MODELLING UNCERTAINTY AND RELIABILITY FOR WATER RESOURCE ASSESSMENT IN SOUTH AFRICA

Report to the
Water Research Commission

by

J Glenday[1,2], J Tanner[1], A Rebelo[3], P Holden[4], P Metho[5], D Gwapedza[1], S Gokool[6], & F Jumbi[2]

[1] Rhodes University, Institute for Water Research
[2] South African Environment Observation Network
[3] Agricultural Research Council
[4] University of Cape Town, African Climate & Development Initiative
[5] Zutari (Pty) Ltd
[6] University of KwaZulu-Natal, Centre for Water Resources Research

WRC Report No. 3149/1/24
ISBN 978-0-6392-0627-1

May 2024

Obtainable from
Water Research Commission
Bloukrans Building, Lynnwood Bridge Office Park
4 Daventry Street
Lynnwood Manor
PRETORIA

orders@wrc.org.za or download from www.wrc.org.za

This is the final report of WRC project no. C2022/2023-00967.

EXECUTIVE SUMMARY

BACKGROUND, AIMS, & OBJECTIVES

Hydrological modelling has become central to water resources management and catchment management in South Africa. With ever-growing pressure on the nation's water supply systems, basing decisions on reliable estimates of surface and groundwater resource availability is critical. At the same time South Africa's meteorological and hydrological monitoring infrastructure has declined severely in recent decades (Bailey and Pitman, 2015). These data are critical for the validation of models, hence declining monitoring infrastructure is increasing the uncertainty of model outputs. The water sector leans on models to fill in data gaps and to forecast the future. Given the risks associated with making decisions based on model outputs, it is critical to quantify the uncertainty within these outputs. As reliance on modelling grows, so too should attention to model uncertainty: quantifying it, finding practical ways to reduce it, and accounting for it in decision making.

There are multiple sources of uncertainty in catchment modelling: the hydrometric data used to drive and calibrate models; the parameters that describe catchment properties; and the model structure (mathematical representation of hydrological processes). Three WRC-supported research projects have quantified uncertainties in hydrological modelling for case studies typical of the South African context in terms of data availability and modelling tools. These studies demonstrated that uncertainty levels are case specific, but can often be alarmingly high. Projects K5-1838: *Incorporating uncertainty in water resources simulation and assessment tools in South Africa* (Hughes *et al.*, 2011) and K5-2056: *Implementing uncertainty analysis in water resources assessment and planning* (Hughes, Mohobane and Mallory, 2015), demonstrated an approach for quantifying and reducing uncertainty due to parameters and input data. Project K5-2927, *Critical catchment hydrological model inter-comparison and model use guidance development*, showed that models with differing structures can have comparable streamflow accuracy in calibration at the same time as having differing internal process representation, and so go on to predict very different outcomes to one another when they are applied to scenarios of change (Glenday *et al.*, 2022).

Despite this work, and a growing body of international research on uncertainty in hydrological modelling, it has remained common practice in the South African water sector to not attempt to quantify uncertainty in hydrological modelling outputs. As such, the current project focuses on participatory research into model uncertainty across the modelling sector, with a particular focus on modeller decisions and structural uncertainty. This was complemented by activities to foster discussion across the modelling community on how to practically address these issues.

Aims:

1. To improve shared understanding of model structural uncertainty, and the role of modeller decision-making, and their potential scale of impact, across the hydrological modelling community
2. To empower the community to identify practical strategies for assessing, communicating, and ideally reducing modelling uncertainty

Objectives:

1. Design and host an open **'model-a-thon' activity** in which participants from across the water sector model the same catchment area, given the same input and calibration data, and apply the model to the same alternative scenario.

2. Assess and discuss the diversity of modelling approaches, structures, and output predictions in the ‘model-a-thon’ activity as means of collective scoping of the issue of structural uncertainty.
3. Host **synthesis engagements** to discuss the implications of the findings around modelling uncertainty for water resources studies and initiate visioning of how these studies can be done in a way that accounts for and, where feasible, aims to reduce uncertainty.
4. Synthesise the engagement outcomes into a **policy brief**.
5. Supplement and publicise the **‘wiki’ website** about modelling tool capabilities (https://hydromodel-sa-wiki.saeon.ac.za/), initiated during project K5-2927, to increase awareness of modelling options and accessibility of tools.
6. Promote use of the **Stack Exchange** online question-and-answer platform (Stack Exchange Earth Sciences site: https://earthscience.stackexchange.com/) to facilitate information sharing across the modelling sector.

MODEL-A-THON

The goal of the ‘model-a-thon’ was to explore how differently individual modellers are likely to set up and calibrate a catchment model when given the same brief and input data, and to understand how much these differences influence the modelled predictions and what this means for uncertainty. The activity was open to anyone interested, with the additional goals of connecting the modelling community and fostering discussion around uncertainty. An effort was made to attract participants with a variety of experience levels and who use different modelling tools, with the aim of having multiple users of some of the commonly used tools participate. The activity was publicised via the South African Hydrological Society (SAHS) and launched with a workshop at the SAHS Conference 2022.

Each participant built and ran a model of the same case study catchment for both a baseline case and a scenario of change, using the modelling tool of their choice (*APPENDIX A.2 – Model-a-thon instructions document provided to participants*). Participants were provided with observed streamflow data with which to calibrate their baseline model if they chose to do so. The variability in the change predictions across the models that achieved sufficient performance against observed streamflow gives an indication of model structural uncertainty. The case study used was the Two Steams experimental catchment on the Mvoti River in KwaZulu-Natal, selected based on data availability and land cover. The change scenario was the removal of mature wattle plantation, covering ~60% of the catchment, and replacement with sugar cane.

The activity succeeded in obtaining a sufficient number and diversity of submissions for useful analyses. There were 43 sign-ups to participate. While there were only 18 finalised models submitted, these were built using five different modelling tools (WRSM-Pitman, SPATSIM-Pitman, ACRU, SWAT, and MIKE-SHE, with some variation in versions) and there was more than one entry for each tool. The modellers submitting had a range of experience levels from beginner to experienced, although the majority (11/18) rated themselves as intermediate. There was a huge range in the spatial discretisation across the models, from representing the catchment as a single lumped unit, to having 7,382 units (10 m grid cells) for which hydrological processes were calculated. The majority manually calibrated their models’ parameter values, while three participants also made use of automated calibration tools. Four did not attempt to calibrate, submitting models with their a-priori parameter value selections. These uncalibrated models had the lowest performance against the observed streamflow, even when the modeller was an expert, highlighting the danger in assuming default values will provide adequate outputs.

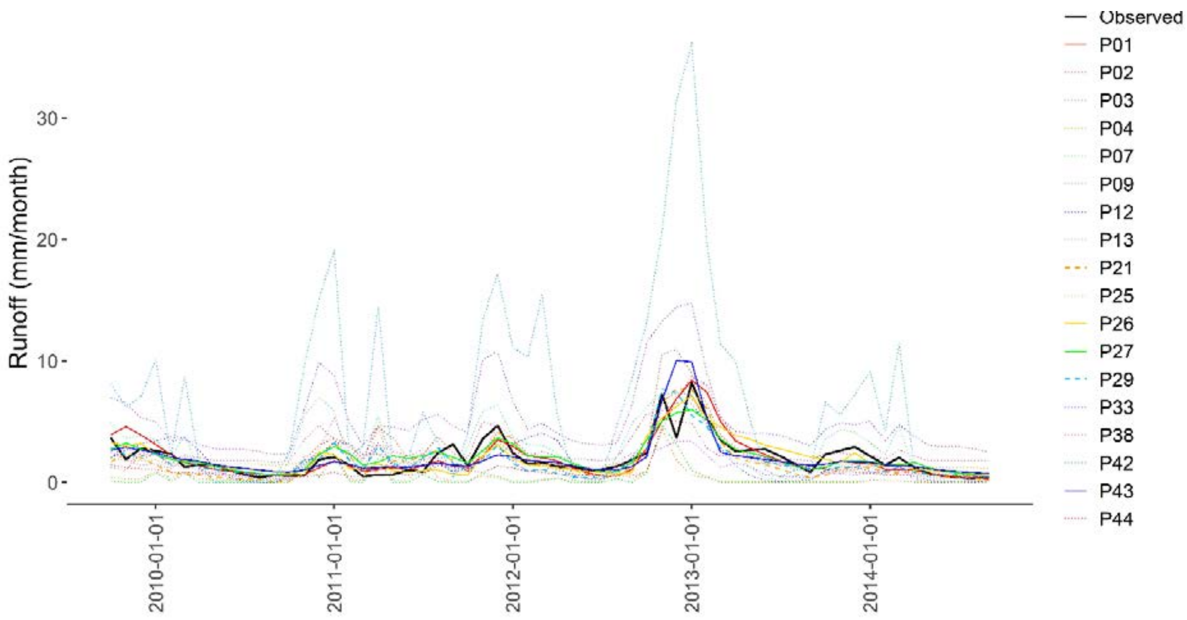


Figure ES-1 Monthly observed and modelled hydrographs for all model submissions for the calibration period (2009/10/01 to 2014/09/30)
(the four models meeting all five performance criteria are shown with solid lines)

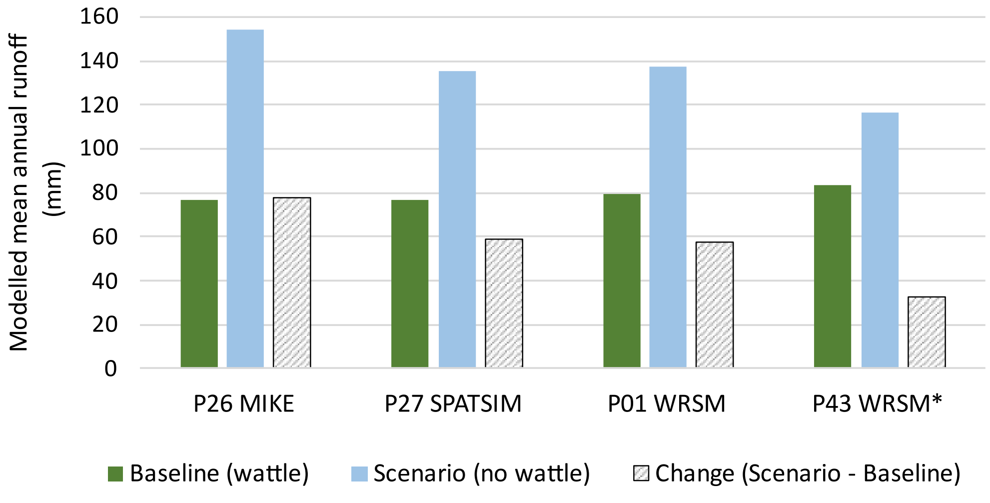


Figure ES-2 Predicted mean annual runoff for both land cover cases for the assessment period (2003-10 to 2014-09) from models meeting streamflow calibration criteria
(P43 was found to model runoff generation mechanisms inaccurately)*

The model-a-thon served as a powerful demonstration of uncertainty due to model structure and modeller decisions. Only four models met all of the calibration criteria (goodness-of-fit vs observed streamflow) to be accepted for further use in scenario impact prediction (Figure ES-1, Figure ES-2). These four were built using different tools (SPATSIM, WRSM, MIKE-SHE), had different structures (lumped, modular/semi-distributed, gridded/distributed), and were calibrated either manually or using automated tools. While they all produced streamflow outputs that would be considered satisfactory according to generally accepted metrics, when they were used to model the land cover change scenario, they predicted very different streamflow impacts. Modelled impacts of removing the wattle plantation ranged from a 39% to a 101% increase in mean annual streamflow (Figure ES-1). This highlighted that even when there is observational streamflow data to calibrate and verify models, when they are applied to predict the impacts of different scenarios there can easily be a very high degree of uncertainty in the output.

The study also demonstrated that modelling uncertainty can be notably reduced using additional information about catchment processes, beyond observed streamflow data. This can identify models that may be producing reasonable streamflow output for the calibration period, but are doing so using an unrealistic representation of internal flows and storages. In this case, one of the models producing acceptable streamflow results was predicting that more runoff would be generated as surface flow than subsurface flow in total. Isotope tracer research has shown that this is not the case in Two Streams (Everson *et al.*, 2014; Watson, 2015; Ngubo, 2019). When this model was excluded from the accepted set, uncertainty in the scenario impact prediction dropped notably: the remaining models predicted mean annual streamflow increases of 72% to 101% with wattle removal (Figure ES-1). This information roughly halved the uncertainty in the change prediction.

Looking at the opposite case, when there is less data about a catchment, for example if the catchment is ungauged, uncertainties can be far higher. In the model-a-thon, if there had been no observed streamflow or process information to use as a basis to accept or reject models, all 18 models submitted could have been considered potentially feasible. If this were the case, the prediction of change in mean annual streamflow with wattle removal would have ranged from -75% (a decrease in mean flow) to +291% (almost a threefold increase in mean flow)!

POLICY BRIEF

Research findings on hydrological modelling uncertainty were synthesised with the help of water sector professionals to develop a policy brief that accounts for the practical hurdles faced in this field. It is intended that this brief be used to promote awareness and gather support for the activities needed to mainstream uncertainty analyses in hydrological modelling in the sector. Two online discussion sessions were held to better understand how different stakeholders view and experience the issue of hydrological modelling uncertainty and how they would recommend it be handled going forward. A total of 37 people were invited, including: consultants, researchers, and academics who specialise in hydrological modelling; the Water Research Commission (WRC); and employees of the Department of Water and Sanitation (DWS) and Catchment Management Agencies (CMAs) who routinely use catchment modelling output for decision-making, commission modelling studies, and/or run models in-house in operational contexts. Although only twelve attended the sessions in the end, the invitations elicited widespread expression of interest in the topic. Those attending spanned consulting, academia, and a range of roles across DWS, from supply planning to flood and dam safety, and from chief engineers to scientific managers and modellers to early career candidate engineers. This made for fruitful discussion. The resulting policy brief was distributed via SAHS and to those invited to the discussions in December 2023.

The full policy brief has been included in this report in *APPENDIX B.3 – Policy brief on uncertainty in hydrological modelling*. Recommendations brought up in the discussion sessions were distilled and grouped into overarching or programmatic recommendations and technical recommendations on quantifying and reducing uncertainty.

Overarching or programmatic level recommendations:

1. **Build capacity:** Promote theoretical and operational understanding of analyses and decision-making under uncertainty, through university curricula and certified short-courses.
2. **Standardise efforts:** Develop context-appropriate standard practices for estimating, reducing and communicating hydrological uncertainty for decision-making, e.g. checklists of assessments depending on risk levels, data availability, and scale.
3. **Invest in technology:** Include facilities in models that automate sensitivity analyses, uncertainty analyses, calibration, and water balance assessments across a range of storages and fluxes. These functions can be programmed into existing modelling tools or internationally developed tools that meet these needs can be harnessed.
4. **Support data collection & open access data:** Long-term, continuous, spatially-distributed data on rainfall, evaporative demand, and streamflow need to be made accessible. Data on other storages and fluxes (groundwater, soil moisture, evapotranspiration) can further resolve uncertainties. Calibrate remote sensing products locally with field data before relying on these as alternative sources. Government-funded hydrological modelling efforts, with associated input and output databases, should be easily available.
5. **Make budget available:** Ensure researchers and consultants have enough time and budget to quantify and reduce uncertainty in hydrological modelling, and account for it in further analyses and decision-making. Resources are also needed for the process of establishing standardised approaches and reviewing these over time.

Technical recommendations, approaches proposed for quantifying and reducing hydrological modelling uncertainty (noting that further engagement is needed to recommend specific methods for different case types – decision making settings, risks, scales, data availability, etc.):

1. **Identify and document sources of uncertainty as well as assumptions and subjective decisions:** This could be a baseline requirement in all modelling projects, and consider input data, calibration data, parameters, and model structure.
2. **Conduct sensitivity analyses:** This determines which uncertain datasets and parameters will have the greatest impact on the model outputs.
3. **Apply more than one modelling tool and/or model structure:** This would be particularly advised when informing high risk and high cost decisions.
4. **Validation** – **not just streamflow:** Apply a suite of model ‘reality-checks’ targeting a variety of hydrological processes during the calibration and acceptance of models and parameter ranges (e.g. evapotranspiration, runoff ratios, surface vs subsurface flow dominance, groundwater recharge).
5. **Propagate uncertainty:** At a minimum, identify a reasonable high flow and low flow dataset to carry through to further analyses, such as stochastic yield modelling.

It is critical to note that while these recommendations require funding to achieve, for which there is heavy competition, if implemented, these steps can result in considerable savings from avoiding the costs associated with inappropriate water management and emergency response measures that may otherwise ensue. Indeed, the financial sector already operates with clear measures of uncertainty and makes effort to quantify uncertainties for this reason. In addition, technology is rapidly making many parts of this process easier and easier. Significant progress can be made by building the capacity in the water sector to take advantage of these developments.

ONLINE RESOURCES FOR THE HYDROLOGICAL MODELLING COMMUNITY

Project activities included working on, and promoting, two online resources that facilitate information exchange and capacity building across the hydrological modelling community: an editable 'wiki' website on hydrological modelling tools (https://hydromodel-sa-wiki.saeon.ac.za/) and the pre-existing online 'question-answer' (Q&A) platform 'Stack Exchange', specifically the Earth Science subsite (https://earthscience.stackexchange.com/). The "HydroModel SA wiki" site was initially designed in WRC project K5-2927 on modelling tool intercomparison (Glenday *et al.*, 2022). A focus of the site's content is building awareness of different modelling strategies and the options available across different modelling tools, which can also help new modellers entering the field. While the site architecture and basic content were created in a previous project, the current project engaged in further content development and publicising the site. The current project also promoted use of Stack Exchange Q&A platform by the hydrological modelling community to make it more efficient for those with modelling expertise to assist newer users and easier for new users to find the information they are looking for. An introductory guide to using Stack Exchange (*APPENDIX C.1 – Stack Exchange guidelines document*) was produced and distributed via SAHS, and the project team initiated posting and answering relevant questions about commonly used models on the platform.

KNOWLEDGE DISSEMINATION & CAPACITY BUILDING

Several project activities served to disseminate knowledge around hydrological modelling and uncertainty within the water sector: the model-a-thon and related discussions, the discussion sessions which informed the policy brief, and the online community resources promoted in the project. The South African Hydrological Society (SAHS) email list and communication channels were used to share information about project activities, and to share the policy brief across the hydrology community. In addition, academic researchers and lecturers, consultants, and DWS employees engaged in modelling work were contacted directly to encourage participation in the model-a-thon and/or the policy brief discussion sessions. Additional knowledge dissemination took place through presentations at the SAHS conference in 2022 and more are planned for the 2024 conference. A journal article on the model-a-thon activity has been drafted for submission for publishing in 2024. Those who engaged in project activities, project team included, gained a deepened understanding of modelling approaches and uncertainty assessment, valuable skills for the South African water sector.

In addition, building capacity within the project team can be seen as a notable contribution because the team members were all early-career hydrologists from a range of institutions. The team included three postdoctoral researchers and four early career researchers. Another team member started the project as an MSc student and by the second year had completed her degree and was a junior hydrologist at a consulting firm. The team members are connected to various universities (Rhodes University, UCT, UKZN, UWC, Stellenbosch University), research institutions (SAEON, ARC), and consulting firms. Their involvement and learning in this project will positively impact their respective institutions through sharing with colleagues and students, and incorporation of new knowledge into their further work.

CONCLUSIONS & RECOMMENDATIONS

Previous research, and the current project's model-a-thon study, make a clear case for mainstreaming uncertainty assessment in hydrological modelling studies in the South African water sector. The policy brief generated through this project put forward a number of recommendations for how this could be achieved. The engagements in this project were initial steps and further collaborative work needs to be done across water resource managers, consultants, and researchers to establish accepted, standardised practices for quantifying and considering modelling uncertainty that are tailored to different 'case types.' Case type refers to the types of decisions being made, or assessments being done, with the modelling outputs (e.g. national prioritisation, regional supply planning, dam operations, supply allocation, individual licensing, flood risk management, catchment land cover management), their scales, associated risks, and the types, amounts, and quality of data readily available for the area. This will benefit from a review of how to make best use of different types of data and information to inform, calibrate, and validate modelling, from gauge data to remote sensing to hydrological processes studies that have been conducted across different settings. ***While it will be a collaborative effort, the process of establishing accepted uncertainty assessment practices will require a dedicated programme with at least one institutional champion to facilitate and ensure progress is made, with the WRC and DWS being natural candidates.***

Approaches for assessing modelling uncertainty from two different sources – data, parameters, and model structure – need to be developed for future iterations of the **national water resources studies**, as a specific 'case type', for the output of these studies to be used wisely and to make progress towards uncertainty reduction. This will require use of, or development of, software tools that make these assessments practical, which has presented a barrier to parameter and data uncertainty quantification in the studies to date (Glenday *et al.*, 2022). The current project, and project K5-2927 comparing modelling tools, clearly highlight that model output uncertainty due to uncertainty about the model structure is also often large and typically ignored. However, applying multiple modelling structures nationwide may not be out of reach, given that the WRSM, SPATSIM, ACRU, and SWAT modelling tools have been, or are being applied at this scale through other WRC projects, e.g. (Hughes, 2005; Schulze, 2007; Bailey and Pitman, 2015; Le Roux *et al.*, 2023; Schutte *et al.*, 2023). If data, assumptions, and calibration approaches were discussed and harmonised across the modelling teams that already have some national scale set-ups, significant progress could be made in future assessments that consider structural uncertainty. As with mainstreaming uncertainty analyses approaches, ***the process of developing and conducting national water resources studies also requires a programmatic home to progress and achieve its potential value for resource management.***

ACKNOWLEDGEMENTS

The funding of the project by the Water Research Commission and the contribution of the team members is acknowledged gratefully.

The following members of the Reference Group are acknowledged for their valuable contributions and guidance:

PROJECT REFERENCE GROUP MEMBERS

Mr W Nomquphu (Chair)	Water Research Commission (WRC)
Ms Penny Jaca (Coordinator)	Water Research Commission
Prof. J Ndiritu	University of the Witwatersrand
Prof. S Lorentz	SRK Consulting / University of KwaZulu-Natal (UKZN)
Dr E Kapangaziwiri	Council for Scientific and Industrial Research (CSIR)
Dr J Le Roux	University of the Free State (UFS)
Dr A Watson	Stellenbosch University
Dr T Sawunyama	Inkomati-Usuthu Catchment Management Agency (IU-CMA)
Dr J Hallowes	EkoSource Insights
Prof. M Van der Laan	Agricultural Research Council (ARC)
Dr B Mwaka	Department of Water and Sanitation (DWS)
Dr Piotr Wolski	University of Cape Town (UCT)

Those who participated in the model-a-thon activity:

Mr Bennie Haasbroek
Mr Christiaan Schutte
Dr David Gwapedza
Mr Edward Smit
Dr Elias Moussoulis
Ms Harriette Adhiambo-Okal
Prof. Johan van Tol
Dr Julia Glenday
Mr Mark Horan
Prof. Michael Van der Laan
Ms Penisoh Metho
Mr Peter Wasswa
Dr Shaeden Gokool
Mr Sinetemba Xoxo *(student prize winner!)*
Dr Stefanie Schutte

Those who participated in the facilitated discussions on hydrological modelling uncertainty to inform the policy brief

Ms Louise Dobinson
Mr Stephen Mallory
Prof. Jay le Roux
Prof. John Ndiritu
Mr Joshua Rasifudi
Ms Jenny Pashkin
Ms Celiwe Ntuli
Mr Ernest Oakes
Mr Mehari Frezghi
Ms Siphindile Shoba
Mr Mongezi Gxamza

We would also like to thank interested parties who were not able to attend the policy brief discussion sessions, but who sent input via email: Mr Bennie Haasbroek and Mr James Cullis.

The team would like to thank staff members at Rhodes University Institute of Water Research who assisted with support, administration, and technical logistics, particularly Juanita McLean and David Forsyth.

In addition, this project was made possible through support and co-funding from the project team and their respective institutions: Rhodes University, South African Environmental Observation Network, Agricultural Research Council, University of Cape Town, University of Kwa-Zulu Natal. The work presented here represents time spent above and beyond that covered by the Water Research Commission.

Last, but not least, the project team would also like to give gratitude for the invaluable support of family and friends!

This page was intentionally left blank

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

ACRU	Agricultural Catchment Research Unit
AET	Actual evapotranspiration
ARC	Agricultural Research Council
AWC	Available water holding capacity
AWS	Automatic weather station
CI	Canopy interception
CMA	Catchment Management Agency
CPD	Continuing professional development
CUP	Calibration and Uncertainty Program
CWRR	Centre for Water Resources Research
DEM	Digital Elevation Model
DHI	Danish Hydrologic Institute
DWS	Department of Water and Sanitation
ET	Evapotranspiration
FAO	United Nations Food and Agriculture Organisation
FC	Field capacity
FDC	Flow duration curve
GIS	Geographic Information Systems
GRA	Groundwater Resource Assessment
GW	Groundwater
GW-SW	Ground water-surface water
HAMSA	HydrologicAl Model for South Africa
HRU	Hydrological Response Unit
IAP	Invasive alien plants
IWR	Rhodes University, Institute of Water Resources
Kc	Crop coefficient (evapotranspiration coefficient)
KZN	KwaZulu-Natal
LAI	Leaf Area Index

MAE	Mean absolute error
MAP	Mean annual precipitation
MAR	Mean annual runoff
NLC	National Land Cover
NRF	National Research Foundation
NSE	Nash-Sutcliffe Efficiency
PBIAS	Percent Bias
PE	Pan evaporation
PET	Potential evapotranspiration
R^2	Correlation coefficient
RMSE	Root mean square error
RU-HEC	Rhodes University Human Ethics Committee
SACNASP	South African Council for Natural Scientific Professions
SAEON	South African Environmental Observation Network
SAHS	South African Hydrological Society
SANCIAHS	South African National Committee of the International Association of Hydrological Sciences
Sat SM	Saturation soil moisture
SAWS	South African Weather Service
SHE	Système Hydrologique Européen
SM	Soil moisture
S-pan	Symon's pan
SPATSIM	Spatial and Time Series Information Modelling
SRO	Surface runoff
SW	Surface water
SWAT	Soil and Water Assessment Tool
SZ	Saturated zone
UCT	University of Cape Town
UKZN	University of KwaZulu-Natal
UTM	Universal Transverse Mercator

UWC	University of the Western Cape
UZ	Unsaturated zone
VZ	Vadose zone
WP	Wilting point
WR2012	National Water Resources Study 2012
WRC	Water Research Commission
WRSM	Water Resources System Model
WRYM	Water Resource Yield Model

This page was intentionally left blank

CHAPTER 1. PROJECT BACKGROUND

1.1 MOTIVATION: UNCERTAINTY IN HYDROLOGICAL MODELLING AND THE NEED FOR GREATER AWARENESS

Hydrological modelling has become central to water resources management and catchment management in South Africa. With ever-growing pressure on the nation's water supply systems, basing decisions on reliable estimates of surface and groundwater resource availability is critical. At the same time South Africa's meteorological and hydrological monitoring infrastructure has declined severely in recent decades (Bailey and Pitman, 2015). The sector leans on models to fill the many gaps in observational data and to forecast the future under changing land cover, climate, and water management. The outputs of hydrological models are then carried through to water supply systems modelling, flood line and impact modelling, and used in other decision-making contexts. As reliance on modelling grows, so too should attention to model uncertainty: quantifying it, finding practical ways to reduce it, and accounting for it in decision making.

There are multiple sources of uncertainty in catchment modelling: uncertainty in the hydrometric data, used to drive and to calibrate models; uncertainty in the model parameters, which account for catchment properties; and uncertainty in the model structure itself, the mathematical representation of hydrological processes. Uncertainties in these elements propagate into uncertainty in the model outputs, such as the predicted streamflow or groundwater storage. This uncertainty is reduced with increasing data and information about the catchment, which can allow modellers to identify and exclude unrealistic parameter values and models. Uncertainty ranges for model outputs can be quantified by modelling across the range of inputs, parameter values, and model structures that can be considered potentially realistic for a catchment, given the data and information available.

It is important to note that catchment hydrological modelling is a step in the water resources modelling process, and that additional, important uncertainties also arise in the other steps. For example, supply system modelling often entails assumptions about future water demands across sectors, given population growth, agricultural and industrial trends, infrastructure degradation or improvement, technological developments, etc. The current study focuses only on uncertainty in catchment hydrological modelling, however there is a need for quantitative uncertainty estimation and consideration throughout the process.

1.1.1 Research on uncertainty in hydrological modelling

Three WRC-supported research projects have quantified uncertainties in hydrological modelling for case studies typical of the South African context in terms of data availability and modelling tools. These studies demonstrated that uncertainty levels are case specific, but can often be alarmingly high. Projects K5-1838: *Incorporating uncertainty in water resources simulation and assessment tools in South Africa* (Hughes *et al.*, 2011) and K5-2056: *Implementing uncertainty analysis in water resources assessment and planning* (Hughes, Mohobane and Mallory, 2015), developed an approach for quantifying and reducing uncertainty due to parameters and input data. This was tested in the SPATSIM-Pitman modelling tool across many case studies, including propagation of hydrological model uncertainty into yield prediction. When modelling ungauged catchments, parameter uncertainty was shown to result in estimates of mean annual streamflow that varied by 150% to 400% (Hughes *et al.*, 2011; Kapangaziwiri, Hughes and Wagener, 2012; Hughes, Mohobane and Mallory, 2015). (Hughes, Kapangaziwiri and Sawunyama, 2010) found that some ungauged cases would have such wide confidence intervals around predictions that these would be unusable for water resources decision-making.

Project K5-2927: *Critical catchment hydrological model inter-comparison and model use guidance development* (Glenday *et al.*, 2022), focused on structural uncertainty, applying a set of commonly used modelling tools (WRSM-Pitman, SPATSIM-Pitman, ACRU4, SWAT, and MIKE-SHE) to a set of case studies. This work demonstrated that models with differing structures can have comparable, acceptable accuracy in calibration, but these models can then predict very different outcomes to one another when they are further applied to scenarios of change. For example, when used to predict the impact of removing timber plantations from the riparian zone of a catchment in Kwa-Zulu Natal, one model predicted mean annual streamflow would only increase by 4%, while another model predicted an increase of 24% (Glenday *et al.*, 2022). These models used the same input data, were structured based on the same understanding of catchment properties, and were calibrated against the same streamflow dataset. This highlights high levels of uncertainty in change prediction due to model structure, over and above data and parameter uncertainty, even in cases where there is flow gauge data with which to assess and calibrate models. We can reasonably assume this uncertainty would be far greater in ungauged cases.

In practice, the various sources of uncertainty – input data, calibration data, model structure, and model parameter values – are linked to each other through various decisions that modellers must make in the modelling process. Modellers decide if and how to use various data and information from different sources, how to structure and parameterise a model of the catchment, and how to test and refine that model, including deciding when the model's performance is fit-for-purpose. Different individuals can make these decisions differently. Decisions are made across contexts of available time, funding, data, capacity and experience, and computing power, among other constraints, and ideally consider the ultimate uses of the model outputs. Often the first decision made is which modelling software tool to use, and this is most frequently chosen based on the modeller's familiarity with a tool (Addor and Melsen, 2019; Glenday *et al.*, 2022). The choice of tool influences the data used and how it is processed; the model structure, spatial discretisation, and process representation; and how parameter values are selected, including the calibration techniques applied (Breuer *et al.*, 2009; Holländer *et al.*, 2009; Huisman *et al.*, 2009; Glenday *et al.*, 2022). However, even when using the same modelling tool, individual modellers may model the same catchment differently, even when provided with the same data and information.

1.1.2 Hydrological modelling uncertainty is not often quantified or considered

Despite the growing number of research studies, including local ones, on modelling uncertainty and how to account for it and reduce it, it has remained common practice in the South African water sector to not attempt to quantify uncertainty in hydrological modelling outputs. While uncertainties in the input data and information informing parameter choices is generally acknowledged in applied modelling studies, the quantitative implications on modelled streamflow are rarely explored or presented. It is also typical to only apply one model structure, providing no estimation of structural uncertainty. As such, single model output datasets from single model structures, single sets of parameters, and single input datasets are commonly presented to decision-makers, rather than a range of likely output values that considers the uncertainties in the modelling elements. Without a sense of how uncertain a model output value is, there is no way for the decision-maker to consider the potential risks of using it to inform a decision. This has the potential to result in inappropriate decisions around water supply management or flood risk reduction, with negative consequences for lives, livelihoods, and ecosystems.

A salient and important example is the national water resources studies. Conducted periodically since the 1950s, these studies have provided foundations for water management and catchment management at various scales. Uncertainties related to both data and modelling are openly acknowledged in the studies (Pitman, 2011; Hughes, 2013; Bailey and Pitman, 2015); however, uncertainty bounds for the resulting resources estimates are not provided, likely due to time and resource constraints. Variations of the Pitman model (Pitman, 1973), a conceptual monthly-timestep rainfall-runoff model, were used for national water resources studies in 1981, 1994 (WR90), 2008 (WR2005), and 2015 (WR2012), resulting in the development of WRSM-Pitman software. This software does not have facilities for automated exploration of the feasible parameter space, automated calibration, or stochastic modelling to account for uncertainties about input data or parameters. As such, modelling for these studies has been done deterministically, presenting a single modelled timeseries for each quaternary catchment. Calibration has been done manually through trial-and-error parameter value adjustments, which is labour intensive, but made feasible by the level of experience of the modellers. For quaternaries with no gauge (or reservoir) at their outlet, the parameter values used were based on calibrated values from gauged catchments in the same region.

The most recent study, WR2012, made a step in the right direction regarding uncertainty. It presents categorical rankings of the quality of the observational data used, maps that show gauge locations vs the quaternary boundaries, and some calibration statistics for the 612 gauged catchments used to parameterise the 1,947 quaternaries (Bailey and Pitman, 2015). Of the gauge records used, 58% were considered 'reliable' (Bailey and Pitman, 2015). Combining data availability, data quality, and model calibration accuracy indices together can help users understand where the modelled predictions are likely to be the most or least uncertain across the country. However, there is still no indication of what the resulting magnitudes of this uncertainty may be.

A first step in addressing this issue is to grow a more concrete awareness in the hydrological modelling community and the water sector about uncertainties in hydrological modelling. Although most involved in the field are aware that modelling and model outputs are uncertain, the actual magnitudes of this uncertainty and the real-world, practical implications of this are rarely quantified, acknowledged, or acted upon. As such, this project focused on participatory, quantitative research into model uncertainty across the modelling sector, with a particular focus on structural uncertainty and modeller decisions. These elements of uncertainty are less frequently addressed in research and practice and can be the most substantial contributors (Butts *et al.*, 2004; Højberg and Refsgaard, 2005; Clark *et al.*, 2008; Mendoza *et al.*, 2015, 2016; Mockler *et al.*, 2016; Mockler, O'Loughlin and Bruen, 2016; Troin *et al.*, 2018; Melsen *et al.*, 2019; Knoben *et al.*, 2020; Moges *et al.*, 2020). This was complemented by activities to foster discussion across the modelling community on how to practically address these issues, as well as the promotion of resources that can assist modellers in understanding and comparing model structures available across software tools.

1.2 AIMS & OBJECTIVES

To begin addressing issues around insufficient consideration of hydrological modelling uncertainty, this project had the following aims and objectives:

Aims:

1. To improve shared understanding of model structural uncertainty, and the role of modeller decision-making, and their potential scale of impact, across the hydrological modelling community
2. To empower the community to identify practical strategies for assessing, communicating, and ideally reducing modelling uncertainty

Objectives:

1. Design and host an open 'model-a-thon' activity in which participants from across the water sector model the same catchment area, given the same input and calibration data, and apply the model to the same alternative scenario.
2. Assess and discuss the diversity of modelling approaches, structures, and output predictions in the 'model-a-thon' activity as means of collective scoping of the issue of structural uncertainty.
3. Host synthesis engagements to discuss the implications of the findings around modelling uncertainty for future water resources studies and initiate visioning of how these studies can be done in a way that accounts for and, where feasible, aims to reduce uncertainty.
4. Synthesise the engagement outcomes into a policy brief.
5. Supplement and publicise the 'wiki' website about modelling tool capabilities (https://hydromodel-sa-wiki.saeon.ac.za/), initiated during project K5-2927, to increase awareness of modelling options and accessibility of tools
6. Promote use of the Stack Exchange online question-and-answer platform (Stack Exchange Earth Sciences site: https://earthscience.stackexchange.com/) to facilitate information sharing across the modelling sector.

1.3 REPORT STRUCTURE

The activities and outcomes of the project regarding its objectives are documented in this report in separate sections on: the model-a-thon activity, the development of the policy brief, and the development of shared online resources (wiki and Stack Exchange).

For each of these elements, more detailed coverage and products produced have been included as appendices.

CHAPTER 2. MODEL-A-THON

2.1 ACTIVITY GOALS

The goal of the 'model-a-thon' was to explore how differently individual modellers are likely to set up and calibrate a catchment model when given the same starting brief and input data, and to understand how much these differences influence the modelled predictions and what this means for uncertainty. Each participant built and ran a model of the same case study catchment for both a baseline case and a scenario of change using the modelling tool of their choice. Participants were provided with observed streamflow data with which to calibrate their baseline model if they chose to do so. The variability in the change predictions across the set of models that achieved sufficient performance in baseline calibration gives an indication of potential model structural uncertainty. An effort was made to attract participants with a variety of experience levels and who use different modelling tools, with the aim of having multiple users of some of the commonly used tools in the group.

In addition to being a quantitative research exercise, the activity also served as a reflective exercise for the community of practice. Participation in the activity, and the discussions around it, aimed to foster additional discussion about modelling approaches, uncertainty, and challenges faced more broadly. The activity was purposefully launched at an in-person event attached to the South African Hydrological Society (SAHS) conference, allowing modellers from various institutions to meet one another, and become aware of one another's work and research.

2.2 ACTIVITY PREPARATION

The main tasks completed in preparing the model-a-thon activity were:

- Selection of an appropriate case study catchment and scenario.
- Preparation of the necessary datasets for participants to set-up and calibrate their models.
- Development of a submissions process to allow participant anonymity and with facilities to receive potentially large modelling files.
- Development of the activity instructions and submission survey, including testing the activity by the project team and revising as needed.

2.2.1 Case study catchment selection

The process of selecting a case study catchment to be used in the model-a-thon is covered in *APPENDIX A.1 – Review of South African research catchments for model-a-thon case study site selection*. An effort was made to find sites with readily available and share-able long-term rainfall and streamflow data with relatively simple land cover composition, no major dams or diversions, well described properties, and observational data on other hydrological processes (such as evapotranspiration, soil moisture, groundwater levels) as additional checks on model realism. Several South African research catchments were assessed in light of these criteria, and although none met all the more specific criteria, a site deemed to be a reasonable compromise across these was selected.

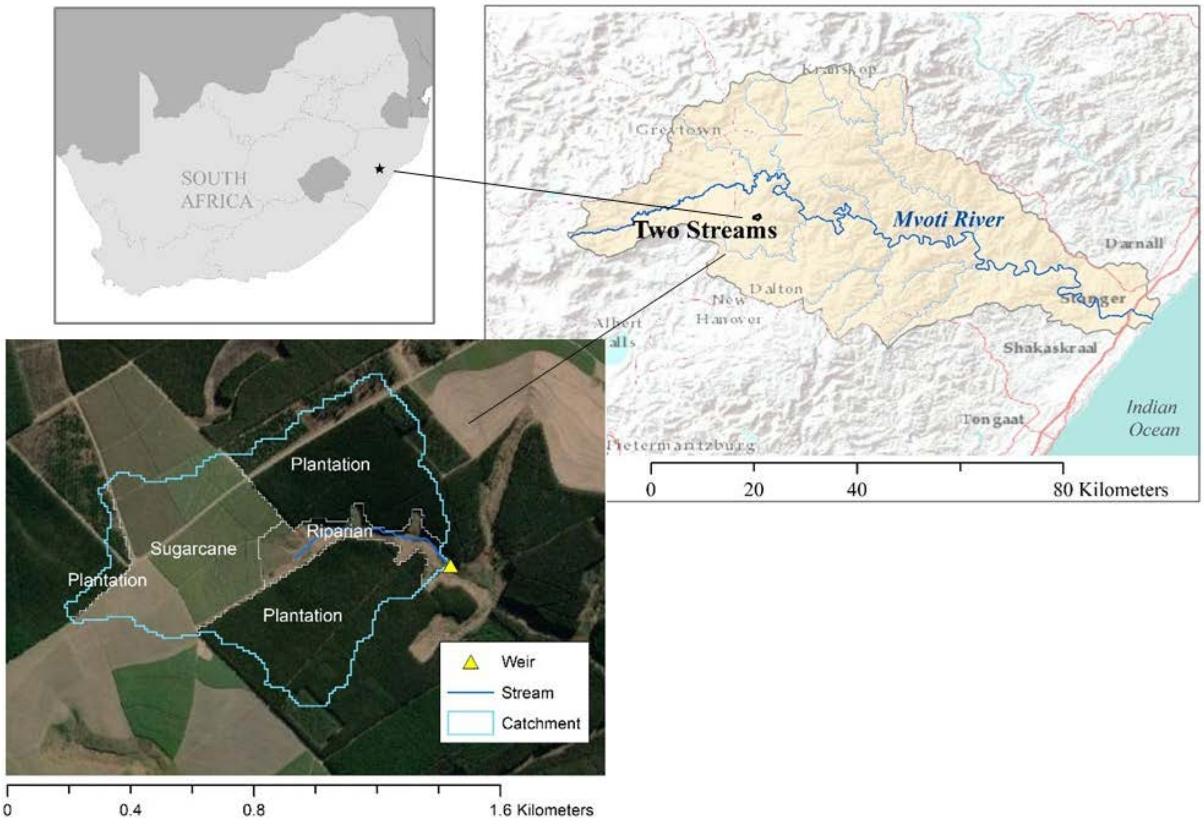


Figure 2-1 Two Stream catchment area, location within South Africa and the larger Mvoti River catchment

The catchment selected for the activity was the Two Streams experimental catchment (Figure 2-1), an instrumented headwater stream of the Mvoti River in the midlands of KwaZulu-Natal (KZN). The Two Streams area was selected for its land cover and data availability from previous hydrological research (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014, 2018; Ngubo, Demlie and Lorentz, 2022). The catchment has had sizeable coverage of contrasting land cover types which were relatively stable in their areas and properties over a five-year period for which climate and streamflow data are also available (2010-2014). These cover types were sugar cane (31% of the catchment in this period), mature black wattle (*Acacia mearnsii*) plantation (59%), and riparian vegetation (9%). For this reason, streamflow data from this period could be used to calibrate baseline models representing this cover distribution. Participants could then model a hypothetical scenario in which all the wattle is cleared and replaced with sugar cane using parameters derived in their baseline models. No new parameter values that were not tested in the calibration phase would need to be introduced for the scenario.

Drawbacks of this case study were that the Two Streams catchment is relatively small (0.7382 km^2) and that a five-year calibration period can be considered short relative to weather variability patterns. However, it is not uncommon to only have this amount of data for model calibration in practical settings. A larger catchment with more long-term data that was available and sharable, which also had data on multiple hydrological processes and had multiple contrasting land cover types that were fairly stable over time, could not be found in the time available for this project.

2.2.2 Data preparation

The datasets and information that participants were given to set-up their models are listed and described in Table 2-1 and Table 2-2 and the catchment data and properties are further described in the activity instructions document, included as *APPENDIX A.2 – Model-a-thon instructions document provided to participants*. In an effort to keep the time needed to complete the exercise to a minimum, some input data was provided in generic file formats as well as in the formats and file types needed for several of the most commonly used modelling tools: WRSM, SPATSIM, ACRU, SWAT, and MIKE-SHE. For some inputs, pre-formatting was not done as the processing will depend on how the individual participant choses to set up their model. All spatial data was provided in the same UTM (Universal Transverse Mercator) projection because this facilitates direct distance and area calculations in GIS software.

Streamflow, rainfall, and other climate parameter timeseries data for Two Streams were provided by UKZN and SAEON, as processed daily data and raw instrument data. Through various projects, automatic weather stations (AWS) were installed roughly 2 km from the catchment, at an elevation similar to the catchment average, one near the Seven Oaks farm (active since 1999) and one in an open grass area conforming to FAO guidelines for estimating reference PET (active since 2007); two tipping bucket rainfall gauges were installed at different elevations within 500 m of the catchment (active 1999/06-2009/08); and an additional AWS was installed on a mast above the tree canopy near the centre of the catchment (active since 2012) (Everson et al., 2007, 2014, 2018; Clulow, Everson and Gush, 2011). Comparing data for overlapping periods, no definitive elevation gradients were detected. This was not unexpected given the catchment size and elevation range. Estimated gap-free catchment-average datasets were prepared for the period 2000/01/01 to 2014/12/31, based on the relationships between stations for periods of data overlap. It should be noted that daily catchment average values were informed by data from several gauges for some times, and by only one or two gauges at others, making the catchment-scale rainfall input more uncertain at these times.

Table 2-1 Data on the case study catchment provided to the model-a-thon participants

Category	Timeseries	Spatial	Values / info	Data sources	Notes
Streamflow (catchment outlet)	yes daily & monthly	yes shapefile – point (outlet)		**V-notch weir, water level logger & flow data:** UKZN & SAEON (Everson *et al.*, 2007, 2014, 2018; Clulow, Everson and Gush, 2011)	
Rainfall (catchment scale)	yes daily & monthly	yes shapefile – point (centroid)		**Weather station & tipping bucket data:** UKZN & SAEON (Everson *et al.*, 2007, 2014, 2018; Clulow, Everson and Gush, 2011)	No strong rainfall gradient was found based on the gauge data available. As such a timeseries of estimated catchment average rainfall was prepared using the various gauges. Some tools require station locations for inputs (e.g. SWAT): catchment centroid provided as a 'dummy' station.
Reference PET, FAO-56 method (catchment scale)	yes daily & monthly	yes shapefile – point (centroid)		**Weather station data:** UKZN & SAEON (Everson et al., 2007, 2014, 2018; Clulow, Everson and Gush, 2011)	Timeseries of estimated catchment averages for temperature, humidity, solar radiation, and wind were prepared based on the weather station near the centre of the catchment, patched with those close by when needed. Ref. PET values need to be consistent with (calculated from) the temperature, wind, solar radiation, and relative humidity data provided because some models calculate PET internally from these inputs. Some tools require station locations for inputs (e.g. SWAT): catchment centroid provided as a 'dummy' station.
Temperature, min & max (catchment scale)	yes daily	yes shapefile – point (centroid)			
Relative humidity (catchment scale)	yes daily	yes shapefile – point (centroid)			
Solar radiation (catchment scale)	yes daily	yes shapefile – point (centroid)			
Wind speed (catchment scale)	yes daily	yes shapefile – point (centroid)			

Category	Timeseries	Spatial	Values / info	Data sources	Notes
A-Pan & S-Pan evaporation, Ref. PET equivalent (catchment scale)	yes daily, monthly, month-of-year averages			Calculated directly from the reference PET dataset (see above)	A-pan: Standard ACRU input, also needed for some modules in WRSM S-pan: Standard WRSM & SPATSIM input
Topography & elevation		yes rasters (DEM)	yes Ave, min, max elevation; Ave, min, max slope	ALOS-PALSAR RTE DEM rescaled from 12 m to 10m resolution Average elevation and slope calculated from DEM for catchment polygon (see below)	DEM rescaled to 10 m resolution to simplify the process of potentially using different model grid cell sizes in MIKE-SHE (new cell sizes should be multiples of the input data's resolution.) Raster needs to extend beyond the catchment boundary in case the user wishes to coarsen the grid. DEM provided for bounding rectangle.
Catchment area		yes shapefile – polygon	yes Catchment area	Delineated from the DEM & catchment outlet point	SWAT delineates the model catchment internally from the input DEM & MIKE-SHE needs the catchment to be consistent with DEM to not lose flow out the sides. Catchment delineated in QWAT (uses TauDEM tools) to ensure all participants are using the same area.
River network		yes shapefile – polyline	yes Channel rough estimate of width, depth, & bed material	Drainage lines delineated from the DEM provided; Approx channel properties from Naiken personal comm.	When delineating the channel (done in QSWAT, see note above), aerial photography was used to determine channel extent and a 0.3 km^2 flow accumulation threshold for channel definition.
Land cover (baseline & alternative scenario)		yes shapefile – polygon (cover types) raster (integer ID for cover types)	yes Descriptions by type: dominant species, LAI & root	Manual mapping from aerial imagery, informed by cover layer from eZemvelo KZN Wildlife; Property values from literature (see Appendix B) & ACRU Compoveg database	Parameter values from the ACRU 'Compoveg database' provided as an option. Rasters with integer ID numbers for each type, aligned with DEM raster, provided with table linking ID to type for SWAT input. Polygons generated from raster without smoothing grid cell boundaries to ensure areas match across data types.

Category	Timeseries	Spatial	Values / info	Data sources	Notes
Soils		yes shapefile – polygon (soil types) raster (integer ID for soil types)	yes By location type: depth to bedrock, layers & thickness By layer: texture, porosity, field capacity, wilting point, saturated conductivity	Local studies & mapping (Everson *et al.*, 2007; Van Tol *et al.*, 2007; Clulow, Everson and Gush, 2011) & ACRU Autosoils database	Values from soil pits and augering profiles were averaged for the samples in each type's polygon. Parameter values from the ACRU 'Autosoils database' provided as an option. Rasters with integer ID numbers for each type, aligned with DEM raster, provided with table linking ID to type for SWAT input. Polygons generated from raster without smoothing grid cell boundaries to ensure areas match across data types.
Geology & aquifers		no (all main formations & layers are thought to be present throughout catchment, so no spatial extents of formations needed*)	yes Conceptual diagram of layering and text description; estimated range of groundwater depths (in different positions); By layer: average thickness, specific storage /yield; saturated conductivity (vertical & horizontal)	Local studies & mapping; (Everson *et al.*, 2014; Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022) GRA II report properties for quaternary (DWAF, 2006)	* Some tools (e.g. MIKE-SHE using certain options) require gridded spatial data of the upper and lower elevations of each layer to be represented. This was not pre-prepared for users as there are many decisions that would need to be made in doing so that different users may do differently

Table 2-2 Input timeseries data file formats required by modelling tools that were pre-prepared for model-a-thon participants

Category	Model	Timestep	Units	Input file types to prepare	Notes
Streamflow (outlet, observed)		daily	m^3/s (ave), m^3 (tot), mm	.csv (comma separated values)	
		monthly	Mm^3, mm	.csv (comma separated values)	
	SPATSIM-Pitman	monthly	Mm^3	.txt	100x the observed flow volume as Pitman tool users were asked to use 100x catchment area
	WRSM-Pitman	monthly	Mm^3	.obs file (.txt file, change extension)	Row for each water year: year, 12 columns for monthly value (Oct start), column for total
	MIKE-SHE	daily	m^3/s (ave)	.dfs0 (generate in MIKE-SHE software)	
	SWAT	daily	m^3/s (ave)	SWAT format .txt file	one column, start date in first row
Rainfall (catchment average)		daily, m	mm	.csv (comma separated values)	
		monthly	mm	.csv (comma separated values)	
	ACRU4	daily	mm	ACRU “composite Y2K” ASCII .txt file	Composite file includes several climate timeseries
	SPATSIM-Pitman	monthly	mm	.txt	
	WRSM-Pitman	monthly	% of MAP	.ran file (.txt file, change extension)	Row for each water year: year, 12 columns for monthly value (Oct start), column for total
	MIKE-SHE	daily	mm	.dfs0 (generate in MIKE-SHE software)	
	SWAT	daily	mm	SWAT format .txt file	one column, start date in first row; file name match ‘dummy’ station @ centroid (need .txt meta with ID, lat/long, elevation)

Category	Model	Timestep	Units	Input file types to prepare	Notes
Reference PET (catchment average, FAO-56 method)		daily	mm	.csv (comma separated values)	
		monthly	mm	.csv (comma separated values)	
	MIKE-SHE	daily	mm	.dfs0 (generate in MIKE-SHE software)	
	SWAT	daily	mm	SWAT format .txt file	one column, start date in first row; file name match ‘dummy’ station @ centroid (need .txt meta with ID, lat/long, elevation)
A-pan evaporation (catchment average, calc from refPET)		daily	mm	.csv (comma separated values)	
		monthly	mm	.csv (comma separated values)	
	ACRU4	daily	mm	ACRU “composite Y2K” ASCII .txt file	Composite file includes several climate timeseries
	WRSM-Pitman	month-of-year	mm	.csv	12 month-of-year values starting in Oct. Values to be pasted or typed into WRSM software interface – external file format irrelevant
S-pan evaporation (catchment average, calc from refPET)		daily	mm	.csv (comma separated values)	
		monthly	mm	.csv (comma separated values)	
	SPATSIM-Pitman	monthly	% of mean annual	.txt	
	WRSM-Pitman	month-of-year	mm	.csv	Values to be typed into WRSM software interface – external file format irrelevant
Max & Min Temperature (catchment average)		daily	ºC	.csv (comma separated values)	
	ACRU4	daily	ºC	ACRU “composite Y2K” ASCII .txt file	Composite file includes several climate timeseries
	SWAT	daily	ºC	SWAT format .txt file	two column, start date in first row; file name match ‘dummy’ station @ centroid (need .txt meta with ID, lat/long, elevation)

Category	Model	Timestep	Units	Input file types to prepare	Notes
Windspeed	SWAT	daily	m/s	SWAT format .txt file	
Relative humidity	SWAT	daily	(ratio)	SWAT format .txt file	one column, start date in first row; file name match ‘dummy’ station @ centroid (need .txt meta with ID, lat/long, elevation)
Solar radiation	SWAT	daily	MJ/m^2	SWAT format .txt file	

Short (several day) gaps in the streamflow data were patched based on a seasonally derived relationships between flow changes and an antecedent precipitation index. Equipment failures in late 2014, and in 2015 and 2016 lead to gaps of over a month. As such, streamflow data was only provided to participants for 2009/10/01 to 2014/09/31 to cover a five water-year period in which the wattle trees were close to mature and the streamflow data had no major gaps. Data was screened for quality. Days with flow rates in the highest 20% were checked to ensure there was a corresponding rain event. This process highlighted days which had anomalously high flows. These were tracked to errors in the processing of the raw sub-daily data from the instrument and were corrected.

Given the timeseries data available, it was decided that the assessment period for comparing the baseline and change scenario model outputs would be 2003/10/01 to 2014/09/31. This means there is climate data for a three year 'warm-up' period for those who choose to include one. Participants were not instructed to include a warm-up, but were been provided the data to do so if they choose. A longer assessment period than the calibration period was chosen to cover a wider variety of climate conditions in the land cover change impact prediction.

2.2.3 Activity logistics

An online sign-up form was developed for the activity that automatically assigned participants a participant number and mailed them their participant number and the instructions document. Participant numbers were used to anonymise submissions in analyses and reporting. The case study database was made available to participants through an online drive hosted by Rhodes University Institute of Water Resources (IWR). Participants only had download rights to this drive such that no one could alter the database. A similar drive was created for participants to submit their models and outputs. Participants only had upload rights to the submission drive such that no one can open or alter anyone else's submissions.

Other logistics of running the event were handled manually by the project team. A dedicated email account was established to allow the project team to offer assistance to participant. Assistance was expressly limited to questions regarding interpretation of the instructions, accessing the data provided, and technical issues with the online submission process, and did not extend to technical or theoretical questions about modelling software or modelling strategy.

An online submission survey was developed to gather information about the participants' modelling set-ups and approaches as well as their experience with the activity. This survey was not provided to participants until they had completed their submissions and as some of the questions could potentially have prompted them to change their modelling strategy. It was decided not to wait until everyone had submitted to release the survey because participants may forget things if weeks pass between completing the activity and receiving the survey. As such the project team regularly checked the online submissions drive and to emailed participants the survey link once they had submitted their models.

2.2.4 Attracting participants

Effort was made to advertise the model-a-thon widely in the hydrology community to attract a sizeable participant group that would ideally include users of several different modelling tools and a mix of experience levels. It was publicised through a scoping survey and emails sent to the South African Hydrological Society (SAHS) general distribution list and to all those registering for the SAHS 2022 conference. Targeted emails were also sent to various known modelling experts. Communications highlighted the benefits of participation, including contributing to research, community building, student prizes, acknowledgements, paper co-authorship opportunity, and South African Council for Natural Scientific Professions (SACNASP) Continuing Professional Development (CPD) points.

2.3 ACTIVITY LAUNCH, SUBMISSION PERIOD, AND FEEDBACK

The model-a-thon activity was launched on 11th October 2022 at an in-person briefing session held at the SAHS inaugural conference. An online briefing was also held on 18th October 2022, which was recorded and made available online (https://youtu.be/PeOAKEbAz9E). Between the launch and submission time, participants could ask questions via a model-a-thon email account that was attended to by project team members in rotation. When a participant submitted their models and output, a project team member would then send them a submission survey regarding their model set-up and experience with the exercise.

The initial due date for submitting models was the 18th November 2022, giving participants five weeks to complete the activity, however several extensions had to be given to get a sufficient number of submissions. There were 43 sign-ups by 38 individuals intending to complete the activity (participants were allowed to do the activity multiple times using different modelling tools), but only seven models were submitted by the initial due date. A two-week extension to 5th December was publicised. Four additional models were submitted by this date. The remaining people who had signed up but not submitted models were then surveyed to ascertain whether they would commit to completing the activity if given an extension to mid-January 2023. This was done to determine if it was worthwhile for the project team extend the submission date again, and thereby delay the outcomes analysis phase of the project. Nine out of thirteen people who responded to this survey indicated they would complete the activity if given this extra time and so the extension was granted. This extension resulted in eight additional submissions, resulting in a total of nineteen.

After the submissions had been analysed, as described below (Section 2.4), an online feedback session was held on 15th March 2023. All those who signed up to participate in the model-a-thon and the project reference group were invited. The session was recorded and made available online (https://youtu.be/C1f1SeZ_0gE). Participants were asked to reflect on the main 'take-home' messages they thought were demonstrated by the results. In April an invitation was distributed to those who completed the model-a-thon to be part of the drafting process of a journal article. Those who review and contribute to the paper outline and drafts of the article will be listed as co-authors alongside the project team.

2.4 ANALYSIS OF SUBMISSIONS

2.4.1 Sign-up and submission survey data: participant population & modelling approach

Information given by participants in the sign-up form and submission survey were used to look at the modelling tools used, modeller experience levels, and modelling strategies. Modellers were asked to indicate their level of experience with the tool they used to complete the activity using one of six categories. These were lumped into three broader categories for analyses due to the numbers of participants: beginner (1 – first time use during model-a-thon, 2 – only used during a course), intermediate (3 – used in one independent project, 4 – used in multiple projects), experienced (5 – use on regular basis, 6 – teach and/or develop the tool). Participants were also asked about the level of discretisation in their models in terms of the number of spatial units and about their calibration strategy (none, manual, automated with user limited parameters and value ranges, automated unlimited – using tool's full potential parameter value space). Participants were also asked to indicate how long they spent on the activity. The diversity of modelling strategy, structure, and performance was assessed across and within experience level and tool groups, with the level of assessment being guided by the number of submissions.

2.4.2 Streamflow prediction performance and acceptability criteria

The streamflow outputs of participants' baseline models were compared to the observed streamflow dataset for the calibration period (2009/10/01-2014/09/30) and assessed based on several 'goodness-of-fit' performance statistics. Statistics were calculated using monthly timeseries for all submissions, while daily statistics were also calculated for the daily timestep models. Five statistics were used to determine whether a model would be considered acceptably 'behavioural' for further use in scenario outcome prediction. This implies that a model was considered to be a reasonable representation of the system. These statistics were: percent bias (PBIAS), coefficient of determination (R^2), Nash-Sutcliffe Efficiency (NSE) of untransformed and log transformed flow (Nash and Sutcliffe, 1970), and Kling-Gupta Efficiency (KGE) (Gupta *et al.*, 2009). Equations, interpretation, and thresholds of acceptability for these statistics are described in Table 2-3. Thresholds of acceptability were informed by commonly applied performance assessment categories (Moriasi *et al.*, 2007, 2015; Kling, Fuchs and Paulin, 2012). Additional performance statistics were calculated for further exploration of model fit to the observed dataset, including error in the mean, mean absolute error (MAE), root mean square error (RMSE), and the error in average flows and slopes of specific portions of the flow duration curve (FDC): high flows (0-10% exceedance probability), medium-high (10-40%), mid-range (40-60%), medium-low (60-90%), and low flows (90-100%).

Table 2-3 Model performance statistics and thresholds of acceptability applied to select sufficiently behavioural models

Statistic	**Equation & interpretation**	**Value range** (& optimal value)	**Acceptability threshold** (& minimum performance class*#)
Percent Bias (PBIAS)	$PBIAS = 100\% \times \frac{\sum_{t=1}^{T}(Q_m^t - Q_o^t)}{\sum_{t=1}^{T} Q_o^t}$ Average magnitude & direction of the model's prediction error.	$-\infty\ to\ \infty$ optimal = 0	**|PBIAS| ≤ 25%** satisfactory*
Coefficient of determination (R^2)	$R^2 = \left[\frac{\sum_{t=1}^{T}(Q_o^t - \bar{Q}_o)(Q_o^t - \bar{Q}_m)}{\sum_{t=1}^{T}(Q_o^t - Q_m^t)^2 \sum_{t=1}^{T}(Q_m^t - \bar{Q}_m)^2}\right]^2$ Degree of co-linearity between model & observed data. Shows the match in temporal pattern of values, but not necessarily in their magnitude.	$0\ to\ 1$ optimal = 1	**R^2 ≥ 0.6** satisfactory**
Nash Sutcliffe Efficiency (NSE)	$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_o^t - Q_m^t)^2}{\sum_{t=1}^{T}(Q_o^t - \bar{Q}_o)^2}$ *(Nash and Sutcliffe, 1970)* Model's residual variance or error ('noise') vs variance in measured data ('information'). Shows fit of pattern and magnitude.	$-\infty\ to\ 1$ optimal = 1 NSE = 0 means model is no better than using mean of the observed data as the prediction for every timestep	**NSE ≥ 0.5** satisfactory*
NSE of log transformed data (NSE log)	*Calculated as above, using log transformed observed and modelled datasets.* Error in high flow values is weighted highly in NSE, especially as values are squared. Log transformation of the data increases the relative weight of low flows.		**NSE log ≥ 0.5** satisfactory*
Kling-Gupta Efficiency (KGE)	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{s_m}{s_0} - 1\right)^2 + \left(\frac{\bar{Q}_m}{\bar{Q}_0} - 1\right)^2}$ *(Gupta et al, 2009)* Combined index of the linear correlation, match in variation, and match in mean (or bias) between the modelled and the observed modelled datasets. Shows fit of pattern and magnitude.	$-\infty\ to\ 1$ optimal = 1	**KGE ≥ 0.5** satisfactory#

Where Q_o^t is the observed value for timestep *t*, Q_m^t is the modelled value for timestep *t*, $\bar{Q}_o$ and $\bar{Q}_m$ are the means of the observed and modelled values for timesteps t=1 to T respectively, S_o and S_m are the standard deviations of the observed and modelled values for timesteps t=1 to T respectively, and r is Pearson's correlation coefficient between the Q_o^t and Q_m^t datasets for t=1 to T.

Performance class references: *(Moriasi et al., 2007), **(Moriasi et al., 2015), #(Kling, Fuchs and Paulin, 2012)

2.4.3 Predicted change in streamflow with the land cover change scenario

The predicted change in streamflow between the baseline and the wattle removal scenario was assessed for all submissions; however, only the subset of models with acceptably behavioural streamflow prediction for the baseline were used to look at structural uncertainty. It was assumed that any model in this accepted set might reasonably be used for prediction in an applied setting. As such, the range of predictions across this set gives an indication of the potential magnitude of model structural uncertainty for this case. Predicted change in streamflow due to wattle removal was assessed in terms of magnitude and percentage change in mean annual runoff (MAR) for the modelled period, as well as the absolute and relative changes predicted for different parts of the hydrograph: high flows (0-10% exceedance probability), medium-high (10-40%), mid-range (40-60%), medium-low (60-90%), and low flows (90-100%).

Modelled streamflow for the baseline and wattle removal scenario were compared for the period 2003/10/01 to 2014/09/30, a longer time period than the calibration period. The calibration period was limited by the availability of streamflow data for a time when the land cover distribution and properties were stable. A longer assessment period was chosen for the land cover impact assessment in order to include potential catchment responses to a wider variety of weather conditions.

2.4.4 Baseline and scenario modelled water balances

For the subset of models meeting the streamflow prediction performance criteria, the modelled catchment water balances were assessed for both the baseline and the wattle removal scenario. To keep the activity simple, the model-a-thon participants were only requested to submit their models' streamflow outputs and model project files (files needed to run the model). As such, for most cases, the project team needed to re-run the model to obtain outputs for the other hydrological fluxes. Different modelling tools predict and output different fluxes at different scales. For all models, mean annual total evapotranspiration (ET), surface runoff, interflow (or shallow subsurface runoff), groundwater runoff, and storage change predicted for the catchment could be obtained. When possible, total ET was further subdivided into canopy interception, ET drawn from soil, and ET drawn from groundwater and predicted ET for the different land cover types were obtained and compared.

For each accepted model, the catchment water balance fluxes for the baseline case were compared to those for the wattle removal scenario to see what combination of predicted flux changes resulted in the predicted streamflow change. It was expected that models with greater differences in their predicted baseline water balances would then predict more different streamflow responses to one another when used to model the impacts of a land cover change. It was expected that these differing streamflow responses would be driven by different patterns of change in across the contributing fluxes in the contrasting models.

2.4.5 Modelled water balance fluxes compared to field observations

Modelled hydrological fluxes were compared to various observational datasets for the Two Streams catchment, again focusing on models with acceptable streamflow performance. The nature of the comparisons was constrained by various mismatches in scale between modelled units and output and the field observations**.** Measurements of total ET, canopy interception, and transpiration for the wattle plantation (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014; Clulow *et al.*, 2022); total ET from sugarcane (Wiles, 2006); soil moisture and groundwater levels (Everson *et al.*, 2014; Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022); and stream, rain and groundwater stable isotope concentrations (Watson, 2015; Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022) were used (Table 2-10). Data were collected for different studies at different time periods with varying weather conditions and so modelled outputs for the matching time intervals were used for comparison. For data collected in the wattle plantation, only data from 2009/10 onwards, when the plantation was mature, were used to compare to modelled output. Data collected for one or more full year was used to estimate an annual total or a mean annual value for a flux which was compared to the equivalent modelled output. As noted above, not all models explicitly represent, or produce outputs for, all fluxes.

Most ET observation in Two Streams has focused on the wattle plantation. Some data exists for sugarcane, but there are not comparable field observations for the riparian vegetation. Wiles 2006 conducted a one year scintillometry campaign to estimate ET from a sugarcane field at Two Streams from 2004/10 to 2004/09. Wattle ET was measured using both scintillometry and an eddy covariance (EC) flux tower for several years during the relevant period (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014). Clulow et al., 2022 harmonised these datasets to generate an ET timeseries for 2003 to 2013. Data for 2009/10 to 2013/09 from this set was used here. Wattle transpiration was estimated using heat pulse velocity sapflow datasets from twelve trees across four sites (upper and lower slope, north and south-facing), which were scaled by sapwood and stem density to get a rate for the wattle area (Everson *et al.*, 2014). Wattle canopy interception was estimated using above-canopy rainfall measurements compared to below-canopy throughfall troughs (Everson *et al.*, 2014); however, this method ignores stemflow and so may be an overestimate. A short sampling period in 2013 suggested stemflow could be significant, as much as 39-50% of rainfall (Everson *et al.*, 2014). Using another approach, when estimated transpiration is subtracted from total ET, the remainder is canopy interception and soil moisture evaporation. Direct evaporation of soil moisture is likely small in the plantation due to shade and thick litter, so this amount should be mostly canopy interception.

Soil moisture data and groundwater level data were used to estimate the change in subsurface water storage at the catchment scale over time. Because different models delineate different subsurface storages and linkages between them in different ways, only total storage was considered. The change in soil profile water storage down to 2.4 m for 2009/10 to 2013/09 was estimated by Everson et al., 2014, based on six sets of soil moisture probes distributed across the wattle plantation. Probes installed to 4 m depth were active for a shorter period and suggested less moisture and less change at this depth compared to layers above (Everson *et al.*, 2014). Groundwater levels were monitored in four boreholes spatially distributed across the catchment (Everson *et al.*, 2014; Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022). The average change in groundwater depth between 2009/10 to 2013/09 across the boreholes was scaled by a range of specific storage values estimated for the aquifer, 0.005-0.03 (Ngubo, 2019), to estimate the change in groundwater storage. The resulting total storage change estimate is acknowledged to be very rough and does not include sugarcane or riparian soil data. It was used as an indicator of the direction and order of magnitude of the actual storage change for the period.

Watson 2015 and Ngubo 2019 collected stable isotope (^{2}H and ^{18}O) concentration data for rainfall, soil water, borehole water, and streamflow in the Two Streams catchment to determine the relative contributions of different flow pathways to streamflow. Watson 2015 concluded that streamflow was dominated by groundwater outflow, noting times with concentration overlaps between rainfall, soil water, and groundwater, complicating separation. There was some indication of hydraulic lift in the wattle, with soil isotope concentrations approaching groundwater values in some dry periods (Everson *et al.*, 2014; Watson, 2015). Ngubo 2019 conducted further monthly sampling and used this to estimate proportional contributions of ‘event’ versus ‘pre-event’ water to streamflow over time. ‘Event’ water has the chemical signature of the recent rain event and so would have followed a direct route to the stream, such as surface flow, quick macropore or fracture flow, and rain falling directly on the stream. ‘Pre-event’ contributions reach the stream via delayed pathways through soil and/or groundwater stores that result in differing chemical signatures to the most recent rain event. Two years, 2012/10 to 2014/09, of this data were used to estimate the mean annual contribution of direct runoff to total annual flow for the period. This was compared to modelled contributions of surface runoff to streamflow. However, because it is possible for interflow to contribute to 'direct' flow in some instances, and differences in how 'interflow' is modelled, the 'event' flow contribution was also compared to modelled surface flow plus interflow.

2.5 ACTIVITY OUTCOMES AND MODELLING RESULTS

There were eighteen usable submissions for the model-a-thon. Although there were nineteen submissions, one set of baseline and scenario models submitted by a first-time user of a modelling tool unfortunately had to be excluded from the analyses. During results analysis it was found that the models in this submission were not set-up to represent the land cover distributions specified for the exercise.

2.5.1 Participant population

The activity succeeded in attracting modellers across various experience levels and users of a variety of modelling tools (Figure 2-2). Of the eighteen models submitted, five were built with ACRU (ACRU3 or 4), five with SWAT (SWAT2012 or SWAT+), three with WRSM-Pitman, three with SPATSIM-Pitman, and two with MIKE-SHE. These are the five catchment hydrology modelling tools that are most commonly used in South Africa according to a 2021 survey from (Glenday *et al.*, 2022). The majority of submissions (11/18, or 61%) were from intermediate level users of a tool, while five were from experienced users (use on regular basis, teach or develop the tool) and two were from beginners. Fortunately, there were multiple submissions using each major tool and submissions from multiple experience levels for each tool (Table 2-4).

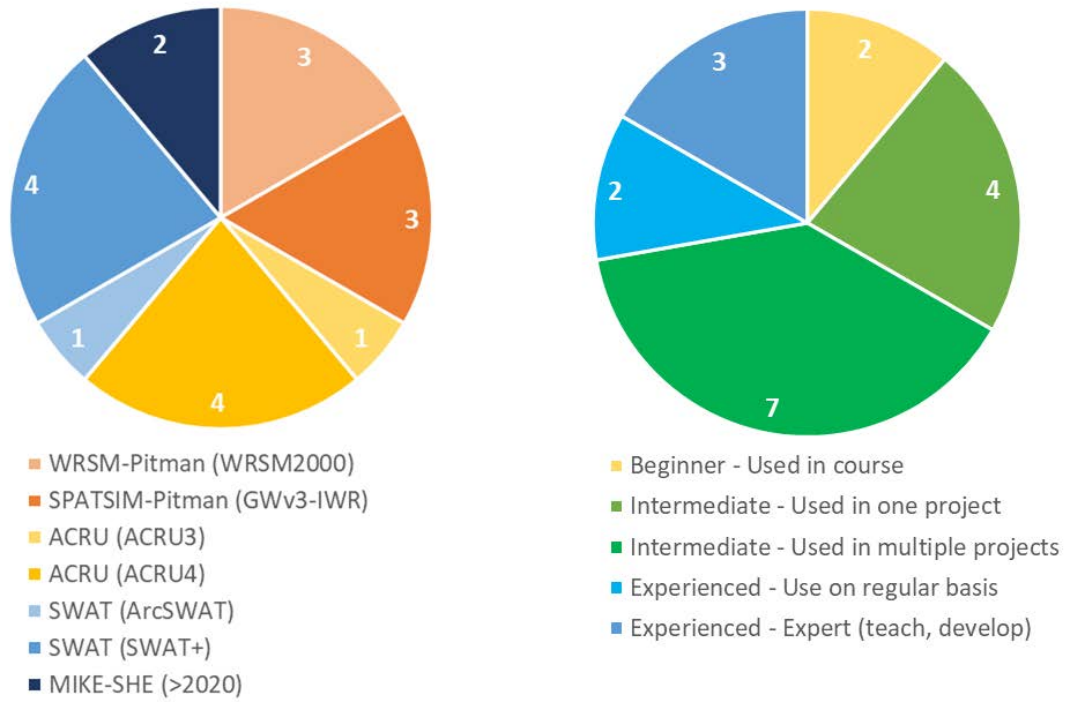


Figure 2-2 Numbers of submissions using different modelling tools (left) and numbers of submissions from users of different experience levels with the tool applied (right)

Table 2-4 Numbers of sign-ups and submissions by modelling tool and user experience level (tool listing sorted by number submissions then sign-ups)

Tool & experience level	Sign-ups n	Sign-ups % of total	Submissions n	Submissions % of total	Submissions % of tool subs	Submission rate vs sign-up (%)
SWAT	**12**	**28%**	**5**	**28%**		**42%**
beginner	*5*	*12%*	*2*	*11%*	*40%*	*40%*
intermediate	*6*	*14%*	*3*	*17%*	*60%*	*50%*
experienced	*1*	*2%*	*0*			*0%*
ACRU	**9**	**21%**	**5**	**28%**		**56%**
beginner	*0*		*0*			-
intermediate	*5*	*12%*	*3*	*17%*	*60%*	*60%*
experienced	*4*	*9%*	*2*	*11%*	*40%*	*50%*
SPATSIM-Pitman	**5**	**12%**	**3**	**17%**		**60%**
beginner	*2*	*5%*	*0*			*0%*
intermediate	*2*	*5%*	*2*	*11%*	*67%*	*100%*
experienced	*1*	*2%*	*1*	*6%*	*33%*	*100%*
WRSM-Pitman	**5**	**12%**	**3**	**17%**		**60%**
beginner	*0*					-
intermediate	*4*	*9%*	*2*	*11%*	*67%*	*50%*
experienced	*1*	*2%*	*1*	*6%*	*33%*	*100%*
MIKE-SHE	**8**	**19%**	**2**	**11%**		**25%**
beginner	*2*	*5%*	*0*			*0%*
intermediate	*3*	*7%*	*1*	*6%*	*50%*	*33%*
experienced	*3*	*7%*	*1*	*6%*	*50%*	*33%*
Other: HBV, HEC-HMS, Goldsim	**4**	**9%**	**0**			**0%**
beginner	*4*	*9%*	*0*			*0%*
intermediate	*0*					-
experienced	*0*					-
TOTAL	**43**		**18**			**42%**

It would have been ideal to have multiple submissions for each major tool and experience level combination, but it is likely that the hydrological modelling community is not large enough for that to be easily attained, particularly through a volunteering activity. Even if everyone who signed up initially had managed to complete the activity, there were too few highly experienced users signing up for some tools (SWAT, SPATSIM, WRSM) and few to no beginners for the others (Table 2-4). Having multiple submissions per tool allows for assessment of the diversity of strategies, structure, and performance across models built using the same tool, however there were too few per tool in this case to warrant statistical analyses of these factors. Indications of potential correlations were noted with caution as there is insufficient data to verify their existence or strength.

Modellers in academia and research dominated the potential and actual participant group, accounting for 74% (28/38) and 86% (12/14) of the people signing-up and submitting models respectively. Post-graduate students were included in this grouping and accounted for more than a third of all those signing up (13/38, 34%) and submitting models (5/14, 36%). Four people working in government agencies signed up, but none completed the activity, and six people in the private sector (consultancies and other companies) signed up, of whom two submitted models.

Higher proportions of beginner model users and MIKE-SHE users who signed up did not complete the activity than for other groups. Overall, 37% of individuals who signed up submitted models. As some submitted multiple models, this accounted for 42% of intended models. For most tools and experience levels 40-60% of sign-ups submitted (Table 2-4). However, only 15% (2/13) of those who indicated that they were beginners with a tool completed the activity. Within this group, none of the seven who stated this would be their first time using a tool completed the activity. Six had used the tool in a course, of which two submitted. Across tools, MIKE-SHE had the greatest rate of attrition: eight sign-ups, two submissions. Of the tools in this set, MIKE-SHE typically takes the longest to set up and run (Glenday *et al.*, 2022), which may be why more users did not complete the activity.

2.5.2 Modelling approaches

A variety of approaches were applied across the participants when modelling the Two Stream case study (Table 2-5). There was a huge range in the spatial discretisation across the models, from representing the catchment as a single lumped unit to having 7,382 units (10 m grid cells) for which hydrological processes were calculated. Spatial units in distributed (gridded) models, are often parameterised by cover class and soil type. As such, the level of spatial discretisation is not a clear proxy for the level of parameterisation, but is used here as a rough indicator of model complexity. The majority of participants (11, 61%) manually calibrated their models' parameter values, while three made use of automated calibration tools. Four did not attempt to calibrate, submitting models with their a-priori parameter selections. There were no obvious associations between experience level and approach within a given tool and there were insufficient participant numbers using each tool to be able to explore this.

Spatial discretisation choices generally appeared to be associated with the modelling tool used and the nature of the data provided for the exercise. SPATSIM uses subcatchment-scale units and all the SPATSIM models used a single unit for this case. The MIKE-SHE users chose to model with a 10m grid (7,382 cells), which was the resolution of the gridded datasets provided. The maps provided had three land cover types and four soil types, with one cover type occurring on multiple soil types while the other two were largely associated with one soil. The ACRU models all had either three or four hydrological response units (HRUs), in line with these datasets. Most chose to use three cover type HRUs and so would have generalised soil properties as needed.

There was more structural diversity across the WRSM models and much more across the SWAT models. WRSM uses subcatchment units with the option of adding special sub-units within subcatchments to represent areas of vegetation of particular interest: plantations, irrigated areas, invaded areas. Both are termed 'modules' in the software. All three WRSM modellers used different numbers of 'modules' (2, 3, and 4), reflecting different approaches to representing the land cover distribution. SWAT uses HRUs, like ACRU4. However, unlike ACRU, SWAT software automates HRU generation by overlapping maps of subcatchments, vegetation types, soil types, user determined slope classes (optional), and, in SWAT+, an optional floodplain vs upland area delineation. Users can then choose to lump combinations that only make up very small areas or not. The SWAT users in the model-a-thon made notably different decisions in the HRU delineation process, with all five models having different numbers of HRUs: 14, 15, 32, 700, and 4,316.

Approaches used to parameterise models did appear to be influenced by the facilities of the of modelling software tool used, however different individuals using the same tool made different choices. The four who did not try to calibrate their models were using SWAT and ACRU. Both tools have in-built suggested parameter databases for a variety of vegetation and soil types in their interfaces, which is not the case for the other tools in the set (Glenday *et al.*, 2022). However, others using SWAT and ACRU did try to calibrate parameters and two of those who did not calibrate noted that this was due to a lack of time to work on the activity, rather than this being their preferred or typical approach. The three who used automated calibration tools were using SPATSIM and MIKE-SHE, both tools which have in-built routines for exploring a user-defined parameter space. Not all SPATSIM users in the activity choose to use this facility. ACRU and WRSM do not have this capability and while there are automated calibration tools for SWAT, these have been built outside of the software and their use is generally taught separately (Glenday *et al.*, 2022).

Most participants (11, 61%) used one day or less to complete the activity. The time used to complete the activity can reflect the modellers' familiarity with the tool, the complexity of the tool, the run time of the model, as well as the time the individual was able to, or chose to, dedicate to model improvement. In some cases (ACRU, SWAT, MIKE-SHE), more experienced uses took less time to complete, however several noted time availability constraints in the submission survey. The ACRU and WRSM users generally took less time to complete the activity. Those using MIKE-SHE and beginner users of SWAT generally used more time than others, 2-4 days. The three modellers using automated calibration tools took longer than most others: one used two days (MIKE-SHE) and two used four days (one with SPATSIM, one MIKE-SHE). There were too few participants to indicate that these observations would be generalisable.

Table 2-5 Calibration approach, spatial discretisation (baseline model), and time used by modelling tool and user experience level (tools sorted by typical level of spatial discretisation)

Tool & experience level	n	n applying parameter calibration approach:			Number of spatial units in model (range)	Days used to complete activity (range)
		none	manual	automated		
All tools	**18**	**4**	**11**	**3**	**1-7382**	**0.25-4**
beginner (SWAT)	*2*	*1*	*1*		*14-32*	*2-3*
intermediate	*11*	*2*	*7*	*2*	*1-7382*	*0.75-4*
experienced	*5*	*1*	*3*	*1*	*1-7382*	*0.25-2*
SPATSIM-Pitman	**3**		**2**	**1**	**1**	**1-4**
intermediate	*2*		*1*	*1*	*1*	*1 4*
experienced	*1*		*1*		*1*	*2*
WRSM-Pitman	**3**		**3**		**2-4**	**1**
intermediate	*2*		*2*		*2-3*	*1*
experienced	*1*		*1*		*4*	*1*
ACRU	**5**	**2**	**3**		**3-4**	**0.25-1.5**
intermediate	*3*	*1*	*2*		*3-4*	*1-1.5*
experienced	*2*	*1*	*1*		*3-4*	*0.25-0.5*
SWAT	**5**	**2**	**3**		**14-4316**	**0.75-3**
beginner	*2*	*1*	*1*		*14-32*	*2-3*
intermediate	*3*	*1*	*2*		*15-4316*	*0.75-1*
MIKE-SHE	**2**			**2**	**7382**	**2-4**
intermediate	*1*			*1*	*7382*	*4*
experienced	*1*			*1*	*7382*	*2*

2.5.3 Streamflow prediction performance

There was a wide range in monthly streamflow prediction performance across the set of models, with those for which no parameter calibration was attempted having the poorest results (Figure 2-3, Table 2-6). Observed mean annual runoff for the calibration period was 23.9 mm, a mean monthly runoff of ~2 mm. Modelled mean monthly runoff values ranged from 0.4 to 6.9 mm, 18% to 347% of the observed. PBIAS magnitudes ranged from 1% to 251%, R^2 from 0.05 to 0.7, NSE from -28 to 0.7, and KGE from -3.7 to 0.8. Most models (13/18) had a general overprediction bias (positive PBIAS), but each had a different pattern of over- and underprediction of high, medium, and low flows. Four models met all five acceptability criteria. Two models met four criteria, both having below-threshold NSE log values. Five did not meet any of the criteria, of which four were the uncalibrated models.

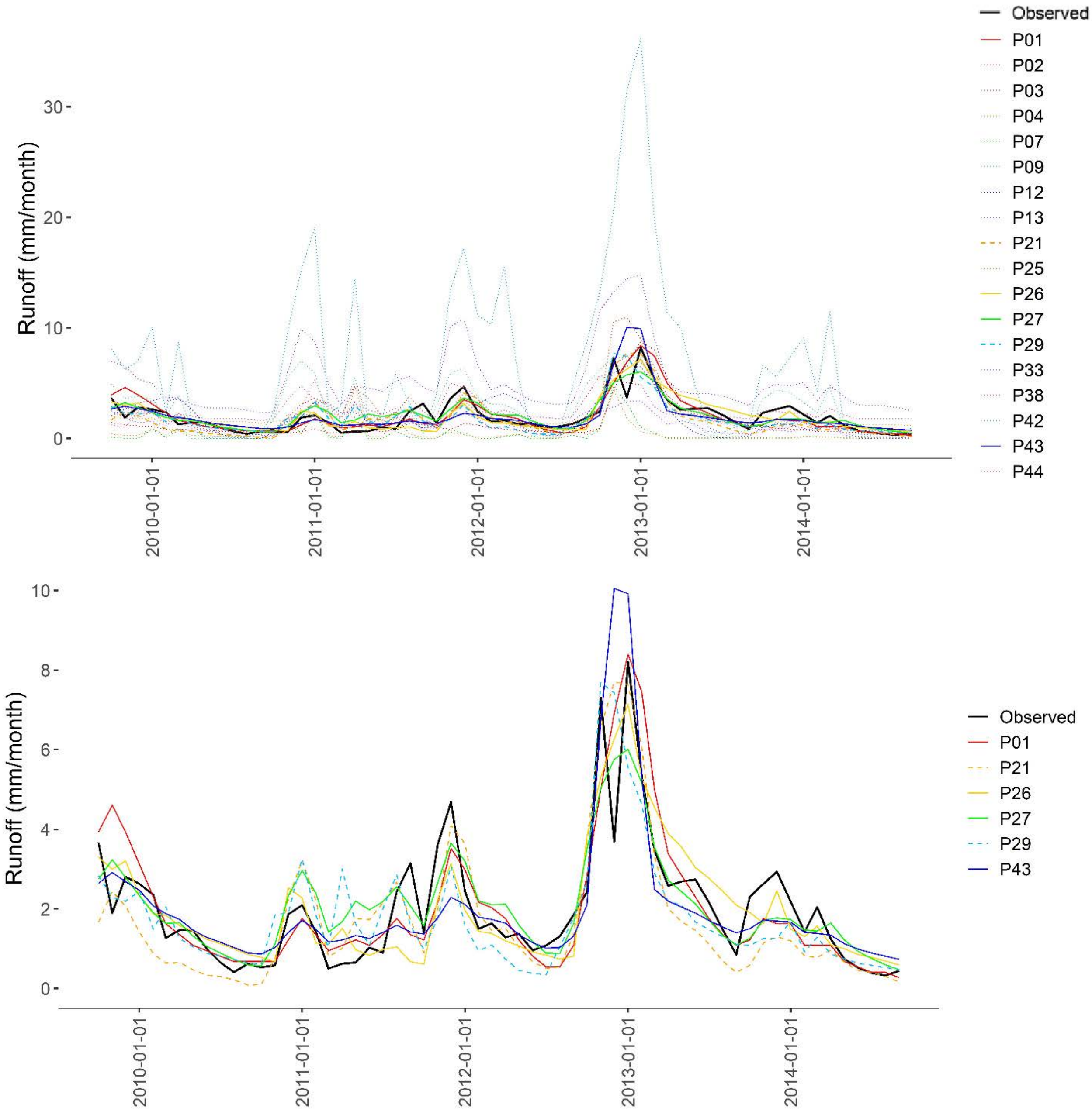


Figure 2-3 Monthly observed and modelled hydrograph for all models (top) and higher performance models (bottom) for the calibration period, 2009/10/01-2014/09/30 *(models meeting 5/5 performance criteria shown with solid lines, 4/5 with dashed lines, fewer with dotted lines)*

Table 2-6 Metadata and performance statistics for monthly streamflow prediction for all model submissions *(performance statistics meeting the acceptance criteria are highlighted in green with the three best in the set shown in bold font and the best underlined; models were sorted by the number of criteria met and then by descending NSE)*

Dataset	Modelling metadata				Monthly runoff (mm) statistics, calibration period				Performance statistics					Number of acceptability criteria met
	Tool	Experience with tool	Days used	Calibration approach	Mean	Min	Max	Std. dev	PBIAS	R^2	NSE	NSE log	KGE	
Observed					**2.0**	**0.3**	**8.2**	**1.5**						
P27	SPATSIM	Intermediate	4	Automated	2.1	0.5	6.0	1.2	25%	**<u>0.73</u>**	**<u>0.73</u>**	**0.66**	0.74	**5**
P26	MIKE-SHE	Intermediate	4	Automated	2.0	0.6	7.1	1.5	16%	**0.71**	**0.70**	0.60	**<u>0.83</u>**	**5**
P01	WRSM	Intermediate	1	Manual	2.0	0.3	8.4	1.7	**5%**	**0.72**	**0.65**	**<u>0.73</u>**	**0.81**	**5**
P43	WRSM	Experienced	1	Manual	2.0	0.7	10.1	1.8	19%	0.63	0.50	**0.64**	0.74	**5**
P29	ACRU	Intermediate	1.5	Manual	1.8	0.3	7.7	1.5	**7%**	0.63	0.59	0.48	**0.77**	**4**
P21	SPATSIM	Intermediate	1	Manual	1.7	0.1	7.7	1.7	-13%	0.69	0.57	0.13	0.75	**4**
P04	SWAT	Beginner	3	Manual	2.0	1.1	7.1	1.3	38%	0.64	0.64	0.38	0.74	3
P38	MIKE-SHE	Experienced	2	Automated	2.0	0.4	7.5	1.7	21%	0.53	0.41	0.35	0.72	2
P44	ACRU	Intermediate	1	Manual	1.4	0.4	8.7	1.8	-29%	0.60	0.31	0.12	0.60	2
P12	SWAT	Intermediate	0.75	Manual	1.8	0.0	6.6	1.7	**<u>1%</u>**	0.43	0.25	-15.4	0.64	2
P33	SWAT	Intermediate	1	Manual	1.3	0.1	3.4	0.9	-21%	0.37	0.15	-0.38	0.31	1
P03	ACRU	Experienced	0.5	Manual	2.3	0.6	10.9	2.2	40%	0.57	0.12	0.30	0.51	1
P13	SPATSIM	Experienced	2	Manual	5.2	2.3	14.7	3.0	243%	0.62	-5.06	-1.69	-0.87	1
P02	WRSM	Intermediate	1	Manual	2.6	1.6	5.7	1.0	90%	0.40	0.26	-0.05	0.43	0
P09	SWAT	Beginner	2	*None*	2.6	0.0	7.4	2.3	49%	0.40	-0.49	-34.6	0.31	0
P07	ACRU	Intermediate	1	*None*	0.4	0.0	4.8	0.8	-80%	0.23	-0.91	-461	-0.08	0
P25	ACRU	Experienced	0.25	*None*	0.5	0.0	4.7	0.9	-59%	0.05	-1.01	-8.76	-0.14	0
P42	SWAT	Experienced	1	*None*	6.9	0.0	36.2	7.6	251%	0.54	-27.6	-3.25	-3.67	0

There were no obvious relationships between model performance and modelling tool, level of discretisation, or modeller experience level; however, calibration effort did appear to improve performance. The sample size was too small to explore associations with rigor. Models meeting four or more acceptability criteria were built using a variety of tools, and hence structures: SPATSIM, MIKE-SHE, WRSM, and ACRU. Models with no calibration attempts had the poorest performance, while models for which the modellers spent more time and for which automated calibration tools were applied generally had the highest performance (Table 2-6). Poor performance of models with no calibration shows that the use of default parameters, and/or a-priori value selections of experienced users, with no testing or refinement, resulted in models with very poor performance in this case.

Of the models meeting all acceptability criteria, no one model had the best performance across all statistics, each showing differing strengths. Model P27 had the highest NSE, 0.73, and R^2, 0.73, but its PBIAS was relatively high, 25%, showing a reasonable prediction of flow pattern, but frequent overprediction. Of those meeting all criteria, P01 had the lowest PBIAS, 5%, and highest NSE-log, 0.73, better predicting lower flows. P26 had the highest KGE, 0.83, which looks at pattern and magnitude together, but the lowest NSE-log of those accepted, reflecting poorer prediction of low flows. The hydrographs (Figure 2-3) illustrate that none of the models captured the steep drops in wet season flow seen in March 2011 and December 2012. These followed summer peaks due to high ET demand relative to rainfall in these warm months.

Looking at daily streamflow prediction, none of the daily models met more than one of the performance criteria (Table 2-7). Three had bias magnitudes under 25% and two had KGE over 0.5. The best daily R^2 values achieved were 0.3-0.4, NSE ~0.3, NSE-log 0.3-0.4, and KGE 0.4-0.6. All models predicted more frequent flow peak responses to rainfall events than were observed, while underpredicting magnitudes of major summer peaks (Figure 2-4). P26 and P29 had the best overall daily performance, better capturing medium to low daily flows and recessions than the others. These were also among the best performers for monthly prediction. P03 best captured the largest peak flows, giving it the highest daily NSE and R^2 in the set, but it greatly overpredicted medium and low flows, giving it an unacceptably high PBIAS (57%).

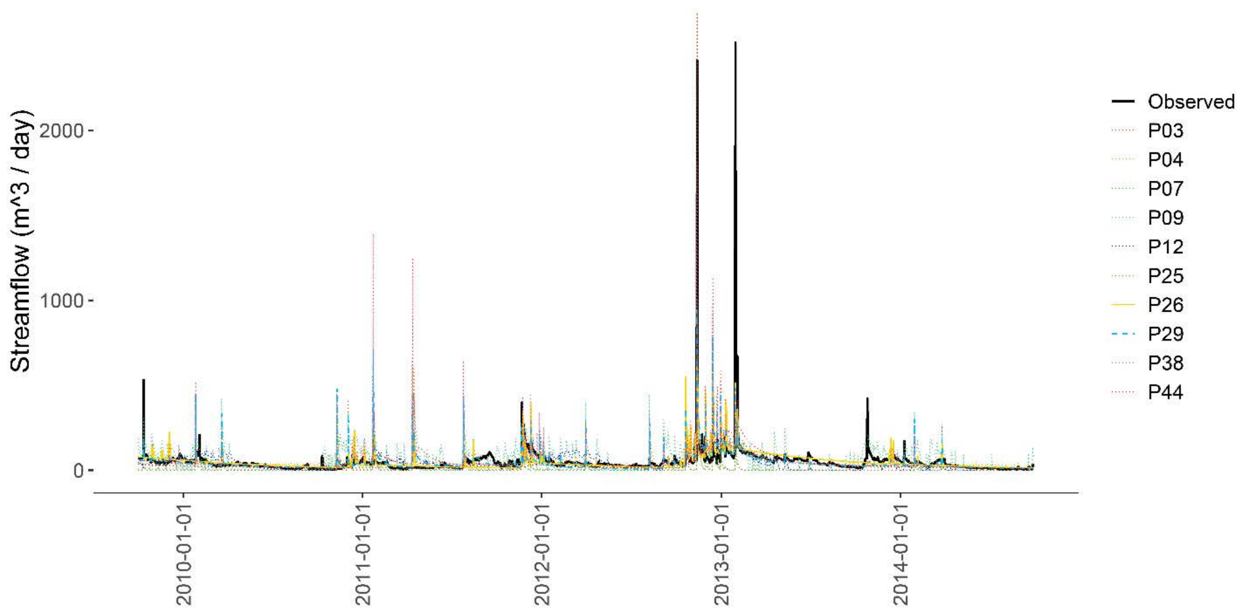


Figure 2-4 Daily observed and modelled hydrograph for all daily timestep models (top) for the calibration period, 2009/10/01-2014/09/30 *(models meeting monthly performance criteria shown with solid lines)*

Table 2-7 Metadata and performance statistics for daily streamflow prediction for daily timestep model submissions *(performance statistics meeting the acceptance criteria are highlighted in green with the three best in the set in bold font and the best underlined; models sorted by combined performance)*

Dataset	Modelling metadata				Daily streamflow (m^3/day) statistics, calibration period				Performance statistics					Number of acceptability criteria met
	Tool	Experience with tool	Days used	Calibration approach	Mean	Min	Max	Std. dev	PBIAS	R^2	NSE	NSE log	KGE	
Observed					**48**	**1.5**	**2515**	**90**						
P29	ACRU	Intermediate	1.5	Manual	43	7.6	943	57	**19%**	**0.30**	**0.29**	**0.34**	**0.42**	1
P26	MIKE-SHE	Intermediate	4	Automated	48	13	611	51	31%	**0.31**	**0.31**	**<u>0.38</u>**	0.38	0
P03	ACRU	Experienced	0.5	Manual	56	12	2693	90	57%	**<u>0.44</u>**	**<u>0.32</u>**	0.19	**<u>0.63</u>**	1
P38	MIKE-SHE	Experienced	2	Automated	48	10	1811	81	32%	0.29	0.15	**0.30**	**0.53**	1
P04	SWAT	Intermediate	3	Manual	49	24	659	41	62%	0.28	0.27	0.22	0.28	0
P44	ACRU	Intermediate	1	Manual	34	9.4	395	45	**-17%**	0.20	0.17	0.13	0.20	1
P12	SWAT	Intermediate	0.75	Manual	44	0.0	184	43	**<u>11%</u>**	0.09	0.06	-130	0.13	1
P07	ACRU	Intermediate	1	*None*	8.7	0.0	1011	46	-81%	0.22	0.03	-759	-0.09	0
P25	ACRU	Experienced	0.25	*None*	12	1.0	947	52	-58%	0.12	-0.10	-8.40	-0.08	0
P09	SWAT	Beginner	2	*None*	64	0.0	899	85	63%	0.16	-0.17	-444	0.32	0

2.5.4 Predicted changes in streamflow with the wattle removal scenario

Using the four models with acceptable baseline streamflow to estimate the impacts of removing the wattle plantation resulted in a wide range of predicted changes (Table 2-8), indicating a high level of structural uncertainty. These models met all five criteria for monthly streamflow prediction for the five-year calibration period. To consider a greater diversity of weather conditions, scenarios were modelled for an eleven-year period (2003-10-01 to 2014-09-30) that included wetter years than the calibration period. Accepted baseline models predicted similar mean annual runoff (MAR) for the calibration period: 23.8-24.7 mm MAR, a range of 0.9 mm or 4% of the mean across the set. Their baseline MAR predictions for the scenario assessment period were more divergent: 77-84 mm MAR, a range of 7 mm or 9% of the mean. When the wattle removal scenario was applied, all four predicted an increase in MAR, however the magnitudes varied greatly, ranging from 33 to 78 mm, or a 39% to 101% increase compared to baseline MAR (Table 2-8, Figure 2-5).

These accepted models all predicted an increase in the variability of monthly flows with wattle removal, but each predicted a different pattern of change across high, medium, and low flows (Table 2-8, Figure 2-6). P26 predicted that the medium flows would have the highest relative (%) increases, while P27 predicted this for the high flows and P01 for low flows. P43 predicted increases in high flows, but decreases in lower flows, and resulting in the lowest predicted increase in MAR in the set. Changes in low flows were assessed using the median of the 91-100% exceedance probability flows. Predicted changes in these values ranged from a 21% *decrease* to a 96% *increase* (-0.2 to 1.1 mm).

If additional models had been considered acceptable, the range in change prediction would have been much larger. For example, if NSE of log-transformed flows had not been included in the performance criteria, models P21 and P29 would have been accepted (Table 2-8) and the range of predicted MAR increases would have been 18 to 78 mm or 17% to 101%. P21 predicted relatively similar changes to others in the set, but P29 predicted a much smaller increase. If all eighteen models were seen as usable, reflecting an ungauged basin case with no other reality checks imposed on model output, predicted changes in MAR would have ranged from -28 to 117 mm, -75% to 291%.

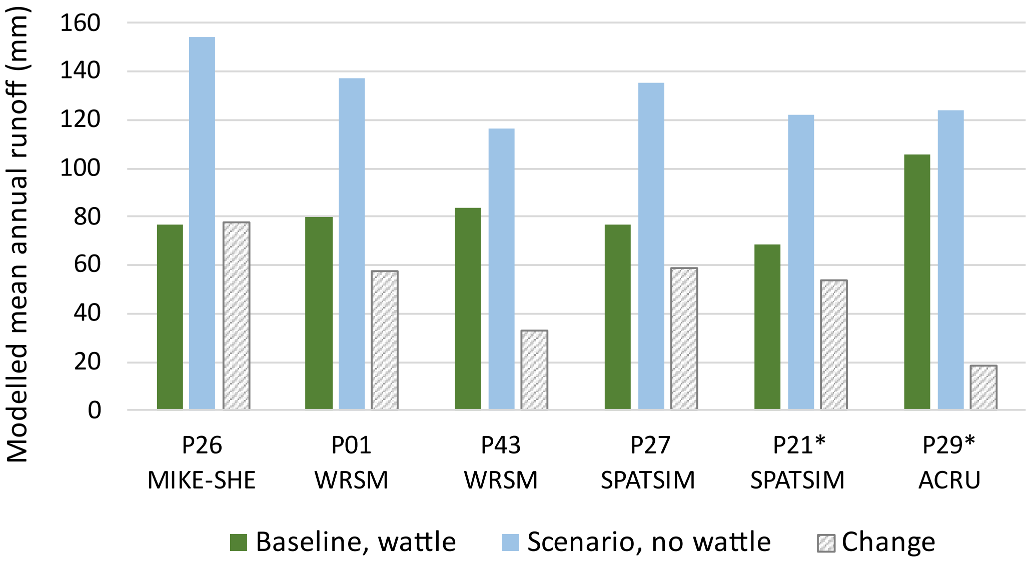


Figure 2-5 Modelled mean annual runoff for both land cover cases for the assessment period (2003-10 to 2014-09) for higher performance models *(* = model met 4/5 performance criteria, others met 5/5)*

Table 2-8 Ranges of modelled changes in mean annual runoff and in high, medium, and low flows with the wattle removal scenario applied to the assessment period (2003-10-01 to 2014-09-30)

Model	Tool	Mean annual runoff, MAR (mm/yr)		Change in MAR		Change in monthly std. deviation		Relative change (%) in monthly extreme flows & mid-points of sections of the flow duration curve *(ranges in exceedance probability (% EP))*						
		Wattle, Baseline	**No wattle, Scenario**	**mm/yr**	**%**	**mm/mon**	**%**	**Max mon flow**	**High flow** *(0-10%)*	**Med-high** *(11-40%)*	**Medium flow** *(41-60%)*	**Med-low** *(61-90%)*	**Low flow** *(91-100%)*	**Min mon flow**
Accepted models														
P26	MIKE-SHE	77	154	78	101%	6	83%	59%	92%	104%	125%	106%	97%	79%
P27	SPATSIM	77	136	59	76%	8	42%	30%	326%	44%	*66%*	*81%*	*122%*	*177%*
P01	WRSM	80	137	57	72%	9	86%	100%	38%	48%	88%	145%	187%	200%
P43	WRSM	84	116	33	39%	8	50%	34%	46%	-9%	-12%	3%	-21%	-28%
Models met 4 of 5 criteria														
P21	SPATSIM	68	122	54	78%	7	41%	34%	175%	82%	99%	123%	154%	188%
P29	ACRU	106	124	18	17%	2	10%	7%	13%	20%	24%	26%	42%	*3%*
All models: mean & range														
Mean (n = 18)		76	110	35	52%	5	73%	79%	81%	38%	40%	39%	57%	71%
Minimum		38	10	-28	-75%	-3	-59%	-67%	-51%	-92%	-97%	-96%	-100%	-28%
Maximum		192	206	117	291%	31	785%	957%	508%	106%	125%	145%	187%	200%

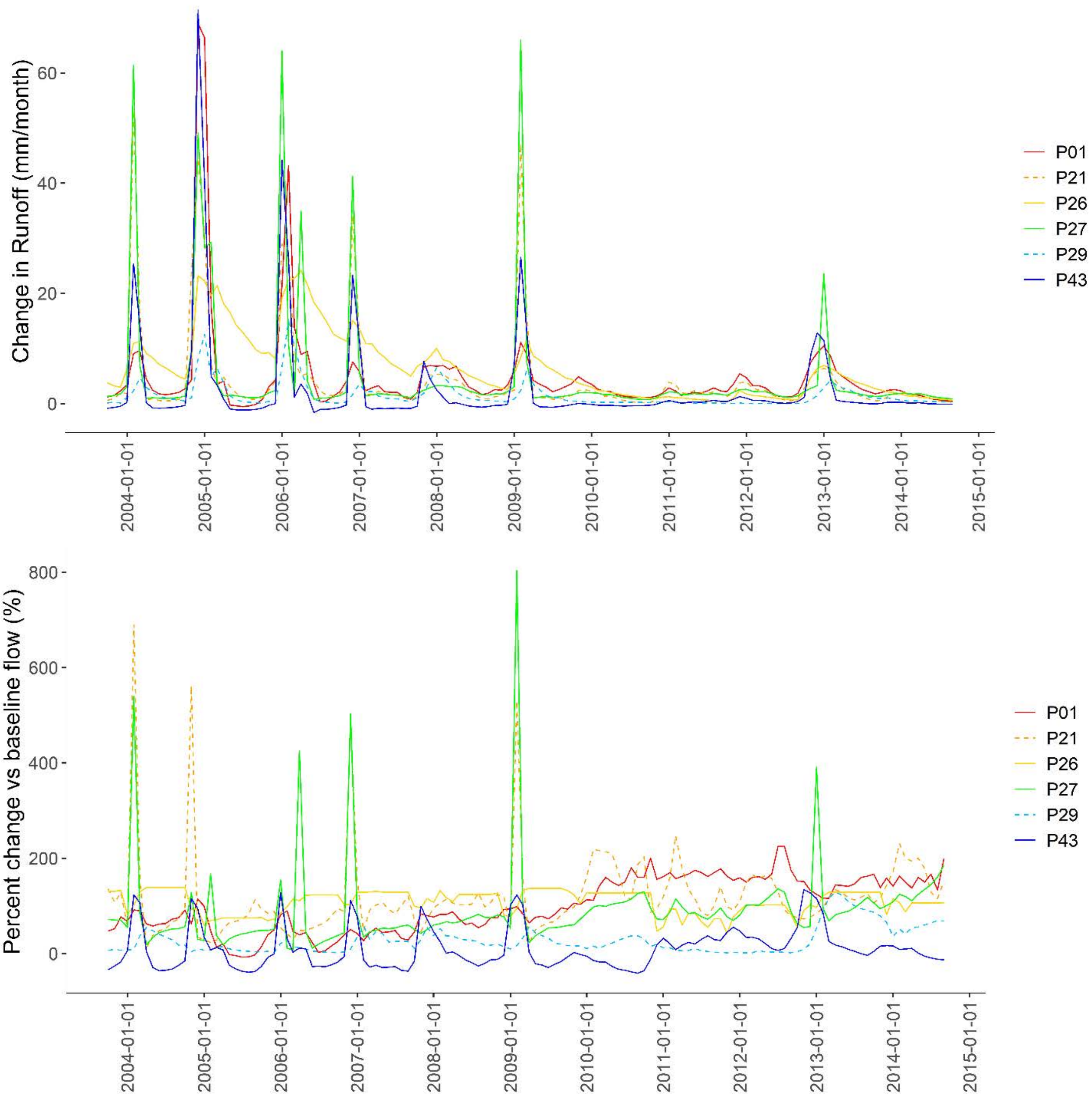


Figure 2-6 Monthly absolute (top) and relative (bottom) predicted changes in runoff due to wattle removal compared to flows with baseline land cover shown for higher performance models *(models meeting 5/5 performance criteria shown with solid lines, those meeting 4/5 shown with dashed lines)*

The daily timestep models with higher monthly flow performance statistics, P26 and P29, predicted very different degrees of change in peak and high daily flows. A key reason for using daily models is to look at peak flows and floods that manifest at daily or sub-daily timescales. None of the models were considered sufficiently behavioural in predicting streamflow at the daily timestep; however, in applied settings, daily models can end up being accepted for use based on their monthly performance. Both models predicted that removing wattles would have bigger relative impacts on medium flows than on the highest daily flows; however, P26 also predicted notable increases in high flows. P26 predicted an 83% increase (an added 454 m^3/day) in the median of the highest flows, the 0-10% exceedance probability flows, while P29 only predicted a 17% increase (173 m^3/day).

Models built using the same modelling software tool did not necessarily predict similar relative magnitudes or patterns of change, despite using mostly the same process algorithms. The two higher performance models built using SPATSIM, P27 and P21, did have similar predictions of the relative change in MAR (76 and 78%) and both predicted the highest monthly flows would have the largest relative increases (Table 2-8, Figure 2-5, Figure 2-6). However, P13, also built with SPATSIM, differed notably, predicting a 52% increase in MAR with medium flows having the largest relative increases. Of the tools used, SPATSIM is the most prescriptive in terms of model structure. These models all used a single subcatchment and so only really differed in parameterisation. The WRSM or ACRU models had somewhat different discretisation and unit connections across models, in addition to differences in parameterisation. The two higher performing WRSM models, P01 and P43, differed markedly in predicted MAR changes (72% vs 39%) and patterns change across high and low flows (Table 2-8, Figure 2-5, Figure 2-6). Looking across the five ACRU models, predicted MAR increases ranged from 5% to 87%, with some predicting that the greatest relative increases would be in the low flows, while others predicted this for the medium and medium-high flows instead. The SWAT models had a wide range in HRU numbers, and the predicted MAR changes ranged from -75% to 291%, with all different patterns of changes across the hydrograph in the set.

2.5.5 Baseline and scenario models' predicted water balances

Looking at mean annual water balances from the models with acceptable streamflow performance, all four predicted relatively similar total ET and runoff with the baseline land cover; however, they varied notably in their predictions of processes contributing to these major net fluxes and how the removal of the wattle plantation would impact them (Figure 2-7). With baseline land cover, all models predicted 91-92% of rainfall would evaporate. WRSM, used for P01 and P43, does not output ET fluxes, restricting the analyses. P26 (MIKE-SHE) and P27 (SPATSIM) predicted similar total ET for very different reasons. P26 predicted only 4% of rainfall would be lost to canopy interception, while P27 predicted 21%, but P27 predicted much less ET from soil and groundwater (566 mm) than P26 (702 mm). All four predicted higher ET from wattle compared to sugarcane and riparian vegetation, and a net decrease in catchment total ET with removal of the wattle. They also all predicted that the ET per unit area from the non-wattle vegetation would be slightly higher in the scenario where the wattle had been removed, due to increased access to water. Predictions of the difference between mean annual ET from wattle and from non-wattle vegetation (with the wattle removed) ranged from 52 to 175 mm. P27 predicted the largest difference and P43 lowest. P27 predicted such a large drop in interception due to wattle removal that the ET from soil was predicted increase as a result (Figure 2-7). The wattle's high interception limited ET from soil and groundwater to a large degree in this model.

Despite similar baseline MAR, the contributions of different runoff generation pathways were very different across the models (Figure 2-7, Table 2-9), explaining their diverging predictions of how the hydrograph would change with wattle removal (Figure 2-6, Table 2-8). P43 predicted that 87% of MAR would come from surface flow, while P26 predicted there would be no surface flow and the 84% of MAR would come from deeper groundwater. Removing the wattle trees was predicted to increase runoff via all pathways in most models, though by varying amounts (Figure 2-7, Table 2-9). P26 predicted most of the increase in streamflow would be due to an increase groundwater outflow. P43 and P27 predicted that increases in surface runoff would dominate. P01 predicted a small *decrease* in surface runoff, with an increase in interflow dominating the change.

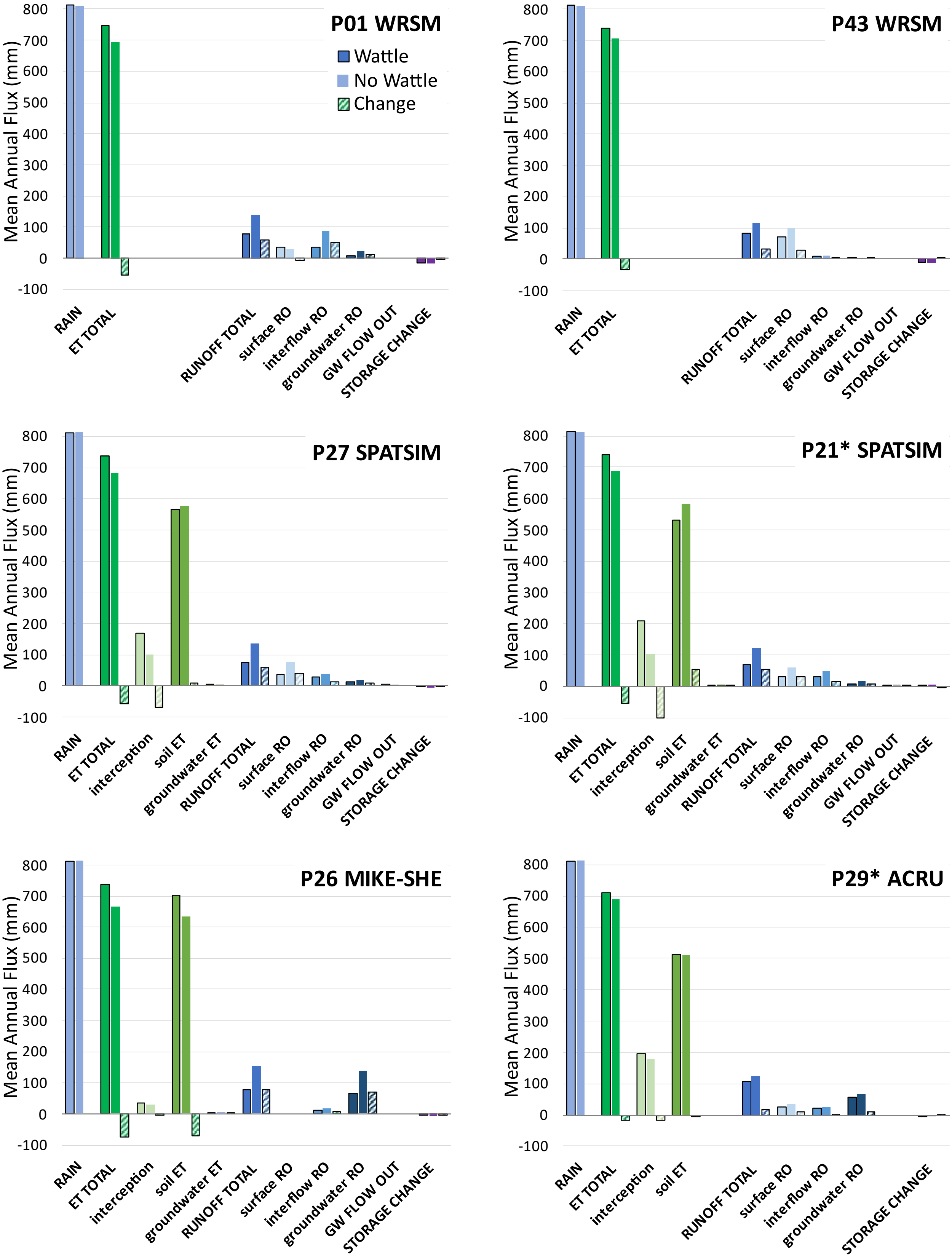


Figure 2-7 Mean annual water balance fluxes with baseline land cover vs the wattle removal scenario as predicted for the assessment period (2003-10 to 2014-09) by the higher performing models *(* = model met 4/5 performance criteria, vs others that met 5/5)*

Table 2-9 Modelled mean annual water balances for 2003-10 to 2014-09 for the baseline and land cover change (wattle removal) scenarios

	P26 MIKE-SHE			**P01 WRSM**			**P43 WRSM**			**P27 SPATSIM**			**P21* SPATSIM**			**P29* ACRU4**		
Flux	**Wattle**	**No Wattle**	***Change***	**Wattle**	**No Wattle**	***Change***	**Wattle**	**No Wattle**	***Change***	**Wattle**	**No Wattle**	***Change***	**Wattle**	**No Wattle**	***Change***	**Wattle**	**No Wattle**	***Change***
Mean annual (mm)																		
Rain	**813**	**813**		**813**	**813**		**813**	**813**		**813**	**813**		**813**	**813**		**813**	**813**	
ET total	**739**	**665**	***-74***	**747**	**694**	***-52***	**741**	**707**	***-33***	**738**	**680**	***-58***	**741**	**688**	***-53***	**709**	**690**	***-19***
Interception evap	35	29	*-5*							170	101	*-68*	208	101	*-107*	196	179	*-17*
ET from soil	702	632	*-70*							566	576	*11*	531	585	*54*	513	512	*-1*
ET from GW	1.7	3.6	*1.9*	6.5	15	*8.2*	0.4	0.1	*-0.3*	2.2	2.2	*0.0*	2.4	2.4	*0.0*			
RUNOFF (RO) TOTAL	**77**	**154**	***78***	**80**	**137**	***57***	**84**	**117**	***33***	**77**	**136**	***59***	**68**	**122**	***54***	**106**	**124**	***18***
Surface RO	0	0	*0*	37	30	*-7*	73	101	*28*	37	76	*39*	30	59	*29*	27	36	*8*
Interflow RO	12	18	*6*	35	87	*52*	9.1	13	*4*	28	40	*12*	31	47	*17*	22	24	*2*
Groundwater RO	65	136	*72*	7.4	20	*12*	1.7	2.6	*1*	12	20	*8*	7.8	16	*8*	56	64	*8*
Groundwater flow out	0	0	*0*	0	0	*0*	0	0	*0*	5.3	5.3	*0.0*	5.1	5.1	*0.0*	0	0	*0*
Storage change	-2.9	-7.0	*-4.1*	-14	-19	*-5.0*	-12	-11	*-0.4*	-2.0	-3.0	*-1.0*	2.8	2.2	*-0.6*	-2.3	-1.8	*0.4*
Flux % of rainfall																		
ET total	**91%**	**82%**		**92%**	**85%**		**92%**	**91%**		**91%**	**84%**		**91%**	**85%**		**87%**	**85%**	
Interception evap	4%	4%								21%	12%		26%	12%		24%	22%	
ET from soil	86%	78%								70%	71%		65%	72%		63%	63%	
ET from GW	0.2%	0.4%		1%	2%		0%	0%		0.3%	0.3%		0.3%	0.3%				
RUNOFF (RO) TOTAL	**9%**	**19%**		**10%**	**17%**		**10%**	**14%**		**9%**	**17%**		**8%**	**15%**		**13%**	**15%**	
Surface RO	0%	0%		5%	4%		9%	12%		5%	9%		4%	7%		3%	4%	
Interflow RO	1%	2%		4%	11%		1%	2%		3%	5%		4%	6%		3%	3%	
Groundwater RO	8%	17%		1%	2%		0%	0%		1%	2%		1%	2%		7%	8%	
Groundwater flow out	0%	0%		0%	0%		0%	0%		1%	1%		1%	1%		0%	0%	
Storage change	-0.4%	-0.9%		-2%	-2%		-1%	-1%		-0.2%	-0.4%		0.3%	0.3%		-0.3%	-0.2%	

With baseline land cover, all accepted models predicted a relatively small net loss of stored water (soil moisture and groundwater) over the period, equivalent to a 2-14 mm loss per annum. P01 and P43 predicted the largest losses (12-14 mm). Most models predicted a slight increase in storage loss when the wattle was removed. This was due to increased infiltration and recharge without the wattle in the relatively wet warm-up period. This resulted in greater storage at the start of the assessment period, that contributed to interflow, groundwater runoff, and ET over the simulation. This was not the case for P43. P43 predicted much more surface flow and less infiltration and recharge.

As with streamflow prediction, the water balance analyses illustrated that models built using the same tool can predict similar water balances, but this is not always the case. P21 and P27, both lumped models built in SPATSIM, had relatively similar water balances and change predictions when compared to the other models (Figure 2-7, Table 2-9). However, P01 and P43, both built with WRSM but with differing modular set-ups, had water balances that differed from one another as much as models built with different tools, despite the fact that they share the same basic process algorithms.

2.5.6 Modelled ET compared to field observations

Comparing modelled fluxes to field observations (Table 2-10), it appears that all four models underrepresented the wattle's access to water and underpredicted ET as a result. Modelled mean annual wattle ET was 32-39% less than the measured rate (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014; Clulow *et al.*, 2022). This appeared to be largely driven by underpredicting transpiration: modelled values were 32-36% less than sapflow measurements (Everson *et al.*, 2014). Canopy interception measurement was compromised by technical difficulties and stemflow (Everson *et al.*, 2014); however, the range of estimates suggests that P26 underpredicted interception while others may have captured it more realistically. Although soil and groundwater storage changes could not be fully quantified from available data, it seems likely that the models also underpredicted storage losses. Given measured soil moisture to 2.4 m depth and groundwater levels (Everson *et al.*, 2014; Ngubo, 2019), a mean annual decline in storage of 25-40 mm for 2009/10 to 2013/09 could be seen as a low estimate, while the models predicted declines of 5 to 19 mm/a. Underprediction of both ET and storage loss suggests that models did not allow the wattle to transpire enough stored water. Looking at the temporal variation of observed and modelled wattle ET, the largest underestimations were in the dry winters (Figure 2-8) when wattle would have likely drawn on deeper water stores.

In contrast, these models may have over-predicted sugarcane ET, although field and model data for comparison is more limited. Wiles 2006 estimated 630 mm of sugarcane ET for 2004/2005. This was a wet year: 1215 mm of rain, 1.4 times the mean. For this year, P26 predicted 715 mm ET from sugarcane. P01, P27, and P43, built with WRSM and SPATSIM, did not output ET for the non-wattle cover types separately. Sugarcane made up 78% of non-wattle area and 22% was riparian vegetation. P26 predicted 593 mm riparian vegetation ET, resulting in 688 mm spatially averaged ET for the non-wattle area. P01, P27, and P43 predicted 698-766 mm for this area. If they had the same riparian ET as P26, their sugarcane ET would be 728-815 mm. If their riparian ET equalled potential reference grass ET (1062 mm), their sugarcane ET would be 596-683 mm, with only P27 having a value below Wiles' 630 mm. If wattle ET was underpredicted and sugarcane ET was overpredicted, the predicted changes in ET with wattle removal may be too low. This would suggest the predicted streamflow increases may also be underestimated by the models, with the magnitude of underestimation depending on where the additional wattle ET is drawn from relative to runoff generation.

Table 2-10 Observed and modelled hydrological fluxes in the Two Streams catchment

Observation	Period	Method	Value	Model output for matching period					
				P26 **MIKE-SHE**	**P01** **WRSM**	**P43** **WRSM**	**P27** **SPAT-SIM**	**P21*** **SPAT-SIM**	**P29*** **ACRU 4**
<u>Sugarcane</u>									
Total ET									
Mean annual (mm)	2004/10-2005/09 (1 yr)	scintillometry *(Wiles 2006)*	**630**	715	755*	766*	698*	680*	747
***<u>A. mearnsii</u>* <u>(wattle) plantation</u>**									
Total ET									
Mean annual (mm)	2009/10-2013/09 (4 yr)	scintillometry & EC flux tower, *(Clulow et al., 2011, Everson et al., 2014)*	**1130**	688	700	695	773	795	685
Interception evaporation									
% of rainfall	2008/04-2011/03 (3 yr)	above canopy rain gauge vs troughs below	**28%**[#]	5%			28%	37%	30%
Mean annual (mm)			***185***[#]	35			189	247	202
% of rainfall	2012/10-2013/09 (1 yr)	total ET (EC) – transpiration (sapflow) *(Everson et al., 2014)*	***10%***[^]	*5%*			*28%*	*37%*	*30%*
Mean annual (mm)			**81**[^]	42			217	284	229
Transpiration									
Mean annual (mm)	2012/10-2013/09 (1 yr)	heat pulse sapflow *(Everson et al., 2014)*	**1123**	749			697	649	560
<u>Storage change (soil & groundwater)</u>									
Mean annual (mm), soil	2009/10-2013/09 (4 yr)	soil moisture probes (2.4 m)	**-25**	-5	-19	-16	-11	-3	-7
GW store		borehole loggers *(Everson et al., 2014, Ngubo 2019)*	**-2 - -15**[+]						
<u>Streamflow composition (runoff pathways)</u>									
% direct runoff event water [@]	2012/10-2014/09 (2 yr)	stable isotope sampling *(Ngubo 2019, 2022)*	**23%**	0% 19%	5% 86%	75% 96%	2% 79%	0% 85%	19% 35%
% delayed flow [@]			**77%**	81% 100%	14% 95%	4% 25%	21% 98%	15% 100%	65% 81%

* *value for non-wattle cover: sugar & riparian*
[#] *likely an upper bound: method not account for stemflow (e.g. 2013 stemflow = ~39-50% of rainfall)*
[^] *likely an upper bound: includes soil water evaporation (total ET – transpiration = interception + soil evap)*
[+] *range due to differing aquifer specific storage estimates*
[@] ***direct runoff*** *portion compared to modelled* ***surface runoff and to surface runoff + interflow;***
delayed *portion compared to modelled* ***groundwater runoff and to groundwater runoff + interflow***

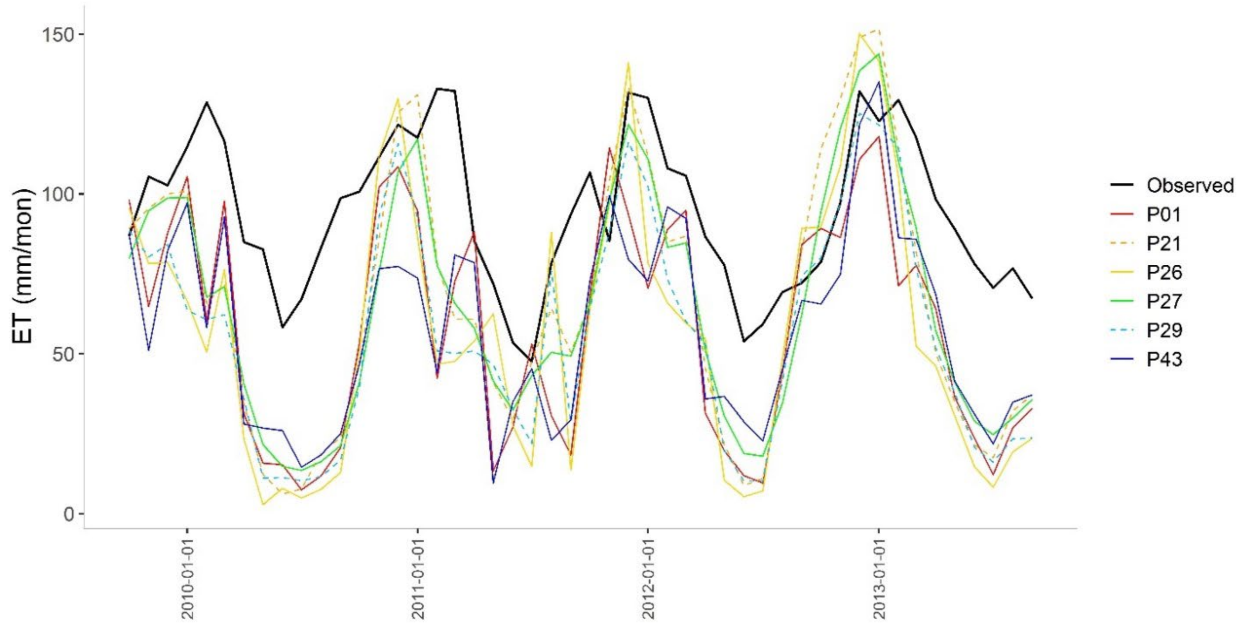


Figure 2-8 Monthly observed (Clulow et al., 2022) and modelled ET from the wattle plantation in Two Streams *(models meeting 5/5 performance criteria shown with solid lines, those meeting 4/5 shown with dashed lines)*

These errors in ET prediction could be seen as grounds for rejecting all the models for use in this specific case. The reason for modelling the catchment was to determine the likely impacts of removing the wattle plantation and replacing it with sugarcane. It is not common to have this level of ET observation to compare models to. Modellers may have built and/or parameterised their models differently if they had been given this observed ET data. However, they were provided with information about the wattle's roots being observed down to at least 8 m and the deep clayey unsaturated zone that would have a reasonably high water holding capacity. It may be the case that structural and/or parameter value limitations imposed, or suggested, by the modelling software tools meant that these properties were not adequately represented in the models despite being known. For example, WRSM, SPATSIM, and ACRU all make provision for vegetation to access groundwater stores when located in the riparian zone specifically. In this case the wattle was not in the riparian zone, however appeared to be accessing water stored at depth (Everson *et al.*, 2014; Watson, 2015).

2.5.7 Modelled streamflow generation pathways compared to field observations

With the exception of P43, the models with acceptable streamflow prediction performance also represented the observed dominance of subsurface flows in contributing to streamflow (Table 2-10). Stable isotope studies indicated that mean annual streamflow is dominated by subsurface flow contributions (Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022), likely with significant amounts from deeper groundwater (Everson *et al.*, 2014; Watson, 2015). Soil water signatures showed overlap with rain and groundwater at times, complicating separations (Everson *et al.*, 2014; Watson, 2015). As such the estimated 22% contribution from direct runoff of event water to total runoff for 2012/10-2014/09, based on Ngubo 2022, may be an upper bound. Because 'direct runoff' can include preferential interflow, and because the conceptualisation of interflow varies notably across models (Glenday *et al.*, 2022), both modelled surface flow and surface plus interflow were compared to this direct runoff estimate and both modelled groundwater flow and interflow plus groundwater flow were compared to the delayed flow estimate. Only P26 predicted strongly groundwater dominated runoff

(81%), but P27 and P01 predicted subsurface flow dominance, looking at interflow and groundwater flow. P43 predicted that runoff was dominated by surface flows, which appears to be unrealistic.

If P43 is not considered to be a reasonable representation of Two Streams because of its surface flow dominance, the range of potentially realistic change predictions for the wattle removal would be greatly reduced. P43 predicted a much smaller change in flow than the other three models that had acceptable streamflow prediction performance. If P43 is excluded, the range of predicted increases in MAR with wattle removal narrows considerably, going from 33-78 mm, or a 39-101% increase, to 57-78 mm, or a 72-101% increase (Figure 2-5). This is a notable reduction in the prediction uncertainty. It is not typical to have isotope tracer studies to determine the flow pathways contributing to streamflow in a catchment. However, other approaches like hydrograph baseflow separation using numerical filters or visual ones, field observation of the presence/absence of surface flow, and/or reasoning based on terrain slope, cover, slope and rainfall intensity can be used to determine if a catchment is likely to have significant contributions from surface runoff.

2.6 CONCLUSIONS AND LESSONS LEARNED

The model-a-thon results showed that the uncertainty in modelled predictions due to the differences how models are structured and parametrised by different people can be very high, but this prediction uncertainty can be reduced with greater use of observational data. In this case study, if all eighteen model submissions were considered potentially reasonable representations of the Two Streams catchment, the predicted changes in mean annual runoff (MAR) with the removal of the wattle plantation ranged from -75% to +291%. If only the models with acceptable streamflow performance were considered, the range of predictions narrowed considerably, but may still be considered large in a management context. The decision about what constitutes ‘acceptable performance’ also had a big impact on the prediction range. If the seven models with satisfactory monthly NSE (≥0.6) are included, predicted MAR increases ranged from 17% to 101%. If only the four models with acceptable performance across all five streamflow prediction criteria are used, the prediction range narrows to 39% to 101%. If information about the catchment’s dominant flow paths was considered, only three models were considered acceptable representations of the system and the prediction range narrowed further to 72% to 101%. These three models ranged from lumped to gridded and were built using different tools: WRSM, SPATSIM, and MIKE-SHE. Uncertainty estimates in this study did not include consideration of uncertainties in input or calibration data, nor the variability in how different users would process raw data to prepare model inputs.

Overall, the model-a-thon succeeded in demonstrating aspects of modelling uncertainty in a way that engaged the modelling community. The activity brought together modellers across a variety of experience levels and using a range of modelling tools. The results highlighted a number of important issues for modelling practice and use of modelling outputs in decision-making. There were a number of limitations to the activity and the uncertainty study which could be addressed in future research, potentially including additional model-a-thons. Key take-home messages and suggestions are summarised below.

2.6.1 Take-home messages from the modelling outcomes

- Even for a relatively simple catchment, for which the input data was pre-prepared and assumed to be certain, the differing approaches used by different modellers can result in very different model predictions.

- If only one model is applied to a use case, the model structural uncertainty in the predictions will go unrecognised and unaccounted for and this uncertainty can be very high. Testing and applying more than one model structure to consider the range of predictions is likely to be worth the added effort, particularly when predictions will inform high-risk or high-cost decisions. This would be an additional step to considering data and parameter uncertainty.

- Multiple model structures can have satisfactory streamflow prediction performance, but at the same time they are predicting different combinations of contributing hydrological fluxes. As a result, models predicting similar streamflow for a baseline case can predict very different responses when a change scenario is imposed. This is more likely to be the case when the observed streamflow record that was used to assess model performance is short relative to weather variability, and/or when fewer performance metrics are used to determine if a model is an acceptable representation of the catchment.

- Having additional information and observational data regarding different hydrological fluxes in a catchment, such as data on ET and contributing fluxes, runoff generation mechanisms, soil and groundwater storage, can be used to build more realistic models from the outset and identify models that are not realistic representations of the system, even if they managed to produce reasonable streamflow output. Doing this can greatly reduce model prediction uncertainty.

- Model performance in this case was not associated with the modelling tool used or the level discretisation, but did appear to be associated with the effort spent on calibration. This supports the development of tools that assist modellers in exploring the reasonable parameter value space efficiently. This also supports the allocation of adequate time for model building and testing in applied settings.

- Use of default parameter values or expert a-priori parameter values estimates, without any additional evaluation of the model's outcomes, produced models with very poor streamflow prediction performance. This highlights the need to identify and apply ways of reality-checking models of ungauged catchments.

- In this case study, all the models that achieved acceptable streamflow performance underpredicted ET from the wattle plantation by not allowing the vegetation sufficient access to stored water. This is particularly problematic when the goal was to model the impacts of vegetation cover change. This further supports the need for more field measurements of ET and local calibration of remote-sensing-based ET models to expand the availability of reliable data. It also suggests a potential need to evaluate how the commonly used modelling tools that were used here can represent catchment water storage and which water storages vegetation can access for transpiration in the models compared to field observations.

2.6.2 Limitations of the study and suggestions for further exploration of modelling uncertainty in practice

This study constitutes a very limited exploration of model uncertainty in practice, in part due to the volunteer, crowd-source approach of holding a model-a-thon. However, this approach served other important goals. It served to broaden interest, awareness, and discussions about uncertainty amongst the modelling community and brought modellers from different institutions into contact. Also, a main aim of the study was to see how different individuals would model the same area. As such it would not have helped for the project team to build a variety of models, as these may not represent what modellers across the community do in practice. An alternative could have been to pay a set of individuals to do the modelling, however this would notably limit the number of modellers involved given the project budget, defeating the purpose to some degree.

In an effort to get a useful number of models using the volunteering, participatory, model-a-thon approach, the exercise was kept very simple to reduce the time needed to complete it. This limited the aspects of uncertainty explored. Using this model-a-thon approach also meant that there was not control over the number of models submitted in total or across different tools and experience levels. Given the limitations of the case study and the submissions received, it is possible that the resulting prediction uncertainty from acceptable models is an underestimate; there may be other model structures and parameterisations that would also meet the calibration criteria, but diverge even further in change impact predictions.

To inform further research into model prediction uncertainty, key limitations in the uncertainty analysis from this activity, and suggestions for overcoming these, are summarised below:

- **Single case study:** Without conducting a similar exercise using additional case study catchments, one cannot conclude that the degree of prediction uncertainty observed in this case is widely representative. It illustrates a potential level. Conducting similar experiments for other catchments with different properties (climate, geomorphology, soil, cover, management, data availability, etc.) would be of benefit.

- **Simple land cover and management:** The case study was very simple in that there were only three land cover types to consider and no water management activities like dams, irrigation, transfers, etc. to represent. The diversity across model structures (and structure linkages to the tool being used), their performance, and their scenario predictions may be even greater for complex cases than observed here. This could be explored with additional case studies with good management data available and modellers willing to spend a bit of extra time to incorporate this in model construction.

- **Small catchment:** The small size of the Two Streams catchment (~0.7 km^2) was a recognised limitation. Some of the modelling tools commonly used in South Africa were developed for, and tested with, larger areas, particularly those with monthly timestep algorithms based on the Pitman model (WRSM, SPATSIM). In this activity modellers were permitted to use any tool they wished and still chose to use these models and achieved reasonable performance. However, it would be worth doing similar uncertainty studies with larger catchments.

- **Data uncertainty:** Uncertainties in the input climate data and calibration streamflow data, due to instrument inaccuracies, technical problems, user error, placement, etc., were not considered in this study. It would likely be difficult to address this in a model-a-thon setting, however estimated uncertainties in each of these datasets could be used to create stochastic sets to determine the impact this has on model prediction individually and in combination.

- **Uncertainty due to input data selection and preparation:** Over and above the uncertainties in weather station and gauge data, there are also uncertainties introduced by how individual modellers decide to use these data. In this case modellers were given pre-prepared catchment-scale climate input data. Different modellers may make different assumptions about the spatial distribution of climate variables that impact predictions, particularly when there are large spatial gradients across a catchment. Similarly different modellers may choose to use different land cover and soil maps from different sources and simplify these in different ways for model input. Many of these choices were already made in this exercise. Impacts of these choices on model predictions were not explored here to simplify the activity and to put focus on the impact of model structure on prediction uncertainty. To look at the impact of input preparation separately from model structure, one could use sets of inputs generated by different modellers to drive a fixed set of model structures. In practice, the way these inputs are prepared can be tied to the model structure and options in a modelling tool. It would be useful to determine the degree to which this is the case and the impacts it has.

- **Streamflow dataset for calibration:** The dataset available for model calibration in this case was relatively short, five years, and covered a period that was drier than the longer-term average. This is not uncommon in applied settings. Demonstrating the impact of longer calibration datasets on modelling uncertainty could garner further support for monitoring networks, however it should be noted that this has been investigated by several prior studies that have found that the degrees of improvement with data length are case specific, however longer datasets generally improve performance and to a greater degree for more variable and drier catchments (Brath, Montanari and Toth, 2004; Li *et al.*, 2010; Arsenault, Brissette and Martel, 2018; Shen, Tolson and Mai, 2022). An additional concern in this case was potential errors in the streamflow dataset which would have a greater impact given its shorter length with fewer large flow events. Effort was made to quality check the data here, however it is possible some of the peak flow readings were inaccurate leading to the poor performance of all the daily models when evaluated against this data. In future studies, longer and more highly vetted streamflow datasets would be beneficial.

- **Calibration approaches and uncertainty sources & quantification:** In this study modellers were permitted to calibrate their models using an approach of their choosing. This was done to explore the variability in this choice across modellers, however this limits the degree to which different sources of uncertainty are explored and fully quantified. To systematically look at parameter uncertainty and structural uncertainty separately, all models would need to be subject to the same calibration routine. The parameter sets providing acceptable performance for a given structure could be used to look at prediction uncertainty due to parameter uncertainty. The calibration approaches chosen also serve as further source of modelling uncertainty. Model structures with acceptable parameterisations could then be used to look at prediction uncertainty due to structural uncertainty. In the current study, model structures that were submitted may have been able to perform better than they did with further exploration of the parameter space. As such parameter uncertainty wasn't systematically explored here. The modelling tools commonly used in South Africa make it impractical to ask model-a-thon participants to calibrate their models in a standardised way and it would also be challenging for a research team to implement this.

- **Ungauged case and levels of process data provision:** As many catchments of interest are ungauged, it would be useful to explore prediction uncertainty due to different modeller's approaches for a case in which no streamflow data was provided. This was somewhat explored here given that some modellers did not attempt to calibrate their models, but only a few did this, representing a limited selection of tools and structures. A future study could be done with different rounds of a model-a-thon in which participants model without streamflow data provided and submit their models and then are given streamflow data with which to calibrate. It would also be interesting to add another round in which other process data, such as the ET observations available for this case study, are additionally provided. This would require more time from participating modellers.

- **Low number of models with acceptable performance:** Only four of the submitted models had acceptable monthly streamflow performance based on the five criteria selected, despite the thresholds for acceptance being fairly low (i.e. 'satisfactory', rather than 'good'). Only three of these were accepted when considering runoff generation mechanisms, and all could have been rejected based on ET prediction. It is likely that there are other model structures built with the same tools as those submitted that would be able to achieve similar or better performance, perhaps even some of the structures that were submitted, given further calibration effort. It would be useful to know if additional acceptable model structures would increase the range of prediction outputs further, or if additional predictions would have fallen inside the range covered by these. Obtaining more acceptable models would likely mean attracting more model-a-thon participants.

- **Representative of practice in applied modelling settings:** This study was intended to explore modelling practice across individual modellers similar to how they would approach modelling in real applied settings for research and/or consulting. As already described, some of the decisions they would be making in applied settings had been controlled already in this case, however it is possible that the effort and amount of time spent on this activity for some participants does not represent what they would ordinarily do. Some participants expressed this directly in the submission survey, noting they had to do a rushed job due to limited availability. It is possible that similar effort levels may happen in quick turnaround consulting settings, but the degree to which this is true was not actually explored here. In addition, most participants were from academia with few people in the private sector completing the activity. A way to overcome this would be to run a modelling activity as a real consulting project, but having multiple firms all complete it as they would do other jobs. The City of Copenhagen did this for a groundwater modelling project and found highly divergent approaches and outputs across the firms (Refsgaard *et al.*, 2006).

2.6.3 Additional suggestions for further model-a-thon activities

There are several reasons to hold additional model-a-thon activities in the future. Such activities could serve to continue bringing the modelling community of practice together, to motivate modellers to hone their skills and approaches, to deepen the appraisal of modelling of uncertainty, and/or to potentially gauge changes in modelling practice across the community over time, if enough people participate on a regular basis. As described in the section above, it would be difficult overcome all the limitations of the current study in terms of uncertainty analyses using this type of model-a-thon approach. Some questions would need a more controlled type of study. However, it would be useful to apply this approach to further cases and it may be possible to get more participants than this activity achieved. All of the participants in this model-a-thon indicated in the submission survey that they would participate in another model-a-thon in future. Several who signed up to this event but did not submit, contacted the team to express that they would have liked to complete it and were limited by time constraints, rather than interest or other barriers.

To assist potential future model-a-thons, some suggestions have been compiled based on the experience gained during this activity. Some suggestions stem from participants who gave input in the submission survey and the feedback session. These should be viewed in addition to relevant points already made in section 2.5.4 above (e.g. using case studies with different properties and management issues, larger catchments, catchments with longer streamflow records, etc.). A main goal would be to get more participants and ideally several using each of the main modelling tools.

- **Timing of the activity:** The model-a-thon was conducted at the end of the calendar year, initially intended to be completed October to November 2022, but extended through January 2023. This was done to kick off the activity at the SAHS conference; however, some participants indicated that this was not an ideal time of year given both the end of the academic year and projects that may be closing at the end of the calendar year. Other timings should be considered in future.

- **Duration of the submission period:** In the end the activity ran for over three months, due to the multiple extensions granted to get more submissions. The initial plan was to allow participants five weeks from the activity release to submit their models, but this appeared to be too little for many people to find sufficient time to work on this. This may have been due to the specific time of year, however future activities may need to consider this.

- **Activity complexity and time needed to complete:** Participants took anywhere from a few hours to four days to complete this activity. Most took about a day and for the most part the models prepared in less time were not calibrated and had poor performance. As described, this example was kept as simple as possible and many inputs were pre-prepared. More complex case studies will likely require more volunteered time, hence a longer submission window and perhaps more incentives to participate.

- **Ring-fenced work time at kick-off:** It would likely assist participants to complete the activity if they have a half-day or full day set aside to start work on their modelling when the event kicks off. This would ideally be done in person, potentially as a side event to a conference or other gathering. However, this would incur both venue costs and potentially travel and accommodation for participants. It could also be done at university departments, research institutions, or consultancies on their own premises. Another option would be to try to get participants to set aside specific times in their calendars well in advance to work on the activity on their own. If this was successful for a sufficient number of participants, it may be work to limit the submission period rather than extending it.

- **Linking with courses and institutions:** A way to get more beginner modellers would be to have people attending modelling courses, either as part of a university course or an externally run short course, to do the model-a-thon activity as part of their learning. This would require consultation with those running the courses and flexibility around the timing over which the activity was run to be able to work with courses on different tools that occur at different times of year. A way to attract more intermediate and advanced modellers would be to have model-a-thon activities specifically adopted by and run at university departments, research institutions, and consultancies, ideally with a group kick-off and a day or half day set aside for participants to start work in person on the premises.

- **Targeted contacts and follow-ups:** Some more experienced modellers were specifically contacted to participate in this activity. All participants were sent reminders about submission due dates and extensions with some individual follow ups for those who said they were still intending to submit. This likely assisted with getting the number of submissions finally obtained.

- **Incentives:** The current activity earned participants SACNASP CPD points, recognition, prizes for students (Takealot gift vouchers), and the opportunity to be co-authors on a research article (conditional on engagement in drafting). While the activity did garner good interest with 43 sign-ups, there were few consultants participating and the rate of attrition was high. Additional incentives, such as additional or bigger prizes, could assist in getting more submissions.

- **Participation and advertisement beyond South Africa:** One way to get greater numbers of participants would be to advertise the activity internationally – regionally or globally. This would change the meaning of the results, as it would no longer be focused on the modelling community in South Africa, but it could still be focused on South African modelling questions and levels of data availability. This would require a greater degree of event support and would result in submissions using models the team running the event may not know how to work with. The activity could be constrained to a certain suite of modelling tools.

- **Additional 'market' research:** To keep the activity as short and simple as possible to promote completion, the submission survey given to participants was kept brief and largely focused on the participant's modelling process. Participants were given the opportunity to make suggestions regarding the activity, but were not prompted for specifics. As such, further targeted surveys in the modelling community could be used to inform things like timing, setting, and incentives when planning further activities.

CHAPTER 3. SYNTHESIS POLICY BRIEF

3.1 GOALS OF THE ENGAGEMENT AND POLICY BRIEF

Quantifying and reducing uncertainty in catchment hydrological modelling has been the subject of a significant body of research globally, with notable studies having been conducted locally focusing on South African models and settings (Sawunyama and Hughes, 2007; Hughes, Sawunyama and Kapangaziwiri, 2008; Hughes, Kapangaziwiri and Sawunyama, 2010; Hughes *et al.*, 2011; Sawunyama, Hughes and Mallory, 2011; Kapangaziwiri, Hughes and Wagener, 2012; Glenday *et al.*, 2022). However, lessons learned from this work do not appear to have been mainstreamed into modelling practice in South Africa's water sector. Reasons for this could include awareness, capacity, technical, and other practical barriers to implementation, as well as motivation in relation to competing issues and constraints.

As a step towards addressing this, this project synthesised uncertainty research findings with the help of water sector professionals to develop a policy brief on modelling uncertainty that accounts for the practical hurdles faced in the field. It is intended that this brief can be used to promote and gather support for the activities needed to consider and to reduce model prediction uncertainty in the sector.

The goals of the engagement around the policy brief and of the brief itself were to:

- Hear from a variety of stakeholders across the modelling process, those building and running models as well as those using the outputs in operational settings and decision making.
- Understand stakeholders' views, challenges, and levels of engagement regarding uncertainty in hydrological modelling. (Is it seen as an important problem?)
- Consolidate ideas around what is needed to enable uncertainty analyses and consideration of uncertainty in decision making.
- Consolidate elements of 'good' practice for handling uncertainty in terms of quantifying it, reducing it where possible, communicating about it and accounting for it explicitly when making decisions.

3.2 ENGAGEMENT APPROACH

The primary engagement informing the policy brief took the form of two online facilitated discussion sessions. A decision was made to conduct more targeted online engagements rather than one large workshop, with the hope that this would make it easier to attend and allow greater depth of engagement by working with smaller groups at a time. The aim was to have 10 to 15 participants per session with break-away sessions having a maximum of 5 participants to talk through sets of discussion questions. Invitations were sent to consultants, researchers, and academics who specialise in hydrological modelling, as well as parties in the Department of Water and Sanitation (DWS) and Catchment Management Agencies (CMAs) who routinely use catchment modelling output for decision-making, commission modelling studies, and/or run models in-house in operational contexts.

The discussion sessions were held on the 10th and 12th July 2023. An initial round of invitations was sent out in April 2023. Follow-up invitations and reminders were sent in June 2023, inviting additional parties to try to secure at least ten participants for each session. A total of 37 people received invitations to the sessions, with 20 confirming attendance.

The sessions were comprised of:

- Welcome and introductions across all participants, including capturing participant's initial impressions of modelling uncertainty.

- A **background knowledge-sharing** presentation summarising local research findings on uncertainty in hydrological modelling (see *APPENDIX B.1 – Presentation on uncertainty in hydrological modelling for policy brief discussion sessions*)

- **Small group discussion on the implications** of modelling uncertainty for water management, followed by feedback to the full group, using the following questions as guides:
    - *What are some practical implications of these levels of uncertainty?*
    - *Does this concern you? Does it matter in practice?*
    - *Can you think of examples from your own experience where you have had to deal with model output uncertainty? Or where it would have helped to do so?*
    - *Do you think it would be valuable for DWS, and others in the water sector, to allocate resources to include uncertainty assessment in hydrological analyses?*

- **Small group discussion on recommendations** for accounting for modelling uncertainty in water management, followed by feedback to the full group, using the following questions as guides:
    - *How should we be approaching quantifying & reducing uncertainty?*
    - *How should we be approaching communicating uncertainty & considering it in decision-making?*
    - *What is needed to achieve this?*

- A brief recap of participant impressions and discussion on the **way forward for drafting the policy brief**

Sessions were recorded and the project team synthesised the points made and ideas shared in these discussions, alongside research findings on hydrological modelling uncertainty, into a draft policy brief. This included coming up with statements that identify what the group understands to be most salient issues regarding modelling uncertainty and listing potential and preferred responses. Drafting entailed several rounds of editing and discussions of format and layout in order to clarify and express key ideas in an approachable manner and length for a policy brief format.

Participants in the discussion sessions were sent a draft of the brief in October 2023 and given the opportunity to provide feedback and decide if they would like to be acknowledged by name on the brief itself. It is anticipated that the finalised policy brief will be made available online by December 2023, publicised and distributed via SAHS and the WRC.

3.3 DISCUSSION SESSIONS SUMMARY

A more detailed account of the discussion sessions can be found in *APPENDIX B.2 – Consolidated notes from the policy brief discussion sessions*, including attendees, key points raised across major topics, and a full list of recommendation ideas. An abridged summary is provided here. As demonstrated by the word-clouds in Figure 3-1 below, participants perceive modelling uncertainty as complex and challenging issue to engage with, however they left the discussions with a sense that progress could be made in this area through ongoing collaboration.

A total of twelve people attended the discussion sessions alongside project team members. This group comprised two private consultants from different firms; three academics from different universities; and seven people working at DWS across a range of departments, from supply planning to flood and dam safety, and a range of roles and experience levels, from chief engineers to scientific managers and modellers to early career candidate engineers. While the total number of attendees was less than we had targeted, as several parties who had confirmed were unable to attend last minute, the group present represented a diversity of key stakeholders and there was fruitful, in-depth discussion.

There was general acknowledgment of the potential serious implications of not quantifying uncertainty, and communicating this, in the event that single model outputs provided are inaccurate. Attendees noted that this could lead to inappropriate management of water supply systems, allocations, infrastructure design, and flood risk reduction, with the potential to result in avoidable harm to lives, livelihoods, and infrastructure. The potential to result in inappropriate land cover management decisions was raised by the project team rather than the other attendees. However, issues of uncertainties in model representation of land cover impacts was raised by attendees in the context of 'naturalising' models of catchments which have numerous anthropogenic alterations.

Starting impressions of modelling uncertainty **Reflections on the discussion session**

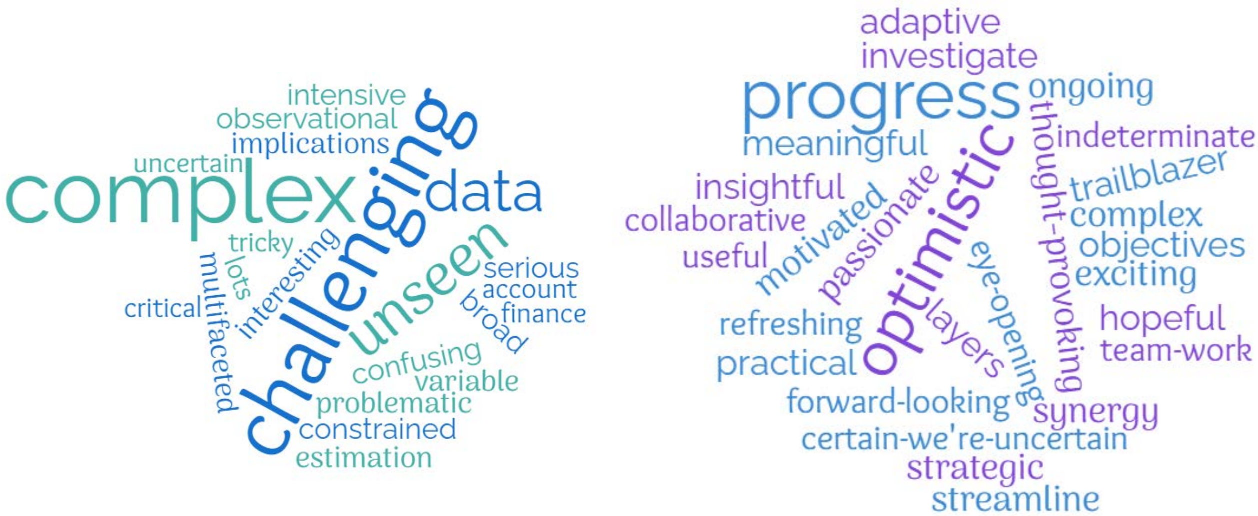


Figure 3-1 Word-clouds of all the one-word starting impression of modelling uncertainty (left) and one-word closing reflection on the discussion session (right) provided by the discussion participants (font size reflects the number of mentions)

It was noted that there are well-accepted and commonly applied practices to account for some elements of uncertainty in water resources modelling and flood modelling in South Africa. This is predominately through stochastic modelling of supply system yields and flood probabilities based on long-term streamflow timeseries. This aims to account for the inherent variability in weather and hence runoff, when considering future streamflow. However, while the variability in rainfall over time is a large contributor, there are other major sources of added uncertainty that the current approach does not account for. Hydrological modelling uncertainty is one of these. Single hydrological models (single tool and structure, single parameter set, single input climate dataset) are typically used, with historical climate data, to generate the long-term streamflow timeseries that are then used to determine the variability statistics for the stochastic yield analyses. In addition to the unquantified uncertainty in the modelling of the past (due to data, parameters, structure), this approach also does not account for future climate changes, land cover changes, and water management and use changes. Fortunately, the existing stochastic technique does mean that many in the water sector are familiar with calculating and communicating uncertainty probabilistically and with decision-making that accounts for uncertainty, at least in some settings. Unfortunately, the existing approach is often perceived to be a sufficient consideration of uncertainty overall, although this has been shown not to be the case (Hughes *et al.*, 2011; Hughes, Mohobane and Mallory, 2015).

Participants noted several barriers to implementing uncertainty assessments in their hydrological modelling work. These included not having time, resources, and capacity to do the additional work and not being requested by clients or end-users to do it, lack of automated routines to make analyses efficient (e.g. parameter testing, sensitivity analyses, and calibration, water balance assessments, etc.) within frequently used software tools, lack of experience or stakeholder trust in other tools, data and data access limitations. There was also concern about communicating uncertainty in modelling outputs to stakeholders, depending on the audience, for fear of them losing trust in the process and the modellers, rather than seeing this as a necessary step in the process and due diligence.

In light of the issues discussed, various types of recommendations were proposed, ranging from overarching programmatic recommendations, such as initiating capacity building programmes, to more technical recommendations on uncertainty quantification methods. A summary of these is presented in section 3.4 below. In the session closing participants expressed that they found the discussions and policy brief initiative to be a move in the right direction, although it was the first of many steps, discussions, and collaborative work still needed to see the recommendations practically implemented.

3.4 KEY RECOMMENDATIONS AND POLICY BRIEF

The resulting policy brief has been included in *APPENDIX B.3 – Policy brief on uncertainty in hydrological modelling*. The brief was drafted to present the issue of ignoring uncertainty in hydrological modelling as a problem with serious consequences, but one that can be reasonably be addressed. Because modelling uncertainty is a complex and multi-faceted concept, with prediction uncertainty coming from many contributing sources (i.e. data, parameters, structure), effort was put into making the background explanations and demonstration of the issue as tangible as possible, presenting concrete examples from this and previous projects (Hughes *et al.*, 2011; Hughes, Mohobane and Mallory, 2015; Glenday *et al.*, 2022). This included presenting one of the case studies in the form of a hypothetical narrative presenting a case in which a decision maker is given with a single model output with no uncertainty quantification compared to a case in which they are given the range of predicted values across well-performing models.

The many recommendations put forward during the discussion sessions are given in full in *APPENDIX B.2 – Consolidated notes from the policy brief discussion sessions*. These were grouped and distilled for the purpose of the policy brief. In brief the recommendations were presented in two groupings, overarching or programmatic recommendations, and technical recommendations, given in a separate box on “Approaches for quantifying and reducing modelling uncertainty.”

Overarching or programmatic level recommendations:

- **Build capacity:** Promote theoretical and operational understanding of analyses and decision-making under uncertainty, through university curricula and certified short-courses.
- **Standardise efforts:** Develop context-appropriate standard practices for estimating, reducing and communicating hydrological uncertainty for decision-making, e.g. checklists of assessments depending on risk levels, data availability, and scale.
- **Invest in technology:** Include facilities in models that automate sensitivity analyses, uncertainty analyses, calibration, and water balance assessments across a range of storages and fluxes. These functions can be programmed into existing modelling tools or internationally developed tools that meet these needs can be harnessed.
- **Support data collection & open access data:** Long-term, continuous, spatially-distributed data on rainfall, evaporative demand, and streamflow need to be made accessible. Data on other storages and fluxes (groundwater, soil moisture, evapotranspiration) can further resolve uncertainties. Calibrate remote sensing products locally with field data before relying on these as alternative sources. Government-funded hydrological modelling efforts, with associated input and output databases, should be easily available.
- **Make budget available:** Ensure researchers and consultants have enough time and budget to quantify and reduce uncertainty in hydrological modelling, and account for it in further analyses and decision-making. Resources are also needed for the process of establishing standardised approaches and reviewing these over time.

Approaches for quantifying and reducing modelling uncertainty

Further engagement is needed to recommend specific methods for different cases. Some suggestions are given here:

- **Identify and document sources of uncertainty as well as assumptions and subjective decisions:** This could be a baseline requirement in all modelling projects, and consider input data, calibration data, parameters, and model structure.
- **Conduct sensitivity analyses:** This determines which uncertain datasets and parameters will have the greatest impact on the model outputs.
- **Apply more than one modelling tool and/or model structure:** This would be particularly advised when informing high risk and high cost decisions.
- **Validation – not just streamflow:** Apply a suite of model 'reality-checks' targeting a variety of hydrological processes during the calibration and acceptance of models and parameter ranges (e.g. evapotranspiration, runoff ratios, surface vs subsurface flow dominance, groundwater recharge).
- **Propagate uncertainty:** At a minimum, identify a reasonable high flow and low flow dataset to carry through to further analyses, such as stochastic yield modelling.

It is critical to note that while these recommendations would require some funding to achieve, for which there is heavy competition, if implemented these steps are likely to result in considerable savings from avoiding the costs associated with inappropriate water management and emergency response measures that may otherwise ensue. In addition, technology is making many parts of this process easier and easier, such that significant progress can be made by building the capacity in the sector to take advantage of these developments.

3.5 CONCLUSIONS AND WAY FORWARD

The process of drafting the policy brief by engaging with parties across the water sector was recognised by participants as a notable step towards integrating model uncertainty assessment into common practice. However, it was clear that it is an early step and further work and engagement are needed. During the process of inviting participants to the discussion sessions it was clear that there is a good amount of interest in the topic, though not all interested were able to attend. It is hoped that this interest extends to the resulting policy brief and that it is used to encourage implementation of the recommendations across institutions.

Through the discussions and their synthesis, it became clear that a larger, ongoing, participatory process of standardisation of practices around uncertainty assessments is needed, and that this would require some sort of institutional champion for it to be sustained. Recommended approaches for different settings would need to be tested and reviewed on a regular basis for practicality and to keep updated with technological advances over time. This would require more prolonged and widespread engagements across DWS and other implementors, consultants, and researchers than could be achieved in this small project, and the need for ongoing revisions in the long-term make the process ill-suited to only being supported by short-term research projects. No conclusion was reached about which institution would be most suited or willing to drive such an effort; however, DWS, the WRC, and potentially SACNASP would be relevant options to drive or partner it.

In addition to further laying out recommended methodological approaches for quantifying and reducing modelling uncertainties, there is also need for more reflection on specific methods for systematically considering uncertainty in decision-making processes, as well as an appraisal of the situations where these are or are not currently in place. While this was touched upon, the scope of the current project and discussions focused on the preceding step of quantifying and reducing uncertainties, and less on specific steps for considering the uncertainty bounds of estimates when planning or decision making in different contexts.

CHAPTER 4. ONLINE RESOURCES FOR THE HYDROLOGICAL MODELLING COMMUNITY

Identifying ways to improve modelling practice over time, and helping these to be taken up, will require ongoing engagement and exchange of ideas across hydrological modellers and users of model outputs. Hopefully the number of hydrological modellers across South Africa is growing. However, the 'modelling community' appears to be somewhat disparate, with variable degrees of communication across institutions and institution types, government, consultancies, and academia. Modellers trained in different institutions may be well-versed in the use of a particular modelling software tool, often associated with certain conceptual approaches, with a locally supported community of users. Unless extra effort is made, they may have limited exchanges with those using other tools and approaches from which there could be mutual benefit.

There are emerging opportunities and available tools that can help facilitate a greater degree of exchange across researchers and practitioners in hydrological modelling, and this project tried to leverage some of these. Project activities included working on, and promoting, two online resources that facilitate information exchange and capacity building across the community: an editable 'wiki' website on hydrological modelling tools and the pre-existing online 'question-answer' platform 'Stack Exchange' (https://stackexchange.com/), further described in sections below. In addition, the recently formed South African Hydrological Society (SAHS), a formalisation and revitalisation of South African National Council of the International Association of Hydrological Sciences (SANCIAHS), provides email, website, and social media channels that can be used to spread the word about resources and events.

4.1 WIKI

A modelling guidance 'wiki' website (https://hydromodel-sa-wiki.saeon.ac.za/) was initially designed in WRC project K5-2927 on modelling tool intercomparison (Glenday *et al.*, 2022). A 'wiki' website, is a website that a user community can edit and add to online on an ongoing basis (e.g. Wikipedia, https://en.wikipedia.org/). The "HydroModel SA wiki" site is intended to be a community resource that allows users to participate in developing it over time by suggesting edits and discussing content via 'discussion' pages. A focus of the site's content is building awareness of different modelling strategies and the options available across different modelling tools, which can also help new modellers entering the field. The content describes process representation strategies and options available across different commonly used modelling tools, allowing users to compare these side by side. While the site architecture and basic content were created in a previous project, the current project engaged in further development and publicising the site.

4.1.1 Further site content development

Additional content was developed and added to the wiki site:

- content directing users to the Stack Exchange question-answer site (see Section 4.2 below) to ask for, and provide, technical modelling assistance
- making the methods for users to suggest edits and additions to the wiki and engage in discussions on the wiki clearer and more accessible

- packaging information about modelling approaches for selected use cases onto targeted pages for ease of access: modelling riparian zones in which the vegetation has additional water access, modelling wetlands, modelling irrigation from different sources, modelling catchments with many small farm dams.

In addition, content created that had been created for the wiki site during the previous project (K5-2927), which had not been formatted for web-viewing and uploaded was also added to the site on the ‘process representation’ pages.

4.1.2 Publicising the resource

The HydroModel SA wiki site was publicised during a presentation at the South African Hydrological Society (SAHS) inaugural conference, 10-12 October 2022. The presentation covered the structural review across different commonly used modelling tools that was undertaken for project K5-2927 (Glenday *et al.*, 2022), highlighting the wiki website as place to find the information produced. The site was further publicised during the model-a-thon feedback in 2023 and via SAHS through the society’s email list and social media in early 2024, to coincide with the start of university terms.

There were delays in completing the additional site content, and hence in publicising the resource compared to timing put forward in the project’s inception report. This was in large part due to the model-a-thon timeline being longer than originally planned, both due to more data mining and preparation work being needed than anticipated and participants requiring longer periods to finish the exercise, delaying the output processing phase. This resulted in less project team time available for the wiki in total and in scheduling conflicts, delaying the content completion.

4.1.3 Recommendations for resource longevity

To maintain the wiki site in the longer term will require a team dedicated to periodic reviews and updates, particularly as software tools develop and change and new tools and resources become available. The site is currently set up such that any user can suggest edits and start or add to discussion strings, however the main content is editable by the project team. The site could be opened up to allow anyone who registers to freely edit the material, as is done for Wikipedia, however this approach requires a large active community to ensure timely self-correction if someone accidently or purposefully adds incorrect or otherwise harmful material to the site. For the near-term future it is suggested that a group of particularly interested parties have administration rights reviews and updates the site on a quarterly basis. This could be incentivised by institutional support for the time spent, recognition for the work, and/or CPD points awarded by SACNASP.

4.2 STACK EXCHANGE

Stack Exchange is a free, online, question answer platform that has developed a set of websites with different focus topics. This project promoted the ‘Earth Science’ Stack Exchange site (https://earthscience.stackexchange.com/) as a support tool for hydrological model users in South Africa. The site allows anyone who creates a user profile to post a question or answer existing questions. Questions and answers (Q&A) are tagged with keywords and ranked for helpfulness by users. Anyone can search for and read existing question answer strings online, even without creating a user profile or logging in. Stack Exchange Q&A strings often appear in general web search results, such as in Google searches, making the information on the platform very easy to find.

Use of this platform can strengthen the hydrological modelling community by making it more efficient for those with modelling expertise to assist newer users and easier for new users to find the information they are looking for. Building a presence on the site for various hydrological modelling software tools may also facilitate engagement with overseas users of the tools, further extending capacity and exchange. For locally developed modelling tools, for which there is often limited documentation and limited expert capacity to provide support to new users, Stack Exchange provides a free and easy mechanism to share the knowledge base in an accessible format.

4.2.1 Initiating relevant content

There needs to be a critical mass of activity and Q&A strings about hydrological modelling on the Stack Exchange platform to facilitate wider uptake. The project team has contributed to this by creating StackExchange.com profiles and adding questions and answers to the platform based on their previous experiences using different modelling tools. Starting in August 2022, the team has posted 18 Q&A strings about the use of various catchment modelling tools on the https://earthscience.stackexchange.com/ site: four on ACRU4, four on WRSM-Pitman, three on SWAT, and seven on MIKE-SHE. All of these strings have had ten or more external views, with some having 30-40 views and two of the MIKE-SHE posts having over 100 views to date.

4.2.2 Publicising the resource

To facilitate community uptake, instructional content has been created which covers what Stack Exchange is and its benefits; how to search to see if one's question has already been asked and answered; how to create a log-in, post new questions and tag them appropriately, post answers to questions, and promote helpful answers. This user guidelines document has been included in *APPENDIX C.1 – Stack Exchange guidelines document*.

Stack Exchange use was promoted using several communication channels, targeting the broader hydrology community and also specifically contacting those who teach hydrological modelling and software. The resource will only become helpful if advanced users answer questions on the platform, and so it is particularly important to get some who teaching modelling on board. It also has potential to lessen their workload over time. The platform was publicised during a presentation on modelling tool intercomparison given at the SAHS 2022 conference in October 2022. The use guidance material was also put onto the HydroModel SA wiki site (see Section 4.1). The guidance document was also distributed to the SAHS email distribution list and social media in November 2023. The guidance material will be re-posted via SAHS in early 2024 as reminder for the new academic year. The project team also personally emailed people known to teach hydrological modelling across various institutions, providing them with the guidance material and offering to give them a live online introduction to the platform in January or February 2024 if they would like one.

4.2.3 Recommendations for resource longevity

Once there is a critical mass of regular Stack Exchange Earth Science site users in the hydrological modelling community, and a helpful volume of information posted on the site, it will become a self-sustaining resource. People looking for assistance will easily find it on their own when doing web searches on their queries and may then use it to post questions that have not already been answered. In the interim it is suggested that the resource is actively promoted for the next few years. This could simply be done by re-posting the guidance documentation via SAHS communication channels twice a year to make new members aware of the resource and remind others who may not have taken up its use yet.

CHAPTER 5. CONCLUSIONS, RECOMMENDATIONS, AND IMPLICATIONS FOR FUTURE NATIONAL WATER RESOURCES STUDIES

5.1 CONCLUSIONS

The core research activity in this project, the model-a-thon, served as a powerful demonstration of hydrological modelling uncertainty issues that are often ignored in management settings. In this case, several different models – with different structures, produced using different tools, built by different individuals – all produced streamflow outputs that would be considered satisfactory according to generally accepted metrics. However, when these models were used to model a land cover change scenario, replacing a wattle tree plantation with sugar cane, they predicted very different streamflow impacts, ranging from a 39% to a 101% increase in mean annual streamflow. This highlighted that even when there is observational streamflow data to calibrate and verify models, when they are applied to predict the impacts of different scenarios there can easily be a very high degree of uncertainty in the output. The study also demonstrated that this uncertainty can be reduced using additional information about catchment processes. This can identify models that may be producing reasonable streamflow output, but are doing so using an unrealistic representation of internal flows and storages. However, when there is even less data about a catchment (e.g. it is ungauged), uncertainties in model predictions may be even higher than those found here.

Despite a growing body of global and local research on uncertainties in hydrological modelling, it has remained common practice in applied contexts for model outputs to be presented as single values or timeseries without an indication of the uncertainty in the prediction. Without a quantitative indication of uncertainty, decision makers are ill-equipped to consider risks when using these values to inform decisions. Other elements of uncertainty in water resources management, such as predictions of future rainfall given its inherent variability, are more routinely considered in current practice by using stochastic streamflow inputs into supply system yield models. However, the uncertainties in the hydrological modelling, that is used as the statistical basis for these stochastic inputs, is not routinely included in this process and can be highly significant (Hughes, Mohobane and Mallory, 2015). Fortunately, these existing stochastic modelling efforts may suggest that further incorporating hydrological modelling uncertainty may not entail major adjustments or learning curves in some parts of the water sector.

It is hoped that the participatory nature of both the model-a-thon activity and the development of the policy brief awakened productive conversations on modelling uncertainty across the water sector. The policy brief puts forward programmatic level recommendations for improving and mainstreaming uncertainty consideration in the sector, as well as some more technical recommendations for quantifying and reducing uncertainty. It was clear that these were initial steps towards seeing the recommendations implemented, and further collaborative work needs to be done across water resource managers, consultants, and researchers to establish more standardised practices for quantifying modelling uncertainty and considering this in decision making, tailored to different case types. This process needs at least one dedicated institutional champion to ensure progress is made, with the WRC and DWS being natural candidates.

5.2 RECOMMENDATIONS SUMMARY

Each chapter of this report has recommendation sections for the topic covered, i.e.:

- recommendations for further model uncertainty research (section 2.6.2) and model-a-thon activities (2.6.3);
- recommendations put forward in the policy brief (3.4) and for continued efforts to see these implemented (3.5); and
- recommendations for maintaining the HydroModel SA wiki (4.1.3) and further promoting the Stack Exchange question and answer site (4.2.3).

As such, these more specific points are not fully repeated here – the reader is referred to the sections mentioned.

The main overarching recommendations stemming from this project are:

- Establish institutional support for the process of developing, and reviewing, standard practices for quantifying, reducing, and considering uncertainty in hydrological modelling for different case types. Case types refer to different decision-making contexts (water supply systems, flood risk reduction, catchment land cover management, etc.), data availability levels (ungauged systems, well monitored systems, etc.), system size, etc. This will require further engagement across managers, consultants, and researchers to come up with initial procedure suggestions, pilot these, and review them over time.

- Invest in capacity building in hydrological modelling, including uncertainty analyses, via academic programmes as well as through learning opportunities for those already active in the water sector. Widespread implementation of best practice around modelling uncertainty will require relatively deep understanding across many practitioners. This will also require either use of new software tools, or further developing existing tools, to automate uncertainty analyses. This requires well capacitated teams with skills in hydrology, modelling theory, and software programming. Accessible online resources, such as the wiki and Stack Exchange question and answer platform that were promoted in this project, can assist in building capacity across the community given support for maintenance.

- Maintain and expand hydrological and meteorological field observation data collection, ideally across a range of fluxes and storages (i.e. in addition to weather stations and stream gauges, strategically monitor groundwater levels, soil moisture, ET, flow source tracers, etc.), and ensure data is made freely and easily accessible. This project and others have clearly demonstrated how model output uncertainty can easily reach untenable levels when there is little observational data to constrain the uncertainties about structure, parameters, and inputs. Having data about a variety of hydrological processes and storages is highly valuable, especially when streamflow records are short or non-existent. Remote sensing data, while very helpful for looking at spatial distributions, requires local to regional calibration against field observations (collected over a range of weather conditions) to provide reliable quantitative estimates. Inaccessible data, or data that is difficult or time consuming to access, due to findability, formatting, or cost, essentially negates the high potential value of this data for improving management and hence lives and livelihoods.

- Further participatory research activities across the hydrological modelling community, similar to this project's model-a-thon, can be used to further multiple aims of growing awareness, building and assessing capacity, taking stock of current practice, furthering research, and building a more connected community of practice

5.3 IMPLICATIONS FOR FUTURE NATIONAL WATER RESOURCES STUDIES

Every iteration of the national water resources study has added advancements to the modelling approach used, balancing developments in technology, data, process understanding, and management concerns against practical and resource constraints. Findings from this project and other local research on modelling uncertainty (Hughes *et al.*, 2011; Hughes, Mohobane and Mallory, 2015; Glenday *et al.*, 2022) make a strong case for including uncertainty analyses in future water resources studies, amongst other potential improvements. Developing an uncertainty estimation approach for the next national water resources study would be a helpful test case in the development of more standardised practices for modelling uncertainty assessments for different settings. Providing likely value ranges for resources estimates, rather than single values with unknown degrees of error, would better equip decision-makers to consider risks. Quantifying uncertainty in the runoff estimates for every quaternary catchment across the country would highlight priority areas for growing monitoring networks and/or conducting further research to reduce high uncertainties in areas of particular concern or projected growth.

A major barrier to implementing systematic uncertainty analyses in previous water resources studies has been the capabilities of the software tool used, WRSM-Pitman (most recent version: WRSM2000). As described more fully in (Glenday *et al.*, 2022), this software does not have automated facilities to batch-run many versions of a model with different sets of input parameter values to explore the feasible parameter space, do sensitivity analyses, and calibrate parameters. The process of inputting and changing parameters manually is also very labour intensive in the current interface. WRSM2000 also does not have any automated facilities to easily batch-run versions of a model with different input climate timeseries or other inputs, to assess the impacts of using different data sources and/or data processing methods, or of considering stochastic variability or climate change scenarios. In addition, obtaining modelled output for various catchment storages and fluxes, beyond streamflow, is also inefficient in the software. As such, systematic parameter and input uncertainty analyses, parameter calibration, and internal process reality checks using the current version of WRSM software at the national scale is not practically feasible.

The most recent SPATSIM version of the Pitman model (Hughes, 2019) does have automated facilities that make uncertainty analyses, calibration, and process reality checks more practical; however, the development of SPATSIM-Pitman and WRSM-Pitman have diverged and there are also useful features in WRSM that are not in SPATSIM. WRSM has a more modular structure that allows for a variety of connection options between subcatchment areas, irrigated subareas, dams, and river reaches to enable explicit representation of various water management arrangements that cannot easily be directly considered in SPATSIM. SPATSIM's current uncertainty modelling routines rely, in part, on the more uniform subcatchment structures and connections the tool allows, compared to WRSM. It should be noted that there are modelling software tools which do include stochastic modelling capabilities for inputs and parameters, while also supporting more flexible and complex structures, such as MIKE-SHE and SWAT. This demonstrates that this issue is not an unsurmountable barrier for local Pitman-model-based tools, if given programming resources and support.

The current project, and a growing body of research (Butts *et al.*, 2004; Georgakakos *et al.*, 2004; Højberg and Refsgaard, 2005; Clark *et al.*, 2008; Mockler *et al.*, 2016; Mockler, O'Loughlin and Bruen, 2016; Troin *et al.*, 2018; Glenday *et al.*, 2022), highlight the large, and often dominant role of model structural uncertainty, which requires the testing of multiple model structures to assess and address. The commonly used local modelling tools, WRSM, SPATSIM, and ACRU, do not allow notably different model structures to be easily built within the same software. As such, incorporating modelling structural uncertainty in the national water resources studies using these tools would require conducting modelling with multiple software tools at the national scale. This appears to be a feasible proposition given that WRSM, SPATSIM, and ACRU have all already been set up and run for every quaternary catchment in the country for multiple different projects in the past (Hughes, 2005; Schulze *et al.*, 2005; Schulze, 2007; Bailey and Pitman, 2015; Schutte *et al.*, 2023). In addition, there is an existing initiative to prepare databases of basic SWAT inputs at a national scale (Le Roux *et al.*, 2023) and to set-up the SWAT model at a national scale through the HydrologicAl Model for South Africa (HAMSA) project: https://www.waterresearchobservatory.org/hamsa. These modelling efforts have used different sets of inputs and assumptions for different purposes. As such their existing outputs are not directly comparable to one another. Leveraging these efforts further to improve future national water resources studies would require bringing the different teams involved together to harmonise various elements of the modelling, such as the input data used, representation and calibration strategies, and how parameter and input uncertainties would be handled across the different model structures.

There is a clear need for further dialogues across managers, consultants, and researchers on the methods to apply in future national water resources studies. These studies should take advantage of both new developments in methodologies, software, computing power, and data sources, as well as the knowledge base that has been built up through many decades of prior research and modelling efforts. Options need to be evaluated and decisions made across a myriad of issues, such as: which modelling tools and model structures to use; what data sources to use and how to process the data to get model inputs and datasets to calibrate against; how to calibrate models and criteria for accepting models as satisfactory; how to quantify uncertainty across sources (model structure, parameters, input data, calibration data); how models, inputs, and outputs should be presented and shared and who will host the database; and who should do various aspects of the work. Advancements could be prioritised and phased into the process over time. As with the process of standardising practice around uncertainty assessments, the process of bringing relevant parties together to plot a course for future national water resources studies, and implement this, will require a responsible institutional lead and programmatic support to facilitate and ensure progress over the long term.

REFERENCES

Addor, N. and Melsen, L.A. (2019) 'Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models', *Water Resources Research*, 55(1), pp. 378-390. Available at: https://doi.org/10.1029/2018WR022958.

Arsenault, R., Brissette, F. and Martel, J.-L. (2018) 'The hazards of split-sample validation in hydrological model calibration', *Journal of Hydrology*, 566, pp. 346-362. Available at: https://doi.org/10.1016/j.jhydrol.2018.09.027.

Bailey, A.K. and Pitman, W.V. (2015) *Water Resources of South Africa 2012 Study (WR2012)*. K5/2143/1. Pretoria, South Africa: Water Research Commission (WRC).

Brath, A., Montanari, A. and Toth, E. (2004) 'Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model', *Journal of Hydrology*, 291(3), pp. 232-253. Available at: https://doi.org/10.1016/j.jhydrol.2003.12.044.

Breuer, L., Huisman, J.A., Willems, P., Bormann, H., Bronstert, A., Croke, B.F.W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A.J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindström, G., Seibert, J., Sivapalan, M. and Viney, N.R. (2009) 'Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM). I: Model intercomparison with current land use', *Advances in Water Resources*, 32(2), pp. 129-146. Available at: https://doi.org/10.1016/j.advwatres.2008.10.003.

Butts, M.B., Payne, J.T., Kristensen, M. and Madsen, H. (2004) 'An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation', *Journal of Hydrology*, 298(1), pp. 242-266. Available at: https://doi.org/10.1016/j.jhydrol.2004.03.042.

Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T. and Hay, L.E. (2008) 'Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models', *Water Resources Research*, 44(12), p. W00B02. Available at: https://doi.org/10.1029/2007WR006735.

Clulow, A.D., Everson, C.S. and Gush, M.B. (2011) *The long-term impact of Acacia mearnsii trees on evaporation, streamflow, and ground water resources*. WRC Report TT505/11. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Clulow, A.D., Kunz, R.P., Gokool, S., Kaptein, N., Schutte, S., Schulze, R., Horan, R., Everson, C.S., Thornton-Dibb, S.L.C., Horan, M., Toucher, M. and Germishuizen, I. (2022) *The expansion of knowledge on evapotranspiration and stream flow reduction of different clones/hybrids to improve the water use estimation of SFRA species (i.e. Pinus, Eucalyptus, and Acacia species)*. WRC Report K5/2791. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

DWAF (2006) *National Groundwater Resource Assessment, version 2 (GRA II)*. DWAF Report. South Africa: Department of Water Affairs & Forestry (DWAF).

Everson, C., Clulow, A., M, B., Watson, A., C, N., Bulcock, H., Mengistu, M., S, L. and Demlie, M. (2014) *The long term impact of Acacia mearnsii trees on evaporation, streamflow, low flows and ground water resources. Phase II: Understanding the controlling environmental variables and soil water processes over a full crop rotation*. WRC Report 2022/1/13. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Everson, C.S., Gush, M., Moodley, M., Jarmain, C., Govender, M. and Dye, P. (2007) *Can effective management of riparian zone vegetation significantly reduce the cost of catchment management and enable greater productivity of land resources*. WRC Report K5/1284. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Everson, C.S., Toucher, M., Vather, T., Mfeka, S., Lawrence, K., Horan, R., Govender, J., Gray, B., Ramjeawon, M. and Aldsworth, T. (2018) *Continued hydrometeorological monitoring at the Two Streams Catchment for the clear felling of the Acacia mearnsii stand.* WRC Report K5/2780. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Georgakakos, K.P., Seo, D.-J., Gupta, H., Schaake, J. and Butts, M.B. (2004) 'Towards the characterization of streamflow simulation uncertainty through multimodel ensembles', *Journal of Hydrology*, 298(1), pp. 222-241. Available at: https://doi.org/10.1016/j.jhydrol.2004.03.037.

Glenday, J., Gokool, S., Gwapedza, D., Holden, P., Jumbi, F.T., Metho, P., Rebelo, A.J. and Tanner, J.L. (2022) *Critical catchment model inter-comparison and model use guidance development*. WRC Report Project K5-2927. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F. (2009) 'Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling', *Journal of Hydrology*, 377(1), pp. 80-91. Available at: https://doi.org/10.1016/j.jhydrol.2009.08.003.

Højberg, A.L. and Refsgaard, J.C. (2005) 'Model uncertainty – parameter uncertainty versus conceptual models', *Water Science and Technology*, 52(6), pp. 177-186.

Holländer, H.M., Blume, T., Bormann, H., Buytaert, W., Chirico, G.B., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G. and Flühler, H. (2009) 'Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data', *Hydrology and Earth System Sciences*, 13(11), pp. 2069-2094. Available at: https://doi.org/10.5194/hess-13-2069-2009.

Hughes, D., Sawunyama, T. and Kapangaziwiri, E. (2008) 'Incorporating estimation uncertainty into water resource development planning in ungauged basins in southern Africa', *Proceedings of the 10th National Hydrology Symposium of the British Hydrological Society* [Preprint].

Hughes, D.A. (2005) *SPATSIM, an Integrating Framework for Ecological Reserve Determination and Implementation*. WRC Report TT 245/04 (Project K5/1160). Gezina, Pretoria, South Africa: Water Research Commission (WRC) & Rhodes University.

Hughes, D.A. (2013) 'A review of 40 years of hydrological science and practice in southern Africa using the Pitman rainfall-runoff model', *Journal of Hydrology*, 501, pp. 111-124. Available at: https://doi.org/10.1016/j.jhydrol.2013.07.043.

Hughes, D.A. (2019) 'SPATSIM v3 & IWR version of the Pitman model'. IWR Rhodes University, Grahamstown, South Africa.

Hughes, D.A., Kapangaziwiri, E., Mallory, S.J.L., Wagener, T. and Smithers, J. (2011) *Incorporating uncertainty in water resources simulation and assessment tools in South Africa*. WRC Report WRC Report No. 1838/1/11. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Hughes, D.A., Kapangaziwiri, E. and Sawunyama, T. (2010) 'Hydrological model uncertainty assessment in southern Africa', *Journal of Hydrology*, 387(3), pp. 221-232. Available at: https://doi.org/10.1016/j.jhydrol.2010.04.010.

Hughes, D.A., Mohobane, T. and Mallory, S.J.L. (2015) *Implementing uncertainty analysis in water resources assessment and planning: report to the Water Research Commission*. WRC Report 2056/1/14. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Huisman, J.A., Breuer, L., Bormann, H., Bronstert, A., Croke, B.F.W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A.J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindström, G., Seibert, J., Sivapalan, M., Viney, N.R. and Willems, P. (2009) 'Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) III: Scenario analysis', *Advances in Water Resources*, 32(2), pp. 159-170. Available at: https://doi.org/10.1016/j.advwatres.2008.06.009.

Kapangaziwiri, E., Hughes, D.A. and Wagener, T. (2012) ‘Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa’, *Hydrological Sciences Journal*, 57(5), pp. 1000-1019. Available at: https://doi.org/10.1080/02626667.2012.690881.

Kling, H., Fuchs, M. and Paulin, M. (2012) ‘Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios’, *Journal of Hydrology*, 424-425, pp. 264-277. Available at: https://doi.org/10.1016/j.jhydrol.2012.01.011.

Knoben, W.J.M., Freer, J.E., Peel, M.C., Fowler, K.J.A. and Woods, R.A. (2020) ‘A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments’, *Water Resources Research*, 56(9), p. e2019WR025975. Available at: https://doi.org/10.1029/2019WR025975.

Le Roux, J., Mararakanye, N., Mudaly, L., Weepener, H. and van der Laan, M. (2023) *Development of a South African National Input Database to Run the SWAT Model in a GIS*. WRC Report No. 3053/1/22. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Li, C., Wang, H., Liu, J., Yan, D., Yu, F. and Zhang, L. (2010) ‘Effect of calibration data series length on performance and optimal parameters of hydrological model’, *Water Science and Engineering*, 3(4), pp. 378-393. Available at: https://doi.org/10.3882/j.issn.1674-2370.2010.04.002.

Melsen, L.A., Teuling, A.J., Torfs, P.J.J.F., Zappa, M., Mizukami, N., Mendoza, P.A., Clark, M.P. and Uijlenhoet, R. (2019) ‘Subjective modelling decisions can significantly impact the simulation of flood and drought events’, *Journal of Hydrology*, 568, pp. 1093-1104. Available at: https://doi.org/10.1016/j.jhydrol.2018.11.046.

Mendoza, P.A., Clark, M.P., Mizukami, N., Gutmann, E.D., Arnold, J.R., Brekke, L.D. and Rajagopalan, B. (2016) ‘How do hydrologic modelling decisions affect the portrayal of climate change impacts?’, *Hydrological Processes*, 30(7), pp. 1071-1095. Available at: https://doi.org/10.1002/hyp.10684.

Mendoza, P.A., Clark, M.P., Mizukami, N., Newman, A.J., Barlage, M., Gutmann, E.D., Rasmussen, R.M., Rajagopalan, B., Brekke, L.D. and Arnold, J.R. (2015) ‘Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts’, *Journal of Hydrometeorology*, 16(2), pp. 762-780. Available at: https://doi.org/10.1175/JHM-D-14-0104.1.

Mockler, E.M., Chun, K.P., Sapriza-Azuri, G., Bruen, M. and Wheater, H.S. (2016) ‘Assessing the relative importance of parameter and forcing uncertainty and their interactions in conceptual hydrological model simulations’, *Advances in Water Resources*, 97, pp. 299-313. Available at: https://doi.org/10.1016/j.advwatres.2016.10.008.

Mockler, E.M., O’Loughlin, F.E. and Bruen, M. (2016) ‘Understanding hydrological flow paths in conceptual catchment models using uncertainty and sensitivity analysis’, *Computers & Geosciences*, 90, pp. 66-77. Available at: https://doi.org/10.1016/j.cageo.2015.08.015.

Moges, E., Demissie, Y., Larsen, L. and Yassin, F. (2020) ‘Review: Sources of hydrological model uncertainties and advances in their analysis’, *Water*, 13, p. 28. Available at: https://doi.org/10.3390/w13010028.

Moriasi, D., Gitau, M., Pai, N. and Daggupati, P. (2015) ‘Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria’, *Transactions of the ASABE (American Society of Agricultural and Biological Engineers)*, 58, pp. 1763-1785. Available at: https://doi.org/10.13031/trans.58.10715.

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D. and Veith, T.L. (2007) ‘Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations’, *Transactions of the ASABE*, 50(3), pp. 885-900. Available at: https://doi.org/10.13031/2013.23153.

Nash, J.E. and Sutcliffe, J.V. (1970) ‘River flow forecasting through conceptual models part I — A discussion of principles’, *Journal of Hydrology*, 10(3), pp. 282-290. Available at: https://doi.org/10.1016/0022-1694(70)90255-6.

Ngubo, C.Z. (2019) *An investigation of the impacts of Acacia mearnsii plantations on secondary aquifer systems within the Two-Streams catchment, KwaZulu-Natal, South Africa*. MSc dissertation. University of KwaZulu-Natal.

Ngubo, C.Z., Demlie, M. and Lorentz, S. (2022) 'Investigation of hydrological processes and the impacts of Acacia mearnsii plantations on groundwater in secondary aquifers: Case study at the two-stream research catchment, South Africa', *Journal of Hydrology: Regional Studies*, 40, p. 101018. Available at: https://doi.org/10.1016/j.ejrh.2022.101018.

Pitman, W.V. (2011) 'Overview of water resource assessment in South Africa: current state and future challenges', *Water SA*, 37(5), pp. 659-664.

Refsgaard, J.C., van der Sluijs, J.P., Brown, J. and van der Keur, P. (2006) 'A framework for dealing with uncertainty due to model structure error', *Advances in Water Resources*, 29(11), pp. 1586-1597. Available at: https://doi.org/10.1016/j.advwatres.2005.11.013.

Sawunyama, T. and Hughes, D. (2007) 'Assessment of rainfall and evaporation input data uncertainties on simulated runoff in southern Africa', *IAHS Publication, IUGG Proceedings*, 313, p. 98.

Sawunyama, T., Hughes, D.A. and Mallory, S.J.L. (2011) 'Evaluation of combined contribution of uncertainty sources to total output uncertainty in water resource estimation in South Africa', *IAHS Publication, IUGG Proceedings*, 347.

Schulze, R.E. (2007) *South African Atlas of Climatology and Agrohydrology.* WRC Report 1489/1/06. Pretoria, South Africa: Water Research Commission.

Schulze, R.E., Warburton, M.L., Hewitson, B.C., Engelbrecht, F., Horan, M., Jack, C., Tadross, M., Vogel, C., Boardley, S., Dlamini, D., Maharaj, M., Lumsden, T., Reid, P., Gray, R. and Massey, R. (2005) *Climate change and water resources in Southern Africa*. WRC Report 1430/1/05. Gezina, Pretoria, South Africa: Water Research Commission (WRC), p. 54.

Schutte, S., Schulze, R.E., Clark, D.J., Kunz, R.P., Wolski, P., Lumsden, T., Smithers, J.C., Stuart-Hill, S.I., Gwena, K.R., Thornton-Dibb, S.L.C., Horan, M., Toucher, M. and Jele, Z. (2023) *A National Assessment of Potential Climate Change Impacts on the Hydrological Yield of Different Hydro-Climatic Zones of South Africa*. WRC Report No. 2833/1/22. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Scott, D.F., Prinsloo, F.W., Mehlomakulu, M. and Simmers, A.D.A. (2000) *A Re-analysis of the South African Catchment Afforestation Experimental Data*. WRC Report WRC No 810/1/00. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Shen, H., Tolson, B.A. and Mai, J. (2022) 'Time to Update the Split-Sample Approach in Hydrological Model Calibration', *Water Resources Research*, 58(3), p. e2021WR031523. Available at: https://doi.org/10.1029/2021WR031523.

Troin, M., Arsenault, R., Martel, J.-L. and Brissette, F. (2018) 'Uncertainty of Hydrological Model Components in Climate Change Studies over Two Nordic Quebec Catchments', *Journal of Hydrometeorology*, 19(1), pp. 27-46. Available at: https://doi.org/10.1175/JHM-D-17-0002.1.

Van Tol, J.J., Kunene, B.T., Freankel, C.H., Bezuidenhout, J.C., Clark, J. and Le Roux, P.A.L. (2007) *Hydropedological survey report of the Two Streams Catchment*. unpublished report contributing to WRC K5/1748 & others. Bloemfontein, South Africa: University of the Free State.

Watson, A. (2015) *The use of environmental isotopes, soil water measurements and soil water modelling to understand tree water use of an Acacia mearnsii (black wattle) stand in KwaZulu-Natal*. MSc thesis. University of KwaZulu-Natal.

Wiles, L. (2006) *Total evaporation estimation from sugarcane using the scintillation technique*. MSc thesis. University of KwaZulu-Natal.

APPENDIX A.1 – REVIEW OF SOUTH AFRICAN RESEARCH CATCHMENTS FOR MODEL-A-THON CASE STUDY SITE SELECTION

SELECTION CRITERIA

The initial aim was to select a case study catchment that met the following criteria:

- Data needed (see below) is freely share-able.
- **Streamflow data** for the catchment outlet for ten years or more, reliable (i.e. weir with no frequent capacity exceedance, etc.), daily timestep or finer, covering a period with relatively constant land cover and management conditions.
- **Weather station data (rainfall, temperature, humidity, wind speed, solar radiation)** for ten years or more overlapping with the streamflow data timeseries, daily timestep or finer, station located within the catchment (may consider within 20 km)
- **Rainfall gauge data** for multiple elevations in the catchment (may consider surrounding, within 20 km), with overlapping data at least 5 years, to assess spatial rainfall distribution.
- **Maximum size** of two quaternaries, **minimum size** may be less than one quaternary.
- Available descriptions of vegetation types, soil types, and aquifer types.
- No major irrigation, dams, or water diversions in the catchment
- A limited number of major land cover classes, to limit complexity and time needed, but ideally more than one cover type, to facilitate change scenario modelling
- No major contribution of groundwater to streamflow that is likely coming from aquifers that extend beyond the catchment's surface flow boundaries.
- Available 'auxiliary' hydrometric dataset(s) spanning multiple years to assess the realism of the modelled water balance for at least one component beyond streamflow:
    - o soil moisture timeseries for multiple locations in the catchment or previously analysed satellite estimates of soil moisture,
    - o groundwater data from boreholes and/or piezometers in several locations,
    - o estimates of actual evapotranspiration from flux towers, surface renewal systems, and/or calibrated analyses of remote sensing data.

Building on learnings from the previous model intercomparison study (K5-2927 (Glenday *et al.*, 2022)), an effort was made to select a case-study catchment for which observational data is available to assess the models' prediction of one or more water balance components beyond streamflow. A key limitation of the previous study's case study modelling exercise was that, while it was identified that different models predicted different balances of processes (e.g. surface flow, interflow, groundwater outflow, evapotranspiration, retention in soils and aquifers), there was insufficient data at hand to determine if any of the water balance predictions was more realistic than the others.

Table A.1-1 Comparison of potential model-a-thon case study catchments

Catchment	Area (km^2)	Period with rain & flow data	Land cover in periods with data	Auxiliary data available
Weatherly	1.57	1998-2010 (12 yr)	• 1998-2002 (5 yr) grassland & wetland • 2003-2010 (7 yr) 40% *young* pine & eucalyptus *(growing)* + 60% grass & wetland	• soil moisture • shallow & deep GW levels • flow path tracer data
Two Streams	0.73	1999-2014 (16 yr) & 2019-2022 (3 yr)	• 1999-2000 (1 yr) 66% mature wattle + 34% sugar cane (SC) • 2000-2004 (4 yr) 59% mature wattle + 9% riparian veg (RV) + 31% SC • 2004-2006 (2 yr) 59% fallow + 9% RV + 31% SC • 2006-2009 (5 yr) 59% *young* wattle + 9% RV + 31% SC • **2010-2014 (5 yr) 59% mature wattle + 9% RV + 31% SC** • 2019-2022 (3 yr) 59% *young* eucalypt + 9% RV + 31% SC	• soil moisture • shallow & deep GW levels • flow path tracer data • scintillometry AET (wattle, sugarcane)
Cathedral Peak II	1.9	1948-1993 (46 yr)	• 1951-1962 (12 yr) 75% *young* pine + 25% 'grassland' *(steep rocky…)* • 1963-1981 **(19 yr) 75% mature pine** + 25% 'grassland' *(steep rocky…)*	(use VI for grass part)
Cathedral Peak VI	0.67	1954-1993 (40 yr) & 2012-2022 (13 yr)	• 1954-2022 grassland & wetland, *biennial burn*	• soil moisture • shallow & deep GW levels • flux tower AET
Cathedral Peak IX	0.64		• 1954-2022 woodland + grassland *(woody invaded)* – fire exclusion	(use VI for grass part)

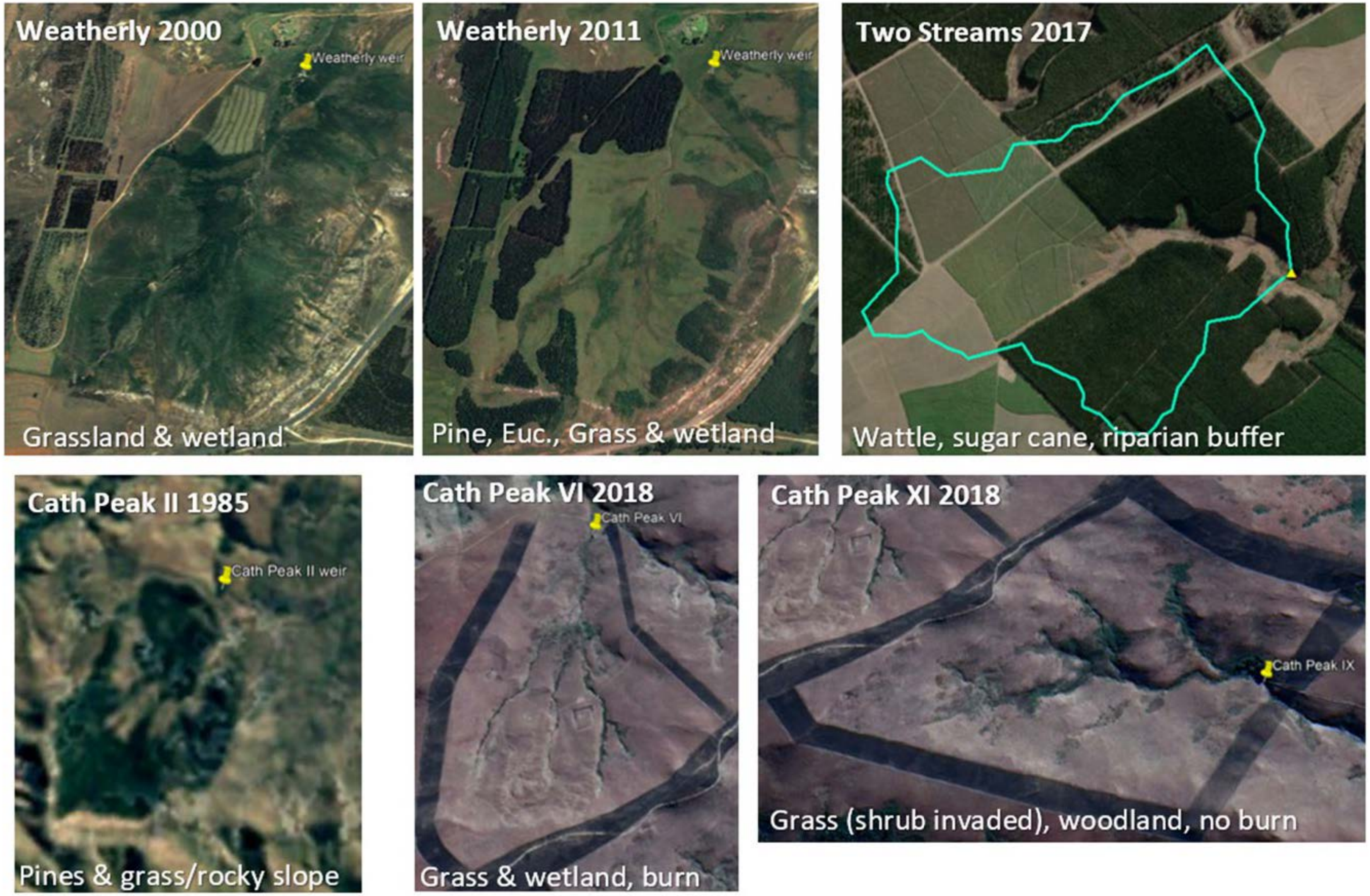


Figure A.1-1 Aerial imagery (Google Earth) of the instrumented research catchments considered for the model-a-thon case-study demonstrating their land cover distributions

CATCHMENTS ASSESSED

There are few catchments in the country which have the desired auxiliary process data readily available, those which have been the subject of hydrological research efforts in the past. Potential candidates assessed included the following instrumented research catchments: the Weatherly catchment in the Eastern Cape, Cathedral Peak catchments in KZN, and the Two Streams catchment in KZN. The Jonkershoek experimental catchments in the Western Cape were not considered because of the dominant contribution from the fractured rock aquifer at the headwaters, which may represent flow from a larger aquifer area than would be defined by the topographically defined surface water catchments and this is challenging in to represent in many catchment hydrological models.

An assessment of the catchments considered is presented in Table A.1-1 with aerial photos demonstrating their landcover distributions shown in Figure A.1-1.

SITE SELECTION & RATIONALE

The Two Streams catchment was the only one that met the criteria of having sizeable areas of contrasting land cover types which were relatively stable in area and properties over time, in this case sugar cane and black wattle plantation (*Acacia mearnsii*). This was needed for including a land cover change scenario in the activity such that the cover types in the change scenario were also present in the baseline land cover distribution, i.e. were present during the calibration period. The wattle in Two Streams was planted in late 2006 and monitoring of tree size and leaf area index (LAI) indicated a tapering of growth by late 2009 (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014).

The other catchments considered all had some drawbacks in terms of their land cover during the monitored periods. The Weatherly catchment had seven years of monitoring data once areas of the catchment had been planted with pines and eucalyptus in 2003, however (Scott *et al.*, 2000) show that pines can take at least this long after planting to show a levelling of their impacts on streamflow. This means that there was effectively little to no data for the site after the pines, known to mature more slowly than wattle, would have reached a more stable or mature state. While Cathedral Peak catchment II had sizeable areas of both mature pine plantation and grassland in the 1963-1981 monitoring period, the grassland was only in areas that were too steep and rocky to plant pines. This means that the grassland area present is in the baseline cover period would not be a reliable representation of the grassland that could exist if the pines were cleared. Cathedral Peak XI is a fire exclusion experiment and has areas that are densely encroached by woody species and other grassland parts that are less so, however these areas are difficult to delineate consistently over time.

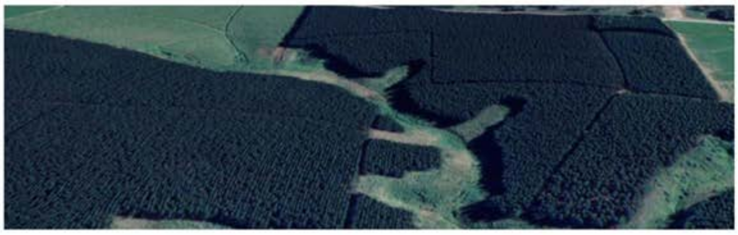
An option would have been to have participants model two different gauged catchments, Cathedral Peak II and VI, to obtain calibrated parameters for pine and grassland areas and then model the case of the pines being cleared from Cathedral Peak II. However, this was considered to be too complex and too much work to ask of volunteering participants.

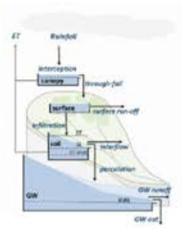
The disadvantages in using Two Streams as the case study were the duration of the period with stable land cover and streamflow and climate data, only five years compared to the desired ten or more, and the small catchment size, 0.7382 km^2. The relatively short calibration period can be somewhat problematic in that a more limited range of climate and flow conditions are represented and so the model set up will not have been tested over the range of conditions one may wish to apply it to. This is a common issue in many applied modelling cases and represents another source of uncertainty in prediction. One would actually need a case study with longer dataset to estimate the degree of uncertainty introduced by only using a shorter period. The Cathedral Peak catchments offered a longer observation period, but not the land cover distribution desired.

In terms of catchment size, the other research catchments considered were also relatively small, so would not have conferred much advantage. Some of the modelling tools have not necessarily been designed for modelling at this scale, particularly in terms of the way the subsurface storages and their outflow rates are conceptualised. A small headwater catchment may have considerable “underflow”, meaning that water recharging subsurface stores within the delineated catchment area may only contribute to streamflow further downslope than the location of the outlet point in question. ACRU4, for example, does not have a way to explicitly represent groundwater leaving the catchment without contributing to streamflow or to a riparian zone: all water entering a “baseflow store” is subject to its proportional outflow algorithm. This becomes less problematic for larger catchment areas for which a greater proportion of the recharged water would become streamflow within the catchment. The more spatially lumped, monthly timestep Pitman-based modelling tools WRSM and SPATSIM, actually do include ways for groundwater to leave catchments without contributing to streamflow. Small catchment size does pose another issue when using these tools in that their software is designed to work with larger values, million cubic metres of outflow per month to a few decimal places. This can be overcome by synthetically inflating the catchment size in the model and then rescaling the final model output. This instruction was given for users of those tools, see: *APPENDIX A.2 – Model-a-thon instructions document provided to participants.*

It is acknowledged that any selected case study will have some specific characteristics of which may confer an advantage to one software tool over another in terms of their capabilities in light of a specific issue. However, experienced users may be aware of ways to overcome certain issues in their modelling tool of choice, for example by having developed approaches to implicitly represent a given process within that tool’s structural options. Ideally this activity should be done with a wider variety of case studies with different characteristics to get an idea of model structural uncertainty across different contexts, but that is beyond the scope of this project and would require too much time and effort from the volunteering participants for a single engagement. The results of the model-a-thon are to be interpreted with the understanding that they are strongly impacted by the choice of case study catchment.

APPENDIX A.2 – MODEL-A-THON INSTRUCTIONS DOCUMENT PROVIDED TO PARTICIPANTS





Catchment Hydrology Model-a-thon 2022: Activity instructions & case study catchment info

J.Glenday, S.Gokool, D.Gwapedza, F.Jumbi, P.Metho, A.Rebelo, J.Tanner

WRC project 2022/23-00967: *Modelling uncertainty and reliability for water resource assessment in South Africa*

Contents

1 What is the 'model-a-thon'?

The "model-a-thon" is like a "hackathon" for hydrological modelling. Participants build and run a model of the same case study catchment, given the same data on climate, observed streamflow, land cover, and other catchment properties. As a second step, participants model this catchment with a new land cover distribution to predict the impact of the cover change. Participants can use whatever modelling software they would like and submit both their models and streamflow results.

The aim is to engage the modelling community and to explore how differently, or how similarly, individuals may model the same catchment, and the implications this can have for model outputs and change prediction. Individuals will differ in the modelling software tools that they use as well as in the specific ways they chose to set-up a model using a particular tool. Looking at the range of impacts predicted across participant models for the cover change scenario can give an indication of the potential uncertainty in prediction due to model structure uncertainty *

We hope this will be a fun and interesting challenge for the modelling community. Anyone who has used a hydrological model in the past is welcome to participate, whether you are a new user or an expert model developer. The activity has been designed, and data prepared, to minimise the amount of time participants will need to set-up their models. There will be a hosted online discussion of the outcomes. We also aim for this to be a community-building activity for those interested in modelling in South Africa. This activity is part of WRC research project *Modelling uncertainty and reliability for water resource assessment in South Africa*, 2022/2023-00967.

Participants will remain anonymous in data analyses in that participants' names will not be linked to their individual submissions in any analyses or reporting. Participants will be acknowledged for taking part in the activity in all reporting, unless they wish otherwise (to be specified in the submission survey). This activity has received ethical clearance from the Rhodes University Human Ethics Committee (RU-HEC). If you have any queries in this regard you may contact Ethics Coordinator Dr Janet Hayward (Janet.Hayward@ru.ac.za).

* The model structure uncertainty results will be strongly influenced by the specific choice of case study and the data provided for it. This will be considered interpretations of output. If this activity goes well, and there is participant interest, there may be future model-a-thons using a variety of different case studies, allowing broader conclusions to be drawn.

1.1 Benefits of participation

- **Challenge yourself, test your skills, build your experience**
- **Contribute to scientific research**: Take part in a novel study that builds understanding of modelling uncertainties.
- **Help the hydrology sector** to keep improving modelling practice.
- **Network with other modellers** at the kick-off events and output discussion.
- **Acknowledgements & certificate of completion:** Participants who complete the activity will be given a certificate of completion. For students and early career hydrologists, this can contribute to your CV. Participants who complete the activity will be acknowledged by name

in all reports and publications from the activity, unless they wish otherwise (NB: submission data itself will remain anonymised)

- **CPD Points:** Scientists registered with SACNASP can obtain 1.4 points for completing the model-a-thon activity.
- **Recognition for best performing models:** With the participants' permission, public recognition on the SAHS website and society communications will be given for models with the best performance statistics in the submissions set.
- **Student prizes for best performing models:** Student participants whose models have the best performance statistics out of the set of student submissions will be given public recognition and with the top five to be awarded prizes: takealot gift vouchers with values between R300-R800 (voucher value distribution to be determined by model performance statistics distribution!)
- **Co-authorship opportunity:** Participants who complete the model-a-thon and then choose to engage further in article drafting sessions and engagements will have the option of being co-authors in a publication on the outcomes.

2 Activity instructions

Instructions & data released: 11th October 2022

Model & results submissions due: 5th December 2022

This activity has been designed to take roughly one day to complete in most cases. Participants may wish to, or need to, use more time than this.

You should be able to do the modelling using only the data and information provided. Participants are not required, or expected, to do additional research or data mining to inform the modelling. However, if there are other information sources or databases you would typically consult in your modelling projects, you are welcome to use them, as long as the catchment size, climate input, and land cover distribution provided are adhered to.

For clarification questions regarding the instructions or datasets, or for data access or submission technical assistance, participants can contact the model-a-thon organisers by email at: model.a.thon@gmail.com.

2.1 Sign-up and receive a participant number

- When you **signed up** to participate (https://forms.gle/LEAmtr9h75wdR79Y9) you should have been sent a **participant number**. This number will be used to keep your submission anonymous in data analyses and reporting.
- If you wish to complete the activity multiple times, using a different modelling software tool each time, you are free to do so. We'd be very grateful to have the additional samples! To use multiple tools, *please sign up to participate multiple times* such that you receive a unique participant number for each of your model submissions.

2.2 Access the case study database

The case study database for the model-a-thon is available to participants through an online drive that can be accessed through the following web address:

https://iwr.ru.ac.za/MURRA/

User name: Downloader
Password: Model@Thon_1

- We recommend reading through this instructions document and the catchment description before starting to download data.
- A metadata file describing the files and file structure in the database is provided in the main folder and has the file name: **TwSt_Database_metadata.xlsx**
- To download: Click on the boxes next to folder names and/or next to individual files to add a "check mark". At the top of the page you will see the option to download what you have selected. You may choose to download the entire database, or look through the metadata file to help select specific files you wish to download.

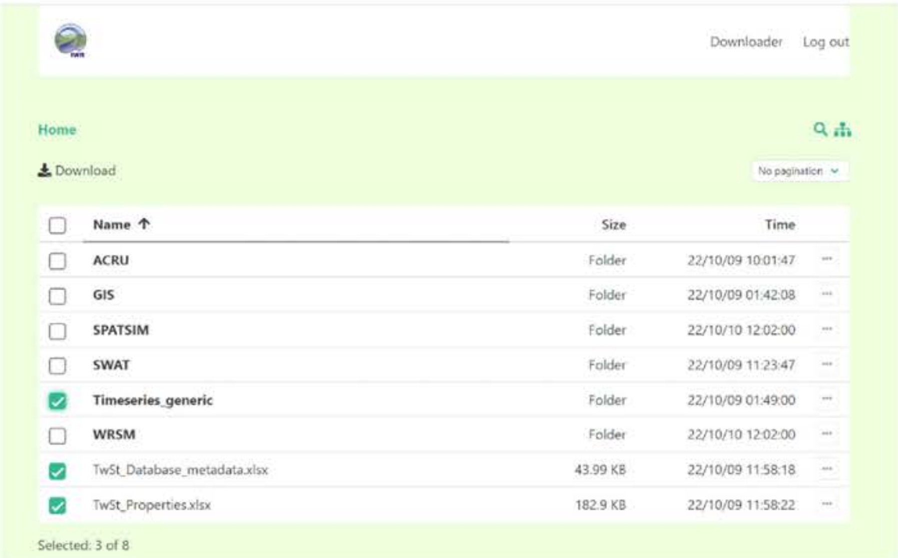


Figure 1 Screenshot of the online Model-a-thon database drive, accessible using the "Downloader" log-in. Once some items have been "checked" in the boxes to the left, the option to download them appears at the top of the screen.

2.3 Model the catchment with baseline land cover

2.3.1 Review the site description & database

- A site description and details regarding the input data, including modelling tool specific considerations has been provided in **Sections 3 & 4** of this document.
- An observed streamflow timeseries, climate data timeseries, and spatial and property data on topography, land cover, soil, and hydrogeology are provided in the case study database. **TwSt_Database_metadata.xlsx** contains a list of file names and details for time-series data, property table data, and GIS data.
- To reduce model set-up time, some data have been made available in multiple formats, timesteps, and units, both generic and as needed by some of the commonly used hydrological modelling tools in SA (ACRU, WRSM-Pitman, SPATSIM-Pitman, SWAT, MIKE-SHE).

2.3.2 Build a baseline model

- Using a catchment modelling tool of your choice, build a model of the case study catchment with its **baseline land cover**. This was the land cover present during the period when the streamflow dataset was collected.
- **Rules**: You are generally free to design your model as you see fit based on the information provided about the catchment. However, **every participant needs to use the same catchment size, climate input, and land cover type distribution. Do not adjust rainfall and evaporation timeseries**. The catchment-averaged rainfall and evaporative demand inputs that have been provided need to be used without adjustment to allow comparison across the models. ***Models which do not adhere to these rules will not be able to be included.***

2.3.3 Obtain modelled streamflow timeseries:

Run your model to obtain a predicted runoff or streamflow timeseries for the catchment outlet, having either a daily or a monthly timestep.

At a minimum, your output timeseries needs to cover the following period:
2003/10/01 to 2014/09/30

You may run your model for a longer time period than this date range if you wish to or need to. You do *not* need to cut your model output timeseries to this specified date range yourself. Please submit your full run-length outputs. The process of cutting all the submitted flow timeseries to the same analysis period will be automated by the model-a-thon organisers.

This 11 water-year period will be used to assess predicted changes between the baseline and the alternate cover scenario for all submissions. It is longer than the period for which observed streamflow data has been provided. Streamflow has been provided for a five-year period during which the catchment's vegetation properties were relatively stable.

2.3.4 File preparation for submission

You will need to submit:

- model files (all files needed to run your model)
- model output streamflow timeseries

<u>**Model folder:**</u>
Please put your model files into a folder named "P[participant number]_baseline"
e.g. if your participant number is 8, name your model files folder: "P08_baseline"
(Please use two-digit numbers, so 08 for number 8)

Compress or 'zip' the folder. This folder should contain the model project file and any linked, formatted input files that would be needed for someone to run your model, provided they had the modelling software.

<u>**Output streamflow file:**</u>
Please submit your baseline streamflow output timeseries as a separate file with the following specifications:

Data Format: 'long format', two column dataset: column 1 = dates, column 2 = flow values.

***Header row (column names)*:** column 1 = "Date", column 2 = "Q_[relevant units]", specifying the units of the output flow values.
Different modelling tools provide output in different units.
Column 2 header names can be any of the following:

- **"Q_cms"** for flow in m^3/s (average rate for the timestep)
- **"Q_m3"** for flow as m^3 (total flow volume for the timestep, day or month)
- **"Q_Mm3"** for flow as million m^3 (total flow volume for the timestep, day or month)
- **"Q_mm"** for flow as mm of runoff (depth per unit catchment area, for day or month)

File type: comma separated value (csv)
(You can prepare the file in Excel and 'Save As' a csv)

File name: "P[participant number]_Q_baseline.csv"
e.g. "P08_Q_baseline.csv", if you are participant 8

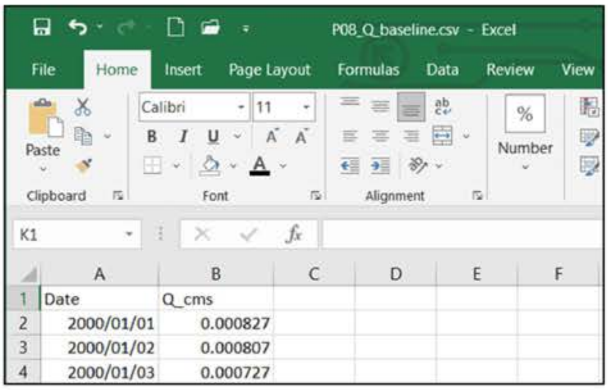


Figure 2 Demonstration of a streamflow timeseries .csv file prepared for submission.

2.4 Model the catchment with land cover change scenario

2.4.1 Modify your model to represent the change in land cover distribution

An alternate land cover scenario has been provided in the form of a new land cover map and land cover distribution table (described below in section 3.6 below). All cover types present in the alternate scenario were also present in the baseline case. The climate inputs must be that same as those used in the baseline model

2.4.2 Obtain modelled streamflow timeseries:

Run your modified scenario model to obtain a predicted runoff or streamflow timeseries for the catchment outlet, having either a daily or a monthly timestep. *The same as for the baseline scenario, at a minimum, your output timeseries needs to cover the following period:*
2003/10/01 to 2014/09/30

If your model output timeseries covers a longer time period, you do *not* need to cut the output to this period. You can submit your full output timeseries.

2.4.3 File preparation for submission

You will need to submit:

- model files (all files needed to run your model)
- model output streamflow timeseries

Model folder:
Please put your model files into a folder named "P[participant number]_scenario"
e.g. if your participant number is 8, name your model files folder: "P08_scenario"

Compress or 'zip' the folder. This folder should contain the model project file and any linked, formatted input files that would be needed for someone to run your model, provided they had the modelling software.

Output streamflow file:
Please submit your scenario streamflow output timeseries as a separate file.
See the file format specifications specified above for the baseline.

File name: "P[Participant Number]_Q_scenario.csv" (e.g. "P08_Q_scenario.csv")

2.5 Upload your submission

You will need to upload four things to complete the model-a-thon activity:

1. Zipped/compressed folder containing your **model files for your baseline model**. *(e.g. "P08_baseline" if you are participant 8).*
2. **Baseline model streamflow output timeseries** as a comma separated value (.csv) file *(e.g. "P08_Q_baseline.csv")*
3. Zipped/compressed folder containing your **model files for your change scenario model**. *(e.g. "P08_scenario").*
4. **Scenario model streamflow output timeseries** as a comma separated value (.csv) file *(e.g. "P08_Q_scenario.csv")*

Submissions must be uploaded to an online drive that can be accessed through the following web address:

https://iwr.ru.ac.za/MURRA/

User name: Uploader
Password: Model@Thon_2

The drive is accessed via the same web address as used for the inputs database. If you are still logged in as "Downloader", you will need to log out and then log in again, this time as "Uploader".

- You will land in a main folder which has subfolders by participant number. Find the folder with your participant number (e.g. P08) and click on the folder name to open it.
- Once the folder is open, you will see an icon on the top of the page which is labelled "Add files". Click on this to get a browser window where you can navigate to your submission files.
- You can select multiple files to upload at once if you wish. Click "Open" and they will upload. There is a size limit of 200 MB per file, however there is a way around this, described below.

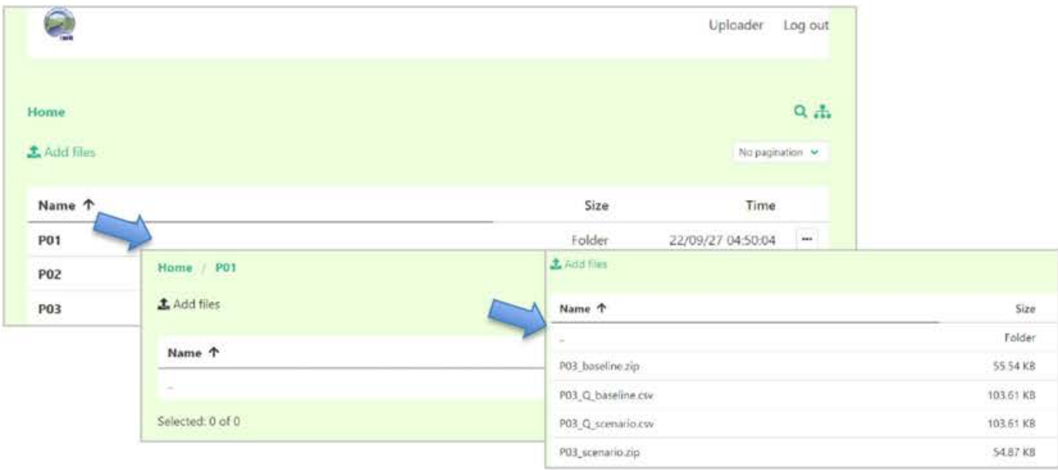


Figure 3 Screenshots of Uploader online drive, in which participants can add submissions into their participant number folder

2.5.1 I've uploaded the wrong file, now what?

You will only have 'upload' capabilities within the submissions drive. This way no participant can open, download, or delete any other participants' submissions. Unfortunately, this also means that if you upload the incorrect file by mistake, you cannot delete it from the drive yourself.

If you think that you have uploaded the wrong thing, simply upload a new version of the file and/or zip folder with the suffix "_1" added at the end of the filename (e.g. a revised upload of the file *P08_Q_baseline.csv* should be named *P08_Q_baseline_1.csv*).
If you find you need to do this another time, add the suffix "_2" to the new file, and so on. The model-a-thon project team will use the file with the highest number suffix added to the file name as your final submission.

2.5.2 Uploading large files or zipped folders (over 200 MB)

The online submissions drive has loads of space, however the server has an upload size limit of 200 MB per file. Once you have prepared your four items to be submitted, please check their sizes. If anything is over 200 MB, please follow these instructions:

1. Download & install **7zip** (free software) from https://www.7-zip.org/download.html

2. Use 7zip to compress the large file (or an already compressed/zipped folder) and split it across several smaller .zip files. Follow these instructions:
   https://www.webhostinghub.com/help/learn/website/managing-files/split-file
   (clicking on each image will enlarge it so you can read the text). Select 20 MB file size.

3. You now have several files named like this (if the original file was named "mybigfile"):
   mybigfile.zip.001
   mybigfile.zip.002
   mybigfile.zip.003
   and so on, up to 99 files

4. Upload these into your participant submissions folder as per the instructions above. As would be the case for any other files you want to upload, you can select all of these .zip files to be uploaded at one time. You do not need to upload them individually.

The project team will use the 7zip programme to "reconstitute" your original file (or your original compressed/zipped folder) from this set of .zip files when processing your submission.

If you need technical assistance, please send your query to model.a.thon@gmail.com

2.6 Complete submission survey

After you have uploaded your submission, you will be sent a short submission survey as a final step in completing the activity.

The results of this survey will be used to:

- better understand the modelling strategies applied by participants, and
- improve activity design if another model-a-thon is held in future.

3 Case study site description: Two Streams catchment

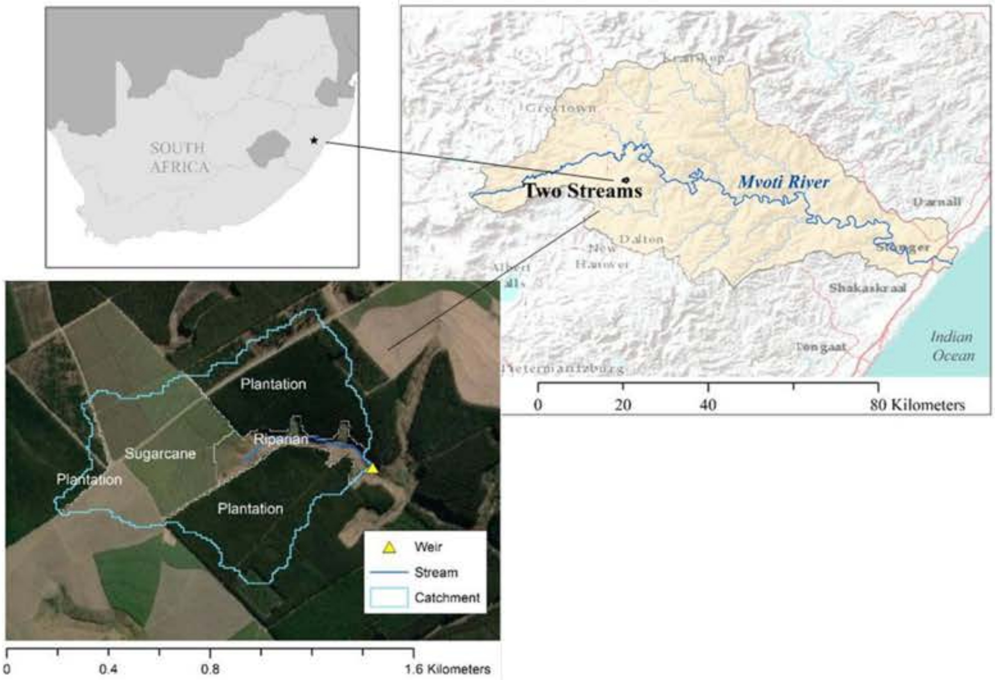


Figure 4 Two Stream catchment area, location within South Africa and the larger Mvoti River catchment

3.1 General catchment description

The case study catchment for this exercise is part of the "Two Streams" site, a research site established by the University of KwaZulu-Natal (UKZN) with assistance and collaboration from the Water Research Commission (WRC), Mondi, Department of Water and Sanitation (DWS, previously DWAF), Centre for Scientific and Industrial Research (CSIR), University of the Free State (UFS), and South Africa Environmental Observation Network (SAEON). It is the ~0.738 km^2 catchment area of a gauging weir on one of the two streams present at the site. It's located in the midlands of KwaZulu-Natal (KZN), ~50 km north-east of Pietermaritzburg and ~17 km south of Greytown. The stream is part of the headwaters of the Mvoti River and lies within quaternary catchment U40C. The site is characterised by rolling hills and has an average slope of 9.2% (5.3°). The catchment covers an elevation range of 76 m, from the weir at 1093 m above sea level (m.a.s.l), up to 1169 m.a.s.l.

Close to two thirds of the catchment area was planted with exotic *Acacia mearnsii* (black wattle) trees in June 2006. The planted area excluded the riparian zone, about 9% of the catchment. The remainder is non-irrigated commercial sugarcane. Regular field monitoring of the *A. mearnsii* plantation showed rapid growth in the first three years and indicated that by late 2009 the rate of increase in height and leaf area index (LAI) had tapered off significantly (Everson et al 2014, 2018). For this reason, observed streamflow has been provided for October 2009 through September 2014 (five water-years), assumed to be a period with relatively constant vegetation properties.

3.2 Climate

The catchment falls in the summer rainfall zone, with monthly precipitation typically peaking in January and being lowest in June. The mean annual rainfall was 850 mm for the fourteen-year period from October 2000 through September 2014. The mean daily maximum and minimum temperatures were 24°C and 12°C for this period, while the means of the annual maxima and minima were 37°C and 0.7°C. The average annual reference potential evapotranspiration (PET) was 1080 mm (FAO-56 method, Allen et al 1998).

Through various projects, automatic weather stations (AWS) were installed roughly 2 km from the catchment, at an elevation similar to the catchment average, one near the Seven Oaks farm (active since 1999) and one in an open grass area conforming to FAO guidelines for estimating reference PET (active since 2007); two tipping bucket rainfall gauges were also installed at different elevations within 500 m of the catchment (active 1999/06 -2009/08); and an additional AWS was installed on a mast above the tree canopy near the centre of the catchment (active since 2012) (Everson *et al.*, 2007, 2014, 2018; Clulow, Everson and Gush, 2011). No definitive elevation gradient in rainfall was detected across these instruments, which was not unexpected given the relatively small catchment size and low elevation range.

For the purpose of this exercise, it will be assumed that the rainfall and atmospheric evaporative demand (and hence its drivers: temperature, solar radiation, wind speed, relative humidity) are spatially uniform over the catchment. Estimated catchment spatial average daily and monthly climate timeseries have been provided for each variable for **2000/01/01 - 2014/12/31**, based on the instrument records.

3.3 Streamflow and channel

Streamflow has been monitored since 1999 using a V-notch weir with continuous stage logging equipment, converting stage to flow using rating tables (Everson *et al.*, 2018). Daily and monthly streamflow timeseries have been provided for **2009/10/01 - 2014/09/30**. This five-year period was selected because it represents a time in which the vegetation properties were relatively stable (*A. mearnsii* tree growth rates had tapered, see above) and because there were notable gaps in the streamflow record from late 2014 through 2015 due to equipment failure. During this 2009/10/01 - 2014/09/30 period the mean annual precipitation was 658 mm while the mean annual catchment runoff was estimated to be 23.8 mm, indicating a rainfall runoff ratio of about 3.6%. The mean daily flow rate at the weir was 0.00056 m^3/s, equating to 48 m^3 in a day, with daily flow rates ranging from 0.000017 m^3/s (1.5 m^3/day) up to a max of 0.03 m^3/s (2,514 m^3/day).

IMPORTANT NOTE: *Given the small catchment area and low flows, modelling software tools designed to handle flows in* Mm^3 *per month to a few decimal places, such as* ***SPATSIM-Pitman*** *and* ***WRSM-Pitman****, require special strategies. The mean monthly flow was only 0.0014* Mm^3. ***Please see section 4.4 below*** *for guidance on how to handle this if needed for the tool you are using.*

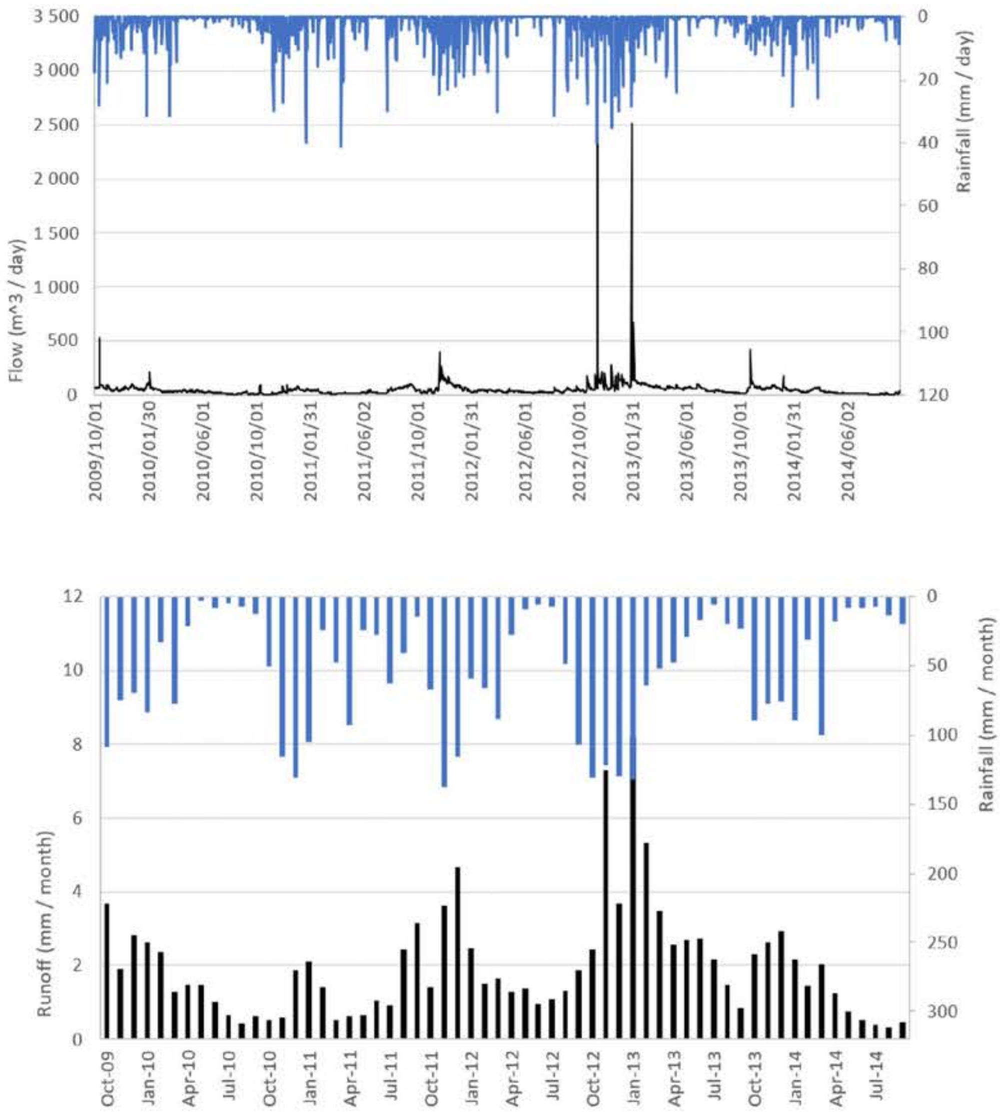


Figure 5 Daily (top) and monthly (bottom) rainfall and streamflow or runoff for the Two Streams catchment, 2009/10/01-2014/09/30

The estimated bank-full stream channel width is less than 1 m in the upper reaches, widening to about 2 m downstream at the weir. The channel depth is roughly 10 cm in the upstream, 30 cm in the middle, and close to 1 m at the downstream end at the weir. The channel has a sediment bed for most of its length, but there is exposed bedrock at the downstream end at the weir.

3.4 Geology and hydrogeology

In the national-scale geological map (1:250000) of the Council of Geoscience, the Two Streams catchment occurs in an area of Natal Group sandstone outcrop; however, boreholes drilled in and around the catchment to depths of 60 m found Pietermaritzburg Formation shales of the Ecca Group overlying a granite basement (Clulow, Everson and Gush, 2011; Ngubo, 2019). No sandstones were encountered. The geology has been deeply weathered by the relatively warm and wet conditions. Weathered material has eroded away along the drainage lines such that the weathered material layer is thickest at hillslope crests, becoming thinner downslope toward the stream channel, which lies on bedrock in some reaches (Figure 6).

Soil pits, auguring, and drilling logs (Everson *et al.*, 2007; Van Tol *et al.*, 2007; Clulow, Everson and Gush, 2011; Ngubo, 2019) showed several main layers of material:

- **Sandy clay loam soils** grading into red-brown clay material. Soils have been sampled to depths of 1-2 m.
- **Red-brown clay material** that extends down to roughly 24 m below ground level (m.b.g.l) at the crest with the thickness decreasing downslope to be about 15 m.b.g.l at the midslope and near 0 m at the stream channel.
- **Dark grey weathered shale** underlying the clay material. The shale layer is estimated to be 10-22 m thick. Near the crests, it extends from roughly 24 m.b.g.l down to 43 m.b.g.l. Further downslope it starts at around 15 m.b.g.l and extends down to 25 m.b.g.l, becoming thinner toward the channel.
- **Basement granite** found from 43 m.b.g.l downwards at the crests, while being exposed or very shallow downslope by the stream channel.

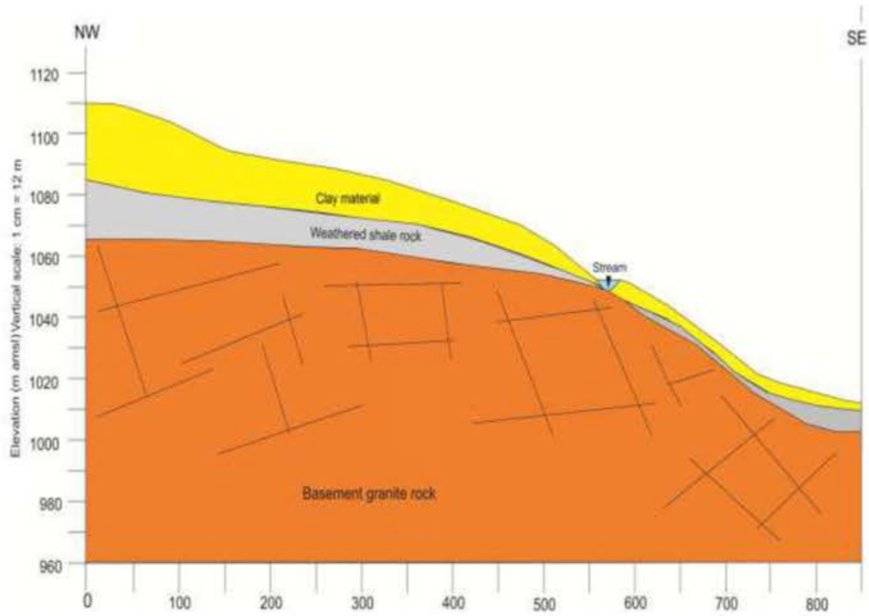


Figure 6 Approximate conceptual cross section of Two Streams catchment from NE to SW (Ngubo 2019)

The groundwater table was observed at 35-45 m.b.g.l. near the crest, 20-30 m.b.g.l. at the midslope, while there is a water table within 2.5 m of the surface in the riparian zone, feeding streamflow (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014, 2018; Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022). At the upland boreholes the water table has been observed to fluctuate by up to 10m over a matter of months in response to wetter and drier periods (Everson *et al.*, 2018).

Water strikes, borehole and piezometer water levels, and resistivity transects indicate that the water table generally lies at depths that fall within the weathered shale layer, presumably in fractures, and in the upper part of the basement granite, likely to be more weathered (Clulow, Everson and Gush, 2011; Ngubo, 2019). In addition, water may also be perched at the base of the clay material, at the interface with the underlying weathered shale bedrock (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014, 2018).

Aquifer properties have been estimated through local borehole pumping tests and through regional studies as shown in Table 1.

Table 1 Aquifer properties estimated by various sources (properties also provided in ***TwSt_Properties.xlxs****)*

Source	Specific yield (unconfined)	Transmissivity (m^2/day)	Saturated Conductivity (m/d)
Local borehole pump testing (Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022)	0.03	0.15 - 0.48	0.04
KZN regional study (DWAF, 1995)	0.005 - 0.0005		0.4 - 7.7
Estimate for U40C quaternary, GRAII (DWAF, 2006)	0.001		
Weathered granite (Brassington, 2006)			0.0003-0.03

3.5 Soils

Soils in the catchment are generally deeply weathered sandy clay loam soils derived from Ecca Group rock. Soils generally have an orthic A-horizon underlain by an apedal red or yellow B-horizon. The soil forms identified in the catchment are predominantly Griffin in the uplands, Hutton on the midslopes on the north-facing side, Clovelly on the midslopes on the south-facing side, and Katspruit in the riparian zone (Van Tol *et al.*, 2007). An approximate spatial distribution of these types has been provided both as a shapefile and a raster.

Litter and organic matter lead to relatively low bulk densities in surface layers. In general, the clay concentrations increase with depth over the profile as the material grades towards the clay layer that overlies the bedrock. Silt deposits near the stream channel differentiate the riparian zone. Soil profiles have been sampled and described to depths of 1-2 m across the catchment area for various studies (Everson *et al.*, 2007; Van Tol *et al.*, 2007; Clulow, Everson and Gush, 2011; Scott-Shaw, 2020). These have been summarised in Table 2. It should be noted that these properties are averages across several sampled pits in each type and are meant to provide 'ballpark' values. In addition, the soils properties provided in the 'AutoSoils' database created for national application of the ACRU model have also been provided (Schulze, 1995).

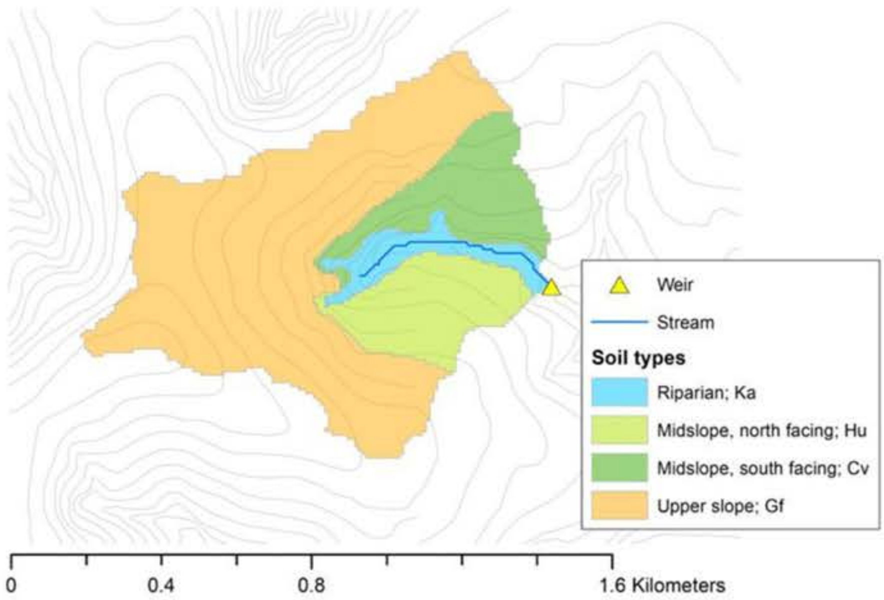


Figure 7 Soil type map for the Two Streams catchment area (approximated from van Tol et al 2007)

*Table 2 Estimated soil properties for the Two Streams catchment (property values also provided in **TwSt_Properties.xlxs**)*

Zone	Dominant Form	Area (ha)	% of catch-ment	Profile Depth (mm b.g.l), RANGE	Layer bottom depth (mm b.g.l), MEAN	Texture	Clay %	Silt %	Sand %	Rock %	Organic %	bulk density (g/cm³)	Sat VWC	FC VWC	WP VWC	PAWC (FC-WP)	K_{sat} (mm/hr)
LOCAL SAMPLING *(Everson et al., 2007; Van Tol et al., 2007; Clulow, Everson and Gush, 2011; Scott-Shaw, 2020)*																	
Riparian	Katspruit	4.69	6%	600-1200	50	silty clay loam	29	55	16	0	12.5	1.00	0.570	0.440	0.272	0.168	151
					350	sandy clay loam	25	19	56	0	7.7	1.16	0.537	0.364	0.228	0.136	172
					1000	sandy clay loam	32	14	54	0	0.4	1.22	0.436	0.288	0.143	0.145	160
Toeslope - Midslope, North-facing (South side)	Hutton	10.31	14%	1200-1800	50	sandy clay loam	25	22	53	1.1	5.9	1.12	0.436	0.288	0.143	0.145	94
					500	sandy clay loam	31	19	50	1.4	3.6	1.19	0.508	0.267	0.152	0.116	127
					1500	sandy clay	37	16	47	1.5	0.9	1.23	0.499	0.258	0.165	0.093	133
Toeslope - Midslope, South-facing (North side)	Clovelly	10.5	14%	1200-1800	100	sandy clay loam	22	25	53	1.1	2.3	1.19	0.430	0.280	0.140	0.140	127
					500	sandy clay loam	29	18	54	1.4	0.8	1.25	0.537	0.364	0.228	0.136	140
					1500	sandy clay loam	32	14	54	1.5	0.3	1.35	0.436	0.288	0.143	0.145	130
Uplands	Griffin	48.31	65%	1400-2000	200	sandy clay loam	29	16	55	1.6	2.3	1.24	0.537	0.364	0.228	0.136	140
					500	sandy clay loam	33	14	53	1.5	0.8	1.35	0.436	0.288	0.143	0.145	130
					1800	sandy clay	36	16	48	1.3	0.3	1.50	0.434	0.310	0.155	0.155	85
AUTOSOILS DATABASE *(Schulze, 1995)*																	
Ad23		73.73	99.9%	1200	300								0.404	0.271	0.170	0.103	
					1200								0.423	0.327	0.209	0.118	
Bb107		0.09	0.1%	760	300								0.449	0.212	0.130	0.086	
					760								0.422	0.244	0.164	0.080	

b.g.l: below ground level, **VWC:** volumetric water content, **Sat:** Saturated, **FC:** field capacity, **WP:** wilting point, **PAWC:** plant available water content,
K_{sat}: Saturated hydraulic conductivity (NOTE: only measured directly in riparian area & north-facing toe-slope)

3.6 Land cover distribution & properties

Baseline: The baseline land cover distribution represents the state of the catchment as it was for the period 2009/10/01 - 2014/09/30 (Figure 4 above, Table 3). During this time 59% of the catchment area was covered by exotic tree plantations, the vast majority (98%) of which were *A. mearnsii* (black wattle) in an intermediate to mature growth phase. Trees were not harvested at all in this period. The riparian zone (9%) of the catchment was not planted with trees and was mostly herbaceous, although a few woody shrubs and young wattle were present. The remaining 32% was non-irrigated commercial sugar cane. The spatial distribution of these cover types has been provided both as a shapefile and a raster.

Change scenario: In the land cover change scenario, it is assumed that all the tree plantation area is *completely cleared* and replaced with sugar cane, while the riparian zone remains unchanged. The vegetation properties of the sugar cane and riparian area are assumed to be the same as they were in the baseline case. The spatial distribution of these cover types has been provided both as a shapefile and a raster.

*Table 3 Baseline and change scenario land cover types and coverages in the Two Streams catchment (table also provided in **TwSt_Properties.xlxs**)*

Cover	Description	BASELINE		CHANGE SCENARIO	
		Area (ha)	% of catchment	Area (ha)	% of catchment
Riparian vegetation	Dominated by herbaceous vegetation (mostly *Setaria megaphylla*, broad leaf bristlegrass), *scattered invasive shrubs and young wattles (<2 m tall)*	6.48	9%	6.48	9%
Sugarcane	Commercial sugar cane, not irrigated	23.72	32%	67.34	91%
Plantation	98% of area: *Acacia mearnsii* (black wattle) trees planted in June 2006, 2% *Eucalyptus grandis*	43.62	59%	0	0%

Some estimated properties of these vegetation types available in literature have been provided in the *TwSt_Properties.xlxs* file, which you may choose to use to inform your modelling. These include values from local studies on the *A.mearnsii* stands at Two Streams (Clulow, Everson and Gush, 2011; Everson *et al.*, 2014, 2018; Clulow *et al.*, 2022) and from various studies on sugarcane (Teruel, Barbieri and Ferraro Jr., 1997; Allen *et al.*, 1998; Smith, Inman-Bamber and Thorburn, 2005; Gilbert *et al.*, 2008; Cardoso *et al.*, 2015).

The vegetation property values provided in *TwSt_Properties.xlxs* include parameter values from the 'Compoveg' database that is built into ACRU3 and ACRU4 software (Schulze, 1995; Clark *et al.*, 2012). These are values are provided by month-of-year. Parameters from this database have been provided for vegetation thought to be appropriate to represent the riparian zone (Natal Mist Belt 'Ngongoniveld', Acocks veld type 45), as well as for sugar cane and wattle plantation. The default type of "crop coefficients" (CAY values) used in ACRU use A-pan evaporation as the evaporative demand. A suggested conversion to crop coefficient values appropriate to FAO-56 grass reference potential evapotranspiration (Kc) has been included (multiply A-pan coefficients by 1.2, Schulze et al 1995). Both sets of values are listed in the *TwSt_Properties.xlxs* file. A recent study, (Toucher *et al.*, 2018), undertook a reassessment of the vegetation property values assigned to indigenous vegetation types in ACRU at a national scale. Parameters from this study for the vegetation type cluster mapped at Two Streams (Grassland cluster 16: KZN midlands, grass 82%, forbs 11%, shrubs 6%, trees 1%) have also been provided.

Notes on sugarcane property values:

Values from the sugarcane studies are given by growth phase starting from planting, or from harvest for a ratoon crop. (Sugarcane can be harvested and allowed to regrow from the remaining stalks for one or two cycles before a new planting, a practice called ratooning.) In the *TwSt_Properties.xlxs* file you will need to scroll beyond the month-of-year values, to the far right of the data sheet to see the growth stage values.

The South African Department of Agriculture Forestry, and Fisheries (DAFF) suggests that rain-fed sugarcane in the KZN midlands should ideally be planted September to October, although autumn planting (March-April) is possible. Cane is generally ready to harvest 12-16 months after planting/harvest+ratoon and harvest is better done in drier periods, which is winter in KZN (DAFF, 2014). At Two Streams the sugarcane area was harvested and replanted or ratooned in different blocks over time, such that there was likely to have been mix of growth stages present at any given time. The ACRU Compoveg database entry for sugar cane gives the same property values for every month of the year. It is likely that this is based on the assumption that a large area of sugar cane is being represented, with different blocks at different growth stages, such that the spatial average for the area over time may be the mid-growth values.

*Table 4 Vegetation property values relevant to cover types in the Two Streams catchment (these values, and **additional,** provided in **TwSt_Properties.xlxs**)*

Cover	Property / Parameter	Source	Source cover type	units	Average value
Riparian vegetation	**Leaf area index (LAI)**	Toucher et al 2018	Grassland cluster 16		1.03
	Root depth	Toucher et al 2018	Grassland cluster 16	mm	1470
	grass PET reference (FAO-56) crop ET coefficient (K_c)	*Toucher et al 2018*	*Grassland cluster 16*		*0.82*
		ACRU Compoveg database	*Natal mist belt (2030119)*		*0.63*
	A-pan reference crop ET coefficient (CAY in ACRU)	Toucher et al 2018	Grassland cluster 16		0.69
		ACRU Compoveg database	Natal mist belt (2030119)		0.52
	max daily interception (interception loss coefficient)	Toucher et al 2018	Grassland cluster 16	mm/rain day	1.88
		ACRU Compoveg database	Natal mist belt (2030119)	mm/rain day	1.34
Sugarcane	**Leaf area index**	Teruel et al 1997	Sugarcane		1.99
		Gilbert et al 2008	Sugarcane		1.66
	Root depth	Smith et al 2005	Sugarcane	mm	1436
	grass PET reference (FAO-56) crop ET coefficient (Kc)	Cardoso et al 2015	Sugarcane		0.99
		ACRU Compoveg database	*Sugarcane (9010102)*		*0.96*
	A-pan reference crop ET coefficient (CAY in ACRU)	*Cardoso et al 2015*	*Sugarcane*		*0.82*
		ACRU Compoveg database	Sugarcane (9010102)		0.8
	max daily interception (interception loss coefficient)	ACRU Compoveg database	Sugarcane (9010102)	mm/rain day	1.8
Plantation: *Acacia mearnsii* (black wattle)	**Leaf area index**	Clulow et al 2022	*A. mearnsii*, intermed-mature		1.75
	Root depth	Clulow et al 2011	*A. mearnsii*, intermed-mature	mm	8000
	grass PET reference (FAO-56) crop ET coefficient (Kc)	Clulow et al 2022	*A. mearnsii*, intermed-mature		1.20
		ACRU Compoveg database	*Wattle (5083203)*		*1.07*
	A-pan reference crop ET coefficient (CAY in ACRU)	*Clulow et al 2022*	*A. mearnsii, intermed-mature*		*1.00*
		ACRU Compoveg database	Wattle (5083203)		0.89
	max daily interception (interception loss coefficient)	ACRU Compoveg database	Wattle (5083203)	mm/rain day	1.94

4 Notes on input preparation for specific modelling tools

All files available in the input database are described in the **TwSt_Database_metadata.xlsx** file. Data sources are also specified in this file. The database includes the data files specifically formatted for a selection of commonly used modelling tools.

4.1 ACRU

For ACRU users, the climate data for the catchment has been pre-prepared in the "ACRU Composite Y2K" txt file format found in the ACRU folder in the project database. Other information required to set up a catchment model in ACRU are available in the *TwSt_Properties.xlsx* file.

4.2 MIKE-SHE

All the raster spatial data prepared for this exercise have been aligned to a 10m grid with orientation specified in the following file in the MIKE folder: *TwSt_MIKE_model domain set-up.xlsx*. Using these specifications in your 'model domain' set-up will avoid resampling spatial data in the model's pre-processing step (which may lead to differences in the areas of the cover types). If you wish to change the model grid cell size in your model, please stick to multiples of 10 and alter the number of X and Y grid cells in your domain accordingly.

Daily climate and observed streamflow time series have been prepared as dfs0 files (*TwSt_refPET.dfs0, TwSt_Rainfall.dfs0, TwSt_ObsFlow.dfs0*) and the DEM as a dfs2 file (*TwSt_DEM.dfs2*) available in the MIKE-SHE folder.

4.3 SWAT

The climate inputs for SWAT, formatted text files of daily rainfall, max and min temperature, windspeed, relative humidity, and solar radiation, as well as station location files, have been prepared and provided in the SWAT/Clim folder. The same formatting can apply in both SWAT2012 and SWAT+, however in SWAT+ set-up the user must specify that the SWAT2012 climate data format is being used during the import step in the SWAT+ Editor. As the values are assumed to apply to the whole catchment area, the "station" location is the centroid of the catchment area polygon. See the *TwSt_Database_metadata.xlsx* file for the file names.

Land cover type and soil type rasters files (projection UTM 36S, resolution 10 m) have been pre-prepared with integer grid codes. To set up a SWAT model using these type maps, the user need only prepare the look up table that links these integer codes to the land cover type and soil type entries in the SWAT databases you are using. You may modify or regroup the soil type map if you wish to do so, or use it as it is, however the land cover type distribution maps must be used as provided.

4.4 Pitman

In both WRSM-Pitman and SPATSIM-Pitman, the observed flow and volumetric flow and storage outputs use units of million cubic metres per month (Mm^3/mon), specifically to two decimal places in WRSM and three decimal places in SPATSIM. Because of the small size and low flow rates of this case study, Mm^3/mon flow values to two or three decimal places would indicate that there was no flow (0.00 Mm^3/mon) in many months.

To avoid this problem, ***please set up your WRSM and SPATSIM models using a catchment area and land cover type areas that are 100 times larger than the actual values: i.e., using a total area of 73.82 km² and a wattle plantation area of 43.62 km²***. All depth unit inputs and outputs and other per area properties will apply unmodified. Any volumetric model outputs will be 100 times greater than expected for the actual catchment. ***You do NOT need to adjust your output data in your submission. The model-a-thon organisers will automate the re-scaling the WRSM and SPATSIM volumetric flow outputs by 100x during submissions processing.***

4.4.1 SPATSIM-Pitman

In the SPATSIM data folder, a text file of monthly rainfall data has been provided cover the dates 2000/10 to 2014/09*, TwSt_rain_month_SPATSIM.txt*. Two formatted evaporation data files have been provided, one for using FAO-56 Penman Monteith method grass reference PET (*TwSt_refPET_month_%ofannual.txt*) and one for using S-Pan evaporation as the evaporative demand input (*TwSt_Span_evap_month_%ofannual.txt*). In both cases the files contain the month-of-year average evaporative demand expressed as a percentage of the mean annual value. The mean annual reference PET was 1080 mm for 2000/10 to 2014/09 and the estimated mean annual S-pan evaporation was 994 mm for this period. An observed flow text file (*TwSt_flowx100_month_SPATSIM.txt*) has been provided with values that are 100 x actual observed flow in Mm^3/mon, with rows for each water year and columns of values for months October to September.

4.4.2 WRSM-Pitman

In the WRSM data folder, a WRSM format rainfall input text file (*TWS.RAN*) has been provided which covers the dates 2000/10 to 2014/09, for which the MAP is 850 mm. Month-of-year average estimated S-pan evaporation for Oct to Sept is provided in *Evap demand_WRSM.csv*. A WRSM format observed flow text file (*TWS100.OBS*) has been provided with values that are 100 x actual observed flow in Mm^3/mon.

5 Frequently asked questions (FAQ)

Are the input climate data and the observed streamflow data free of errors?
No, they are not! The datasets provided are based on data collected by instruments in and around the site. The instruments were regularly checked and calibrated and there has been some data quality control. However, there may be errors introduced by things like instrument accuracy, undetected instrument problems, operator error, spatial distribution of the instruments and spatial averaging of data from different point measurements, etc. For example, there were periods during which not all the rainfall gauges in the area were operating and so the spatial average is based on data from fewer points than at other times.

However, these issues are typical for any modelling effort. The comparison *across submissions* is the focus of this model-a-thon. All participants will be using the same climate input and the outputs will be compared to the same observed streamflow data, making the efforts comparable to one another regardless of the data limitations.

Do I need to stick to the land cover, soil, and aquifer property parameter values or value ranges provided?
No. Various property values have been provided from local studies, national level databases, and other relevant literature sources as points of reference. However, based on your experience with modelling other areas with similar cover, soil types, or geology in your modelling tool of choice, or other relevant information you choose to refer to, you may find different values to be reasonable for use in this case.

Why is the assessment period (2003/10/01 to 2014/09/30) different to the period for which observed streamflow data has been provided (2009/10/01 - 2014/09/30)?
Observed streamflow data has been provided for a period during which the land cover distribution and cover properties were relatively stable. The wattle was planted in the catchment in mid-2006 and the rate of change in tree height and leaf area index (LAI) was observed to taper by late 2009 (Everson et al 2014). Data for this period should be a more "fair" comparison to models in which the land cover types and their properties do not change from year to year.

Participants are asked to run their models for a longer "assessment period" in order to predict how the catchment would likely have responded to a wider range of weather events and conditions with the different land cover distributions (baseline and cover change scenario). This assessment period was selected based on local climate data availability for the site.

How will the submissions be analysed?
Modelled streamflow outputs using the baseline land cover will be compared to the observed streamflow timeseries and standard statistics of 'model fit' will be calculated: MAE, PBIAS, RMSE, NSE, NSE of logged flow, R^2, KGE, errors in flows predicted for five different sections of the flow duration curve: 0-10%, 10-40%, 40-60%, 60-90%, 90-100% exceedance. Statistics will be calculated at a monthly timestep for all submissions, to allow comparison across daily and monthly timestep

models, as well as at a daily timestep where relevant, to compare across the daily timestep model subset.

Modelled streamflow for the assessment period (2003/10/01 to 2014/09/30) for the baseline and the cover change scenarios will be compared to determine the predicted change in mean annual flow, minimums, maximums, and the five sections of the flow duration curve mentioned above. Again this will be done using a monthly timestep for all submissions as well as a daily timestep for the subset of daily timestep submissions.

The range in performance against observed streamflow and the range of change predictions across all submissions will be assessed. This will also be looked at for different subsets of submissions: those from new model users, those from expert model users, those using different modelling tools. Analyses will be dependent on the pool of submissions received.

For a subset of submissions, the modelling files submitted will be used to extract the modelled water balance for the catchment for the assessment period, including modelled canopy interception, actual ET from soil and groundwater, change in soil and groundwater storage, contributions streamflow from surface flow, interflow, and groundwater. This will be done by the model-a-thon organisers (MURRA project team). Modelled water balances across this subset will be compared.

Why do I need to provide all of my model files, instead of just submitting my modelled streamflow output?

For a subset of submissions, selected across tools and performance levels, modelled water balances will be extracted for comparison (see above section on submission analyses). To reduce the time and effort required of volunteering participants, this will be done by the MURRA project team. This requires the original modelling files and, in some cases, re-running the model itself with certain output saving specifications. Interested participants who wish to engage further and work on co-authoring papers from model-a-thon results are welcome to participate in this process. There is also the possibility of looking the parameter ranges used across different models done using the same modelling tool. Also, for cases with anomalous output the project team may wish to look at the original model set-up.

Can I ask other people for assistance with setting up my models for the model-a-thon?

To a certain degree, yes, however we are aiming to have submissions that represent an individual's modelling effort based on their current background and experience with modelling.

If you have questions about the model-a-thon exercise instructions, the input database, or the submissions uploading process, you can get in touch with the model-a-thon organisers via model.a.thon@gmail.com

If you are having technical *modelling software* issues, for example your model won't run, you are getting error messages, you are getting highly nonsensical results, you can ask colleagues, other modelling experts, and/or online help forums for the software for assistance. The event organisers will not be providing this level of support.

6 References

Allen, R.G. *et al.* (1998) *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*. FAO Irrigation and drainage paper 56. Rome, Italy: Food and Agriculture Organization of the United Nations (FAO).

Brassington, R. (2006) *Field Hydrogeology*. London, UK: Open University Press (Geological Society of London, Handbook Series).

Cardoso, G.G. de G. *et al.* (2015) 'Sugar cane crop coefficient by the soil water balance method', *African Journal of Agricultural Research*, 10(24), pp. 2407–2414. Available at: https://doi.org/10.5897/AJAR2015.9805.

Clark, D.J. *et al.* (2012) *ACRU 4 User Manual: User Interface & Tutorials (Volume 3 of Deployment, Maintenance, & Further Development of SPATSIM-HDSF)*. WRC Report 1870/3/12. Gezina, Pretoria: Water Research Commission (WRC).

Clulow, A.D. *et al.* (2022) *The expansion of knowledge on evapotranspiration and stream flow reduction of different clones/hybrids to improve the water use estimation of SFRA species (i.e. Pinus, Eucalyptus, and Acacia species).* WRC Report K5/2791. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Clulow, A.D., Everson, C.S. and Gush, M.B. (2011) *The long-term impact of Acacia mearnsii trees on evaporation, streamflow, and ground water resources*. WRC Report TT505/11. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

DAFF (2014) *Production Guideline: Sugarcane*. Pretoria, South Africa: Department of Agriculture Forestry, and Fisheries (DAFF). Available at: https://www.dalrrd.gov.za/Portals/0/Brochures%20and%20Production%20guidelines/sugar%20cane%20prodocution%20%20guideline.pdf.

DWAF (1995) *Characterization and Mapping of the groundwater resources KwaZulu-Natal Province.* Pretoria, South Africa: Department of Water Affairs & Forestry (DWAF).

DWAF (2006) *National Groundwater Resource Assessment, version 2 (GRA II)*. DWAF Report. South Africa: Department of Water Affairs & Forestry (DWAF).

Everson, C. *et al.* (2014) *The long term impact of Acacia mearnsii trees on evaporation, streamflow, low flows and ground water resources. Phase II: Understanding the controlling environmental variables and soil water processes over a full crop rotation*. WRC Report 2022/1/13. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Everson, C.S. *et al.* (2007) *Can effective management of riparian zone vegetation significantly reduce the cost of catchment management and enable greater productivity of land resources*. WRC Report K5/1284. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Everson, C.S. *et al.* (2018) *Continued hydrometeorological monitoring at the Two Streams Catchment for the clear felling of the Acacia mearnsii stand.* WRC Report K5/2780. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Gilbert, R.A. *et al.* (2008) 'Sugarcane Response to Mill Mud, Fertilizer, and Soybean Nutrient Sources on a Sandy Soil', *Agronomy Journal*, 100(3), p. AGJ2AGRONJ20070247. Available at: https://doi.org/10.2134/agronj2007.0247.

Ngubo, C.Z. (2019) *An investigation of the impacts of Acacia mearnsii plantations on secondary aquifer systems within the Two-Streams catchment, KwaZulu-Natal, South Africa*. MSc dissertation. University of KwaZulu-Natal.

Ngubo, C.Z., Demlie, M. and Lorentz, S. (2022) 'Investigation of hydrological processes and the impacts of Acacia mearnsii plantations on groundwater in secondary aquifers: Case study at the two-stream research catchment, South Africa', *Journal of Hydrology: Regional Studies*, 40, p. 101018. Available at: https://doi.org/10.1016/j.ejrh.2022.101018.

Schulze, R.E. (1995) *Hydrology and Agrohydrology: A Text to Accompany the ACRU 3.00 Agrohydrological Modelling System*. WRC TT69/95. Pretoria, South Africa: Water Research Commission (WRC).

Scott-Shaw, B.C. (2020) *Hydrological modelling assessment and calibration for riparian clearing scenarios within the uMgungundlovu District Municipality, KwaZulu-Natal*. report to Biodiversity Planing & Managment. Hilton, South Africa: NatureStamp.

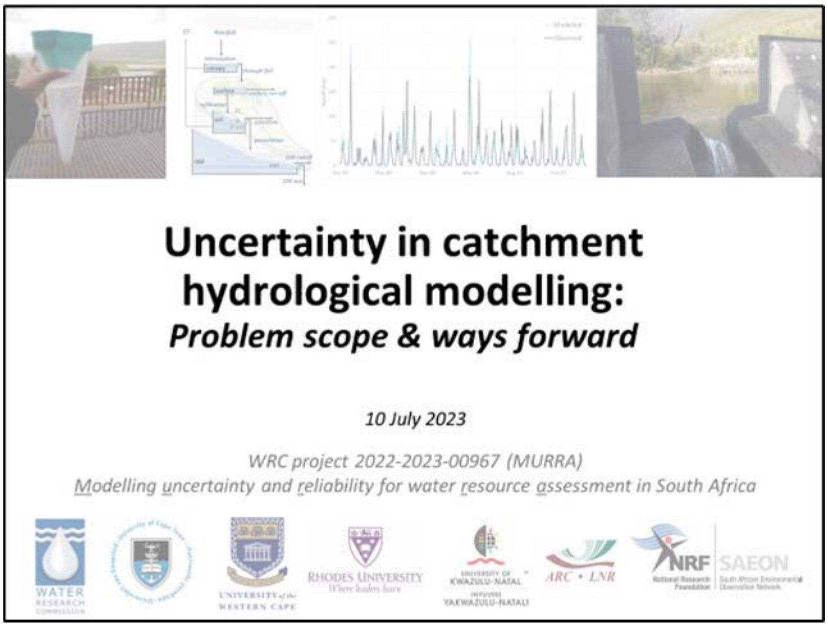
Smith, D.M., Inman-Bamber, N.G. and Thorburn, P.J. (2005) 'Growth and function of the sugarcane root system', *Field Crops Research*, 92(2–3), pp. 169–183. Available at: https://doi.org/10.1016/j.fcr.2005.01.017.

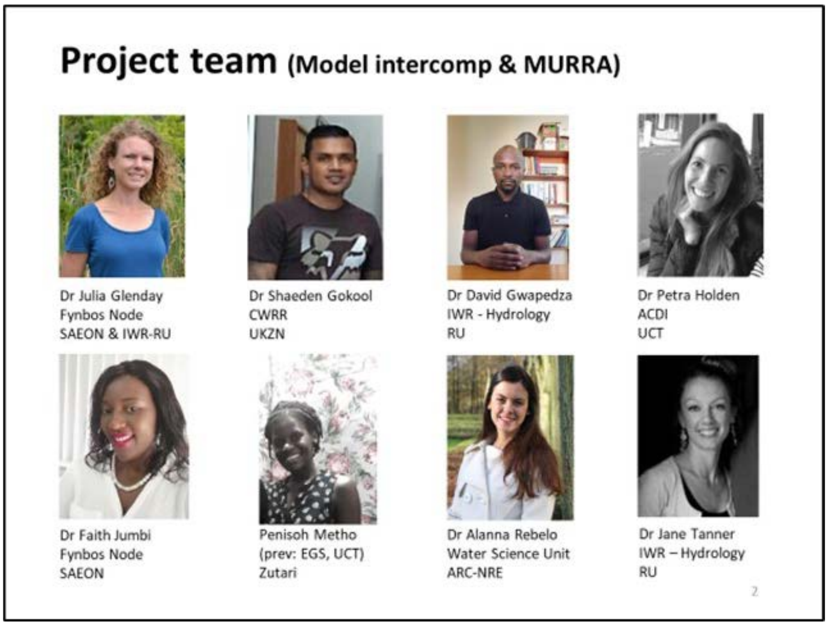
Teruel, D.A., Barbieri, V. and Ferraro Jr., L.A. (1997) 'Sugarcane leaf area index modeling under different soil water conditions', *Scientia Agricola*, 54(spe), pp. 39–44. Available at: https://doi.org/10.1590/S0103-90161997000300008.

Toucher, M. *et al.* (2018) *Resetting the baseline land cover against which streamflow reduction activities and the hydrological impacts of land use change are assessed*. WRC Report K5/2437. Gezina, Pretoria, South Africa: Water Research Commission.

Van Tol, J.J. *et al.* (2007) *Hydropedological survey report of the Two Streams Catchment*. unpublished report contributing to WRC K5/1748 & others. Bloemfontein, South Africa: University of the Free State.

APPENDIX B.1 – PRESENTATION ON UNCERTAINTY IN HYDROLOGICAL MODELLING FOR POLICY BRIEF DISCUSSION SESSIONS





Aim

Find common ground around hydrological modelling uncertainty

- *How big of a problem is it?*
- *What can we / should we do?*

(in different contexts)

Policy Brief

for:

- Modellers
- Users of model outputs
- Funders of modelling work

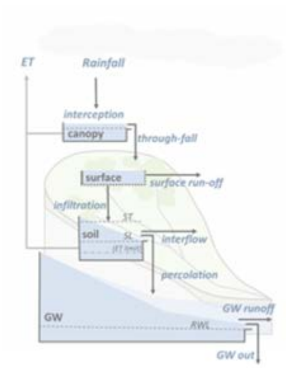
Programme

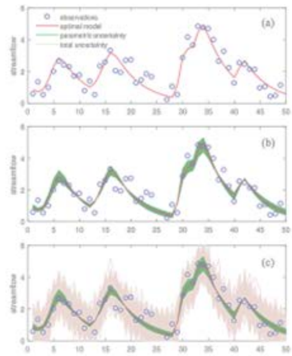
9:00 - 9:15	(15 min)	**Welcome & introductions**
9:15 - 9:40	(25 min)	**Modelling uncertainty presentation**
9:40 - 10:15	(35 min)	**Break-out group discussion 1: Implications (+ feedback)**
10:15 - 10:50	(35 min)	**Round-out group discussion 2: Recommendations (+ feedback)**
10:50 - 11:00	(10 min)	**Recap & way forward**

Introductions

- Name
- Institution
- Role/Position
- What **one word** best captures your thoughts on uncertainty in hydrological modelling in South Africa

Uncertainty in catchment hydrological modelling: *background & local research*





Kavetski, 2018

Error & uncertainty in model predictions

Error: how far model prediction values deviate from observed data values (& how often)

**accepts observed data as accurate…*

Uncertainty: range of values within which we think the real value will fall

i.e. We don't know the real value (ungauged catchment or future scenario), but we are (X%) confident that it will be between value A and value B.

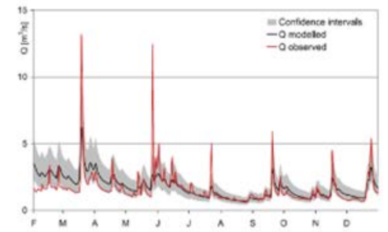


Figure 15. Standard calibration 80% confidence intervals for a multiplicative error model.

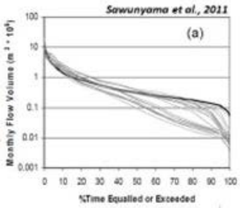
Götzinger & Bárdossy 2008

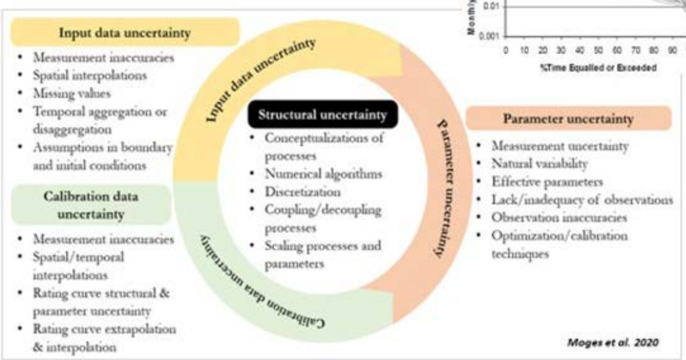
Often use model performance in gauged cases to estimate uncertainty for cases without observed data (ungauged, future).

Sources of uncertainty in model predictions

- **Data**
- **Parameter**
- **Model structure**

Sawunyama et al., 2011





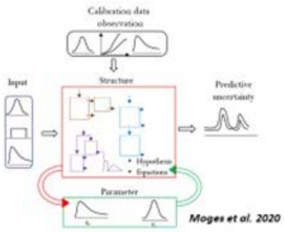
Moges et al. 2020

8

Sources of uncertainty

- **Data**
- **Parameters**
- **Model structure**

climate; streamflow; properties of cover, soils, aquifers; land & water management activities

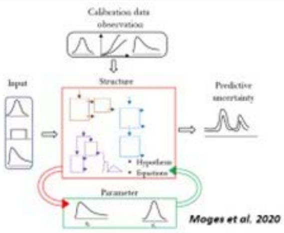


Moges et al. 2020

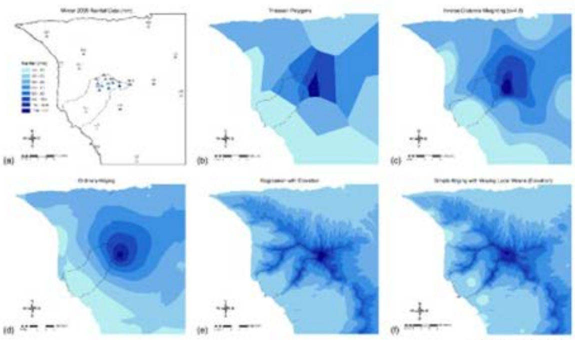
9

Sources of uncertainty

- **Data**
- **Parameters**
- **Model structure**



Moges et al. 2020

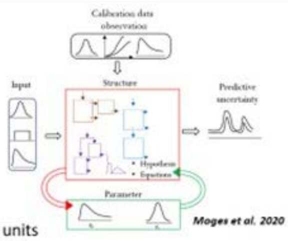

Mair & Fares 2011

10

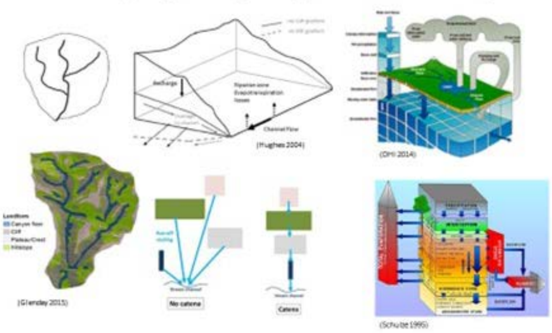
Sources of uncertainty

- **Data**
- **Parameters**
- **Model structure**

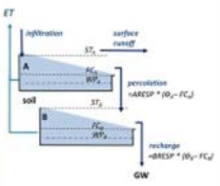
How the catchment are is broken up into modelled units, how units are connected, & algorithms governing water movement through



Moges et al. 2020



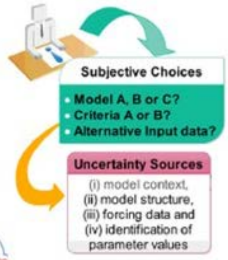
$$C\frac{\partial\psi}{\partial t} = \frac{\partial}{\partial z}\left(K(\theta)\frac{\partial\psi}{\partial z}\right) + \frac{\partial K(\theta)}{\partial z} - S$$



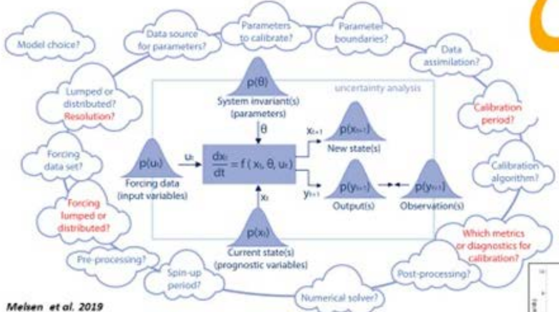
11

Sources of uncertainty

Modeller decisions….

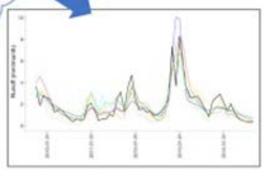


Mockler et al. 2016



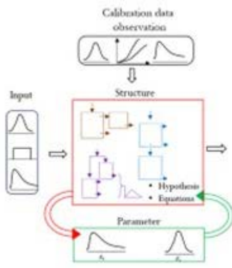
Melsen et al. 2019

Range of different, but potentially feasible outputs….

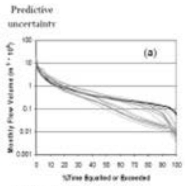


Research on model uncertainty

- Parameter & data uncertainty: more often estimated
- Structural uncertainty: less often
- All sources *can* contribute significant amounts of uncertainty to model predictions
  - Amount of uncertainty & relative contributions by sources = case specific
  - What amount is "significant" = case specific (use-case)



Moges et al. 2020

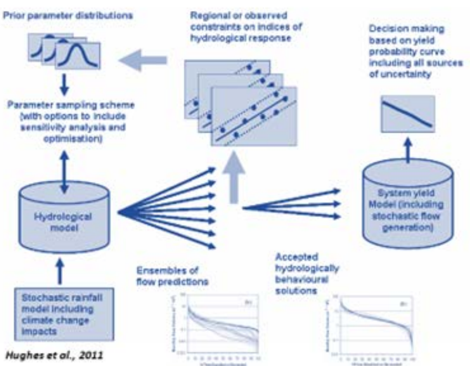


Sawunyama et al., 2011

Local examples

IWR, Hughes et al 2007–2015, uncertainty quantification & reduction framework: *parameter & input data focus*

WRC projects K5-1838 & K5-2056 + associated papers
(Sawunyama and Hughes, 2007; Kapangaziwiri and Hughes, 2008; Hughes, Kapangaziwiri and Sawunyama, 2010; Hughes et al., 2011; Sawunyama, Hughes and Mallory, 2011; Kapangaziwiri, Hughes and Wagener, 2012; Hughes, Mohobane, and Mallory, 2015)



Hughes et al., 2011

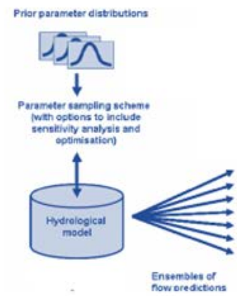
- Select initial parameter ranges based on catchment properties
- Run model using many parameter sets in range & select those with 'reasonable' output
  (if ungauged: use other info, e.g. range of runoff ratios for region, GRAII recharge estimates, etc)
- Range of output from accepted sets = uncertainty range → use for decisions
- Also run with range of climate data & other data inputs…

Parameter uncertainty example

(*Hughes et al., 2010*)

SPATSIM-Pitman, *act as if ungauged (no calibration)*

- Select ranges of parameter values based on catchment physical properties (*Kapangaziwiri and Hughes, 2008*)
- Run 20,000 parameter sets drawn from these ranges (Monte Carlo)
  - X31A: % error in flow ranged from -19 to +9%
  - C12D: % error in flow ranged from -58 to +62%
- **Input all these into reservoir yield model → range in predicted reservoir deficits = uncertainty due to parameter uncertainty:**
  - **X31A: predicted deficits ranged from 3 to 19%** ***(lowest uncertainty)***
  - **C12D: predicted deficits ranged from 6 to 71%** ***(highest uncertainty)***



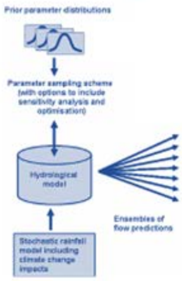
Parameter + data uncertainty example

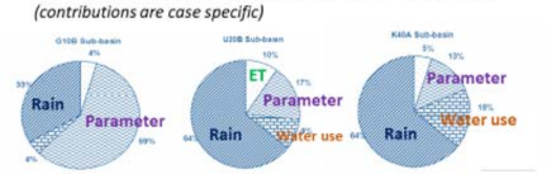
(*Sawunyama et al., 2011*)

SPATSIM-Pitman, *act as if ungauged (no calibration)*

- **Parameter value ranges** selected based on catchment properties, used 3 sets: "best guess" + upper & lower bound
- **Uncertainty in input data – dataset options:**
  Rainfall (catchment scale) x3: WR90 (Midgley et al 1994) vs IDW interpolation of station data vs combo (rescale IDW with W90)
  PET x3: WR90 vs pan-data vs calculated from temperature
  Water use x2: irrigation & afforestation, WR90 vs WR2005+WARMS

Run all combos of parameter sets (x3) & datasets (x3,x3,x2)
→ Uncertainty in modelled flow = high
e.g., G10B mean annual yield ranged from **31 to 79 Mm3/a**
Rain input & parameter uncertainty = dominant sources
(contributions are case specific)



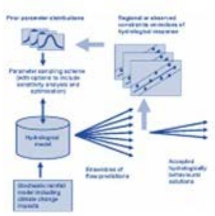


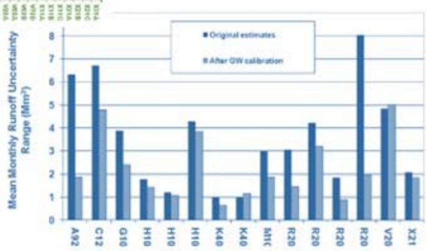
Reducing uncertainty with realism constraints – GW recharge example

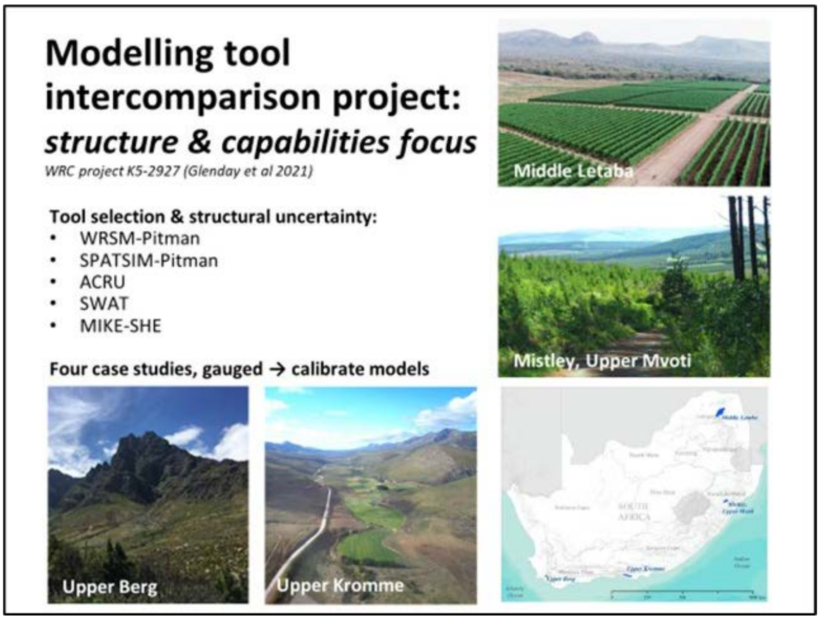
(*Kapangazwiri et al. 2012*)

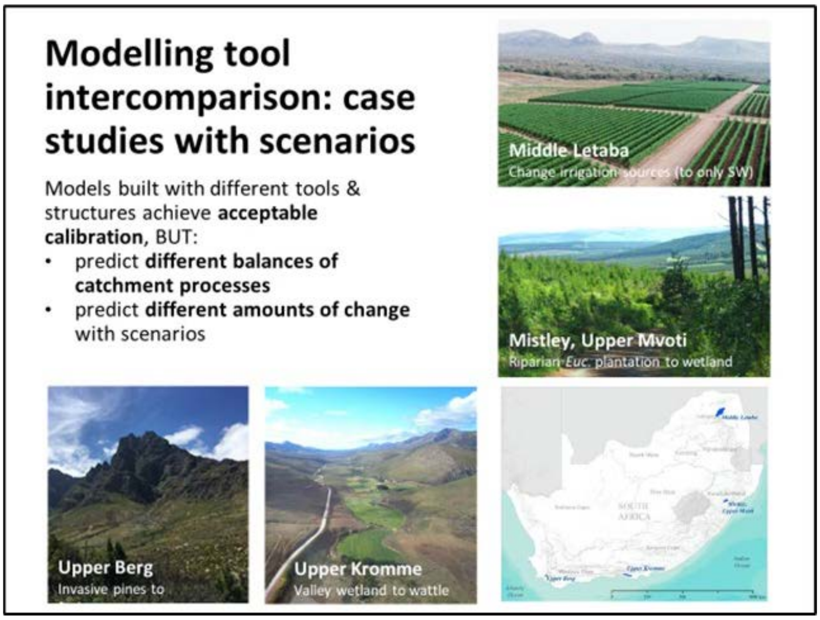
SPATSIM-Pitman, *act as if ungauged*

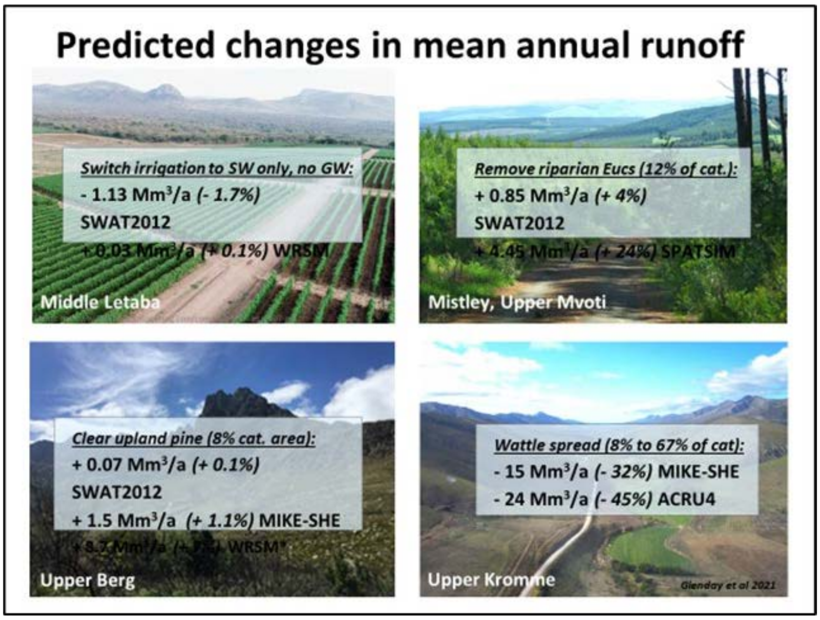
- Select parameter value ranges using catchment properties & run 20,000 sets
- Compare modelled recharge to GRAII estimate range → Select 'behavioural' parameter sets → Reduced uncertainty in flow prediction

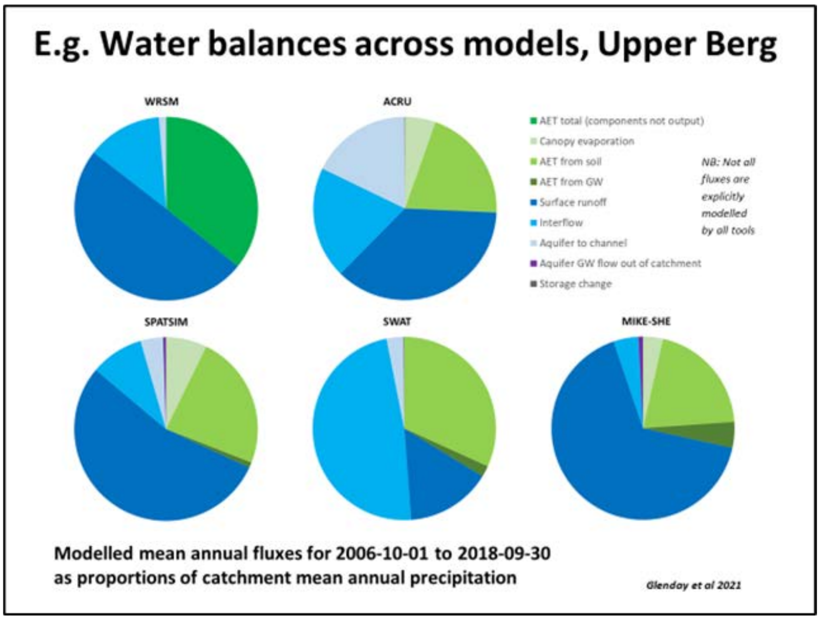


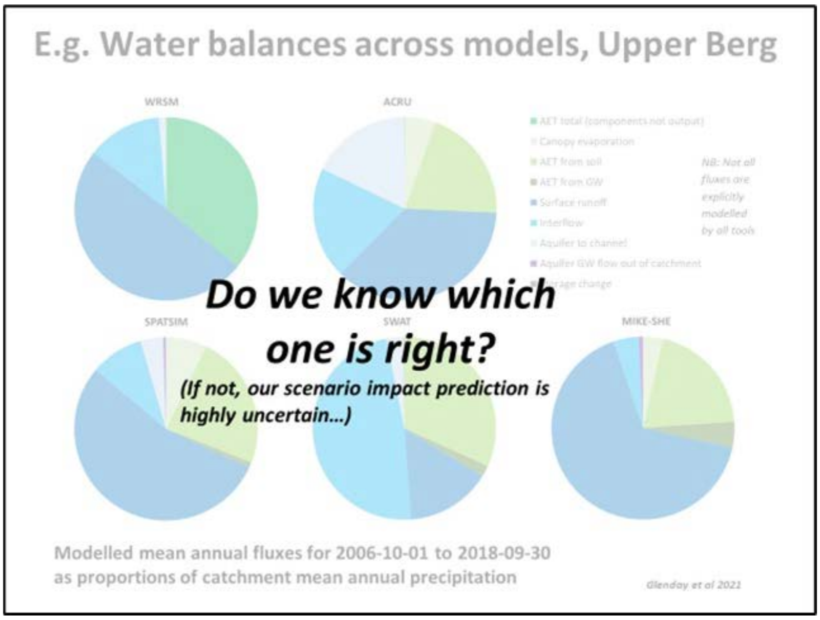


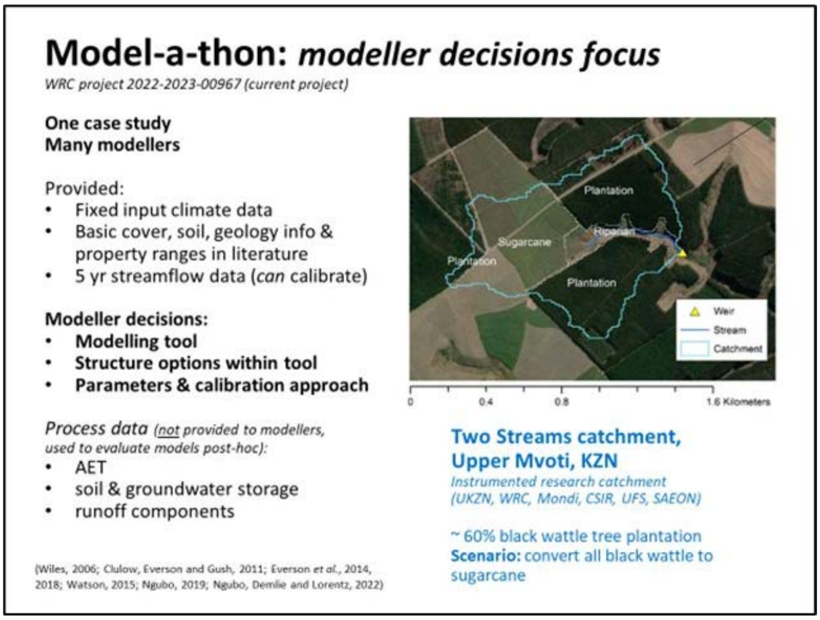

Modelling tool intercomparison project: structure & capabilities focus
WRC project K5-2927 (Glenday et al 2021)
Tool selection & structural uncertainty:
• WRSM-Pitman
• SPATSIM-Pitman
• ACRU
• SWAT
• MIKE-SHE
Four case studies, gauged → calibrate models
Middle Letaba
Mistley, Upper Mvoti
Upper Berg
Upper Kromme


Modelling tool intercomparison: case studies with scenarios
Models built with different tools & structures achieve acceptable calibration, BUT:
• predict different balances of catchment processes
• predict different amounts of change with scenarios
Middle Letaba
Change irrigation sources (to only SW)
Mistley, Upper Mvoti
Riparian Euc. plantation to wetland
Upper Berg
Invasive pines to
Upper Kromme
Valley wetland to wattle


Predicted changes in mean annual runoff
Switch irrigation to SW only, no GW:
- 1.13 Mm³/a (- 1.7%)
SWAT2012
+ 0.03 Mm³/a (+ 0.1%) WRSM
Middle Letaba
Remove riparian Eucs (12% of cat.):
+ 0.85 Mm³/a (+ 4%)
SWAT2012
+ 4.45 Mm³/a (+ 24%) SPATSIM
Mistley, Upper Mvoti
Clear upland pine (8% cat. area):
+ 0.07 Mm³/a (+ 0.1%)
SWAT2012
+ 1.5 Mm³/a (+ 1.1%) MIKE-SHE
Upper Berg
Wattle spread (8% to 67% of cat):
- 15 Mm³/a (- 32%) MIKE-SHE
- 24 Mm³/a (- 45%) ACRU4
Upper Kromme
Glenday et al 2021


E.g. Water balances across models, Upper Berg
WRSM
ACRU
AET total (components not output)
Canopy evaporation
AET from soil
AET from GW
Surface runoff
Interflow
Aquifer to channel
Aquifer GW flow out of catchment
Storage change
NB: Not all fluxes are explicitly modelled by all tools
SPATSIM
SWAT
MIKE-SHE
Modelled mean annual fluxes for 2006-10-01 to 2018-09-30
as proportions of catchment mean annual precipitation
Glenday et al 2021


E.g. Water balances across models, Upper Berg
WRSM
ACRU
SPATSIM
SWAT
MIKE-SHE
Do we know which one is right?
(If not, our scenario impact prediction is highly uncertain...)
Modelled mean annual fluxes for 2006-10-01 to 2018-09-30
as proportions of catchment mean annual precipitation
Glenday et al 2021


Model-a-thon: modeller decisions focus
WRC project 2022-2023-00967 (current project)
One case study
Many modellers
Provided:
• Fixed input climate data
• Basic cover, soil, geology info & property ranges in literature
• 5 yr streamflow data (can calibrate)
Modeller decisions:
• Modelling tool
• Structure options within tool
• Parameters & calibration approach
Process data (not provided to modellers, used to evaluate models post-hoc):
• AET
• soil & groundwater storage
• runoff components
(Wiles, 2006; Clulow, Everson and Gush, 2011; Everson et al., 2014, 2018; Watson, 2015; Ngubo, 2019; Ngubo, Demlie and Lorentz, 2022)
Plantation
Sugarcane
Riparian
Plantation
Plantation
Weir
Stream
Catchment
0
0.4
0.8
1.6 Kilometers
Two Streams catchment, Upper Mvoti, KZN
Instrumented research catchment
(UKZN, WRC, Mondi, CSIR, UFS, SAEON)
~ 60% black wattle tree plantation
Scenario: convert all black wattle to sugarcane

Model acceptance criteria: streamflow fit

Statistic	Equation	Value range & optimal value	Acceptability threshold, minimum performance class*#
Percent Bias (PBIAS)	$PBIAS = 100\% \times \frac{\sum_{t=1}^{T}(Q_m^t - Q_o^t)}{\sum_{t=1}^{T} Q_o^t}$	$-\infty$ *to* ∞ optimal = 0	|PBIAS| ≤ 25% satisfactory*
Coefficient of determination (R^2)	$R^2 = \left[\frac{\sum_{t=1}^{T}(Q_o^t - \bar{Q}_o)(Q_o^t - \bar{Q}_m)}{\sum_{t=1}^{T}(Q_o^t - Q_m^t)^2 \sum_{t=1}^{T}(Q_m^t - \bar{Q}_m)^2}\right]^2$	0 *to* 1 optimal = 1	$R^2 \geq 0.6$ satisfactory**
Nash Sutcliffe Efficiency (NSE)	$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_o^t - Q_m^t)^2}{\sum_{t=1}^{T}(Q_o^t - \bar{Q}_o)^2}$ *(Nash and Sutcliffe, 1970)*	$-\infty$ *to* 1 optimal = 1	NSE ≥ 0.5 satisfactory*
NSE of log transformed data (NSE log)	*Calculated as NSE above, using log transformed observed and modelled datasets (natural log)*	$-\infty$ *to* 1 optimal = 1	NSE log ≥ 0.5 satisfactory*
Kling-Gupta Efficiency (KGE)	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{S_m}{S_o} - 1\right)^2 + \left(\frac{\bar{Q}_m}{\bar{Q}_o} - 1\right)^2}$ *(Gupta et al, 2009)*	$-\infty$ *to* 1 optimal = 1	KGE ≥ 0.5 intermediate#

Where Q_o^t is the observed value for timestep t, Q_m^t is the modelled value for timestep t, $\bar{Q}_o$ and $\bar{Q}_m$ are the means of the observed and modelled values for timesteps t=1 to T respectively, S_o and S_m are the standard deviations of the observed and modelled values for timesteps t=1 to T respectively, and r is Pearson's correlation coefficient between the Q_o^t and Q_m^t datasets for t=1 to T.

Performance class reference: *(Moriasi et al., 2007), **(Moriasi et al., 2015), #(Kling, Fuchs and Paulin, 2012)

Model-a-thon

18 models submitted:

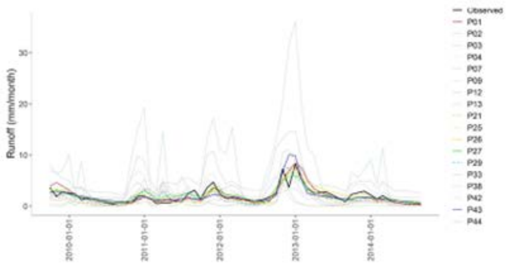
- 3 WRSM
- 3 SPATSIM
- 5 ACRU
- 5 SWAT
- 2 MIKE-SHE

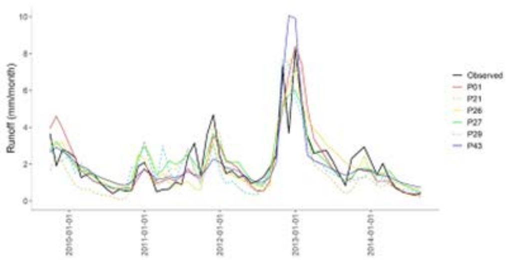
4 met all five streamflow performance criteria
(%Bias, R^2, NSE, NSE log, KGE)

2 others met four of five criteria
(low NSE log = poorer low flow)

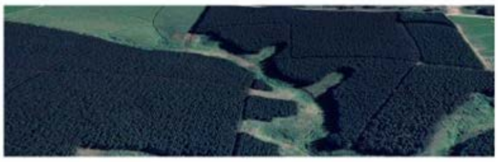
3 of the 4 with 'acceptable' streamflow had subsurface flow dominance (realistic)

1 of the 4 was surface flow dominated = not realistic





Predicted change in mean annual runoff with wattle removal



for 2004-2014, 11 yrs

All 18 models
(no 'reality checks'):

- -28 to +117 mm
- **-75 to + 291%**

6 models, met 4/5 flow criteria:

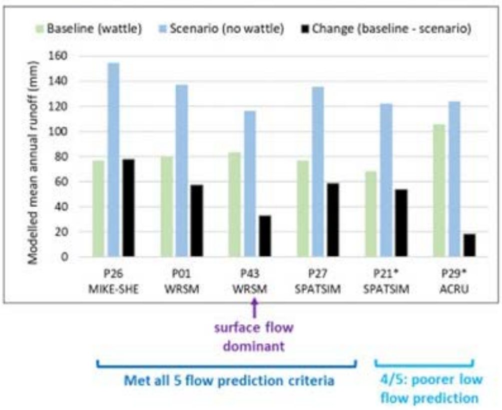
- +18 to +78 mm
- **+17 to +101%**

4 models, met 5/5 flow criteria:

- +33 to +78 mm
- **+39 to +101%**

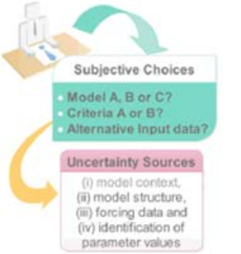
3 models, 5/5 flow criteria & subsurface flow dominant:

- +57 to +78 mm
- **+78 to +101%**

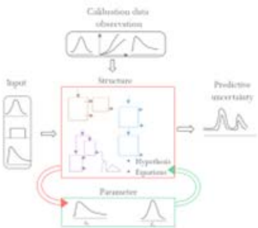


Summary

- **Uncertainty in model predictions** due to uncertainties in data, parameters, & model structures can be **very large**
  ("very large": defined by decision context)
- **Without a specific uncertainty analysis, we will not know how uncertain our model predictions are.**
- **Using one model + one parameter set + one dataset: *hides the uncertainty*** (performance statistics for a single model run are inadequate as a consideration of uncertainty)
- Very different models can predict streamflow to an equally "acceptable" level in calibration/validation (based on typical statistics), but their **representation of processes,** and hence **predictions for alternate scenarios,** can differ notably
- Applying more **criteria of acceptability (realism)** can reduce this divergence and uncertainty
  (need: info & understanding of hydrological processes)



Mockler et al. 2016



Moges et al. 2020

References

Clulow, A.D., Everson, C.S. and Gush, M.B. (2011) The long-term impact of Acacia mearnsii trees on evaporation, streamflow, and ground water resources. WRC Report TT505/11. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

DHI (2019) MIKE SHE Manual, Volume 2: Reference Guide, MIKE 2019. Hørsholm, Denmark: Danish Hydrologic Institute (DHI).

Everson, C. et al. (2014) The long term impact of Acacia mearnsii trees on evaporation, streamflow, low flows and ground water resources. Phase II: Understanding the controlling environmental variables and soil water processes over a full crop rotation. WRC Report 2022/1/13. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Everson, C.S. et al. (2018) Continued hydrometeorological monitoring at the Two Streams Catchment for the clear felling of the Acacia mearnsii stand. WRC Report K5/2780. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Glenday, J. (2015) Modeling the Hydrologic Impacts of Vegetation and Channel Network Change for a Semi-arid, Mountainous, Meso-scale Catchment: the Baviaanskloof, South Africa. PhD dissertation. University of California Santa Barbara.

Glenday, J. et al. (2021) Critical catchment model inter-comparison and model use guidance development. WRC Report Project K5-2927. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Götzinger, J. and Bárdossy, A. (2008) 'Generic error model for calibration and uncertainty estimation of hydrological models', Water Resources Research, 44(12). Available at: https://doi.org/10.1029/2007WR006691.

Gupta, H.V. et al. (2009) 'Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling', Journal of Hydrology, 377(1), pp. 80–91. Available at: https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hughes, D.A. (2004) 'Incorporating groundwater recharge and discharge functions into an existing monthly rainfall–runoff model/Incorporation de fonctions de recharge et de vidange superficielle de nappes au sein d'un modèle pluie-débit mensuel existant', Hydrological Sciences Journal, 49(2), p. null-311. Available at: https://doi.org/10.1623/hysj.49.2.297.34834.

Hughes, D.A. et al. (2011) Incorporating uncertainty in water resources simulation and assessment tools in South Africa. WRC Report WRC Report No. 1838/1/11. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Hughes, D.A., Kapangaziwiri, E. and Sawunyama, T. (2010) 'Hydrological model uncertainty assessment in southern Africa', Journal of Hydrology, 387(3), pp. 221–232. Available at: https://doi.org/10.1016/j.jhydrol.2010.04.010.

Hughes, D.A., Mohobane, T. and Mallory, S.J.L. (2015) Implementing uncertainty analysis in water resources assessment and planning: report to the Water Research Commission. WRC Report 2056/1/14. Gezina, Pretoria, South Africa: Water Research Commission (WRC).

Kapangaziwiri, E., Hughes, D. and Wagener, T. (2012) 'Constraining uncertainty in hydrological predictions for ungauged basins in southern Africa', Hydrological Sciences Journal/Journal des Sciences Hydrologiques, 57, p. 1000. Available at: https://doi.org/10.1080/02626667.2012.690881.

Kapangaziwiri, E. and Hughes, D.A. (2008) 'Towards revised physically based parameter estimation methods for the Pitman monthly rainfall-runoff model', Water Sa, 34(2), pp. 183–191.

Kavetski, D. (2018) 'Parameter Estimation and Predictive Uncertainty Quantification in Hydrological Modelling', in Q. Duan et al. (eds) Handbook of Hydrometeorological Ensemble Forecasting. Berlin, Heidelberg: Springer, pp. 1–42. Available at: https://doi.org/10.1007/978-3-642-40457-3_25-1.

Kling, H., Fuchs, M. and Paulin, M. (2012) 'Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios', Journal of Hydrology, 424–425, pp. 264–277. Available at: https://doi.org/10.1016/j.jhydrol.2012.01.011.

Mair, A. and Fares, A. (2011) 'Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island', Journal of Hydrologic Engineering, 16(4), pp. 371–383. Available at: https://doi.org/10.1061/(ASCE)HE.1943-5584.0000330.

Melsen, L.A. et al. (2019) 'Subjective modeling decisions can significantly impact the simulation of flood and drought events', Journal of Hydrology, 568, pp. 1093–1104. Available at: https://doi.org/10.1016/j.jhydrol.2018.11.046.

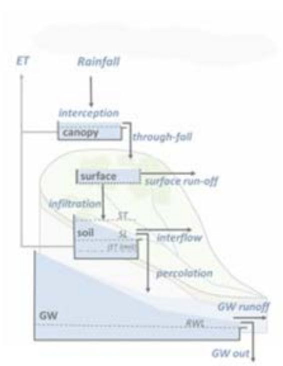
Mockler, E.M. et al. (2016) 'Assessing the relative importance of parameter and forcing uncertainty and their interactions in conceptual hydrological model simulations', Advances in Water Resources, 97, pp. 299–313. Available at: https://doi.org/10.1016/j.advwatres.2016.10.008.

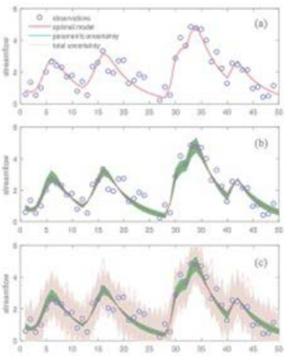
Moges, E. et al. (2020) 'Review: Sources of hydrological model uncertainties and advances in their analysis', Water, 13, p. 28. Available at: https://doi.org/10.3390/w13010028.

Moriasi, D. et al. (2015) 'Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria', Transactions of the ASABE (American Society of Agricultural and Biological Engineers), 58, pp. 1763–1785. Available at: https://doi.org/10.13031/trans.58.10715.

Moriasi, D.N. et al. (2007) 'Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations', Transactions of the ASABE, 50(3), pp. 885–900. Available at: https://doi.org/10.13031/2013.23153.

Questions?





Kavetski, 2018

Break-out Group Discussion 1: *Implications*

- What are some practical implications of these levels of uncertainty?
- Does this concern you? Does it matter in practice?
- Can you think of examples from your own experience where you have had to deal with model output uncertainty? Or where it would have helped to do so?
- Do you think it would be valuable for DWS to allocate resources to include uncertainty assessment in hydrological analyses?

Break-out Group Discussion 2: *Recommendations*

How should we be approaching:

- Quantifying & reducing uncertainty?
- Communicating uncertainty & considering it in decision-making?

What is needed to achieve this?

Way forward

- Policy brief draft: circulate for comment
  - Option to be acknowledged on the brief
- Policy brief final: circulate via the Water Research Commission & the South African Hydrological Society (SAHS)

APPENDIX B.2 – CONSOLIDATED NOTES FROM THE POLICY BRIEF DISCUSSION SESSIONS

SESSION ATTENDENCE

Mon 10th July 2023 – consultants & academics session (C&A)

Attendees (outside project team):

- Louise Dobinson – consultant, water engineer & hydrology, former Zutari, now UK
- Stephen Mallory – consultant, hydrology, planning
- Prof John Ndiritu – Wits University, professor in civil & environmental engineering
- Jane Tanner – Rhodes University-IWR, senior researcher hydrology
- Jay le Roux – University of the Free State, senior lecturer

Starting words on uncertainty: *complex, lots, unseen, challenging (x2), interesting, observational data, finance, variable, constrained, tricky, implications, confusing, account*

Ending words on discussions: *motivated, insightful, eye-opening, passionate, meaningful, progress, useful, exciting, hopeful, refreshing, strategic, optimistic, trailblazer, team-work*

Weds 12th July 2023 – DWS session

Attendees (outside project team):

- Mongezi Gxamza – Chief Engineer, Water Resource Planning Systems
- Jenny Pashkin – Chief Engineer, Water Reconciliation Strategies South Planning area, National Office
- Joshua Rasifudi – Scientific Manager, Integrated Water Resources Studies & National State of Water Report, also PhD student studying uncertainty in modelling and complexity of model structure
- Celiwe Ntuli – Scientific Manager, Water Resource Planning Systems, National Office
- Ernest Oakes – Scientific Manager, Flood Frequency Analysis for Dam Safety, water resource assimilation modelling, National Office
- Mehari Frezghi – Scientist, flood modelling
- Siphindile Shoba – Candidate Engineer

Starting words on uncertainty: *complex (x2), unseen, problematic, challenging, serious, uncertain, multifaceted, data intensive, critical, estimation, broad*

Ending words on discussions: *complex, investigate, certain we're uncertain, thought-provoking, objectives, indeterminate, ongoing, layers, forward-looking, synergy, practical, collaborative, streamline, adaptive*

KEY POINTS RAISED

Using the discussion recordings, key points raised have been summarised and grouped under issue and idea subheadings. Points are not listed in the order in which they were mentioned in the sessions. Note has been kept about which points were raised by participants in general roles/groups (e.g. consulting, DWS). If a group is not listed for a point, this does not signal disagreement, it simply means the idea was not raised. However, if multiple groups raised the same or very similar points this does suggest points of general agreement or shared salience.

Codes denoting which groups expressed a given point:

- *[C&A] = consultants & academics*
- *[DWS] = DWS*
- *[PT] = project team*

Implications of prediction uncertainty

If predicted values provided by models are actually far off from real values, but this is not known or acknowledged (**an uncertainty range for the predicted value *has not* been quantified**), this can result in:

- **Inappropriate water supply planning**: e.g. not being adequately prepared for drought risk, having to do more costly emergency interventions, livelihoods suffer, etc. *[DWS]*
- **Inappropriate water use licence allocations**: e.g. allocating too much water for use and then having shortfalls – livelihoods suffer, environment reserve not met; allocating less for various uses than could have and thereby not giving people opportunities – livelihoods suffer *[DWS]*
- **Under-design infrastructure** (water supply, drainage, crossings): with the risk of failure, causing damage and harm *[C&A, DWS]*
- **Over-design infrastructure** (water supply, drainage, crossings, etc.): with too much spent on it so there is not have enough funding for other things *[C&A, DWS]*
- **Inappropriate decisions about land cover management**, e.g. may decide is or is not worth clearing alien vegetation to increase water supply *[PT]*

In cases where **an uncertainty range for a predicted value *has* been estimated and provided**:

- If uncertainty is quantified and large, may not be able to make decisions based on this output because the potential values span a range for which different decisions would make sense *[DWS, PT]*
    - This could trigger further efforts to reduce uncertainty, which may be possible with more effort: employing additional data sources, gathering additional field data *[PT]*
- If the degree of uncertainty isn't that large, the decision being made may still be the same regardless of using values at the high or low end of the range of estimates (e.g. given a 60% to 80% predicted increase in flows from alien clearing, one would probably still decide to do the clearing). This means the uncertainty is of a level at which it would not impact the decision. This would be a desired target when trying to reduce uncertainty. *[PT]*

Implications of communicating the uncertainty in predictions (particularly when it's large)

- Communicating large uncertainties **can make stakeholders confused and/or lose faith in the study**, the modelling process, the people doing the modelling, the planning exercise, etc. – e.g. If a modelling team presents two different models that predict very different things, and then can't say which of the two is more correct, this can be seen as a failing by stakeholders, rather than as helpful transparency on the part of the modellers *[DWS, PT]*
    - This can instead be framed as showing there is more work to be done: additional work to find ways to narrow the uncertainty, constrain the models and parameters with more realism checks *[PT]*
- For some stakeholders, depending on their background and training, communicating about uncertainty **can lead to more faith in the process by showing due diligence** *[C&A, PT]*

Current state and current practice in applied modelling, highlighting enablers and barriers to uncertainty quantification and communication

Stochastic analyses as a standard practice at DWS

Uncertainties in future flows due to the stochastic nature of climate are currently considered in water supply and flood planning by DWS using the long-term variability in past streamflow over time. This is done using observed streamflow records where a sufficiently long record exists (ideally over 30 years of data), however more often this streamflow is modelled based on the historical rainfall record. Statistics that describe the variability in the streamflow data are then used to generate many (200-1000) stochastic potential future sequences of streamflow that are used to estimate the probabilities of different water levels or dam levels in the future (e.g. what amount of water can be provided from a dam with a 95% assurance of supply given the variability). *[DWS]*

- **Limitation:** This approach, as currently implemented, does not include the uncertainty inherent in the hydrological modelling that produced the historical streamflow timeseries. In the case of an observed historical timeseries, the flow measurement uncertainty should also be considered. It also does not account for climate and land cover changes likely to occur in the future. Additional potential future streamflow timeseries could be included in this stochastic analyses to account for these other uncertainties could (i.e. what would the 95% assurance level yield look like if a different, and equally valid, model structure or parameter set had been used?) *[PT]*
- **Enabler:** This current practice means that many modellers and water managers are already used to working with probabilistic predictions and ranges, seeing box and whisker plots, rather than single deterministic values, which bodes well for communicating about uncertainty further in some cases *[PT]*
- **Barrier:** Some in the water sector may assume that uncertainty has been sufficiently accounted for through this stochastic yield analysis, however, this only accounts for one of many elements contributing to the uncertainty around these estimates. *[PT]*
    - **Diverging views on need for more uncertainty analyses:** Some participants had the perception that the variability in rainfall over time is the biggest source of uncertainty in flow prediction, so it does not add much value to quantify additional uncertainty from other sources. Perception that the current stochastic analyses approach is sufficient *[DWS]*

- This is not always the case (Hughes *et al.*, 2011; Hughes, Mohobane and Mallory, 2015) and this approach does not address the model structural uncertainties that loom large when modelling the impacts of a land cover change or climate change *[PT]*

"Naturalised" models for DWS yield analyses studies

Water resources modelling is commonly done using estimates of 'naturalised' flows, the streamflow expected if there were no major anthropogenic water management activities or land cover changes in the catchment. This allows different scenarios of management and cover to be separately applied in a yield modelling framework. The current approach to obtaining estimated timeseries of naturalised flows is to calibrate a model that includes the water uses and infrastructure relevant to the period of observational data, and then alter this model to represent the catchment without them. This modified naturalised model has not been, and cannot be, assessed against any observational record. There is uncertainty in how the impacts of land and water use are represented in the model and how the model then represents the catchment in a naturalised state, which is not being quantified *[DWS]*

- The uncertainties in land cover change impact predictions that were demonstrated in the model-a-thon and the model intercomparison studies are relevant to the case of 'naturalisation.' These illustrated how differently different models may represent hydrological processes and their interaction with land cover, even when all are calibrated to observed streamflow *[PT]*

Engineering safety margins for designs

Engineering standards specify buffering a design against margins of error in a standardised way (for dams, bridges, drainage, etc.). This can mean adding an industry standard buffer onto values, but this doesn't necessarily standardise the modelling methods to be used, etc. *[C&A]*

- **Enabler:** Precedents for standardised approaches for considering uncertainty that can be built upon further. Depending on the risk levels, this approach may be sufficient in some cases. *[PT]*

Time pressure for decisions

There can be pressure from some clients/stakeholders/ministers/water users to provide answers and make decisions quickly and this doesn't leave enough time to do many checks or uncertainty analyses *[C&A, DWS]*

- **Barrier:** Adding additional uncertainty analyses will require adjusted expectations, process timelines, and funding allocated to processes to cover the time. However various advances are making uncertainty analyses process much faster than before, if we are able to take advantage of these (software, computing power, data access, etc.) *[PT]*

Pressure to provide single value outputs to decision-makers

There can be pressure from some clients/stakeholders/ministers/water users to provide a single output, and also have high confidence in that value, to facilitate decision-making. There is also the need to avoid paralysis around decision-making. *[C&A, DWS]*

- There is some understanding from some of the managers and decision-makers who request a single estimate value that there is actually a range of uncertainty around the value *[DWS]*
- Not assessing the actual magnitude of the uncertainty does not let the decision-maker consider risks and it does not allow someone to make the call that more effort needs to be put in to reduce the uncertainty given the context *[PH]*
- **Barrier:** Some decision-makers appear to be averse to explicitly accounting for uncertainty in the decision-making process. This may be because they are not equipped with strategies for making decisions in the context of uncertainty and this may require targeted capacity building and co-development of standard practices where these do not yet exist. *[PT]*

Data limitations

Barrier: Access to data can constrain how well modellers are able to assess or reduce uncertainty *[C&A]*

Barrier: Limited data on water use and user behaviour can be a large source of uncertainty when looking at highly used systems to either calibrate models of the past, or try to model the future *[DWS]*

- This is often viewed as a main reason why model forecasts may not match what then comes to pass – rather than considering uncertainty in the model structure or other parameters describing the system *[DWS]*

There is general awareness that there is a lot of uncertainty in model outputs that have not been calibrated against local gauge data and where little local climate data exists, although this uncertainty is not often quantified. The fact that there is higher uncertainty around these model estimates vs cases with more data is typically documented and communicated to the decision makers in modelling reports. *[C&A, DWS]*

- **Enabler:** There is already awareness that uncertainty is introduced from data sources, or lack of data. Modellers are used to noting this and decision-makers are used to reading this. *[PT]*
- **Barrier:** There may be a degree of complacency in the practice of highlighting uncertainty issues without quantifying them. Without quantification, decision-makers are not forced, or enabled, to actively consider the degree of uncertainty in their part of the process, which can make their approach simpler, though potentially inappropriate. *[PT]*

Standard practice to only apply one modelling tool and model structure for a project

Clients may request use of a particular tool in the terms of reference (e.g. to use WRSM in a study for DWS, to be consistent with previous studies for a region), a consultancy may have a tool preference, a consultant may only be familiar with one, and/or may only have time to one *[C&A]*

Consulting reports do not generally mention uncertainty due to the model itself. Sometimes do list strengths and potentially weaknesses of the modelling tool *[C&A]*

Perception that the older models perform reasonably well and so there is no need to use others [DWS]

- **Barrier:** Changes in practice are required to consider model structural uncertainty by applying multiple models. There are challenges to developing competencies and trust in new modelling tools – inertia in the system – so there is continued reliance on old ones and not taking advantage of helpful new developments. *[DWS]*

- When two different models have been tried and they give very different answers, the response has been to abandon one of them because it becomes difficult to make decisions based on divergent predictions *[DWS]*
  - Unless there is a physically justified reason to abandon one of the models (one can be shown to be less realistic than the other), the range of outputs between the two is an indication of real structural uncertainty *[PT]*

Barrier: Use of more complex tools is also hindered by data availability [DWS]

Enabler: DWS has already made some efforts to try using other modelling tools in various projects: tried MIKE for some real-time forecasting, and others, but there hasn't been a wide-spread or systematic comparison. There is some openness to trying despite the challenges. *[DWS]*

- Willingness and ability to run training programmes on new tools and methods within DWS (e.g. NatSilt programme) *[PT]*

Limitations due to software tool capabilities and computing power

Enabler: Some modelling software tools make sensitivity analyses and quantifying parameter uncertainty (and sometimes input uncertainty) logistically difficult and very time consuming, while others have automated this process, making it relatively quick and easy to perform. *[PT]*

Barrier: Computing power can be a barrier to uncertainty analyses in complex models with many uncertain parameters *[PT, C&A]*

RECOMMENDATIONS

Overarching / programmatic recommendations

- **Standardise a framework** with approaches to quantify uncertainty, report it, and consider it in decision-making process that are tailored to different case types *[C&A, DWS]*
  - This can include checklists regarding what sources of uncertainty are considered or not considered, the approaches used, reasoning behind decisions, outcomes of analyses *[C&A]*
  - Different sectors, applications, decision types will have different needs, specifications, constraints. It would help to have an idea of what approaches and sources to prioritise in different situations *[C&A]*
  - A minimum requirement should be tracking/documenting the subjective modelling decisions applied in a project to allow improvement in future iterations *[DWS]*
  - Can start developing framework with proposed approaches from modellers that get reviewed and updated over time *[C&A]*
  - DWS can request consultants use the framework – acknowledging that this may add to the time required to do the modelling work *[DWS]*
  - Work needed to inform framework development
    - Try to quantify the amount of time and budget needed to do different levels of uncertainty analyses to help guide proposed approaches *[C&A]*

    - More demonstrations of the practicality of implementing uncertainty analyses using different tools *[DWS]*

- **More training in uncertainty** theory and methods for both modellers and decision-makers (tailored to respective needs) *[C&A, PT]*
  - Have a competency certification verified by SACNASP *[C&A]*
- **To further motivate implementation** of uncertainty assessment: present case studies that quantitatively demonstrate various potential impacts of the levels of uncertainty that have been determined in research: e.g. if a high-end or a low-end value were taken through to a decision-making procedure, carry these scenarios further to estimate potential harm or financial implication quantitatively. *[C&A]*

Recommendations for quantifying uncertainty

- The degree of effort and approach to uncertainty analyses should be tailored to the specific case or to case types *[C&A, DWS]*
- **Input climate data:** Incorporate additional margins of uncertainty on predictions to account for the uncertainties around future climate change *[DWS]*
- **Parameters & input data:**
  - Use sensitivity analyses to determine which parameters and which inputs need to be the focus of the uncertainty analyses *[C&A, DWS]*
  - Use software tools that automate sensitivity analyses and uncertainty analyses. If tools with the process representation and other capabilities that are desired do not have this functionality, make it a priority to develop these functions (either via an add-on or within the software). *[DWS, PT]*
  - Use cloud computing – make this more accessible – to make sensitivity analyses and uncertainty assessment more feasible for complex models *[C&A, PT]*
- **Structural uncertainty:** Apply more than one model, which meet the calibration acceptability criteria, to see the range of predictions *[C&A, DWS, PT]*
  - Use both a simple and a more complex model *[DWS]*
  - In some cases, this could be two different model structures built within the same modelling software tool, e.g. different ways of representing some aspect, different spatial discretisation *[C&A]*
- **Calibration:**
  - Determine model calibration acceptability criteria with respect to the decisions that need to be made with the model output *[PT]*
  - Adjust calibration acceptability criteria with respect to the data available: with shorter observed data timeseries, or more uncertain measurements, acceptability criteria need to be more generous. (e.g. All models are imperfect. Some models or parameter sets may be better at predicting high flows while others are better for low flows. If there is only observational data for a dry period for example, tight calibration to this data could exclude options that may be more accurate in high flow times. These models may then be used to predict for high flow periods.) *[PT]*

- **Yield analyses:** Generate multiple possible modelled streamflow timeseries (using feasible parameter space, using multiple feasible models, using multiple feasible inputs, meeting calibration acceptability) and run these through the stochastic analysis process for yield modelling *[C&A, DWS]*
    - The additional stochastic analyses can be time consuming (especially if models need be naturalised), so would need to motivate regarding need for case types *[DWS]*
        - Look into tools to automate or streamline the implementation of the stochastic analyses in the yield model to make this easier? *[PT]*

Recommendations for reducing uncertainty

- **Uncertainty targets:** Determine target maximum levels of uncertainty for different types of decisions or contexts *[C&A, DWS]*
- **Data:** Maintain and expand monitoring infrastructure to provide long-term observed datasets and cover under monitored area types (e.g. mountain rainfall, etc.) *[C&A, DWS]*
    - Can use existing and new, targeted data collection to try to calibrate and improve remote-sensing based estimation techniques. *[PT]*
    - Improve understanding of groundwater and subsurface processes *[PT]*
    - Ensure data quality checks are done *[C&A]*
    - Accept that we will never have all the data we'd like, and we won't have much more of it in the near future, plus the fact that data and models will never be perfect, so uncertainty analyses and quantification will always be relevant *[PT]*
- **Model reality-check procedures for various hydrological processes:** Application of more multi-criteria calibration & validation methods, including "soft" criteria, to make more use of existing data and information to reality-check models – work further on models to meet these criteria, only keeping those which meet them *[C&A, PT]*
    - Look at models' water balances and internal processes compared to any available, relevant information about these, going beyond looking only at streamflow *[PT]*
    - It becomes especially important to include additional checks when there is only a short observed streamflow record or no streamflow record. *[PT]*

Recommendations for communicating uncertainty

- **Provide training for stakeholders on modelling and uncertainty** at an appropriate level for their needs, i.e. for modellers and for users of output *[PT]*
- Model output should **always be expressed with some uncertainty range** around it [DWS]
    - Uncertainty in model output could be expressed in the form of risk, as done in flood modelling – i.e. probability of occurrence of extreme values [DWS]
- An **assessment of the sources of uncertainty should be presented** with model outputs to facilitate future improvements *[C&A, DWS]*
    - Uncertainty should be communicated alongside suggestions of how it could potentially be reduced *[C&A]*

- **Report uncertainty to stakeholders at an appropriate level** to build trust by showing due diligence *[C&A]*
- **Uncertainty should be communicated alongside risk:** identifying the potential implications of assuming the high and low values in the range. This would be a collaborative effort between modellers and other stakeholders. *[C&A]*

Recommendations for incorporating uncertainty into decision-making

- **Establish frameworks** with check-lists of steps, risk consideration approaches, for accounting for uncertainty that are tailored to various types of decision-making processes in the water sector (e.g. supply infrastructure decisions, operating rules updates, flood set-back decisions, etc.) *[C&A]*
    - Apply a risk framework looking at probabilities of extremes of concern, e.g. consider the higher end high flow estimates for floods, consider lower end low flow estimates for droughts *[PT]*
- Collaboration between modellers and decision-makers to ensure the various specific values and statistics, and their uncertainty bounds, which are provided in the modelling exercise match what is needed for the decision-making process.

APPENDIX B.3 – POLICY BRIEF ON UNCERTAINTY IN HYDROLOGICAL MODELLING

WATER RESEARCH COMMISSION

POLICY BRIEF

Deep waters

How ignoring uncertainty in hydrological modelling increases water security risk

- Streamflow predictions from hydrological models can be highly uncertain. This is due to uncertainties in data and in the modelling process. Uncertainty can be high even when models compare well with observed streamflow.
- Uncertainty in hydrological modelling generally goes unquantified and unaccounted for in decision making. Commonly applied approaches to stochastic yield modelling do not account for this element of uncertainty.
- Ignoring hydrological modelling uncertainty can have serious repercussions such as inappropriate decisions in water supply system planning, water use licensing, catchment land cover and land use management, or flood and drought risk management.
- Poorly informed decision making can result in an increased risk of 'day zero' in cities, decreased water security for agriculture and rural settlements, greater flood damages, and reduced climate change resilience.
- Given programmatic support and capacity building, modellers and decision-makers can work together to standardise practices for quantifying and communicating uncertainty, reducing it when possible, and considering it in decision-making to reduce risk.

Uncertainty in hydrological modelling:

Often large, and largely unseen

There is improved understanding that uncertainty doesn't imply flaw, and not communicating uncertainty is negligent.

Hydrological modelling informs critical decisions in water resource management in South Africa. Hydrological models take inputs of climate data and catchment properties and apply mathematical algorithms to approximate catchment processes, such as evapotranspiration, groundwater storage, and runoff generation.

These models, some of which are locally developed, are used to estimate streamflow for ungauged catchments and to predict flows under different scenarios of climate, land cover, or management (**Box 1**). They provide needed inputs for yield models of reservoirs and supply systems, flood models, and water quality models.

Results from hydrological models inform water resource management as well as decisions influencing land cover management, which can affect livelihoods, food security, ecosystem health, disaster risk reduction, and resilience to climate change.

Despite the importance of hydrological models, it has not been standard practice to quantify and report the full uncertainty in their estimations and predictions. Other elements of uncertainty, such as future rainfall patterns, are routinely considered in water management through stochastic analyses; however, uncertainty in hydrological modelling, a major contributor, is not. Fortunately, decision makers across a range of settings have adapted to dealing with large uncertainties in predictions about the future, in part due to the work of the Intergovernmental Panel on Climate Change.

Sources of uncertainty

01 **Model input data** - *e.g. rainfall, evaporative demand*

02 **Calibration and validation data** - *e.g. streamflow*

03 **Parameter values** - *account for catchment properties*

04 **Model structure** - *mathematical representation of processes*

Modellers make decisions about many uncertain factors (data, model structure, parameters), and different modellers can produce very different results (**Figure 1**).

Even so, it is common for a modeller – researcher or consultant – to use a single model, with a single set of input data and single set of parameter values, to produce a single modelled streamflow output. This represents an informed 'best-guess' of one modeller, influenced by their choices and the data and time available. It gives no quantitative indication of the uncertainty in the output. This approach does not produce a range of values considered reasonable in light of the information available.

What does the research show?

Several research projects have quantified uncertainty in hydrological modelling in South Africa over the last 15 years (**Box 3**). This work has unanimously shown that uncertainty in modelled streamflow is context specific, but can be alarmingly high. When modelling ungauged catchments, parameter uncertainty has been shown to result in different estimates of mean annual streamflow that vary by 150% up to 400%.

Even when there are observed streamflow data to help calibrate model parameters, there can still be high levels of uncertainty, particularly when models are used to predict the impacts of change, such as land cover, management, or climate.

For example, model structural uncertainty has been shown to result in widely different predicted impacts of land cover changes on streamflow. A set of different models, all calibrated, predicted that clearing forestry plantations from a riparian zone could increase mean annual runoff by 4 to 24% (**Figure 1**). This demonstrates how different models of the same catchment can predict similar streamflow to the observed flow for a baseline case, but then predict very different changes in flow to one another when an alternative scenario is applied.

Box 1: Hydrological modelling tools commonly used in South Africa

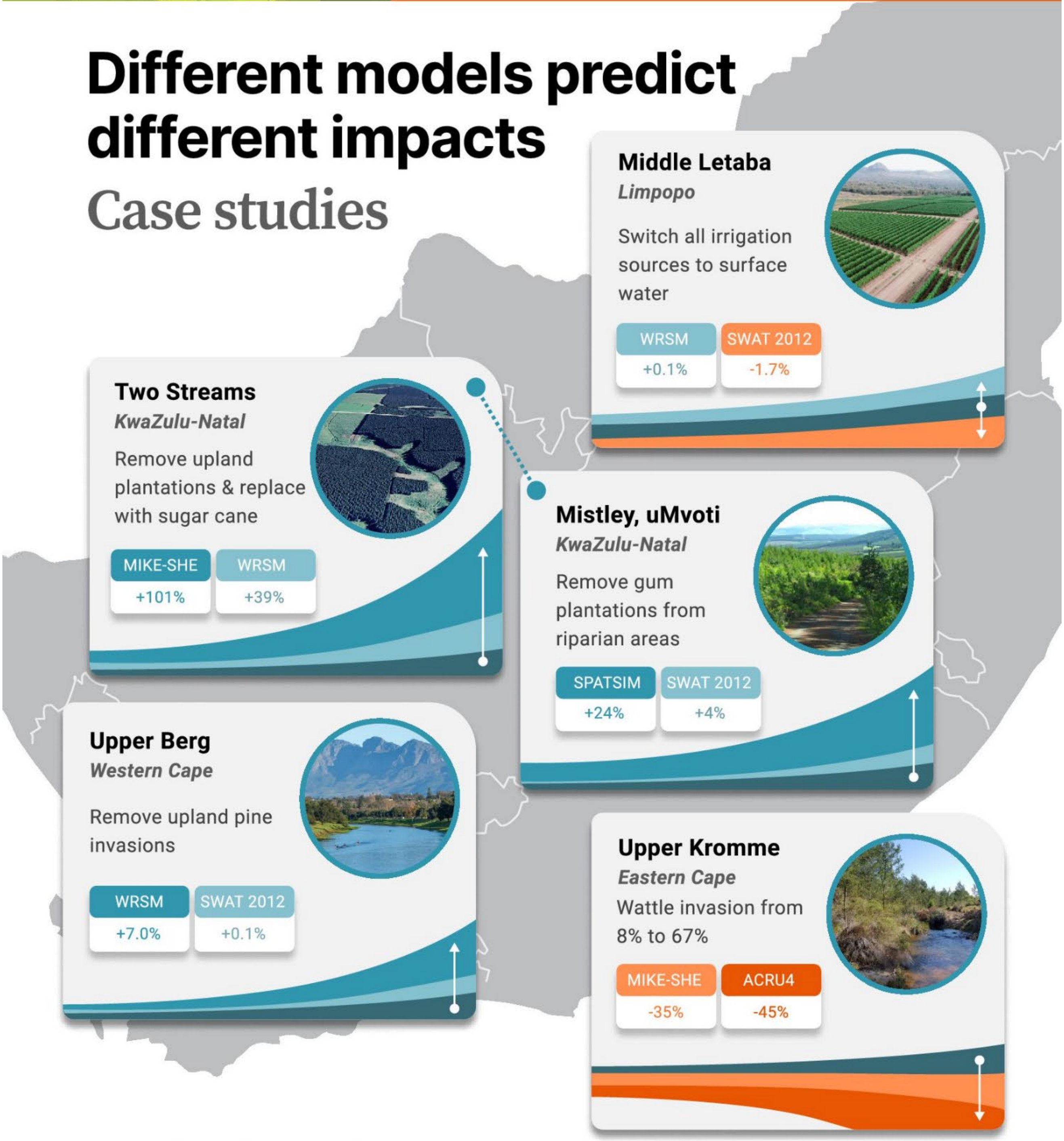
- **WRSM-Pitman:** local, lumped, monthly* model (e.g. used in Water Resources of South Africa, 2012, https://waterresourceswr2012.co.za)
- **SPATSIM-Pitman**: local, lumped, monthly* model (e.g. used in Water Research Commission (WRC) modelling uncertainty research, see Box 3)
- **ACRU**: local, semi-distributed, daily model (e.g. used in WRC national studies on the potential impacts of climate change on water resources https://www.wrc.org.za)
- **SWAT**: international, semi-distributed, daily model (e.g. used in HydrologicAl Model for South Africa (HAMSA) platform, https://waterresearchobservatory.org/hamsa)
- **MIKE-SHE**: international, distributed/semi-distributed, daily/subdaily model (e.g. used by Inkomati-Usuthu Catchment Management Agency, https://riverops.iucma.co.za)

For more information: https://hydromodel-sa-wiki.saeon.ac.za

* Daily versions of these models do exist but are rarely used and have undergone less testing.

Different models predict different impacts

Case studies



Key:
- Baseline mean annual streamflow

Scenario impact (modelled)
- Increased mean annual streamflow
- Decreased mean annual streamflow

Figure 1: Each catchment was modelled using a set of modelling tools (Box 1). All models used the same input data and were calibrated against the same observed streamflow (baseline). They were then used to model alternative scenarios. Different models predicted different impacts. Values listed are the predicted % changes in mean annual streamflow from different models (min & max). The range of values shows the uncertainty in the impact predictions. For more details see Box 3, Glenday et al 2022 and in press [2024].

Ignorance is not bliss:

Repercussions of ignoring hydrological modelling uncertainty

Ignoring hydrological modelling uncertainty can result in inappropriate decisions that have avoidable negative impacts on lives, livelihoods, ecosystems, and infrastructure. Decision makers may be aware that modelled estimates of streamflow are always somewhat uncertain. However, the degree of uncertainty varies from case to case.

If uncertainty is not quantified and communicated, a decision maker will have no indication of how potentially wrong a particular estimate may be, or how to account for this in decision-making. They will not be able to assess the risks involved in taking action, or not taking action, based on the model outputs. They will not be able to decide if additional work to reduce the uncertainty should be commissioned.

A story of how decision making can go wrong

a fictional account based on real model uncertainty results from Glenday et al 2022, see Figure 1 & Box 3

A catchment in KwaZulu-Natal is dominated by private commercial forestry plantations. Legislation requires removing plantations that encroach on riparian (river side) zones. This means 12% of this catchment should be cleared of plantation trees. Should the government invest time and money to ensure that this area is cleared?

A consultancy is commissioned to model the case to help answer this question. Given limited funding and no guidance to use 'best' practices regarding uncertainty, the consultant does not calculate the uncertainty in their results. They provide one result, which suggests that clearing riparian plantations would increase the mean annual streamflow by 4% or 0.9 million m^3. Based on this, a decision-maker decides that the impact is negligible and does not invest time and effort in ensuring the trees are cleared. They instead invest resources elsewhere.

Continue on p. 6

But, had the consultant accounted for uncertainty by using more than one model, this would have changed the results and the decisions made. Instead of one number (e.g. 4%), the consultant would have found and presented a range of potential increase in mean annual streamflow from 4 to 24% (see **Figure 1**). The upper estimate implies 4 million m^3 of water more than the lower one.

If the real impact of clearing plantations in the riparian zone of this catchment lies closer to a 24% increase, this may offer a relatively low-cost approach to increasing water supply. Additional analyses and data could be brought to bear to reduce the uncertainty and determine if this is likely.

Criteria and support for quantification and communication of uncertainty in hydrological modelling, alongside competencies to understand and make decisions in a way that harnesses this uncertainty, can make more informed decisions possible. There are approaches, with different levels of effort and cost, that can be applied to reduce model uncertainties, communicate uncertainties, and make decisions that consider impact thresholds that may lie within the uncertainty ranges established for predictions. On page 8 (**Box 2**), we outline some recommendations on how to do this.

Recommendations:

Shining a light on modelling uncertainty to reduce risk

Based on our research and discussion with practitioners and government officials we recommend the following:

01 Build capacity

Promote theoretical and operational understanding of analyses and decision-making under uncertainty, through university curricula and certified short-courses.

02 Standardise efforts

Develop context-appropriate standard practices for estimating, reducing and communicating hydrological uncertainty for decision-making, e.g., checklists of assessments depending on risk levels, data availability, and scale (see Box 2).

03 Invest in technology

Include facilities in models that automate sensitivity analyses, uncertainty analyses, calibration, and water balance assessments across a range of hydrological storages and fluxes. These functions can be programmed into existing modeling tools. Alternatively, internationally developed tools that meet these needs can be harnessed.

04 Support data collection & open access data

Long-term, continuous, spatially-distributed data on rainfall, evaporative demand, and streamflow need to be made accessible. Data on other storages and fluxes (groundwater, soil moisture, evapotranspiration) can further resolve uncertainties. Calibrate remote sensing products locally with field data before relying on these as alternative sources. Government-funded hydrological modelling efforts, with associated input and output databases, should be easily available.

05 Make budget available

Ensure researchers and consultants have enough time and budget to quantify and reduce uncertainty in hydrological modelling, and account for it in further analyses and decision-making. Resources are also needed for the process of establishing standardised approaches and reviewing these over time.

Resources dedicated to the recommended efforts can be considered a good investment for national government and others, as this will be more than compensated by avoiding the costs associated with inappropriate water supply and flood management. As capacity, tools, computing power, and cloud computing develop, the time and costs associated with uncertainty analyses will also diminish going forward. We must invest in these efforts now to create needed capacity and technology, and position South Africa to benefit from this in future.

Box 2: Approaches for quantifying and reducing modelling uncertainty

Further engagement is needed to recommend specific methods for different cases. Some suggestions are given here:

- **Identify and document sources of uncertainty as well as assumptions and subjective decisions:** This could be a baseline requirement in all modelling projects, and should consider input data, calibration data, parameters, and model structure.
- **Conduct sensitivity analyses:** This determines which uncertain input datasets and parameters will have the greatest impact on the model outputs.
- **Apply more than one modelling tool and/or model structure:** This would be particularly advised when informing high risk and high cost decisions.
- **Calibration and validation - not just streamflow:** Apply a suite of model 'reality-checks' targeting a variety of hydrological processes and conditions (e.g. runoff ratios, evapotranspiration, surface vs subsurface flow dominance, groundwater recharge) during the calibration and acceptance of models and parameter ranges.
- **Propagate uncertainty:** At a minimum, identify reasonable high, middle, and low flow datasets to use in further analyses, such as stochastic yield modelling.

Acknowledgements

Recommendations for this policy brief were co-developed with consultants, academics, and DWS government officials through stakeholder discussion sessions, held as part of WRC project 2022-23-00967: Model uncertainty and reliability for water resources assessment.

We would like to thank the WRC for supporting this work and all those who contributed to the discussions, particularly Professor John Ndiritu, Dr Jay le Roux, Ms Louise Dobinson, Mr Khathutshelo Joshua Rasifudi, and Ms Celiwe Ntuli. Contributions were made as individuals and can not be assumed to represent official views or policies of the institutions of which they are a part.

Contact: j.glenday@saeon.nrf.ac.za

Box 3: References – Research on uncertainty in South African hydrological modelling

This policy brief is informed by several projects supported by the Water Research Commission (WRC) that have looked at hydrological modelling uncertainty in the South African context. (Reports available via WRC Knowledge Hub https://www.wrc.org.za)

- Hughes et al 2011, *Incorporating uncertainty in water resources simulation and assessment tools in South Africa* (K5-1838) and Hughes et al 2015, *Implementing uncertainty analysis in water resources assessment and planning* (K5-2056), developed an **approach for quantifying and reducing uncertainty due to parameters and input data** applicable to gauged and ungauged catchments. The approach was tested in the SPATSIM-Pitman model across many case studies, including carrying the uncertainties through into supply system yield modelling for several examples.
- Glenday et al 2022, *Critical catchment model inter-comparison and model use guidance development* (K5-2927), investigated **model structural uncertainty** using four case study catchments and five modelling software tools (Box 1). The same inputs and calibration data were used for all models. Models with adequate streamflow accuracy were then applied to scenarios (land cover change, irrigation management). A range of impacts was predicted across the different models (see Figure 1).
- Glenday et al [in press, 2024], *Model uncertainty and reliability for water resources assessment* (2022-23-00967), held a "model-a-thon" event to look at the **impacts of individual modeller choices** on streamflow predictions. Individuals modelled a case study catchment under recent land cover and under a cover change scenario, using the same climate inputs, property information, and observed streamflow. Modellers made different decisions across the modelling process, resulting in a wide range of predictions, even from well-calibrated models (see Figure 1, Two Streams example).

How to cite this brief

Julia Glenday, Alanna Jane Rebelo, Petra Holden, Peni Metho, Jane Tanner, David Gwapedza, Shaeden Gokool, Faith Tatenda Jumbi (2023) *Deep waters: How ignoring uncertainty in hydrological modelling increases water security risk [policy brief]*

APPENDIX C.1 – STACK EXCHANGE GUIDELINES DOCUMENT

Stack Exchange, a free, online Question & Answer platform:

Use guidelines for hydrological modellers

Stack Exchange as a resource for hydrological modelling

Have you ever been stuck trying to get a hydrological model to work, not able to find the solution in a manual or from colleagues, and struggling to make contact an expert with the time to assist you?

Are you an advanced modeller who would like to help others more, but don't have much time for meetings or troubleshooting via email?

Are you an instructor who spends a lot of time answering the same modelling questions repeatedly?

Stack Exchange is a well-known, free, online question-and-answer (Q&A) website. It is a readily available resource that we can use as a support system for hydrological modelling. Anyone can create a log-in and ask or answer a question, tagging it for specific themes. Questions and answers on the site are easy to search for and find online. Researchers, students, and practitioners are encouraged to add and tag model-related questions to the platform and to answer each other's questions.

By using a platform such as Stack Exchange, we can strengthen hydrological modelling community by:

(1) making support for hydrological modelling more accessible, open-access, and transparent,

(2) becoming more efficient in providing support, by putting questions and answers on an easily searchable website so that advanced users do not need to answer questions multiple times.

The more people using the site to ask and answer questions, the more useful it will become!

Lecturers, instructors, advanced users

Consider transitioning from answering questions via email to answering them on Stack Exchange. Request that the person asking put their question on the Stack Exchange and email you the link to it. You can enter your response on the site, where anyone can find it. You can even post common Q&A's onto the site yourself in advance and direct students to try looking there first.

Students, newer model users

Try posting your questions on Stack Exchange. You can email a link to the post to lecturers or advanced users and ask them to please answer on the site so that it is available to all. Soon there will be more modelling Q&As on the site so you can search existing posts for help first, before asking.

Everyone

If you've worked through a problem on your own, with a group, or with an instructor in person, do the community a favour and post it as a Q&A on Stack Exchange. One person can post the question and another can answer, or one can post both Q&A. This allows anyone in the world to find the Q&A and so you may even get suggestions of alternative efficient solutions from others online.

How does Stack Exchange work?
A brief overview

StackExchange

Stack Exchange is a network of question-and-answer (Q&A) websites, each with a different focus topic, such as computer programming or earth science (sites list: https://stackexchange.com/sites). The original site, called Stack Overflow, is a Q&A site for computer programming created by Jeff Atwood and Joel Spolsky in 2008.

The **Earth Science** site (https://earthscience.stackexchange.com/) hosts both hydrology and modelling questions. If there are enough users asking and answering hydrology and hydrological modelling questions, a more specific site could be launched in the future.

Stack Exchange sites have features that allow them to be moderated by their user community:

- ***Anyone*** can search for and view existing Q&A strings on the site, no log-in required.

- **Create an account/log-in to post:** One needs to create an account and log in to the site to post a new question or an answer or to add votes, edits, or comments.

- **Tags:** Tags are keywords, or search terms, that users assign to their questions to help others find them. If you want to find out if a question has already been addressed, Google searches often bring up Stack Exchange Q&A posts. You can also search within Stack Exchange sites using tags. Please use "hydrology" and "modelling" to tag your modelling questions.

- **Reputation points:** Users earn points for posting questions and answers and building up points allows you access to more advanced features. Users with more points can: 'upvote' posts, propose new keywords for tagging, put themself up for election to be a moderator, etc. https://earthscience.stackexchange.com/help/whats-reputation

- **Upvoting:** Users with 15+ reputation points can 'upvote' questions and answers that they find useful. There can be many answers posted for one question. When an answer has been 'accepted' by the asker, this answer will be shown at the top of the list. Other answers are listed in order of the number of votes they received. The asker can change which one is their 'accepted' answer as new answers are posted and upvoted by others over time. https://earthscience.stackexchange.com/help/privileges/vote-up

- **Moderation & disputes:** Elected moderators are responsible for managing the site, through activities such as following up on flagged posts, locking and protecting posts, suspending users, and deleting the worst posts on the site. Stack Exchange sites have a "meta" section for users to ask questions about the site and settle disputes in a separate forum to the Q&A.

For more information about how Stack Exchange and the Earth Science site works:

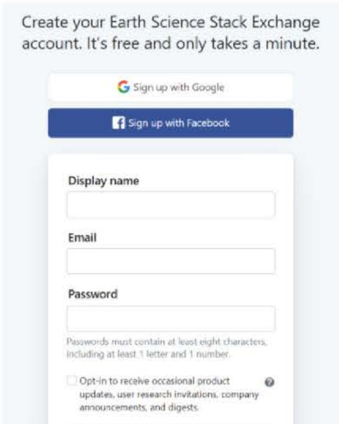
https://earthscience.stackexchange.com/tour

https://earthscience.stackexchange.com/help

https://en.wikipedia.org/wiki/Stack_Exchange

Step by step guide to using Stack Exchange for hydrological modelling support – searching, asking, and answering

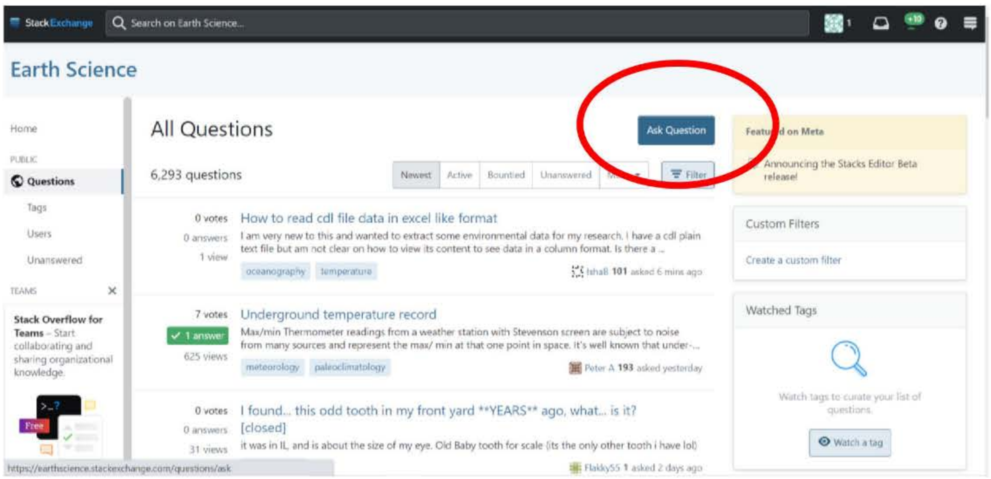
Getting started - create an account



1. Go to the Earth Sciences Stack Exchange site: https://earthscience.stackexchange.com/
2. Create an account with a display name, email, password.
3. Go through the site tour (appears when you create your account): https://earthscience.stackexchange.com/tour
4. Access & edit your profile: Once your profile exists you can edit things like your display name, profile icon, and specify your email preferences (e.g., if you want be notified if someone answers your question, etc). When you are logged in, you can open your profile by clicking on your profile image on the top bar at the right. Here you can also see a log of your questions, answers, reputation points, etc.

Asking questions

1. **Log in to your profile** on the Earth Sciences Stack Exchange site: https://earthscience.stackexchange.com/
2. Before posting a question, **do a search** to see if the question has already been asked. If it has been asked and answered, and you find the posts helpful and you have 15+ reputation points, you can 'upvote' questions and answers using the arrows at the side of the posts.
3. To post a new question: **Click on "Ask Question"**



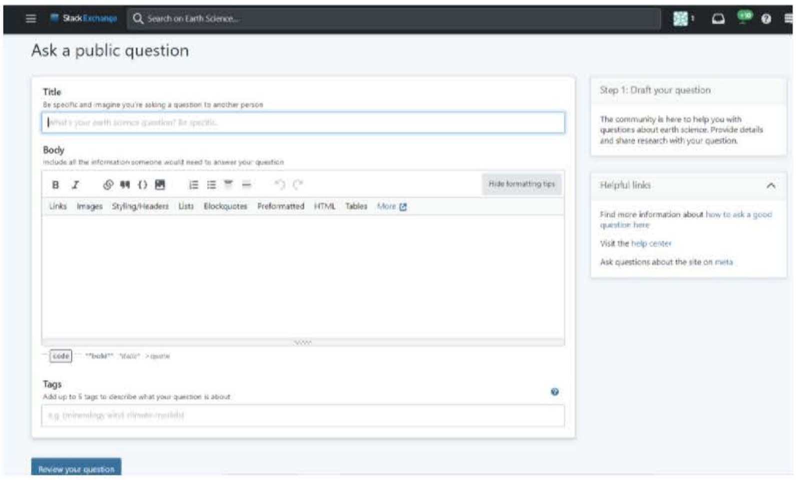
4. **Add a title, body text and tags for your question.** Tips for writing questions for Stack Exchange are given at the end of this guide, and on the Stack Exchange website.

   Use the name of the modelling software tool in your question title and also give the tool name and version in the body of your question (e.g., ACRU4, SWAT+ or SWAT2012, etc).

   Tag the question with **"hydrology"** and **"modelling,"** plus any other relevant tags of your choice.

   Tags must come from the list of tags for the Earth Sciences site. Other potentially relevant tags available include: hydrogeology, water, groundwater, water-vapour, water-table, watershed (Users with 150 reputation points or more can propose new tags.)

   Full tags list: https://earthscience.stackexchange.com/tags



5. **Review & post:** After drafting your question you get to review it; see how it will appear on the site and proofread it one more time. When satisfied, hit "Post."

6. **Send the link to an expert:** Any Stack Exchange user could see and answer your question. However, if you already know an expert or advanced model user who is likely to be able to answer your question, you may wish to contact them directly with a link to your question on Stack Exchange and encourage them to answer it on the site so everyone can access it. This can prevent them from having to answer the same question again for someone else.

7. **Editing after posting:** After you've posted a question, if you realise that you want to change the phrasing or add a clarifying detail (perhaps prompted by the answers and comments coming in response to your post), you can edit your question if you are logged in.

8. **Review the answers, accept, vote:** Once people have posted answers, and you've tried some of them out, you can "accept" an answer that has worked for you. If you have 15+ reputation points you can also "upvote" other useful answers.

Answering questions

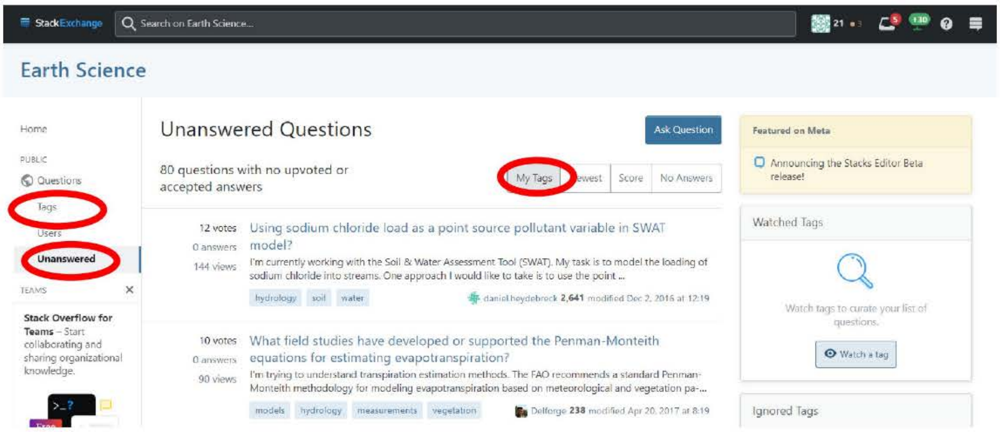
1. **Log in to your profile** on the Earth Sciences Stack Exchange site: https://earthscience.stackexchange.com/

2. **Find a question:** Someone may have contacted you to answer a question that they've asked on the site and sent a link.

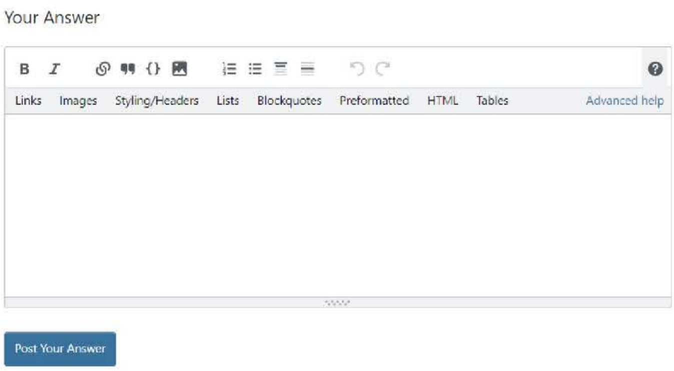
   You can search for questions using keywords of your choice in the search bar (top of page) or you can use the panel on the left to find questions with one or more specified tags (e.g., "hydrology" and "modelling"). Using this panel, you can also choose to look for questions from specific users and for unanswered questions only (note: you can add new answers to ones that have answers already).

   Using "My tags" will find posts related to tags that you have elected to follow (in your account profile settings) and tags you have used previously when posting.

   Click on the title of a question (blue font) to open the post fully.



3. **Post an answer:** At the very bottom of every question post you will find a box titled "Your Answer" where you can enter your answer and post it (if you are logged in). This will be located below any previously posted answers to the question.



Answering your own question: You can post an answer to your own question. You just will not receive any reputation points for your answer, or for it being upvoted.

Tips to asking a good question

Before you post a new question, do a search on the site to make sure your question hasn't been answered already.

When you decide to post a new question:

- **Summarize the problem:** State the problem as clearly and straightforwardly as possible at the very beginning of the post. Add necessary details afterwards.

- **Be specific:**
    - Avoid very broad or open-ended questions that are likely to have many answers or primarily illicit opinions.
    - If the question is specific to a modelling software tool, provide the name of the tool in the title of the question and provide the name and version number in the body of the tool. It may also help to specify the type of computer and operating system you are using.
    - If necessary, break up your query into multiple, more specific questions. These can be linked to one another (you can include a weblink to a previous question in the body of a new question when it is a follow-on issue).

- **Describe what you've tried already:** When appropriate, describe what you've already tried in an effort to solve a problem on your own, and what happened when you did so. This can include describing research you've already done to try to find the answer (you can include citations, links, etc) and why this was insufficient to resolve the issue.

More guidance:

https://earthscience.stackexchange.com/help/asking

Example post:

https://earthscience.stackexchange.com/questions/24081/cannot-initialize-mike-1d-failed-when-solving-for-a-steady-state-solution-wat/24084#24084