters)

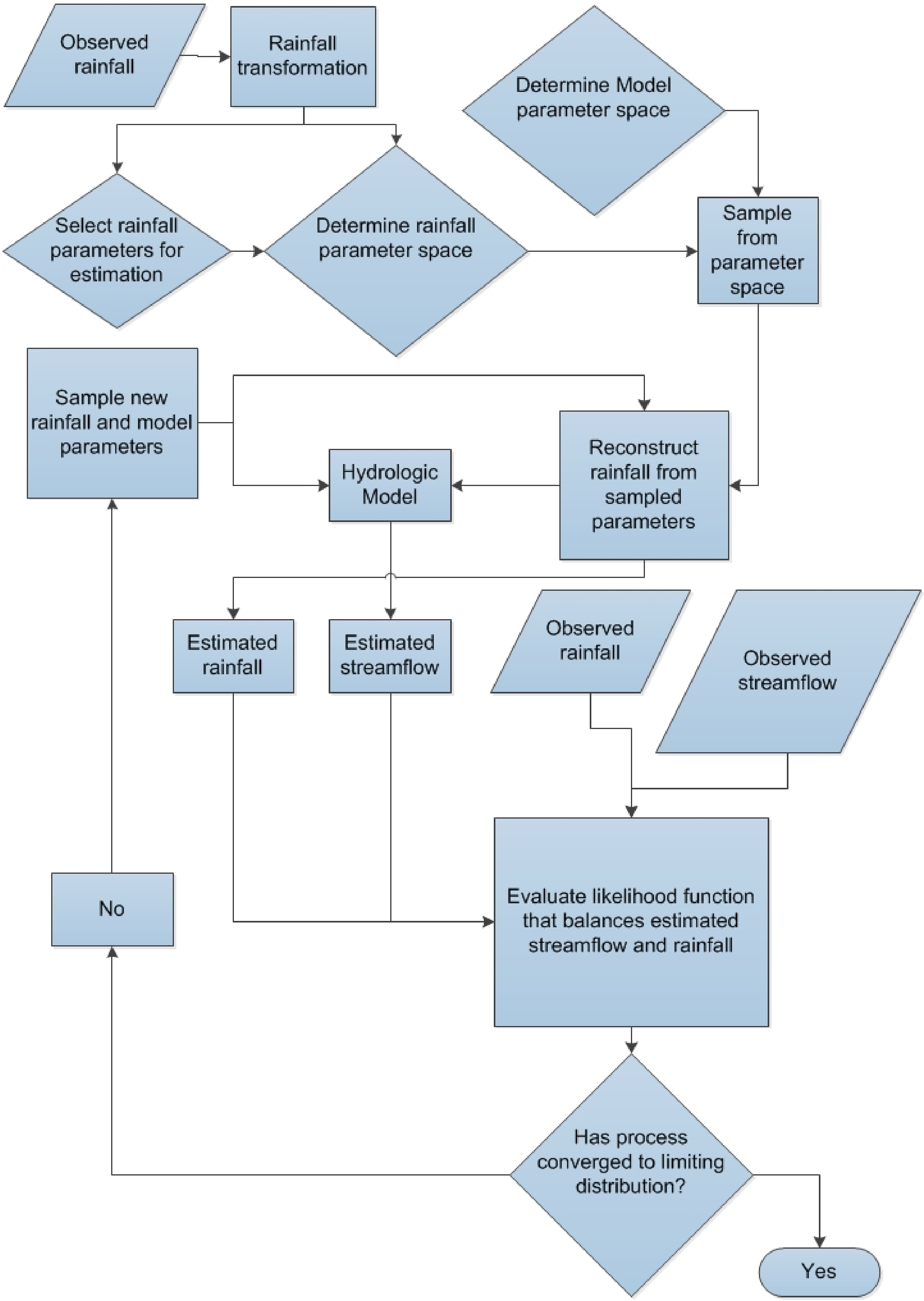

FIGURE 5.5: A representation of the process used to reduce model input data and estimate rainfall for different hydrological models.

was drawn. Rainfall was reconstructed from the parameters before being used as input to the hydrological model. Estimates of both rainfall and streamflow were then evaluated using an objective function that balances streamflow and rainfall. This process is iterated until the sample trajectories were smaller than the $\hat{R}$ convergence diagnostic of 1.2 (*Gelman and Rubin*, 1992). For further detail regarding the rainfall estimation process the reader is referred to *Wright et al.* (2017b). The streamflow simulations generated in the rainfall estimation process were benchmarked against a traditional calibration approach, which assumes no input error and only estimates model parameters. $\mathrm{DREAM_{ZS}}$ was again used as the sampling algorithm with a Gaussian objective function (*Thiemann et al.*, 2001).

5.5.2 Assimilation of remotely sensed soil moisture observations

To assess the compatibility of remotely sensed soil moisture with different hydrological models and rainfall and parameter estimates, the EnKF has been chosen to assimilate remotely sensed soil moisture observations into the hydrological models. A brief description of the EnKF is given in this section; for a more complete discussion the reader is referred to *Reichle et al.* (2002).

Model error was taken into account by running each model with a 32 member ensemble. The ensemble members were generated by adding random multiplicative heteroscedastic Gaussian error to the input rainfall series. The standard deviation (SD) of the Gaussian distribution was equivalent to 10% of the observation. The model was propagated forward in time until an observation was available for assimilation.

When an observation was made available the state vectors were, when possible, transformed to unitless coordinates. Descriptions of the states for each of the models were given in Table 5.2. For the SAC-SMA model this is written as

$$\begin{aligned}\mathbf{X}_{\mathrm{i}}^{\mathrm{SAC-SMA}} = [&\mathrm{UZTWC_i/UZTWM\ UZFWC_i/UZFWM\ LZTWC_i/LZTWM}\\ &\mathrm{LZFSC_i/LZFSM\ LZFPC_i/LZFPM\ ADIMC_i/(UZTWM + LZTWM)}]^{\mathrm{T}},\end{aligned} \tag{5.1}$$

TABLE 5.2: Hydrological model states used in the data assimilation process.

State	Description	
	SAC-SMA	Units
$UZTWC$	Upper zone tension water content	mm
$UZFWC$	Upper zone free water content	mm
$LZTWC$	Lower zone tension water content	mm
$LZFSC$	Lower zone free water supplemental content	mm
$LZFPC$	Lower zone free water primary content	mm
$ADIMC$	Additional impervious area storage content	mm
	PDM	
S	Soil moisture store	mm
Sbf_1	Baseflow store	mm
$SurSto_0$	1st surface store	mm
$SurSto_1$	2nd surface store	mm
	HyMod	
x_1	1st surface store	mm
x_2	2nd surface store	mm
x_3	3rd surface store	mm
x_4	Baseflow store	mm
x_5	Soil moisture store	mm

where $\mathbf{X}_{\mathrm{i}}^{\mathrm{SAC-SMA}}$ is the SAC-SMA state vector at time step i. For the PDM this is written as

$$\mathbf{X}_{\mathrm{i}}^{\mathrm{PDM}} = [\mathrm{S/S_{max}}\ \mathrm{Sbf_1}\ \mathrm{SurSto_0}\ \mathrm{SurSto_1}]^{\mathrm{T}}, \tag{5.2}$$

where $\mathbf{X}_{\mathrm{i}}^{\mathrm{PDM}}$ is the PDM state vector at the time step i. For HyMod this is written as

$$\mathbf{X}_{\mathrm{i}}^{\mathrm{HyMod}} = [\mathrm{x_1}\ \mathrm{x_2}\ \mathrm{x_3}\ \mathrm{x_4}\ \mathrm{x_5/C_{max}}]^{\mathrm{T}}, \tag{5.3}$$

where $\mathbf{X}_{\mathrm{i}}^{\mathrm{HyMod}}$ is the HyMod state vector at time step i. To be compatible with the volumetric soil moisture observations from RS the saturated soil moisture model states need to be scaled by their associated porosity. The dominant soil type in Warwick has been identified as loamy sand (*CSIRO*, 2017), for which the porosity was determined to be 0.45 (*Rawls et al.*, 1982). $\mathbf{X}_{\mathrm{i}}^{\mathrm{SAC-SMA}}$ was transformed to the observation space by $\mathrm{H}^{SAC-SMA}$ (the SAC-SMA transformation matrix), where

$$\mathrm{H^{SAC-SMA}} = [0.45\ 0\ 0\ 0\ 0\ 0]. \tag{5.4}$$

$\mathbf{X}_\mathrm{i}^\mathrm{PDM}$ was transformed to the observation space by H^{PDM}, where

$$\mathrm{H^{PDM}} = [0.45\ 0\ 0\ 0]. \tag{5.5}$$

$\mathbf{X}_\mathrm{i}^\mathrm{HyMod}$ was transformed to the observation space by H^{HyMod}, where

$$\mathrm{H^{HyMod}} = [0\ 0\ 0\ 0\ 0.45]. \tag{5.6}$$

The RS soil moisture observations were assimilated without applying bias correction techniques such as Cumulative Distribution Function (CDF) matching. The innovations in the assimilation routine were defined as

$$\mathbf{innov}_\mathrm{i} = \mathbf{Obs}_\mathrm{i} - \mathrm{H^{model}}\mathbf{X}_\mathrm{i}^\mathrm{model}, \tag{5.7}$$

where $\mathbf{innov}_\mathrm{i}$ and $\mathbf{Obs}_\mathrm{i}$ were the innovations and the observations at the ith time-step, respectively. $\mathrm{H^{model}}$ is the transformation matrix for a selected mode and transforms saturated soil moisture into volumetric soil moisture. $\mathbf{X}_\mathrm{i}^\mathrm{model}$ is the state vector at time step i for a selected model..

5.6 Results and discussion

5.6.1 Estimated rainfall and impact on streamflow forecast

Table 5.3 shows the root mean square error (RMSE) results for the traditional calibration and rainfall estimation approaches, demonstrating that the use of model input data reduction and a dual objective function was able to produce superior streamflow simulations compared to a traditional calibration approach in which only model parameters were estimated. A reduction in RMSE between observed and simulated streamflow was achieved for the two calibration approaches for each model. Yet, not all models achieved the same reduction in RMSE. The observed dif-

TABLE 5.3: Maximum A Posteriori (MAP) and SD of RMSE obtained for a traditional calibration approach and joint calibration and rainfall estimation approach.

	RMSE streamflow $[m^3 s^{-1}]$	
	Traditional	Rainfall
Model	MAP	MAP
SAC-SMA	0.3606	0.2154
HyMod	0.4828	0.4141
PDM	0.4010	0.3452

ference in the reduction in RMSE between models can be due to over-parameterization of a model or a model's inadequate ability to account for complex dynamics within a catchment. To determine this, the rainfall estimates for each model will be analyzed. A cumulative time series of the observed and estimated rainfall is displayed in Figure 5.6. It is worth noting that the mean rainfall estimates from the SAC-SMA model were considerably closer to the observed rainfall than those obtained from the PDM and Hymod. The rainfall estimated using the PDM and HyMod has a significantly smaller variance than the variance in rainfall estimated by the SAC-SMA. The mean rainfall estimates from all of the models were drier than the observed rainfall. The cause of this phenomenon cannot be determined from this study. During the rainfall estimation period the total catchment areal rainfall volume using the IDW method is calculated to be 3205 mm. Of the 5 main rain gauges used, the minimum and maximum total catchment areal rainfall volume is 2764 and 3406 mm, respectively. Combined with the fact that dry rainfall estimates produced superior streamflow to that obtained from using the observed rainfall, this suggests that the spread of rain gauges provides insufficient density to capture the true catchment areal rainfall volume.

5.6.2 Daily mean innovations

Daily innovations for each of the ensemble members were averaged to produce a mean for that time step. The daily mean innovation for each of the $125,000$ rainfall time series were shown in Figure 5.7. Each of the panels represent a different model and RS SM combination. If the modeled soil moisture and RS SM were unbiased it

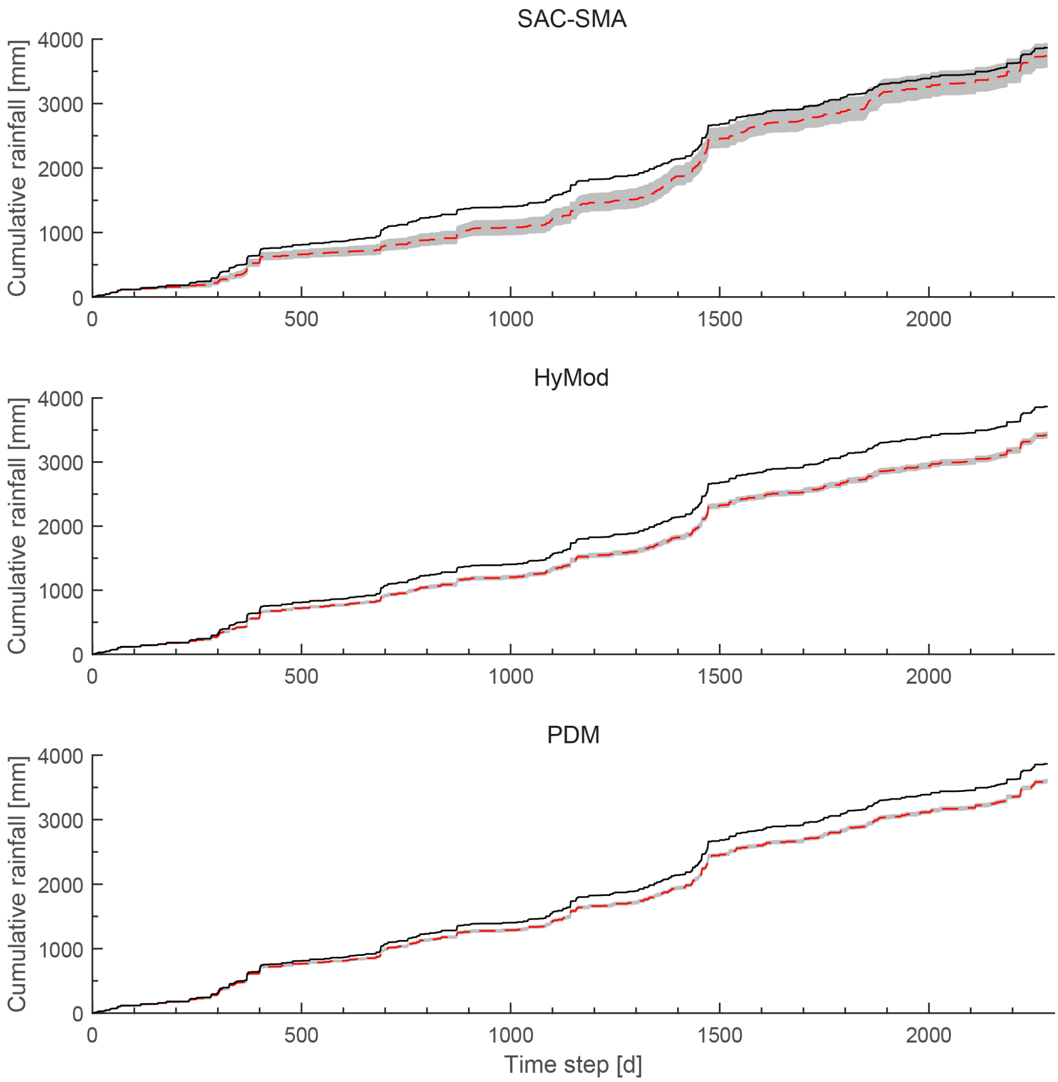

FIGURE 5.6: Cumulative rainfall series for the Warwick catchment. Observed rainfall is plotted using the black line while the mean and 5^{th} to 95^{th} percentile rainfall estimates were represented by the dashed red line and grey shading respectively.

is expected that the innovations fluctuate about zero. Deviations from this demonstrate bias in either the modeled SM and/or the RS SM. When assimilating SMOS RS SM both the SAC-SMA model and HyMod demonstrate low overall bias. This indicates that SM modelled by the SAC-SMA model and HyMod is in agreement with the SMOS RS SM observations. A positive bias, in which the means of the innovation time series were consistently larger than 0, was observed when assimilating the AMSR-E RS SM into both the SAC-SMA model and HyMod. Interestingly, this observation was reversed for the PDM. A low bias is observed when assimilating

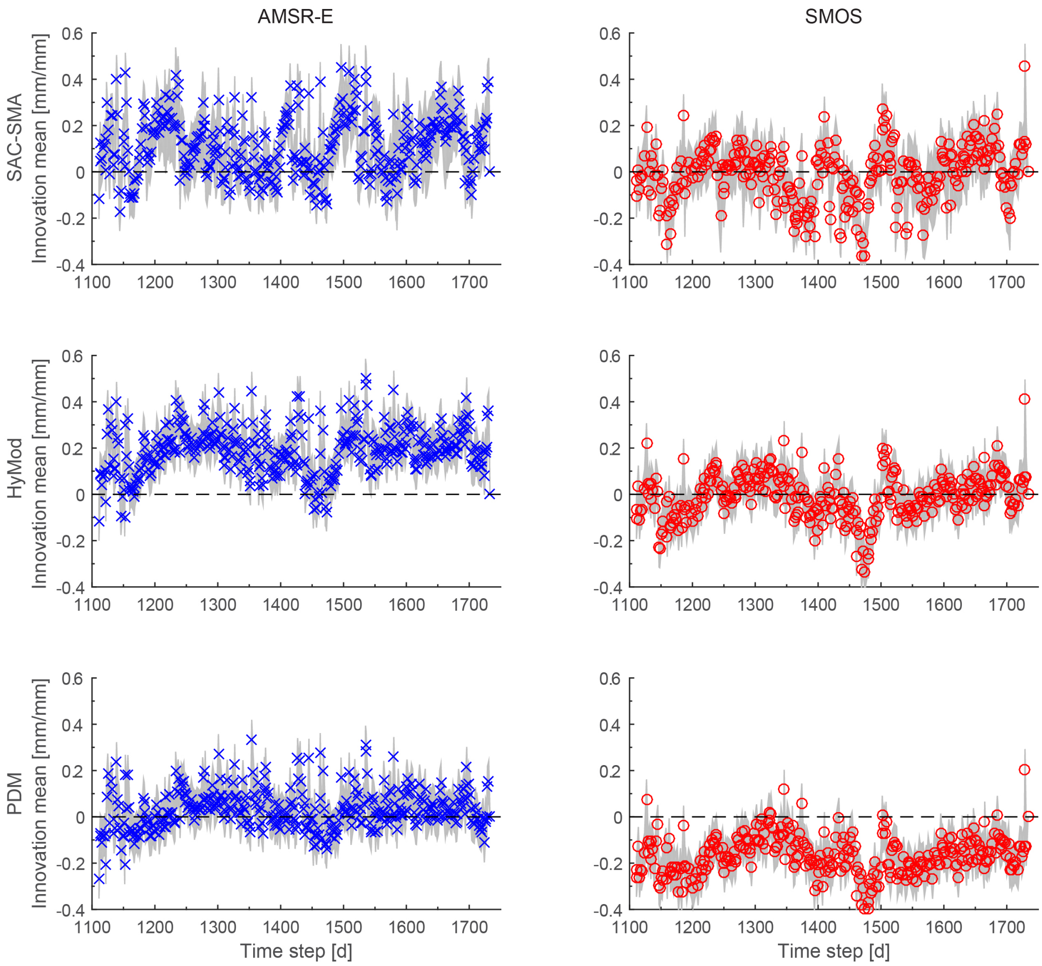

FIGURE 5.7: Daily innovation mean for the ensembles. The symbols represent the mean daily innovation mean for the 125,000 rainfall time series and model parameter sets whilst the grey shading indicates the 5^{th} to 95^{th} percentile daily innovation mean. Each panel represents a different model/RS SM assimilation combination.

AMSR-E RS SM and a negative bias was observed when SMOS RS SM was assimilated. It can be observed in Figure 5.4 that, when compared to the SMOS RS SM observations, the AMSR-E RS SM observations show a low fluctuation about the mean. This suggests that both the SM simulated by the SAC-SMA and HyMod capture the fluctuations in SM that were observed by SMOS. Conversely, as the innovation mean fluctuates approximately about zero when the PDM assimilates AMSR-E soil moisture, the SM simulated by the PDM does not capture the SM fluctuations observed by SMOS. These findings demonstrate that innovations from assimilating unscaled RS SM observations will not automatically lead to biased assimilation results . Biased innovations may be present due to a combination of poor rainfall estimates and soil moisture observations depending on the applied model. Unbiased innovations alone do not guarantee adequate model simulations or rainfall estimates. Areal rainfall obtained from gauged observations does not produce the best streamflow simulations. This suggests that the rainfall observations may not be representative of catchment rainfall.

When evaluating a rainfall-runoff models suitability for forecasting purposes it is essential that the model is able to adequately simulate past streamflow observations. The results demonstrate that good streamflow simulations and unbiased innovations can be obtained from biased rainfall estimates. Consequently, careful consideration needs to be paid towards uncertainty in all components of the water cycle before claims were made that a rainfall-runoff model is able to simulate good streamflow for the right reasons.

5.6.3 Innovation mean for the assimilation period

Over the course of the assimilation period the innovation mean at each time step will ideally fluctuate about zero. The mean of the innovation means for an entire time series is calculated for each of the $125,000$ rainfall time series and parameter sets, models and RS SM product and presented in Figure 5.8. The best representations were centered around 0 and have rainfall volumes closest to the observed rainfall volume over the rainfall estimation period. When assimilating SMOS SM,

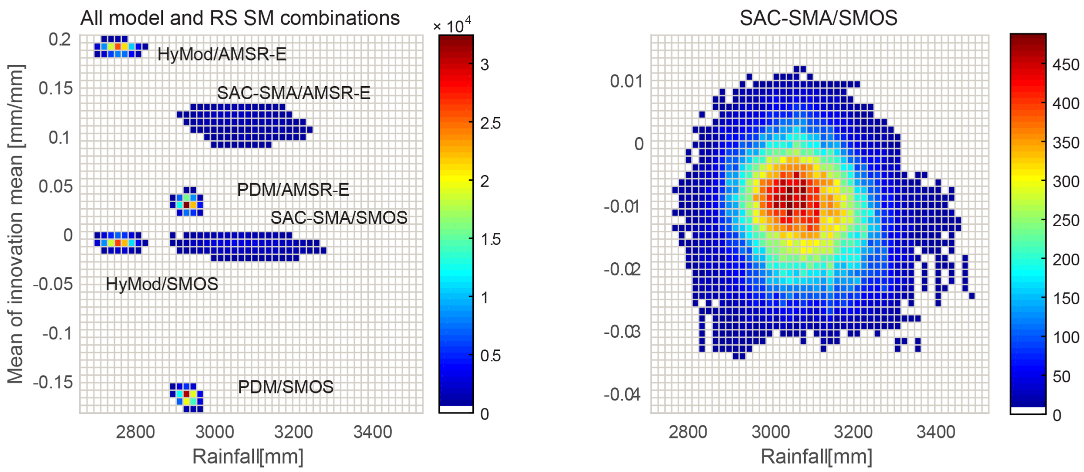

FIGURE 5.8: The left panel is 3-D histogram showing the mean of the daily innovation mean for each of the 125,00 rainfall time series and model parameter sets and the estimated rainfall volume for the estimation period for each of the models and RS SM combinations. The right panel zooms into the 3-D histogram for the case when the SMOS RS SM product is assimilated into the SAC-SMA model.

the innovations from the SAC-SMA model were largely contained between 0.01 and -0.03 (mm/mm). The innovations for the 5 remaining experimental combinations do not have both positive and negative values. Consequently, some bias is present. The rainfall volumes estimated with the SAC-SMA model were contained between 2900 and 3300 mm. This variance in estimated rainfall volume is larger than that shown by the HyMod and PDM rainfall estimates. Further, the extent of rainfall volumes obtained by the SAC-SMA model encompasses the observed rainfall volume of 3205 mm for the Warwick catchment. This suggests that the unbiased rainfall estimates obtained using the SAC-SMA will benefit from the unbiased SM observations from SMOS. Without CDF matching, the SAC-SMA configuration will not benefit from assimilating AMSR-E SM observations. When assimilating SMOS SM observations into HyMod the mean of innovation mean is close to 0. Conversely, the biased rainfall estimates obtained from HyMod demonstrate that the rainfall estimates were unrealistic and that the innovations alone provide insufficient evidence to draw a positive conclusion. This bias is made more evident by the increased discrepancy between the innovations obtained for SAC-SMA and HyMod when assimilating AMSR-E observations, and the innovations obtained for SAC-SMA and HyMoD when assimilating SMOS observations. Conversely, the low bias observed for innovations in the PDM/AMSR-E experiment demonstrates that assimilating a biased soil moisture product into a biased model or model with biased rainfall estimates may still yield good results. To obtain robust streamflow forecasts unbiased models need to be paired with unbiased rainfall observations/forecasts/estimates and unbiased RS SM observations. Uncertainty in all components of the water cycle needs to be considered.The demonstrated methodology can be used as tool to estimate rainfall, or reject models and RS SM observations for a given catchment.

5.7 Conclusions

Previous studies have demonstrated that rainfall estimates obtained via the sole inversion of either streamflow or soil moisture are often unrealistic or lack temporal specificity. This research builds upon a previously developed rainfall estimation methodology by analyzing the rainfall estimates using innovations from the assimilation of RS SM data. The methodology presented can be used by hydrologists to make informed choices regarding model choice and satellite choice. Permutations of estimated rainfall time series, model parameter sets, hydrological models, and RS SM data are analyzed. Rainfall estimates were obtained for the SAC-SMA, HyMod and PDM rainfall-runoff models via a process that involved the dimensionality reduction of input data using the DWT. An objective function that balances estimates of streamflow and rainfall was used in conjunction with the sampling algorithm $\mathrm{DREAM_{ZS}}$ to simultaneously estimate model parameters and rainfall time series. Cumulative plots of the estimated rainfall time series showed that superior streamflow estimates could be simulated with model dependent rainfall estimates, and that all models demonstrated improved streamflow simulations with lower thant observed rainfall time series estimates. Further, the range of estimated rainfall time series was found to be dependent on the model. Data assimilation using the EnKF produced innovations close to 0 when SMOS and AMSR-E RS SM were assimilated into HyMod and PDM respectively. Yet, the rainfall estimates from these models were still discarded as their rainfall volumes during the rainfall estimation period were outside the range of rainfall volumes observed at the gauge. Realistic rainfall estimates and EnKF innovations were obtained with the SAC-SMA and SMOS RS SM. To be considered robust, rainfall estimates obtained via inversion need to produce superior streamflow simulations, be able to simulate soil moisture states that exhibit little to no bias when compared to RS SM observations, and be within an acceptable range of gauge based rainfall observations.

5.8 Acknowledgements

The authors would like to extend their gratitude to the anonymous reviewers for their comments and recommendations. The authors would also like to thank the Bureau of Meteorology for the provision of data. This work was supported by the

Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE) (www.massive.org.au), a Monash University Engineering Research Living Allowance stipend, and a top up scholarship from the Bushfire & Natural Hazards Cooperative Research Centre. Valentijn Pauwels was funded by ARC grant FT130100545.

Chapter 6

Conclusions and further research

6.1 Overview

The research conducted throughout this thesis made steps towards increasing the skill of flood forecasts. Over three tasks this research developed and utilized techniques that enhance the hydrological communities' understanding of uncertainty in rainfall observations and the influence those uncertainties have on streamflow simulations from rainfall-runoff models. As a greater understanding of uncertainty in rainfall observations and the way in which those observations influence streamflow simulations is developed, techniques to condition rainfall forecasts become more reliable. As rainfall forecasts become more reliable, flood forecasting skill increases. For this research to be operationalized further work involving a variety of hydrologic basins and models needs to be conducted. The methodologies described would need to be run for each catchment such that QPF's can be conditioned on rainfall estimates that are known to provide good hydrological simulations.

6.2 Summary of main findings

6.2.1 Hydrologic model input data reduction

Rainfall and its uncertainty can be estimated by describing rainfall observations using parameters and utilizing model inversion techniques. As it is computationally infeasible to estimate a unique parameter for each rainfall observation, there exists a need to be able to efficiently and effectively reduce rainfall observations to a smaller number of parameters. To ensure the robust inversion of rainfall-runoff models to obtain rainfall estimates, methods to reduce the dimensionality of hydrologic model input data were explored. Due to their wide-spread acceptance as transforms for model input data reduction in fields outside of hydrology, the DCT and DWT were used to compress and reconstruct rainfall observations from the MOPEX data set. Succinct descriptions of the DCT and DWT were given along with an outline of possible benefits each of the transform may provide. High- or low-frequency parameters of the DCT can be estimated. Conversely, if the DWT is used, a modeler can choose to estimate a combination of either time or frequency DWT parameters. Using standard simulation performance summary metrics, descriptive statistics, and peak errors to compare the ability of compressed DWT and DCT transform parameters to reconstruct MOPEX rainfall data, it was determined that the DWT was most effective at preserving high-magnitude and transient rainfall events. After analysis of the bias, variance, skewness, and kurtosis, it was demonstrated that rainfall reconstructions from the DWT were closer to the observed rainfall data, and that the DWT was more effective at preserving long term trends. Consequently, it is recommended that the DWT be used as a model input data reduction technique for hydrologic studies that have both short and long time steps.

6.2.2 Rainfall estimation

Since the DWT was found to be more effective than the DCT at preserving rainfall observations, the DWT was used to reduce model input data for the estimation of input uncertainty. While the DWT can be used to reduce the dimensionality of model input data that originates from any measurement instrument, in this study the DWT was only used to reduce the dimensionality of gauge-based rainfall estimates. Using the $\mathrm{DREAM}_{(\mathrm{ZS})}$ sampling algorithm,and a likelihood function that balances input rainfall and streamflow error allowed for multiple configurations of DWT parameters to be used to estimate model parameter distributions and entire rainfall time series. This methodology allows rainfall to be estimated when none was observed.

When compared to the benchmark sole estimation of model parameters, the simultaneous estimation of DWT rainfall and model parameters yielded realistic rainfall estimates and streamflow simulations. The RMSE of these streamflow simulations was improved by a factor of up to 1.78. Consequently, a methodology to realistically estimate rainfall time series has been developed. This methodology was used in a subsequent study that used 3 models to compare rainfall time series, their respective model parameters, and their ability to simulate streamflow and soil moisture observations.

The use of the DWT as a model input data reduction technique in conjunction with model inversion techniques provides an increased understanding of hydrologic uncertainty by providing a technique that both improves streamflow simulations and estimates rainfall input, including when none was observed. The efficiency of the model inversion process along with computational power place an upper limit on the resolution of uncertainty and length of rainfall time series that are able to estimated.

6.2.3 Analysis of rainfall estimates

Past attempts to estimate rainfall through the inversion of streamflow or soil moisture have been either unrealistic or lacked temporal specificity respectively. Consequently, this study filled the need for the development of a methodology that constrains rainfall estimates obtained through the inversion of streamflow observations with soil moisture observations. The developed methodology utilized past studies that reduced model input data and simultaneously estimated model parameter distributions and entire rainfall series. These techniques were applied to the SAC-SMA, HyMod, and PDM rainfall-runoff models. Cumulative plots of estimated rainfall series demonstrated that rainfall estimates are model dependent. Compared to the benchmarks in which only model parameters are estimated, all models were able to obtain superior streamflow estimates when both rainfall time series and model parameters were estimated. Analysis of EnKF innovations obtained from each model when different RS SM products were assimilated demonstrated that innovations close to zero can be obtained for models that have biased rainfall estimates. Rainfall estimates that were close to the gauge based observations were obtained from the SAC-SMA model. Further, innovations either side of zero were obtained when SMOS RS SM observations were assimilated into the model parameter and rainfall time series estimates obtained from the SAC-SMA model. To be considered robust, rainfall estimates obtained via the inversion of streamflow need to be able to; produce superior streamflow simulations, simulate soil moisture states that exhibit little to no bias when compared to RS SM observations, as well as estimate rainfall that is within an acceptable range of gauge based rainfall observations.

It has been successfully demonstrated that rainfall estimates can be constrained by soil moisture observations. Different configurations of models and choice of RS SM product for assimilation demonstrate white noise EnKF innovations. In some situations this is a result of the assimilation of the biased AMSR-E soil moisture into models that obtain biased rainfall estimates. When the unbiased SMOS RS SM product was assimilated into the the SAC-SMA model innovations either side of 0 were obtained, providing an additional element of physical realism to the rainfall retrieval process. The results obtained do not indicate that the assimilation of soil moisture observations restricts the efficacy of the rainfall retrieval process in the presence of model structural inadequacy.

6.3 Opportunities for further research

The provision of improved QPF's is expected to increase flood forecast skill. After being constrained by RS SM observations, the rainfall estimates obtained using model input data reduction and inversion techniques can be considered realistic. Further, in comparison to the gauge based rainfall observations, the rainfall estimates simulate superior streamflow. These rainfall estimates can be used in place of rainfall observations to condition and improve QPFs.

Since the reduction of rainfall dimensionality in hydrology is a relatively new concept there are still many areas that can be explored. Model input data can be reduced using any number of different transforms. As the choice of analysis wavelet influenced how errors were corrected for, the exploration of different wavelet families in the reduction of model input data may be used to extract information about the uncertainties present in the modeling process.

Since, there is a limitation on the number of transform parameters that can be estimated within a feasible time frame the identification of transform parameters that convey the most detail may increase the effectiveness of the estimation process. However, doing this may mean that parameters are only estimated for rainfall events that show the greatest fluctuation in observed rainfall. An alternative approach would be to selectively estimate rainfall parameters at times when the simulated streamflow deviates significantly from the observed streamflow.

For the estimation process to converge in a desirable time frame it is more important that the search space of the parameter estimation problem be kept to a minimum than it is that the number of parameters be kept to a minimum. The use of informed priors places a restriction on the search space and the way it is sampled. Assumptions about rainfall errors can aid in the determination of informed priors. Consequently, it is expected that by developing methods for which informed priors can be used in the rainfall estimation process, a more efficient and effective rainfall estimation process could be developed.

An analysis of rainfall estimates obtained using likelihood functions that do not involve the specification of a multiplicative error structure for rainfall will provide methods for which further knowledge regarding the structure of rainfall errors can be obtained. Developing a likelihood function that involves soil moisture will restrict the possible rainfall time series and model parameters that can be estimated. This will allow for a greater understanding of rainfall and model structural uncertainty.

It is expected that performing this analysis on a variety of models, SM data sets, and catchments with different sizes and climate will provide further validation of the robustness of the techniques developed. Lastly, streamlining of the rainfall estimation and constraint process will create a user friendly approach that promotes further research.

Bibliography

World Meteorological Organization (2008), Guide to meteorological instruments and methods of observation.

Abera, W., L. Brocca, and R. Rigon (2016), Comparative evaluation of different satellite rainfall estimation products and bias correction in the upper blue nile (ubn) basin, *Atmospheric Research*, doi:10.1016/j.atmosres.2016.04.017.

Adamovic, M., I. Braud, F. Branger, and J. Kirchner (2015), Assessing the simple dynamical systems approach in a mediterranean context: Application to the ardèche catchment (france), *Hydrology and Earth System Sciences*, *19*(5), doi: 10.5194/hess-19-2427-2015.

Ahmed, N., T. Natarajan, and K. Rao (1974), Discrete cosine transform, *IEEE Transactions on Computers*, *C-23*(1), 90–93, doi:10.1109/T-C.1974.223784.

Ajami, N., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, *43*(1), doi:10.1029/2005WR004745.

Albergel, C., P. de Rosnay, C. Gruhier, J. Muñoz-Sabater, S. Hasenauer, L. Isaksen, Y. Kerr, and W. Wagner (2012), Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, *Remote Sensing of Environment*, *118*, 215–226, doi:10.1016/j.rse.2011.11.017, cited By 176.

Alvarez-Garreton, C., D. Ryu, A. Western, C.-H. Su, W. Crow, D. Robertson, and C. Leahy (2015), Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: Comparison between lumped and semi-distributed schemes, *Hydrology and Earth System Sciences*, *19*(4), 1659–1676, doi: 10.5194/hess-19-1659-2015, cited By 19.

Amsterdam, V. U., and N. GSFC (2012), LPRM/TMI/TRMM daily l3 day surface soil moisture, ancillary params, and QC.

Anderson, R., V. Koren, and S. Reed (2006), Using ssurgo data to improve sacramento model a priori parameter estimates, *Journal of Hydrology*, *320*(1-2), 103–116, doi:10.1016/j.jhydrol.2005.07.020.

Berthet, L., V. Andréassian, C. Perrin, and P. Javelle (2009), How crucial is it to account for the antecedent moisture conditions in flood forecasting? comparison of event-based and continuous approaches on 178 catchments, *Hydrology and Earth System Sciences*, *13*(6), 819–831.

Beven, K. (2006), A manifesto for the equifinality thesis, *Journal of Hydrology*, *320*(1-2), 18–36, doi:10.1016/j.jhydrol.2005.07.007.

Beven, K., and P. Young (2013), A guide to good practice in modeling semantics for authors and referees, *Water Resources Research*, *49*(8), 5092–5098.

Blazkova, S., and K. Beven (2009), A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resources Research*, *45*(6), doi:10.1029/2007WR006726.

Bowler, N. E., C. E. Pierce, and A. W. Seed (2006), Steps: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled nwp, *Quarterly Journal of the Royal Meteorological Society*, *132*(620), 2127–2155, cited By (since 1996):65 Export Date: 8 October 2014.

Boyle, D., H. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resources Research*, *36*(12), 3663–3674, doi:10.1029/2000WR900207.

Brocca, L., T. Moramarco, F. Melone, and W. Wagner (2013), A new method for rainfall estimation through soil moisture observations, *Geophysical Research Letters*, *40*(5), 853–858, doi:10.1002/grl.50173.

Brocca, L., L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani (2014), Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, *Journal of Geophysical Research: Atmospheres*, *119*(9), 5128–5141.

Brocca, L., C. Massari, L. Ciabatta, T. Moramarco, D. Penna, G. Zuecco, L. Pianezzola, M. Borga, P. Matgen, and J. Martinez-Fernandez (2015), Rainfall estimation from in situ soil moisture observations at several sites in Europe: an evaluation of the SM2RAIN algorithm, *Journal of Hydrology and Hydromechanics*, *63*(3), 201–209, doi: {10.1515/johh-2015-0016}.

Brouwer, R., and R. Van Ek (2004), Integrated ecological, economic and social impact assessment of alternative flood control policies in the Netherlands, *Ecological Economics*, *50*(1-2), 1–21.

Bulygina, N., and H. Gupta (2011), Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, *Water Resources Research*, *47*(5), doi:10.1029/2010WR009614.

Butts, M., J. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *Journal of Hydrology*, *298*(1-4), 242–266, doi:10.1016/j.jhydrol.2004.03.042.

Castelli, F., D. Entekhabi, and E. Caporali (1999), Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance, *Water Resources Research*, *35*(10), 3115–3125, cited By (since 1996):45 Export Date: 19 November 2014.

Chhibber, A., and R. Laajaj (2013), *The Interlinkages between Natural Disasters and Economic Development*, Oxford University Press, New York, doi:10.1093/acprof:oso/9780199841936.003.0003.

Ciabatta, L., L. Brocca, C. Massari, T. Moramarco, S. Puca, A. Rinollo, S. Gabellani, and W. Wagner (2015), Integration of satellite soil moisture and rainfall observations over the Italian territory, *Journal of Hydrometeorology*, doi:10.1175/JHM-D-14-0108.1.

Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby (2004), The schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields, *Journal of Hydrometeorology*, *5*(1), 243–262.

Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *Journal of Hydrology*, *375*(3-4), 613–626.

Crow, W., and D. Ryu (2009), A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, *Hydrology and Earth System Sciences*, *13*(1), 1–16.

Crow, W., M. Van Den Berg, G. Huffman, and T. Pellarin (2011), Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART), *Water Resources Research*, *47*(8), doi:10.1029/2011WR010576.

Crow, W. T. (2007), A novel method for quantifying value in spaceborne soil moisture retrievals, *Journal of Hydrometeorology*, *8*(1), 56–67.

Crow, W. T., G. J. Huffman, R. Bindlish, and T. J. Jackson (2009), Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals, *Journal of Hydrometeorology*, *10*(1), 199–212, export Date: 9 September 2014.

CSIRO (2017), ASRIS - Australian Soil Resource Information System.

Daubechies, I. (1990), The wavelet transform, time-frequency localization and signal analysis, *IEEE Transactions on Information Theory*, *36*(5), 961–1005, doi:10.1109/18.57199.

De Vleeschouwer, N., and V. Pauwels (2013), Assessment of the indirect calibration of a rainfall-runoff model for ungauged catchments in flanders, *Hydrology and Earth System Sciences*, *17*(5), 2001–2016, doi:10.5194/hess-17-2001-2013.

Deloitte Access Economics (2013), Building our nations resilience to natural disasters, doi:http://apo.org.au/node/40437.

Di Baldassarre, G., and A. Montanari (2009), Uncertainty in river discharge observations: A quantitative analysis, *Hydrology and Earth System Sciences*, *13*(6), 913–921, cited By 220.

Draper, C., J. Walker, P. Steinle, R. de Jeu, and T. Holmes (2009), An evaluation of amsr-e derived soil moisture over australia, *Remote Sensing of Environment*, *113*(4), 703–710, doi:10.1016/j.rse.2008.11.011, cited By 203.

Duan, Q., S. Sorooshian, and V. K. Gupta (1994), Optimal use of the sce-ua global optimization method for calibrating watershed models, *Journal of Hydrology*, *158*(3-4), 265–284.

Duan, Q., N. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using bayesian model averaging, *Advances in Water Resources*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014.

Dunne, S., and D. Entekhabi (2005), An ensemble-based reanalysis approach to land data assimilation, *Water Resources Research*, *41*(2), 1–18, cited By (since 1996):58 Export Date: 19 November 2014.

Ebert, E. E., and J. L. McBride (2000), Verification of precipitation in weather systems: Determination of systematic errors, *Journal of Hydrology*, *239*(1-4), 179–202, cited By (since 1996):180 Export Date: 18 November 2014.

Einstein, A., B. Podolsky, and N. Rosen (1935), Can quantum-mechanical description of physical reality be considered complete?, *Physical Review*, *47*(10), 777–780, cited By (since 1996):5087 Export Date: 15 October 2014.

Elliott, J. (1997), Development of an improved real - time flood forecasting model, *Tech. rep.*, Cooperative Research Centre for Catchment Hydrology.

Emergency Management Australia (2005), *What to Do Before, During and After a Flood*, Emergency Management Australia.

Entekhabi, D., E. Njoku, P. O'Neill, K. Kellogg, W. Crow, W. Edelstein, J. Entin, S. Goodman, T. Jackson, J. Johnson, J. Kimball, J. Piepmeier, R. Koster, N. Martin, K. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. Shi, M. Spencer, S. Thurman, L. Tsang, and J. Van Zyl (2010), The soil moisture active passive (smap) mission, *Proceedings of the IEEE*, *98*(5), 704–716, doi:10.1109/JPROC.2010.2043918.

Galantowicz, J. F., D. Entekhabi, and E. G. Njoku (1999), Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed l-band radiobrightness, *IEEE Transactions on Geoscience and Remote Sensing*, *37*(4), 1860–1870, cited By (since 1996):84 Export Date: 19 November 2014.

Gan, T. Y., and G. F. Biftu (1996), Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, *Water Resources Research*, *32*(12), 3513–3524.

Gebremichael, M., and F. Y. Testik (2010), *Microphysics, Measurement, and Analyses of Rainfall*, pp. 1–6, American Geophysical Union, doi:10.1029/2010GM001025.

Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statist. Sci.*, *7*(4), 457–472, doi:10.1214/ss/1177011136.

Gentle, S. N. A., Neil Kierce (2001), Economic costs of natural disasters in australia.

Grayson, R. B., A. W. Western, J. P. Walker, D. D. Kandel, J. F. Costelloe, and D. J. Wilson (2006), Controls on patterns of soil moisture in arid and semi-arid systems, *Dryland Ecohydrology*, pp. 109–127.

Habib, E., G. Lee, D. Kim, and G. J. Ciach (2013), *Ground-Based Direct Measurement*, pp. 61–77, American Geophysical Union, doi:10.1029/2010GM000953.

Hajkowicz, S., and K. Collins (2007), A review of multiple criteria analysis for water resource planning and management, *Water Resources Management*, *21*(9), 1553–1566, doi:10.1007/s11269-006-9112-5.

Hapuarachchi, H., Q. Wang, and T. Pagano (2011), A review of advances in flash flood forecasting, *Hydrological Processes*, *25*(18), 2771–2784.

Heathman, G. C., P. J. Starks, L. R. Ahuja, and T. J. Jackson (2003), Assimilation of surface soil moisture to estimate profile soil water content, *Journal of Hydrology*, *279*(1-4), 1–17, cited By (since 1996):61 Export Date: 19 November 2014.

Herron, N., R. Davis, and R. Jones (2002), The effects of large-scale afforestation and climate change on water allocation in the Macquarie River catchment, NSW, Australia, *Journal of Environmental Management*, *65*(4), 369–381, doi:10.1016/S0301-4797(02)90562-1.

Hino, M. (1986), Improvements in the inverse estimation method of effective rainfall from runoff, *Journal of Hydrology*, *83*(1-2), 137–147.

Hou, A., R. Kakar, S. Neeck, A. Azarbarzin, C. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi (2014), The global precipitation measurement mission, *Bulletin of the American Meteorological Society*, *95*(5), 701–722, doi:10.1175/BAMS-D-13-00164.1.

Houser, P. R., H. V. Gupta, W. J. Shuttleworth, and J. S. Famiglietti (2001), Multiobjective calibration and sensitivity of a distributed land surface water and energy balance model, *Journal of Geophysical Research: Atmospheres*, *106*(D24), 33,421–33,433, cited By (since 1996):18 Export Date: 20 November 2014.

Kavetski, D., G. Kuczera, and S. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. theory, *Water Resources Research*, *42*(3), doi:10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. application, *Water Resour. Res.*, *42*(3), doi:10.1029/2005WR004376.

Kirchner, J. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resources Research*, *45*(2), doi:10.1029/2008WR006912.

Kucera, P. A., E. E. Ebert, F. J. Turk, V. Levizzani, D. Kirschbaum, F. J. Tapiador, A. Loew, and M. Borsche (2013), Precipitation from space: Advancing earth system science, *Bulletin of the American Meteorological Society*, *94*(3), 365–375.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *Journal of Hydrology*, *331*(1-2), 161–177, doi:10.1016/j.jhydrol.2006.05.010.

Kumar, P., and E. Foufoula-Georgiou (1997), Wavelet analysis for geophysical applications, *Reviews of Geophysics*, *35*(4), 385–412.

Labat, D. (2005), Recent advances in wavelet analyses: Part 1. a review of concepts, *Journal of Hydrology*, *314*(1-4), 275–288.

Laloy, E., and J. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing, *Water Resources Research*, *48*(1), doi:10.1029/2011WR010608.

Li, Y., D. Ryu, A. W. Western, and Q. J. Wang (2013), Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags, *Water Resources Research*, *49*(4), 1887–1900, cited By (since 1996):3 Export Date: 16 October 2014.

Li, Y., S. Grimaldi, J. Walker, and V. Pauwels (2016), Application of remote sensing data to constrain operational rainfall-driven flood forecasting: A review, *Remote Sensing*, *8*(6), doi:10.3390/rs8060456.

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

López, P., E. Sutanudjaja, J. Schellekens, G. Sterk, and M. Bierkens (2017), Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products, *Hydrology and Earth System Sciences*, *21*(6), 3125–3144, doi:10.5194/hess-21-3125-2017, cited By 1.

Mallat, S. (1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), doi:10.1109/34.192463.

Mallat, S. (2009), *A Wavelet Tour of Signal Processing*, Elsevier.

Martina, M., E. Todini, and A. Libralon (2006), A bayesian decision approach to rainfall thresholds based flood warning, *Hydrology and Earth System Sciences*, *10*(3), 413–426.

Massari, C., L. Brocca, T. Moramarco, Y. Tramblay, and J.-F. D. Lescot (2014), Potential of soil moisture observations in flood modelling: Estimating initial conditions and correcting rainfall, *Advances in Water Resources*, *74*, 44 – 53, doi:http://dx.doi.org/10.1016/j.advwatres.2014.08.004.

McInerney, D., M. Thyer, D. Kavetski, J. Lerat, and G. Kuczera (2017), Improving probabilistic prediction of daily streamflow by identifying pareto optimal approaches for modeling heteroscedastic residual errors, *Water Resources Research*, *53*(3), 2199–2239, doi:10.1002/2016WR019168.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods (2011), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *Journal of Hydrology*, *400*(1-2), 83–94.

McMillan, H., J. Seibert, A. Petersen-Overleir, M. Lang, P. White, T. Snelder, K. Rutherford, T. Krueger, R. Mason, and J. Kiang (2017), How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications, *Water Resources Research*, doi:10.1002/2016WR020328.

Moore, R. (2007), The PDM rainfall-runoff model, *Hydrology and Earth System Sciences*, *11*(1), 483–499.

Moradkhani, H., and S. Sorooshian (2008), *General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis*, *Water Science and Technology Library*, vol. 63, pp. 1–24, moradkhani, Hamid Sorooshian, Soroosh Summer School on Hydrologic Modelling and Water Cycle 2005 L Aquila, ITALY.

Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resources Research*, *41*(5), 1–17, doi:10.1029/2004WR003604.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *TRANSACTIONS OF THE ASABE*, *50*(3), 885–900.

Nalley, D., J. Adamowski, and B. Khalil (2012), Using discrete wavelet transforms to analyze trends in streamflow and precipitation in quebec and ontario (1954-2008), *Journal of Hydrology*, *475*, 204–228, doi:10.1016/j.jhydrol.2012.09.049.

NWSRFS (2002), *Conceptualization of the Sacramento Soil Moisture Accounting Model*, National Weather Service River Forecast System.

Oudin, L., C. Michel, and F. Anctil (2005), Which potential evapotranspiration input for a lumped rainfall-runoff model? part 1 - can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs?, *Journal of Hydrology*, *303*(1-4), 275–289, doi:10.1016/j.jhydrol.2004.08.025, cited By 62.

Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, and C. Michel (2006), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *Journal of Hydrology*, *320*(1-2), 62–83, doi:10.1016/j.jhydrol.2005.07.016.

Pagano, P. W. Q., T.C. Hapuarachchi (2009), Continuous soil moisture accounting and routing modelling to support short lead- time streamflow forecasting., *Tech. rep.*, CSIRO: Water for a Healthy Country National Research Flagship.

Paniconi, C., M. Marrocu, M. Putti, and M. Verbunt (2003), Newtonian nudging for a richards equation-based distributed hydrological model, *Advances in Water Resources*, *26*(2), 161–178, cited By (since 1996):35 Export Date: 19 November 2014.

Pappenberger, F., and K. Beven (2006), Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resources Research*, *42*(5), doi:10.1029/2005WR004820.

Pauwels, V., and G. De Lannoy (2011), Multivariate calibration of a water and energy balance model in the spectral domain, *Water Resources Research*, *47*(7), doi:10.1029/2010WR010292.

Pauwels, V. R. N. (2008), A multistart weight-adaptive recursive parameter estimation method, *Water Resources Research*, *44*(4), doi:10.1029/2007WR005866.

Pauwels, V. R. N., R. Hoeben, N. E. C. Verhoest, and F. P. De Troch (2001), The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation, *Journal of Hydrology*, *251*(1-2), 88–102, cited By (since 1996):73 Export Date: 19 November 2014.

Peck, E. (1976), Catchment modeling and initial parameter estimation for the national weather service river forecast system, *Noaa tech memo. nws hydro-31*, National Weather Service, Silver Spring, Md.

Pellarin, T., A. Ali, F. Chopin, I. Jobard, and J.-C. Bergès (2008), Using spaceborne surface soil moisture to constrain satellite precipitation estimates over West Africa, *Geophysical Research Letters*, *35*(2), doi:10.1029/2007GL032243.

Pellarin, T., T. Tran, J.-M. Cohard, S. Galle, J.-P. Laurent, P. De Rosnay, and T. Vischel (2009), Soil moisture mapping over west africa with a 30-min temporal resolution using amsr-e observations and a satellite-based rainfall product, *Hydrology and Earth System Sciences*, *13*(10), 1887–1896, cited By 19.

Pellarin, T., S. Louvet, C. Gruhier, G. Quantin, and C. Legout (2013), A simple and effective method for correcting soil moisture and precipitation estimates using AMSR-E measurements, *Remote Sensing of Environment*, *136*, 28–36.

Petersen, A. (1963), The philosophy of niels bohr, *The Bulletin of the Atomic Scientists*, *19*(7).

Polikar, R. (1999), *The story of wavelets*, 192-197 pp.

Queensland Floods Commission of Inquiry (2012), Queensland Floods Commission of Inquiry - Final Report, p. 32.

Queensland Government (2017), Water monitoring information portal.

Raupach, M., P. Briggs, V. Haverd, E. King, M. Paget, and C. Trudinger (2012), Australian water availability project.

Rawls, W. J., D. Brakensiek, and K. Saxton (1982), Estimation of soil water properties, *Trans. Asae*, *25*(5), 1316 – 1320.

Reichle, R., D. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble kalman filter, *Monthly Weather Review*, *130*(1), 103–114.

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, *46*(5), doi:10.1029/2009WR008328.

Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. Franks (2011), Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resources Research*, *47*(11), doi:10.1029/2011WR010643.

Robertson, D., D. Shrestha, and Q. Wang (2013), Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, *Hydrology and Earth System Sciences*, *17*(9), doi:10.5194/hess-17-3587-2013.

Rusjan, S., and M. Mikoš̃ (2015), A catchment as a simple dynamical system: Characterization by the streamflow component approach, *Journal of Hydrology*, *527*, doi: 10.1016/j.jhydrol.2015.05.050.

Sadegh, M., and J. Vrugt (2014), Approximate Bayesian Computation using Markov Chain Monte Carlo simulation: DREAM(ABC), *Water Resources Research*, *50*, doi: 10.1002/2014WR015386.

Samain, B., and V. Pauwels (2013), Impact of potential and (scintillometer-based) actual evapotranspiration estimates on the performance of a lumped rainfall-runoff model, *Hydrology and Earth System Sciences*, *17*(11), 4525–4540, doi:10.5194/hess-17-4525-2013.

Schaake, J., S. Cong, and Q. Duan (2006), The us mopex data set, 307, pp. 9–28.

Schaefli, B., and E. Zehe (2009), Hydrological model performance and parameter estimation in the wavelet-domain, *Hydrology and Earth System Sciences*, *13*(10), 1921–1936.

Seed, A. W., C. E. Pierce, and K. Norman (2013), Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme, *Water Resources Research*, *49*(10), 6624–6641, cited By (since 1996):1 Export Date: 8 October 2014.

Sene, K. (2008), *Flood Warning, Forecasting and Emergency Response*, 1–303 pp., Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-540-77853-0.

Seo, D.-J., A. Seed, and G. Delrieu (2013), *Radar and Multisensor Rainfall Estimation for Hydrologic Applications*, pp. 79–104, American Geophysical Union, doi:10.1029/2010GM000952.

Seuffert, G., H. Wilker, P. Viterbo, M. Drusch, and J. F. Mahfouf (2004), The usage of screen-level parameters and microwave brightness temperature for soil moisture analysis, *Journal of Hydrometeorology*, *5*(3), 516–531, cited By (since 1996):44 Export Date: 19 November 2014.

Shrestha, D., D. Robertson, J. Bennett, and Q. Wang (2015), Improving precipitation forecasts by generating ensembles through postprocessing, *Monthly Weather Review*, *143*(9), 3642–3663, doi:10.1175/MWR-D-14-00329.1.

Sivakumar, B. (2001), Rainfall dynamics at different temporal scales: A chaotic perspective, *Hydrology and Earth System Sciences*, *5*(4), 645–651.

Tang, Y., P. Reed, and T. Wagener (2006), How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrology and Earth System Sciences*, *10*(2), 289–307.

Tebbs, E., F. Gerard, A. Petrie, and E. De Witte (2016), *Emerging and Potential Future Applications of Satellite-Based Soil Moisture Products*, 379-400 pp., Elsevier, doi:10.1016/B978-0-12-803388-3.00019-X.

Testik, F. Y., and M. Gebremichael (Eds.) (2010), *Rainfall: State of the Science*, *Geophysical Monograph Series*, vol. 191, pp. 1–287, doi:10.1029/gm191.

Teuling, A., I. Lehner, J. Kirchner, and S. Seneviratne (2010), Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment, *Water Resources Research*, *46*(10), doi:10.1029/2009WR008777.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resources Research*, *37*(10), 2521–2535.

Thyer, M., G. Kuczera, and B. C. Bates (1999), Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms, *Water Resources Research*, *35*(3), 767–773.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resources Research*, *45*(12), doi:10.1029/2008WR006825.

Vrugt, J. (2016), Markov chain monte carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, *Environmental Modelling and Software*, *75*, 273–316, doi:10.1016/j.envsoft.2015.08.013.

Vrugt, J., and B. Robinson (2007), Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and bayesian model averaging, *Water Resources Research*, *43*(1), doi:10.1029/2005WR004838.

Vrugt, J., and C. Ter Braak (2011), Dream(d): An adaptive markov chain monte carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems, *Hydrology and Earth System Sciences*, *15*(12), 3701–3713, doi:10.5194/hess-15-3701-2011.

Vrugt, J., C. Diks, H. Gupta, W. Bouten, and J. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resources Research*, *41*(1), 1–17, doi: 10.1029/2004WR003059.

Vrugt, J., C. Ter Braak, M. Clark, J. Hyman, and B. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation, *Water Resour. Res.*, *44*, W00B09.

Vrugt, J., C. ter Braak, H. Gupta, and B. Robinson (2009a), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026, doi: 10.1007/s00477-008-0274-y.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, *39*(8), doi:10.1029/2002WR001642.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009b), Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*(3), 273–290.

World Meteorological Organization (2008), Guide to meteorological instruments and methods of observation.

World Meteorological Organization (2011), *Manual on flood forecasting and warning*, WMO.

Wright, A., J. P. Walker, D. E. Robertson, and V. R. N. Pauwels (2017a), A comparison of the discrete cosine and wavelet transforms for hydrologic model input data reduction, *Hydrology and Earth System Sciences*, *21*(7), 3827–3838, doi: 10.5194/hess-21-3827-2017.

Wright, A. J., J. P. Walker, and V. R. N. Pauwels (2017b), Estimating rainfall time series and model parameter distributions using model data reduction and inversion techniques, *Water Resources Research*, doi:10.1002/2017WR020442.

Yee, M., J. Walker, C. Rüdiger, R. Parinussa, T. Koike, and Y. Kerr (2017), A comparison of smos and amsr2 soil moisture using representative sites of the oznet monitoring network, *Remote Sensing of Environment*, *195*, 297–312, doi:10.1016/j.rse.2017.04.019, cited By 0.

Young, P. (2013), Hypothetico-inductive data-based mechanistic modeling of hydrological systems, *Water Resources Research*, *49*(2), 915–935, doi:10.1002/wrcr.20068.