

Unpacking some of the linkages between uncertainties in observational data and the simulation of different hydrological processes using the Pitman model in the data scarce Zambezi River basin.

Denis Hughes¹ and Fabio Farinosi²

¹Rhodes University

²European Commission Joint Research Centre Ispra Sector

September 21, 2020

Abstract

The main objective of this study was to use an uncertainty version of a widely used monthly time step, semi-distributed model (the Pitman model) to explore the equifinalities in the way in which the main hydrological processes are simulated and any identifiable linkages with uncertainties in the available observational data. The study area is the Zambezi River basin and 18 gauged sub-basins have been included in the analyses. Unfortunately, it is not generally possible to quantify some of the observational uncertainties in such a data scarce area and mostly we are limited to identifying where these data are clearly deficient (i.e. erroneous or non-representative). The overall conclusion is that the equifinalities in the model are hugely dominant in terms of the uncertainties in the relative occurrence of different runoff generating processes, although water use uncertainties in the semi-arid parts of the basin can contribute to these uncertainties. The identification of landscape features that suggest the occurrence of saturation excess surface runoff provides some information to constrain the model. Improved independent estimates of groundwater recharge is also identified as a key source of observational data that would help a great deal in constraining the model parameter space and therefore reducing some of the model equifinality.

Unpacking some of the linkages between uncertainties in observational data and the simulation of different hydrological processes using the Pitman model in the data scarce Zambezi River basin.

D.A. Hughes^{1*} and F. Farinosi²

¹ Institute for Water Research (IWR), Rhodes University, Grahamstown, South Africa

² European Commission, Joint Research Centre (JRC), Ispra (VA), Italy

* Corresponding authors email: D.Hughes@ru.ac.za

Abstract: The main objective of this study was to use an uncertainty version of a widely used monthly time step, semi-distributed model (the Pitman model) to explore the equifinalities in the way in which the main hydrological processes are simulated and any identifiable linkages with uncertainties in the available observational data. The study area is the Zambezi River basin and 18 gauged sub-basins have been included in the analyses. Unfortunately, it is not generally possible to quantify some of the observational uncertainties in such a data scarce area and mostly we are limited to identifying where these data are clearly deficient (i.e. erroneous or non-representative). The overall conclusion is that the equifinalities in the model are hugely dominant in terms of the uncertainties in the relative occurrence of different runoff generating processes, although water use uncertainties in the semi-arid parts of the basin can contribute to these uncertainties. The

identification of landscape features that suggest the occurrence of saturation excess surface runoff provides some information to constrain the model. Improved independent estimates of groundwater recharge is also identified as a key source of observational data that would help a great deal in constraining the model parameter space and therefore reducing some of the model equifinality.

Keywords: Processes; Hydrological models; Observations; Uncertainty; Zambezi River basin

INTRODUCTION

Models are typically developed to simulate the response of a system to driving forces in the absence of observations of the response. This is true for many different kinds of models, including environmental models (hydrology, geomorphology, oceanographic, climate, etc.), economic models, health models (drug pharmacokinetics, for example) and others. Models may also be constructed to improve our understanding of the internal dynamics (processes) of the system (Ward, 1985; Fenicia et al., 2008; Beven, 2012), even if there are observations of both the driving forces and the response. The dilemma is that we need some observational data to be able to develop and validate the model structure (McMillan et al., 2011). A further problem lies in the reality that, for many systems, and notably environmental systems, the observational data that are available (including the driving forces) are often deficient in terms of accuracy or representativeness and are therefore uncertain (Beven, 2009; Westerberg and McMillan, 2015). Observational ‘data’ may also refer to different things; some may be ‘hard’ quantitative data (direct measurements), while some may be ‘soft’ qualitative data (Winsemius et al., 2009). While the sources of uncertainty in hard and soft data may be different, both are subject to errors (McMillan et al., 2012) that will potentially influence the development of the model, or the model results (Gan et al., 1997). From a hydrological perspective, models may be developed on the basis of largely soft conceptual data (classical hydrological process theory; Ward, 1984), tested in places (or at times) where hard data are available, and applied in places and times when there are few hard data, through a process of extrapolation (parameter regionalisation, for example) using a combination of hard and soft data (Siebert and McDonnell, 2002).

Models may be constructed in a way that largely ignores the internal processes and concentrate on establishing a quantitative relationship between the inputs (driving forces) and the output responses (Todini, 2011). Alternatively, models may be designed to explicitly simulate the internal processes of the system, using different levels of complexity (Chien and Mackay, 2014). Arguably, the latter require more observational data if we wish to not only validate the responses, but also the realism of the internal process simulations (Kirchner, 2006; Euser et al., 2013). The issue of model complexity has been a recurring theme in the hydrological modelling literature for many years (Hrachowitz et al., 2013), and there have been arguments presented in favour of both simple (or parsimonious) models, as well as more complex models (Jakeman and Hornberger, 1993; Perrin et al., 2001). Arguably, a simple model is easier to apply from a mathematical perspective, particularly if the model time step is short, has many spatial elements, and if automatic calibration (or uncertainty ensemble outputs) methods are to be used. Clearly, a small parameter space defining the hydrological response characteristics of each spatial element will take less time to run, and probably converge to a unique solution quicker (less equifinality; Beven, 2006) than a model with a larger parameter space. For models that are applied with coarse spatial (sub-basins) and time scales (monthly), the issue of model complexity becomes less of a problem from a computer run time perspective, but the issue of equifinality remains. It might also be argued that there is little point in having a complex model structure if it is applied at coarse spatial and temporal scales because all the individual hydrological processes are subsumed in the total sub-basin response characteristics. However, this argument relies on the assumption that the total response cannot be decomposed into the sub-basin scale effects of individual processes. There is evidence to suggest that this argument is false in at least some regions and that it is possible to infer (or hypothesise) the relative effects of different processes from the total basin response (Clarke et al., 2009; Hughes, 2013, 2016). A model that includes, implicitly or explicitly, the range of different hydrological processes can be used to assess the validity of process hypotheses (Gallart et al., 2007; Beven, 2012), through any number of different uncertainty analysis methods (Pechlivanidis et al., 2011). This type of approach would not be possible with a much simpler model structure where processes are lumped together in model algorithms that

are designed to represent the total response, but not individual processes.

The detailed outputs of a more complex model can be compared to any data (hard and soft) that might be available to support the presence and importance, or even partially quantify, specific process activity. This is an important point, given the ever increasing availability of global data sets based on increasingly more sophisticated (and presumably more accurate and representative) methods of collecting and processing remote sensing data that can tell us something about *inter alia* vegetation, evapotranspiration, soil and ground water storage regimes and their variation over space and time (Pekel et al., 2016; Lucey et al., 2020; Sadeghi et al., 2020). The regional context of this contribution is southern Africa, where it is hard enough to maintain even basic hydrometeorological observation networks (rainfall and stream flow), and therefore the likelihood of ever having detailed ground-based observations that might help to resolve questions about process activity is extremely remote. One motivation for more complex models is that we wish to know whether we are modelling the response for the correct reason (Kirchner, 2006), given the many different limitations and uncertainties inherent in the model and available forcing data. Perhaps the key questions are what is the value of this information, how would we benefit from it, and why is it important to generate realistic outputs for the right reason? Apart from the rather esoteric answer that as scientists we want know if we are correct, this knowledge could be valuable for applying the same model in areas that are not gauged. This introduces the other key theme that has perplexed hydrological modellers for a number of decades; what are the best ways of transferring the knowledge about a model and its functioning from gauged basins to ungauged basins? There have been many contributions to this topic and many suggestions for different approaches (Blöschl, et al., 2013; Hrachowitz et al., 2013). Perhaps the two main approaches are those based on parameter regionalisation using basin physical properties and their relationships with calibrated parameter values (Pokhrel and Gupta, 2009), and those based on the regionalisation of basin response indices against which the ungauged basin model outputs can be compared, or constrained (Westerberg et al., 2016; Kabuya et al., 2020; McMillan, 2020). There are also some model packages that provide direct methods of calculating parameter values from basin physical properties. All of these approaches rely upon some observational data associated with the main driving variables (climate), total basin response and landscape characteristics, which will inevitably be subject to uncertainties that will impact on the validity of the parameter estimation methods and simulation results.

The main purpose of this contribution is to unpack some of the uncertainties associated with the observational data as well as the model, and to explore how these uncertainties affect hypotheses about the key hydrological processes that are active within different parts of the basin. The real point is to investigate how this approach might be useful in conjunction with an understanding of the links between process activity and a conceptual interpretation of the landscape to help with parameterising the model in ungauged areas. The term ‘landscape’ is used here to represent the many different characteristics that might influence the dynamics of the runoff response and includes topography, vegetation, soils, geology, drainage pattern, etc. The term ‘conceptual’ assumes that the interpretation could be based on a mixture of both soft (or subjective) and hard (numerical analysis of available data) information. The geographic context is the Zambezi River basin in southern Africa, where different climate zones are represented, where data are typically scarce and often of unknown quality, but where well informed water resource management decisions are required that frequently rely on simulated information. The model is a version (Hughes, 2013) of the Pitman (1973) monthly time-step model that has been widely used in the region and is typically applied at relatively coarse spatial scales (in a semi-distributed, sub-basin structure). However, the principles of the approach are considered to be equally applicable to any other model where individual hydrological processes are represented either implicitly or explicitly.

THE PITMAN MODEL AND PROCESS INTERPRETATION

Most of the original structure (Pitman, 1973), as well as more recent additions (Hughes, 2004; Hughes and Mazibuko, 2019) to the model have been designed to represent processes explicitly (Figure 1), albeit at the sub-basin scale, using approaches that are similar to the probability distributed principle of Moore (1985). Given the rather large parameter space (20 parameters covering the full range of natural hydrological

processes), any form of calibration (manual or automatic) can become a daunting task and experience suggests (Hughes, 2013) that to benefit from the explicit representation of the processes, it is important to understand the conceptualisation of the model algorithms. Figure 1 summarises the main model structure, while the following sub-sections provide a little more detail. The model outputs include some details of the simulations of the individual processes so that these can be compared to any available observational data as well as the total sub-basin output.

Interception and evapotranspiration.

Interception depth is defined by a storage parameter that can vary seasonally, while evapotranspiration losses are dependent upon soil moisture storage, input values of potential evapotranspiration (PET) and a parameter $(0[?]R[?]1$, with lower values implying higher relative actual losses). Spatial and temporal variations in vegetation cover can be readily obtained from satellite imagery such as Leaf Area Index (LAI) or MODIS Normalized Difference Vegetation Index (NDVI) data, but these data do not provide direct measures of interception loss.

Surface runoff.

There are two methods of generating surface runoff in the Pitman model (ISQ and SSQ in Figure 1). The first is effectively a saturation-excess surface runoff process (Hughes and Mazibuko, 2018), while the second is a function only of rainfall depth and represents an infiltration (or adsorption) excess surface runoff process. The key parameter (SSR) is the wetness value ($ST \cdot SSR$) at which this process is initiated. This function was added to account for the presence of relatively flat valley bottom areas (Dambo's) that remain wet during the dry seasons, due to interflow from the surrounding hillslopes (von der Heyden, 2004). These features are relatively straightforward to identify using Google Earth imagery, while Lampitlaw and Gens (2006) refer to quantitative mapping methods using satellite imagery and topographic analysis. Hughes and Mazibuko (2018) demonstrated that the inclusion of this function improved the seasonal distributions of simulated flows in catchments where such landscape features are known to exist, but gave poorer simulations if used in other areas. The second surface runoff function uses a triangular distribution of catchment adsorption rates defined by two parameters and the area under the cumulative frequency curve for a given rainfall depth represents the depth of surface runoff. There are no data sources that can directly help with quantifying these parameters.

Interflow runoff

Interflow runoff depth is determined from a non-linear power relationship (Figure 1; IQ) and there are no observational data that can directly support the determination of the parameters (FT, SL and POW). However, topography and soils data can at least point to the likelihood that interflow is either a dominant or largely irrelevant process. The key signal in the observed stream flow data is the shape of the wet season recession, where slow (or fast) recessions suggest relatively high (or low) proportions of interflow.

Groundwater recharge and discharge to stream flow

The recharge function is the same form as the interflow function (Figure 1) and is routed through a conceptual groundwater storage (influenced by drainage density and storativity parameters), while outflows to the river channel are mostly determined by a transmissivity parameter and the level of storage (used with drainage density to estimate the hydraulic gradient towards the channel). An additional parameter defines the proportion of the sub-basin area that represents the riparian strip from which groundwater can be lost to evapotranspiration. Further details of the structure and algorithms can be found in Hughes (2004). Experience within South Africa suggests that the best information for constraining some of the parameters comes from independent evaluations of groundwater recharge rates and the geological characteristics of the underlying aquifers (DWAF, 2005).

Water use functions

Apart from an option to account for large reservoirs, there are also functions to allow for direct abstractions

from the river, and for storage and abstractions from distributed small dams. The available data for quantifying storages and abstractions in southern Africa is typically almost non-existent (or at least not available), while some global data sets are available to quantify the maximum surface area of water bodies (Pekel et al., 2016; Gonzalez-Sanchez et al., 2020) and areas under irrigation (IFPRI, 2019), but converting these to useful information on patterns of water use is also subject to a great deal of uncertainty (discussed later).

Equifinality between and within the different process representations.

The two surface runoff functions both determine the patterns of moderate to high flows, but they have quite different seasonal distributions because the first is driven by the sub-basin moisture status and rainfall, while the second is driven only by rainfall. Resolving some of the equifinality therefore relies on an assessment of the shape of the wet season stream flow response, or clearly identifying the presence of Dambo type features. The interflow and both surface runoff functions partly determine the shape of the middle part of the flow duration curves and it is never very straightforward to determine the most appropriate parameter combinations. Simulating low flow patterns are associated with equifinalities between the interflow and groundwater recharge functions, within the two functions (the interplay between the scaling (FT and GW) parameters and their respective power parameters (POW and GPOW), Figure 1), as well as between the recharge and the amount lost to riparian evaporation. It is often possible to identify signals in the observed stream flow data that can resolve at least some of these equifinalities, but there almost always remain a quite broad range of plausible parameter sets that produce similar responses. Low flow simulations are also affected by equifinalities between the natural hydrology functions and the impacts of distributed water use.

STUDY AREA and DATA

While there are many gauged sub-basins within the southern Africa region that could be used, the focus is on the Zambezi River basin, largely because this basin has recently been the subject of a model calibration (Hughes et al., 2020) and climate change assessment (Hughes and Farinosi, 2020) study conducted under the auspices of the African Union - NEPAD African Network of Centres of Excellence on Water Sciences and Technology - ACEWATER phase 2 project. The primary objective of this study was to achieve an acceptable calibration of the model across the 76 defined sub-basins (Figure 2) and to investigate the range of uncertainties in the water resources availability in the future. While they were not ignored, there was less focus on the likely realism of the modelled processes or the observational data uncertainties, which are the main concern of this paper.

The Zambezi River basin covers a total area of some 1 350 000 km² and has eight riparian countries (Angola, Botswana, Malawi, Mozambique, Namibia, Tanzania, Zambia and Zimbabwe). The rainfall is highly seasonal and occurs mostly in the summer months between October and March. Annual rainfall amounts vary from about 1 200 mm y⁻¹ in the upper areas of the Shire and Kafue sub-basins, to less than 700 mm y⁻¹ in the semi-arid sub-basins of Zimbabwe (Hughes et al. 2020). There are a number of gauging stations in the basin, some in the headwater areas and others on the main rivers. This study concentrates on 18 headwater gauged sub-basins (Figure 2, Tables 1 and 2) most having records dating back to about 1960. They have been selected to represent the range of climate conditions, as well as the type and range of uncertainties that are expected to exist in the observational data that are available to assist with establishing behavioural model set ups. Table 1 provides the gauging station details, but the remainder of the paper refers to these sites using the model setup sub-area names given in Figure 1 and the first column of Table 1. Additional information about these sub-basins is contained within the results section, where it is considered relevant to the interpretation of the model outputs.

While many of the main tributaries are gauged, the Zambezi River basin is typical of many other parts of southern Africa in that it is largely a data scarce region, particularly with respect to local climate data. Even the available stream flow data contain a number of uncertainties, partly related to possible rating curve problems, and partly related to periods of missing data (Hughes et al., 2020). The original model was forced with the University of East Anglia, Climate Research Unit data (https://crudata.uea.ac.uk/~timm/grid/CRU-TS_2.1.html, accessed during Oct. 2019), available from 1901 to 2017 at a grid scale of 0.5° (Harris et al.,

2014). Additional rainfall data (for the same period and spatial resolution) from the University of Delaware (UNIDEL; Willmott and Matsuura, 2001) were used to assist with identifying key rainfall data uncertainties. Both of these rainfall products are based on extrapolation from sparse ground stations and are expected to contain large uncertainties, particularly in the representativeness of individual monthly rainfall depths. Comparisons between them suggest that in most places they agree quite well, but there are some of the Lake Malawi/Nyasa sub-basins where there are substantial differences in the mean annual rainfall suggested by the two datasets (Table 2).

The potential evaporation (PET) data are based on the LISVAP calculations (Alfieri et al., 2019) using the ERA5 data for 1979 to 2018 (<https://confluence.ecmwf.int/display/CKB/ERA5+data+documentation>, accessed during Oct. 2019), which are also expected to contain a number of uncertainties. However, given that the Pitman model uses a single annual PET depth and a fixed seasonal distribution for each sub-area, the main data uncertainties are expected to be in the mean annual values and the uncertainty range in the model has been set to $\pm 10\%$ of the LISVAP values. Estimates of LAI are expected to be useful for constraining simulated interception depths (annual means, seasonal distributions and even time series values). The major uncertainties are not expected to be in the conversion of LAI into depths of interception for a given climate regime (De Groen and Savenije, 2006; Wu et al., 2019; Navar, 2020). The LAI data (Mao and Yan, 2019) used in the study are long-term (1981 to 2015) monthly means (https://daac.ornl.gov/VEGETATION/guides/Mean_Seasonal_LAI.html, accessed during July 2020) and the seasonal range for all sub-basins used in this study (plotted against their aridity index), as well as some sample seasonal distributions are given in Figure 3. MODIS actual evapotranspiration data (AET) could help with partially resolving some of the annual or long-term water balance (stream flow = rainfall – evaporative losses) uncertainties. However, the MODIS AET data are themselves subject to uncertainties (Velpuri et al., 2017) related, in part, to the availability of local climate data, as well as the interpretation of vegetation reflection signals.

Groundwater recharge data are potentially very useful for resolving some of the equifinalities between simulated interflow and groundwater contributions to stream flow, and estimates for the different geological and climate zones of the basin are available from the British Geological Survey (MacDonald et al., 2012). However, the level of uncertainty is largely unknown as it is not very clear how the estimates were derived. Some remotely sensed soil moisture data were investigated during this study. Although these were not expected to be useful for constraining or checking the simulated soil moisture storage regime (largely due to the shallow depth of penetration of the sensors), it was considered that the data could be useful to identify landscape features (such as Dambos) that have different patterns of near surface moisture storage to other areas and therefore assist with setting the parameter of the saturated surface runoff function. In order to test the validity of our hypotheses, we used here the European Space Agency (ESA) Climate Change Initiative Soil Moisture dataset (ESA-CCI v0.47: <https://www.esa-soilmoisture-cci.org/node/238>, accessed on August 2020) (Dorigo et al. 2017; Gruber et al. 2017, 2019) and the NASA – JPL Soil Moisture Active Passive (SMAP) (respectively Level 4 9km: <https://nsidc.org/data/SPL4SMAU/versions/5>; and Level 2 3km resolution: https://nsidc.org/data/SPL2SMAP_S/versions/2, accessed on August 2020) (Das et al. 2019).

Water use data are notoriously difficult to obtain in most parts of southern Africa, but some indications of agricultural water use can be obtained from GIS analysis of land use data (IFPRI, 2019) to identify areas of irrigation. The uncertainties lie in the accuracy of the remotely sensed land use data as well as any assumptions made about irrigation application rates. Similarly, it is not always clear where the water is obtained from (reservoir, run-of-river or groundwater supplies). There are data available on the maximum surface area of reservoirs, that include quite small farm dams (Pekel et al., 2016; Gonzalez-Sanchez et al., 2020), however, translating the areas into storage volumes is highly uncertain (Hughes and Mantel, 2010; Busker et al., 2019), as is defining the contributing catchment areas of the dams. This issue is particularly relevant to the Zimbabwe sub-basins (Figure 2 and Table 2).

METHODS of ANALYSIS

The main approach to this study has been to use an uncertainty version of the Pitman model to explore

different parameter combinations that generate similarly ‘good’ reproductions of the observed streamflow response. The version of the model used allows for any or all of the parameter inputs to be defined by minimum and maximum values, which are independently randomly sampled (uniform distribution) during each of (typically) 10 000 ensemble runs. The parameter values, a range of summary statistics (e.g. mean monthly values of runoff volume, recharge depth and depth of the four main modelled processes) and goodness-of-fit statistics (objective functions) for each ensemble are part of the model outputs. To avoid a single objective function statistic from dominating the selection of ‘good’, or behavioural simulations a simple combined statistic (CS) is used that combines the Nash coefficient of efficiency values (CE) and the % bias in mean monthly runoff (%Bias), based on untransformed and natural log (ln) transformed values.

$$CS = CE + CE(\ln) + 2 \left| \frac{\%Bias}{100} \right| - \left| \frac{\%Bias(\ln)}{100} \right| \text{Equation 1}$$

The maximum value is 4.0 for a perfect fit, while behavioural ensembles can be selected as those that have CS values greater than (say) 95% of the highest (best fit) value for the whole ensemble set.

The methods are therefore simple, but the process of setting appropriate parameter ranges and interpreting the results is often more complex, particularly when many parameters are set to be uncertain in the same run. Previous experience (Hughes, 2016) therefore suggests that several runs of the model focussing on different groups of parameter interaction (or different process components of the water balance) are frequently necessary to be able to explore the equifinalities in detail. In the context of this Special Issue of the journal, the possible effects of uncertainties in either the forcing climate data or the observed stream flow data are also explored, as well as the value of any other hard or soft observational data (referred to in the previous section) that can be used to resolve some of the equifinalities. The latter would typically be used to either constrain some of the parameter ranges, or exclude ensemble members that do not generate outputs consistent with the data.

RESULTS

There is insufficient space to present the full results for all 18 sub-basins, and some sub-basins are presented in more detail to represent specific elements of uncertainty, while some of the pertinent details of the simulations are presented in Table 3 for all sub-basins. Arguably KAF4 represents the sub-basin with the least amount of uncertainty in the observational data used, and apart from the generic uncertainties in the rainfall, interception and evapotranspiration input data, a key issue is the extent of Dambo occurrence and the effects on saturated surface runoff. The maximum CS value within the 10 000 ensembles is 3.563, very close to the optimal value of 4, and all those (98) greater than 3.38, but with no %Bias or %Bias{ln} values greater than +5.0, were accepted as behavioural. There are very few differences between the minimum and maximum parameter values within the behavioural ensembles compared to the total ensemble set, implying a high degree of equifinality in the model, as is normally the case with a model with so many parameters. The main differences are that the maximum behavioural interception, saturated surface runoff and recharge parameters are somewhat less than the maximum input values. The runoff ratio for the behavioural ensembles lies between 10.2% and 11.3%, while the full ensemble set range is 2.4% and 22.1%. The implication is that the depth of AET (the main determinant of the overall water balance together with rainfall, which is not considered uncertain for this sub-basin) is relatively insensitive to uncertainties in the PET observational data (assumed to be +10% of the available estimates). Further analysis of the simulations of interception were achieved by setting only the interception and evapotranspiration parameters as uncertain. The results confirmed that the overall model fit is almost totally insensitive to the simulated interception depth (in the range of 56.4 to 175.1 mm y⁻¹) and higher interception is compensated for by less effective evapotranspiration from the moisture store (and vice versa). The main impact is a slight shift forwards in time in the seasonal distribution of simulated stream flow for the higher interception depths.

The ranges of mean annual groundwater recharge values are 8.8 to 69.2mm and 7.2 to 154.8mm for the behavioural and total ensemble sets, respectively. The BGS values for this part of the Zambezi are between 109 and 146mm, clearly suggesting that the available observational data are too uncertain to constrain the model. There is a wide range of possible combinations of individual processes within the behavioural ensemble

set, and no clear differences between those with high and low input PET values, suggesting that uncertainties in the PET data have a low impact on the simulation of individual processes. A relatively simple analysis of Google Earth images to approximately quantify the surface area of Dambo features (Figure 4a), suggests that their maximum area is ~15% of the total sub-basin area. Figure 4b shows the relationship between relative moisture content and saturated area calculated by the model for different values of the SSR parameter. The Google Earth observational data suggest that this parameter could be constrained to between about 0.55 and 0.65, allowing for quite high uncertainty in the interpretation of the Google Earth images. This reduces the behavioural ensemble set to 52, but has little impact on the possible combinations of individual processes. Some of the grids for the SMAP_L4 9km, 3 hour, soil moisture data showed characteristics that might be expected from the presence of Dambos (more consistently wet during the wet season and slower drying into the dry season, for example), and most of these could be linked to areas that can be identified as having a high density of Dambos. However, there are other areas where Dambos are clearly visible on Google Earth that do not show the same patterns in the soil moisture data. Part of the problem may be related to the spatial resolution and part to the shallow depth of the observational soil moisture data sample. The ESA product has a resolution of a quarter degree which is too coarse to identify Dambo areas, while the highest resolution data SMAP_L2/Sentinel 1A/B 1 and 3km data, are available only for scattered portions of the basin and every few days, making it difficult to clearly identify signals of the phenomenon investigated. Furthermore, the limited data available for the higher spatial resolution soil moisture data showed very little variation across the sub-basin. Similar conclusions were reached for the other sub-basins and the soil moisture data, in their current stage of development, were not found to be useful for constraining the model or resolving any uncertainties in process simulations.

Figure 5 shows the partitioning of total runoff for two ensembles (low and high recharge) and there are clearly substantial differences in the way in which the model can simulate the observed stream flow response, that are largely independent of any of the observational data uncertainties. No uncertainties in the observed stream flow data have been included, largely because there are no stage-discharge rating data readily available upon which to base quantitative estimates. However, they are expected to be low relative to other sub-basins and the main impact would be simply to increase the number of ensemble members considered to be behavioural. The other conclusions for this site would not substantially change.

KAF11 is similar to KAF4 except that the extent of Dambo features appears to be much less, and there are additional uncertainties associated with water use for mining and irrigation (mostly from direct river abstractions). The patterns within the behavioural ensembles are similar to KAF4, although the SSR parameters are generally much higher, consistent with fewer Dambo features, while the recharge values tend to be higher (42 to 100 mm y⁻¹). The runoff ratio varies between 20.5% and 22.5%, which might reflect the smaller size, and more headwater location, of KAF11 relative to KAF4.

BAR3, BAR4, BAR7 and CHB2 represent the sub-basins of the upper Zambezi River and, apart from BAR3, are mostly underlain by deep Kalahari sand deposits. The results for BAR3 are very similar to KAF4, with behavioural runoff ratios of 9.1% to 10.0%, and recharge range of 24 to 70 mm y⁻¹. The uncertainty in the distribution of process contributions is also similar to that shown in Figure 5. During the initial calibration of the model (Hughes et al., 2020) acceptable simulations for BAR7 and CHB2 could not be achieved. It was also concluded that the observed stream flow data for BAR7 were erroneous as they show much higher low flows than at BAR5, BAR6 and ZAM1 further downstream and below the Barotse floodplain (within BAR5; Figure 2). The application of the uncertainty version of the model suggests that acceptable simulations are obtainable at the sub-basin outlets, while the miss-match with observed data downstream remains a major source of uncertainty in either the observational data, or the model (including the simulation of the wetland impacts of the Barotse floodplain), or both.

The runoff ratios for BAR4 are much lower (5.4% to 6.2%), and surprisingly the behavioural recharge range is only 12 to 38 mm y⁻¹, contrary to the expectation that groundwater would play an important role in the area underlain by Kalahari sands. The model simulates the majority of the low flows as interflow in all of the behavioural simulations, despite all the groundwater parameters having wide enough input

ranges. More consistent with expectations is the low contribution made by saturated surface runoff (no clear indications of Dambos). In contrast, BAR7 is totally dominated by groundwater in the small number of behavioural ensembles (Table 3), with a narrow range of recharge values of 151 to 204 mm y⁻¹, and higher runoff ratios (16.7% to 17.1%). CHB2 has an overall much worse fit to the observed data and very few behavioural ensembles, low runoff ratios (3.8% to 4.2%) and recharge between 19 and 30 mm y⁻¹. It is also more dominated by groundwater outflow contributions (52 to 89% of total flow), and is therefore similar to BAR7. One possible check on the simulations of the sub-areas dominated by Kalahari sands, and especially the high low flows and low high flows suggested by the observed data at BAR7, is to check the downstream simulations at BAR6 (which are also consistent with the observed data at ZAM1 and ZAM2 further downstream). However, this assumes that the dynamics of the Barotse floodplain are simulated appropriately. Unfortunately, the model is not able to simulate the high flows, as well as the delayed peak in the wet season evident from the observed flows at BAR6 (Figure 6), despite quite good simulations for more than 50% of the upstream area (BAR3, BAR4, BAR7), and the fact that all the evidence suggests that most of the ungauged sub-areas (BAR1, BAR2 and BAR5) are unlikely to generate much higher wet season flows (underlain by Kalahari sands). While Figure 6 illustrates that the wetland sub-model is able to account for some of the peak flow delays, this is achieved (as might be expected) at the expense of the peak flows. The uncertainty issues therefore remain unresolved; are the observed data and new simulations at BAR7 behavioural, and the main problem associated with the wetland simulations, or are the observed data and simulations at BAR7 wrong, thus preventing the wetland sub-model from achieving a realistic downstream simulation?

For the semi-arid Zimbabwe sub-basins, the CS values in Table 3 only use the CE and %Bias values because the values based on log transformed flows are often misleading due to the large number of zero and very low flows. The selection of behavioural ensembles is further limited to those that have similar numbers of zero flow months to the observed data. These sub-basins are also impacted by water use (mostly agricultural, but some urban and mining supplies). The estimates from the observational data (see also Hughes and Farinosi, 2020) are assumed to be relatively uncertain and the input parameter ranges have been set at $\pm 20\%$ of the expected values. The real values could also be non-stationary over the gauging period (starting in the late 1950's), adding another source of unknown uncertainty.

The behavioural ensembles for GWA3 do not have substantially different parameter ranges than the full input range, despite there only being 11 ensembles accepted, further reinforcing the high level of equifinality in the model structure. An exception is that the lower range of the input PET values is not included in the behavioural set. The BGS recharge estimates (> 60 mm y⁻¹) are far greater than the range of 2 to 17 mm y⁻¹ simulated by the model. The runoff ratio range is 4.5% to 5.3%, consistent with semi-arid conditions and some water use. GWA4 has quite a large amount of water use and this is reflected in much lower runoff ratios of 1.2% to 1.6%, while the minimum recharge estimates are higher than GWA3 (a range of 9 to 19 mm y⁻¹). There is a weak positive relationship for both GWA3 and GWA4 between the parameters determining low flows (FT, POW, GW and GPOW) and the amount of assumed water use, suggesting some impacts of uncertainty in observational data on water use. For MAZ2 it was not possible to reproduce the observed number of zero flow months (42%) within the ensembles with the best CS values. While the simulated flows dry season flows are very low, they are not actually zero. This is one of the more recent observed stream flow records (2003 to 2017) and this result may be a reflection of hidden uncertainties in some of the other sites related to the non-stationarity of the water use data. The range of runoff ratios is between 12.9% and 13.9%, while simulated recharge is between (38 and 53 mm y⁻¹), both of which can be considered high for this semi-arid sub-basin. This is one of the few sub-basins where a single parameter (GW) has a much reduced range (15 to 20 mm month⁻¹) compared to the full input range (2 to 20 mm month⁻¹). One of the main problems with MAP2 is the fact that the total stream flow record (1951 to 2017) shows a great deal of non-stationarity, with a wetter period up to about 1984 and a much drier period with less frequent and generally lower flows from 1984 onwards. There is some evidence to suggest that the main wet season rainfalls were lower in the later period, while a contributory effect may be changes in land and water use. The extent to which this effects the models interpretation of the dominant processes is difficult to determine

without more reliable information. The main difference between this sub-area and the previous semi-arid ones is the low combined contribution of interflow and groundwater outflow. MAP3 has a range of runoff ratios of 5.0% to 6.9% and recharge depths of 1.8 to 20.5 mm y⁻¹, and appears to be dominated by adsorption excess surface runoff. However, the minimum values for the other processes given in Table 3 are not very representative of all the behavioural ensembles, which tend to have greater proportions of saturated excess surface and groundwater runoff. MAP4 is quite similar to MAP3 with slightly higher runoff ratios, but with a maximum recharge depth of 36 mm y⁻¹ amongst the behavioural ensembles. As with the previous Zimbabwe sub-areas the observed stream flow data are non-stationary with lower overall discharge volumes in the second half of the record (from the mid 1980s).

The Lake Malawi/Nyasa sub-basins are subject to rainfall and observed stream flow data uncertainties to varying degrees (Table 2). MODIS actual evapotranspiration data (AET: 2000 to 2014) has been used to try and resolve some of the uncertainties in the rainfall data, by comparing both CRU and UNIDEL mean annual rainfall data with the values derived from a simple water balance of observed stream flow depth plus MODIS AET depth (Table 4). However, this approach also has to take into account the potential uncertainties in the observed stream flow data (Table 4, ‘Comments’ row), as well as any differences related to the choice of the period used for the water balance checks (determined by the available stream flow data). Despite these additional uncertainties, the decision to use either CRU or UNIDEL rainfall data for RUK2 and RUH2, was quite clear, while either rainfall data set appears to be suitable for RUH1 and NAM1. For RUK1 and RUK3, neither rainfall data set appears to be suitable and both would need to be scaled to achieve a similar value to the water balance derived estimate.

For RUK1 the initial runs with the CRU rainfall data generate behavioural simulations that consistently under-estimate the higher flows in the flow duration curve and have runoff ratios that are over 40%, a very high value even for a topographically steep area. The model was re-run with UNIDEL rainfall scaled to generate a mean annual value of ~1 380 mm y⁻¹ (Table 4), after which the runoff ratio varies from 26.8% to 29.7%, the number of behavioural ensembles increases substantially and high flows are better estimated. It was, however, necessary to adjust the input parameter ranges to account for the much higher rainfall (notably the maximum soil moisture content was increased and the maximum values of the interflow and recharge parameters, FT and GW, were reduced). The two entries in Table 3 for this sub-area indicate that the effects of the input data uncertainties on the modelled processes is evident, with surface runoff normally playing a more important role in the UNIDEL forced simulations. The range of possible recharge depths is also greater in the UNIDEL simulations (29 to 144 mm y⁻¹), compared to the CRU forced simulations (72 to 139 mm y⁻¹). The initial uncertainty model runs for RUK2 did not yield ensembles that had as good statistics as the original manual calibrations (Hughes et al., 2020), suggesting that the manual calibration parameters were a relatively unique combination that could not be found even with 10 000 total ensembles. Reducing the range of some of parameter inputs made a substantial difference and generated 85 behavioural ensembles with runoff ratio and recharge ranges of 11.9% to 14.0% and 13.4 to 40.8 mm y⁻¹, respectively. While most of the full ranges of the input parameters are represented in the behavioural ensembles, the lower estimates of PET were not. Despite increasing the rainfall input to RUK3 (Table 4), the runoff ratios remain extremely high (50.4% to 57.0%), while the simulated recharge is also very high (120 to 305 mm y⁻¹), and even if the UNIDEL rainfall data are used (Table 4), the runoff ratios remain at greater than 40%. The distribution of runoff generation processes remains similar to the other Lake Malawi/Nyasa sub-basins, although the high contribution of intensity excess surface runoff is more consistent across the ensembles than in other sub-basins. There remains a large amount of uncertainty in the input climate data, as well as the response characteristics of this sub-basin.

The runoff ratio range for RUH2 is high at 38.2% to 43.0%, with a recharge range of 65 to 312 mm y⁻¹, and as with some other sub-basins the lower PET estimates do not seem to be valid. Both of these values are high and the estimated recharge is quite close to the BGS estimates (~146 mm y⁻¹). RUH1 is downstream of RUH2 and in order to search for behavioural ensembles independently of the effects of RUH2, the parameters of RUH2 are fixed at values representing one of the best ensembles. RUH1 has the highest number of behavioural ensembles and the highest maximum CS value of all the sub-areas. The range of runoff ratios

is 21.9% to 26.9%, and recharge varies from 16.8 to 163 mm y^{-1} (a wide range that also includes the BGS estimate). Table 3 illustrates that behavioural results can be obtained with a mix of different processes and this is further illustrated in Figures 7 and 8, and Table 5, based on four selected ensemble members. They are those with the lowest and highest proportion of saturated area surface runoff, and the lowest and highest proportion of total low flow processes (interflow and groundwater runoff). Generally, a decrease in saturated area surface runoff is associated with increases in both intensity (adsorption) excess surface runoff and interflow, while groundwater runoff is quite stable across all ensemble members (Figure 8). The range of process proportions amongst the ten best ensemble members is much lower, with both surface runoff processes being $\sim 25\%$, interflow between 6% and 18%, while groundwater runoff is 33% to 40%. A closer inspection of all the simulated years (1991 to 2008), suggests that different behavioural ensemble members perform better in some years than others. To what extent this can be associated with any uncertainties in the accuracy of the input climate or observed stream flow data, or is just part of the overall modelling uncertainty, is almost impossible to resolve without more information. For NAM1 the first noticeable effect is the consistently low values of the saturated surface runoff parameter (SSR) within the behavioural ensembles, and this result is appropriate given clear evidence of Dambo features. The range of runoff ratios is 9.5% to 11.1%, and recharge depths is 4.8 to 30.0 mm y^{-1} , both of which are quite similar to the results for RUK2 (in a similar geographic location). As already noted for some of this group of sub-basins, the lower estimates of PET are not included in the behavioural ensembles.

Tables 2 and 4 both refer to some of the uncertainties in the observed stream flow data for the Lake Malawi/Nyasa sub-basins, but these have largely been ignored in the presentation of the results. The main reason is that the differences between certain parts of the records are far too large to be considered just uncertainties. Figure 9 illustrates the problem in three sub-basins, and they all have substantial periods of missing data within them. The majority of the total record (1957 to 2009) for NAM1 shows a consistent response and it is only the last 9 years where all flows are higher by a factor of ~ 4.7 on average, a figure too high to attribute to any reasons apart from errors. A similar situation arises for RUH1 (total record of 1972 to 2018), where the stream flows for the last 6 years are much higher, after an extended period of 5 years of almost totally missing data. One of the main problems with RUH1 is that almost all of the low flow months are missing after 2012. For RUK3 it is the earlier part of the record where the problems exist, while several years in the middle (1981 to 1994) and the later (2003 to 2018) period are consistent with each other. This may be related to an error in the data records and failure to convert stage observations in feet to metres. This error has been noted in some other early Tanzanian records, and is quite simple to fix if the raw observational data and rating curve information are made available (which they are typically not). However, there is an additional uncertainty issue in RUK3 and the model fit for the later period is far better than the middle period, the former having a maximum CS value of 3.148, compared to 1.223 for the later period (CS = 3.112 for the total period used in the model). The implication is that very different results would be obtained if these two periods were used separately.

DISCUSSION AND CONCLUSIONS

The CS statistic used in this study represents a useful approach for identifying the most behavioural ensemble members. The part of the statistic that includes the % bias objective functions assesses the long-term water balance of the simulations, while the Nash coefficients assess the simulations with respect to the individual monthly stream flows. Table 3 illustrates that there is quite a wide range of maximum CS values and a large part of this variation is expected to be related to spatial variations in the representation of real individual monthly rainfall values, in the absence of enough local data. However, one of the key general observations is that uncertainties in some of the climate inputs and the observational stream flow data are unlikely to be the main effects causing uncertainties in the relative proportions of the main four runoff generation processes. The results suggest that the equifinalities in the model structure will always dominate. While it might be postulated that the behavioural ranges given in Table 3 could contain some outliers, Figure 10 suggests that this is generally not the case. The weighted cumulative frequency is based on the CS values for each ensemble member (i.e. giving slightly more weight to those with better overall fits to the observed data). For RUH1 the ranges of all processes could be reduced slightly as the curves have flattened tops and bottoms. This is not

the case for the semi-arid MAZ2 sub-basin. Clearly, there are only some combinations of different processes that are behavioural (or not), but as noted above (and illustrated in Figure 7) for RUH1, isolating these combinations is not straightforward and simulations of similar total stream flow response can be made up of a number of different combinations, particularly in the wetter sub-basins where all four processes can play a substantial role. While many of the uncertainties will be associated with inadequate representation of the real monthly rainfall variations, additional uncertainties exist in some of the Lake Malawi/Nyasa sub-basins where the rainfall data (both CRU and UNIDEL) can be systematically biased (Table 4). However, even though this may result in unrealistic simulations of the runoff ratio, the distribution of simulated processes (and the associated uncertainties) remains broadly the same, as illustrated by RUK1 in Table 3.

The best way to reduce some of the uncertainties is to add more observational data, even if those data are not directly quantifying the model process components, or are themselves uncertain. In this study MODIS AET data have been used in some sub-basins (Table 4) to resolve some of the long-term water balance uncertainties, largely related to the rainfall data. LAI data have also been used to guide the parameterisation of the interception parameters in the model, but mostly in relative terms across the sub-basins in different parts of the basin, as well as between the warm wet season and cooler dry season. However, there are too many uncertainties in the conversion of LAI values reported in the literature (see for example De Groen and Savenije, 2006; Wu et al., 2019; Navar, J., 2020) into interception depths for different rainfall regimes to allow the LAI values to be used directly to constrain simulated interception depths. The example provided for KAF4 also suggests that the model can compensate for quite large uncertainties in the simulation of interception depth by changing the simulations of soil moisture evapotranspiration without substantially affecting the accuracy of the stream flow simulations.

Some of the example sub-basins show clear evidence of Dambos (KAF4, BAR3, NAM3 and RUK2) on Google Earth imagery, while others show some evidence (KAF11, BAR3, GWA3, MAZ2 and RUH1). A previous study linked these features to the occurrence of saturation excess surface runoff and demonstrated improved simulations when this process was included in the model (Hughes and Mazibuko, 2018). It should therefore be possible to reduce the uncertainty ranges of the ‘Sat. surface’ column in Table 3 by conducting a more detailed analysis (rather than the simple visual assessment used here; Figure 4) of the frequency of Dambo occurrence. Figure 8 illustrates that reducing the saturated area uncertainties, the uncertainties in at least some of the other process simulations could also be reduced. During this study it was thought that remotely sensed soil moisture data might be useful to support this type of analysis. However, while some of the patterns in the observational data could be linked to landscape features, most could not. It is possible that a more detailed investigation of the different soil moisture remote sensing data available might reveal improved linkages with landscape features and hydrological processes, and therefore offer some benefits for setting up models. However, this was rather beyond the scope of this study and should probably be conducted in areas where more ground-truth data are available than in the large sub-basins of the Zambezi River basin. Further development of these earth observation data, however, could represent an important contribution to understanding processes in data scarce areas, as demonstrated recently for the correction of precipitation reanalyses data (Brocca et al. 2019), or for the validation of river flow observation data (Brocca et al., 2020).

Within South Africa there is a national coverage of groundwater recharge estimates (DWAF, 2005), which has proved to be very useful for constraining model simulations and removing some of the equifinality in the simulation of low flow generation processes (Tanner and Hughes, 2013). Unfortunately, the BGS estimates appear to be too uncertain for that purpose in the Zambezi River basin. Some of the uncertainty in the groundwater recharge estimates for this study are also related to the inclusion of a riparian evapotranspiration component in the model. To achieve the same groundwater outflow pattern, it is possible to have relatively high recharge combined with a large riparian area or vice versa. Any data that could limit the range of the riparian loss parameter, would therefore be useful. In theory, it should be possible to use remote sensing data for this purpose (to identify denser vegetation, or enhanced actual evapotranspiration areas close to channels). In practice, this could be quite difficult and time consuming for the large areas covered by the Zambezi sub-basins.

In those basins where water use is expected to have a substantial impact on gauged stream flows (mostly the semi-arid Zimbabwe sub-basins used in this study), the uncertainties in the water use data has an inevitable impact on the simulated processes. This effects mostly low flows (through groundwater recharge and outflow processes) as the water use volumes are relatively small compared to the wet season stream flows. While issues of spatial scale pervade the whole modelling exercise due to the large size of the sub-basins, the water use uncertainties can be exacerbated by the model spatial structure, particularly when the main water uses are from distributed farm dams. For example, in MAZ2 the majority of the water use and farm dams are in the headwater areas, which may be higher runoff areas on the basis of rainfall spatial variations. The proportion of the sub-basin area that contributes to these dams (a model parameter) should therefore take into account the expected ‘sub-grid’ variations in runoff generation, introducing yet another source of uncertainty.

For some of the sub-basins it can be quite easily demonstrated that there are large uncertainties in both the forcing rainfall data, as well as parts of the observed stream flow data records used to evaluate the simulations. Most of these occur within the lake Malawi/Nyasa group of sub-basins (Tables 2 and 4), but tend to have little influence on the distribution of simulated processes. There are, of course, some uncertainties in the rainfall data (including those inherent in the use of a monthly time step) that are impossible to quantify without more local observational data, but these are reflected more in the overall quality of the simulations, rather than the simulation of individual processes. The group of sub-basins above the Barotse floodplain represent a situation where some identified uncertainties in the observed stream flow data could impact on the dominant processes simulated by the model. There are some incompatibilities between the upstream (BAR3, BAR4 and BAR7) and downstream (BAR5 and ZAM1) observed data that cannot be readily accounted for by the impacts of the wetland (Figure 6). The main issue appears to be in the representation of the peak wet season flows, particularly from the quite large sub-basin BAR7. The simulations for BAR7 are acceptable compared to observed stream flows and the process representations (mostly interflow and groundwater outflow) are consistent with the other sub-basins underlain by deep Kalahari sand deposits (Table 3). However, to achieve a match to the downstream observed data, this sub-basin would require much higher wet season peaks generated by surface runoff processes, similar to BAR7 (Table 3). This study was not able to resolve these incompatibilities and further assessments of the hydrological responses in the western headwaters of the Zambezi are strongly recommended.

In terms of the potential benefits to simulating ungauged sub-basins, referred to in the introduction, there is still too much uncertainty in the simulation of individual processes and not enough observational data to support their identification. The use of regionalised indices of the total response of sub-basins, both internationally (Westerberg et al., 2016), as well as for southern Africa (Hughes, 2019; Kabuya et al., 2020), seem to remain the best recommendations for dealing with ungauged sub-basins. This paper therefore reaches similar conclusions to McMillan (2020) that some of the indices (or signatures) are related to multiple processes that are difficult to disentangle. This study suggests that improved, model independent, quantification of groundwater recharge depths offers some potential gains, as does the mapping of landscape features (Dambos and others) that are likely to generate saturation excess surface runoff.

ACKNOWLEDGEMENTS

The work presented in this paper was partially conducted within the activities of the African Union - NEPAD African Network of Centres of Excellence on Water Sciences and Technology - ACEWATER phase 2 project. Contribution from the European Commission, in particular the Directorate-General for International Cooperation and Development (DEVCO) and the Joint Research Centre (JRC), is gratefully acknowledged. The authors would like to thank the Zambezi Watercourse Commission (ZAMCOM) for making available the stream flow information used in the analyses. We are grateful to Dr Sukhmani Mantel of Rhodes University for helping to process some of the soil moisture data.

SOFTWARE AND DATA AVAILABILITY

The Pitman model is available as part of the SPATSIM modelling framework available from

<https://www.ru.ac.za/iwr/research/spatsim/>. Further details about the Pitman model are included in the documentation included with the download (see the Pitman_Guide.pptx file in the SPATSIM_V3/doc folder). The model setup (including the forcing data, parameter sets, simulation results, etc.) can be obtained on request from one of the authors, subject to some restrictions on the distribution of the observed streamflow data.

REFERENCES

- Alfieri L, Lorini V, Hirpa F, Harrigan S, Zsoter E, Prudhomme C, Salamon P. 2019. A global streamflow reanalysis for 1980-2018, *Journal of Hydrology* X, 6. <https://doi.org/10.1016/j.hydroa.2019.100049>.
- Beven K. 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* 320: 18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>.
- Beven KJ. 2009. *Environmental modelling: An uncertain future?* Routledge, Abingdon, UK.
- Beven KJ. 2012. Causal models as multiple working hypotheses about environmental processes. *Comptes Rendus Geoscience* 344: 77–88. <https://doi.org/10.1016/j.crte.2012.01.005>.
- Bloschl G, Sivapalan M, Wagener T, Viglione A, Savenije H. (Eds.). 2013. *Runoff Prediction in Ungauged Basins. Synthesis Across Processes, Places and Scales*. Cambridge University Press, UK.
- Brocca L, Filippucci P, Hahn S, Ciabatta L, Massari C, Camici S, Schuller L, Bojkov B, Wagner W. 2019. SM2RAIN–ASCAT (2007–2018): Global Daily Satellite Rainfall Data from ASCAT Soil Moisture Observations. *Earth System Science Data*, 11 (4): 1583–1601. <https://doi.org/10.5194/essd-11-1583-2019>.
- Brocca L, Massari C, Pellarin T, Filippucci P, Ciabatta L, Camici S, Kerr YH, Fernandez-Prieto D. 2020. River Flow Prediction in Data Scarce Regions: Soil Moisture Integrated Satellite Rainfall Products Outperform Rain Gauge Observations in West Africa. *Scientific Reports* 10 (1): 12517. <https://doi.org/10.1038/s41598-020-69343-x>.
- Busker T, de Roo A, Gelati E, Schwatke C, Adamovic M, Bisselink B, Pekel J-F, Cottam A. 2019. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrol. Earth Syst. Sci.*, 23, 669–690. <https://doi.org/10.5194/hess-23-669-2019>.
- Chien H, Mackay DS. 2014. How much complexity is needed to simulate watershed streamflow and water quality? A test combining time series and hydrological models. *Hydrological Processes* 28(22): 5624–5636. <https://doi.org/10.1002/hyp.10066>.
- Clark MP, Rupp DE, Woods RA, Tromp-van Meerveld HJ, Peters NE, Freer JE. 2009. Consistency between hydrological models and field observations: Linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrological Processes*, 23(2): 311–319. <https://doi.org/10.1002/hyp.7154>.
- Das N N, Entekhabi D, Dunbar RS, Chaubell MJ, Colliander A, Yueh S, Jagdhuber T et al. 2019. The SMAP and Copernicus Sentinel 1A/B Microwave Active-Passive High Resolution Surface Soil Moisture Product. *Remote Sensing of Environment* 233: 111380. <https://doi.org/10.1016/j.rse.2019.111380>.
- De Groen MM, Savenije HHG. 2006. A monthly interception equation based on the statistical characteristics of daily rainfall. *Water Resources Research*, 42: 1–10. <https://doi.org/10.1029/2006WR005013>.
- Dorigo W, Wagner W, Albergel C, Albrecht F, Balsamo G, Brocca L, Chung D et al. 2017. ESA CCI Soil Moisture for Improved Earth System Understanding: State-of-the Art and Future Directions. *Remote Sensing of Environment* 203: 185–215. <https://doi.org/10.1016/j.rse.2017.07.001>.
- DWAF. 2005. *Groundwater Resource Assessment II*. Department of Water Affairs and Forestry, Pretoria, South Africa.
- Euser T, Winsemius HC, Hrachowitz M, Fenicia F, Uhkenbrook S, Savenije HHG. 2013. A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*,

17: 1893–1912. <https://doi.org/10.5194/hess-17-1893-2013>.

Fenicia F, Savenije HHG, Matgen P, Pfister L. 2008. Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research*, 44. <http://dx.doi.org/10.1029/2006WR005563>.

Gallart F, Latron J, Llorens P, Beven K. 2007. Using internal catchment information to reduce the uncertainty of discharge and baseflow predictions. *Advances in Water Resources* 30(4): 808–823. <https://doi.org/10.1016/j.advwatres.2006.06.005>.

Gan, T.Y., Dlamini, E.M., Biftu, G.F., 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modelling. *Journal of Hydrology*, 192(1-4), 81-103. [http://dx.doi.org/10.1016/S0022-1694\(96\)03114-9](http://dx.doi.org/10.1016/S0022-1694(96)03114-9)

Gonzalez Sanchez R, Seliger R, Fahl F, De Felice L, Ouarda TBMJ, Farinosi F. 2020. Freshwater use of the energy sector in Africa. *Appl. Energy*, 270: 115171. <https://doi.org/10.1016/j.apenergy.2020.115171>.

Gruber A, Dorigo WA, Crow W, Wagner W. 2017. Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Transactions on Geoscience and Remote Sensing* 55(12): 6780–92. <https://doi.org/10.1109/TGRS.2017.2734070>.

Gruber A, Scanlon T, van der Schalie R, Wagner W, Dorigo, W. 2019. Evolution of the ESA CCI Soil Moisture Climate Data Records and Their Underlying Merging Methodology. *Earth System Science Data* 11(2): 717–39. <https://doi.org/10.5194/essd-11-717-2019>.

Harris I, Jones PD, Osborn TJ, Lister DH. 2014. Updated high-resolution grids of monthly climatic observations – the CRUTS3.10 dataset. *International Journal of Climatology*, 34 (3): 623–642. <https://doi.org/10.1002/joc.3711>.

Hrachowitz M, Savenije HHG, Blöschl G, McDonnell JJ, Sivapalan M, Pomeroy JW, Arheimer B, Blume T, Clark MP, Ehret U, Fenicia F, Freer JE, Gelfan A, Gupta HV, Hughes DA, Hut RW, Montanari A, Pande S, Tetzlaff D, Uhlenbrook S, Wagener T, Winsemius HC, Woods RA. 2013. A decade of Predictions in Ungauged Basins (PUB) - a review. *Hydrological Sciences Journal*, 58(7): 1198-1255. <https://doi.org/10.1080/02626667.2013.803183>.

Hughes DA. 2004. Incorporating ground water recharge and discharge functions into an existing monthly rainfall-runoff model. *Hydrological Sciences Journal*, 49(2): 297–311. <https://doi.org/10.1623/hysj.49.2.297.34834>.

Hughes DA. 2013. A review of 40 years of hydrological science and practice in southern Africa using the Pitman rainfall-runoff model. *Journal of Hydrology*, 501: 111–124. <https://doi.org/10.1016/j.jhydrol.2013.07.043>.

Hughes DA. 2016. Hydrological modelling, process understanding and uncertainty in a southern African context: lessons from the northern hemisphere. *Hydrological Processes*, 30(14): 2419-2431. <https://doi.org/10.1002/hyp.10721>.

Hughes DA. 2019. Facing a future water resources management crisis in sub-Saharan Africa. *Journal of Hydrology: Regional Studies*, 23. <https://doi.org/10.1016/j.ejrh.2019.100600>

Hughes DA, Farinosi F. 2020. Assessing development and climate variability impacts on water resources in the data scarce Zambezi River basin. Part 2: Simulating future scenarios of climate and development. *Journal of Hydrology: Regional Studies*. Under review.

Hughes DA, Mantel SK, Farinosi F. 2020. Assessing development and climate variability impacts on water resources in the data scarce Zambezi River basin. Part 1: Initial model setup. *Journal of Hydrology: Regional Studies*. Under review.

Hughes DA, Mantel SK. 2010. Estimating the uncertainty in simulating the impacts of small farm dams on streamflow regimes in South Africa. *Hydrological Sciences Journal*, 55 (4): 578-592.

<https://doi.org/10.1080/02626667.2010.484903>.

Hughes DA, Mazibuko S. 2018. Simulating saturation excess surface run-off in a semi-distributed hydrological model. *Hydrological Processes*, 32: 2685-2694. <https://doi.org/10.1002/hyp.13182>.

IFPRI (International Food Policy Research Institute). 2019. Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 1.1. <https://doi.org/10.7910/DVN/PRFF8V>, Harvard Dataverse, V3.

Jakeman AJ, Hornberger GM. 1993. How much complexity is warranted in a rainfall-runoff model? *Water Resources Research* 29(8): 2637-2649.

Kabuya PM, Hughes DA, Tshimanga RM, Trigg MA, Bates P. 2020. Establishing uncertainty ranges of hydrologic indices across climate and physiographic regions of the Congo River Basin, *Journal of Hydrology: Regional Studies*, 30. <https://doi.org/10.1016/j.ejrh.2020.100710>.

Kirchner JW. 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* 42: <https://doi.org/10.1029/2005WR004362>.

Limpitlaw D, Gens R. 2006. Dambo mapping for environmental monitoring using Landsat TM and SAR imagery: Case study in the Zambian Copperbelt. *International Journal of Remote Sensing*, 27(21): 4839-4845. <https://doi.org/10.1080/01431160600835846>.

Lucey JTD, Reager JT, Lopez SR. 2020. Global partitioning of runoff generation mechanisms using remote sensing data. *Hydrology and Earth Systems Science*, 24: 1415-1427. <https://doi.org/10.5194/hess-24-1415-2020>.

MacDonald AM, Bonsor HC, Dochartaigh BEO, Taylor, RG. 2012. Quantitative Maps of Groundwater Resources in Africa. *Environmental Research Letters*, 7(2): 024009. <https://doi.org/10.1088/1748-9326/7/2/024009>.

Mao J, Yan B. 2019. Global Monthly Mean Leaf Area Index Climatology, 1981-2015. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAAC/1653>.

McMillan, H. 2020. Linking hydrological signatures to hydrologic processes: A review. *Hydrological Processes*, 34, 1393-1409. <https://doi.org/10.1002/hyp.13632>.

McMillan HK, Clark MP, Bowden WB, Duncan M, Woods RA. 2011. Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure. *Hydrol. Process.* 25: 511-522. <https://doi.org/10.1002/hyp.7841>.

McMillan H, Krueger T, Freer T. 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrol. Process.* 26: 4078-4111. <https://doi.org/10.1002/hyp.9384>.

Moore RJ. 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30(2): 273-297. <https://doi.org/10.1002/hyp.13632>.

Navar J. 2020. Modeling rainfall interception loss components of forests. *Journ. Hydrol.* 584. <https://doi.org/10.1016/j.jhydrol.2019.124449>.

Pechlivanidis IG, Jackson BM, McIntyre NR, Wheeler HS. 2011. Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global Nest Journal*, 13 (3): 193-214.

Pekel J.-F, Cottam A, Gorelick N, Belward A S. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633): 418-422. <https://doi.org/10.1038/nature20584>.

- Perrin C, Michel C, Andreassian V. 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242(3–4): 275–301. [https://doi.org/10.1016/S0022-1694\(00\)00393-0](https://doi.org/10.1016/S0022-1694(00)00393-0).
- Pokhrel P, Gupta HV. 2009. Regularized Calibration of a Distributed Hydrological Model Using Available Information About Watershed Properties and Signature Measures. IAHS-AISH Publication No. 333: 20–25.
- Pitman WV, 1973. A Mathematical Model for Generating Monthly River Flows from Meteorological Data in South Africa Report No. 2/73. Hydrological Research Unit, University of the Witwatersand, Johannesburg, South Africa.
- Sadeghi M, Gao L, Ebtehaj A, Wigneron J-P, Crow WT, Reager JT, Warrick AW. 2020. Retrieving global surface soil moisture from GRACE satellite gravity data. *Journal of Hydrology*, 584. <https://doi.org/10.1016/j.jhydrol.2020.124717>.
- Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resources Research* 38(11): 1241. <https://doi.org/10.1029/2001WR000978>.
- Tanner JL, Hughes DA. 2013. Assessing uncertainties in surface-water and groundwater interaction modelling - a case study from South Africa using the Pitman model. Chapter 9 In: J. Cobbing, S. Adams, I. Dennis and K. Riemann (Editors), *Assessing and Managing Groundwater in Different Environments*, International Association of Hydrogeologists Selected Papers. CRC Press, Taylor and Francis Group, London UK, 121-134. <https://doi.org/10.1201/b15937>.
- Todini E. 2011. History and perspectives of hydrological catchment modelling. *Hydrology Research*, 42 (2-3): 73-85. <https://doi.org/10.2166/nh.2011.096>.
- Velpuri NM, Senay GB, Singh RK, Bohms S, Verdin JP. 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. *Remote Sens. Environ.*, 139: 35-49. <https://doi.org/10.1016/j.rse.2013.07.013>.
- von der Heyden C J. 2004. The hydrology and hydrogeology of dambos: A review. *Progress in Physical Geography*, 28(4): 544–564. <https://doi.org/10.1191/0309133304pp424oa>.
- Ward RC. 1984. On the response to precipitation of headwater streams in humid areas. *Journal of Hydrology*, 74 (1-2): 171-189. [https://doi.org/10.1016/0022-1694\(84\)90147-1](https://doi.org/10.1016/0022-1694(84)90147-1).
- Ward RC. 1985. Hypothesis-testing by modelling catchment response, II. An improved model. *Journal of Hydrology*, 81 (3-4): 355-373. [https://doi.org/10.1016/0022-1694\(85\)90038-1](https://doi.org/10.1016/0022-1694(85)90038-1).
- Westerberg IK, McMillan HK, 2015. Uncertainty in hydrological signatures. *Hydrol. Earth Syst. Sci.*, 19: 3951–3968. <https://doi.org/10.5194/hess-19-3951-2015>.
- Westerberg IK, Wagener T, Coxon G, McMillan HK, Castellarin A, Montanari A, Freer, J. 2016. Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resour. Res.*, 52: 1847–1865. <https://doi.org/10.1002/2015WR017635>.
- Willmott CJ, Matsuura K. 2001. Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1999), http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html.
- Winsemius HC, Schaeffli B, Montanari A, Savenije HHG. 2009. On the calibration of hydrological models in ungauged basins: a framework for integrating hard and soft hydrological information. *Water Resources Research*, 45. <https://doi.org/10.1029/2009WR007706>.
- Wu J, Liu L, Sun C, Su Y, Wang C, Yang J, Liao J, He X, Li Q, Zhang C, Zhang H. 2019. Estimating Rainfall Interception of Vegetation Canopy from MODIS Imageries in Southern China. *Remote Sens.* 2019, 11: 2468. <https://doi.org/10.3390/rs11212468>

LIST of FIGURES

Figure 1 Structure of the main sub-basin runoff generation components of the Pitman model (the model parameter symbols are shown in *italics*, while the full names are given for the state variables, such as IQ, RCH, S, etc.).

Figure 2 Zambezi River basin, riparian countries and simulated sub-basins (the 19 gauged areas used in this study are shaded in grey). The gauge at BAR6 is used to help resolve some of the uncertainties in the upstream area.

Figure 3 Minimum and maximum mean monthly LAI values (Mao and Yan, 2019) for all sub-basins and some sample seasonal distributions.

Figure 4 A Google Earth image of part ($\sim 2\,000\text{ km}^2$) of the KAF4 sub-basin showing the light coloured Dambo features (a), and the relationship between simulated relative soil moisture content and saturated area for different SSR parameters of the Pitman model (b).

Figure 5 Runoff processes simulated by two equally behavioural ensembles, with low and high groundwater recharge estimates.

Figure 6 Simulated inflows and outflows for the Barotse floodplain sub-basin (BAR5) and observed flows at BAR6.

Figure 7 Observed and simulated (four ensemble members) stream flows for RUH1. The simulations are drawn from the behavioural ensembles with some extremes of different process representations.

Figure 8 Details of the process simulations for the same four ensembles used in Figure 7 for RUH1 (note that differences in the responses for individual years between Figure 7 and 8 are associated with the inclusion of upstream flows from RUH2 in the total stream flow data shown in Figure 7).

Figure 9 Flow duration curves for some of the Lake Malawi/Nyasa sub-basins using different periods of the observed stream flow records.

Figure 10 Weighted (using the CS values) cumulative frequency distributions of process proportions for all the behavioural ensemble members for RUH1 and MAZ2 sub-basins.

Table 1 Details of the gauge station data used in the study.

Sub-basin	River	Location	National Gauge ID	Record length
KAF4	Lunga River	Chifumpa	ZAM4-560	1959-2005
KAF11	Kafue River	Smith Bridge	ZAM4-130	1959-2005
BAR3	Kabompo River	Watopa	ZAM1-950	1958-2016
BAR4	Luanginga River	Kalabo	ZAM2-250	1961-2013
BAR7	Zambezi River	Chavuma	ZAM1-105	1959-2014
CHB2	Linyanti River	Sajawa	BOT 6414	1980-2003
GWA3	Shangani River	Gwayi Confluence Flumes	ZIM A36	1961-2014
GWA4	Gwayi River	Dahlia	ZIM A38	1966-2017
MAZ2	Mazowe River	Mazowe Bridge	ZIM D75	2003-2017
MAP2	Umfuli River	Twyford Weir	ZIM C12	1950-2017
MAP3	Umswesws River	Claw Dam	ZIM C48	1957-2016
MAP4	Sebakwe River	Lower Ziwegwe Dam	ZIM C36	1955-2017
RUK1	Songwe River	Mwandenga	MAL 9B7	1985-2012
RUK2	S. Rukuru River	Mlowe	MAL 7G18	1985-2009
RUK3	Kiwira River	Ipyana	TAN 1RC2A	1986-2017
RUH1	Ruhuru River	Kilonge	TAN 1RB3	1991-2007
RUH2	Ruhuru River	Masigira	TAN 1RB2	1971-2017

Sub-basin	River	Location	National Gauge ID	Record length
NAM1	Bua River	Roadbridge	MAL 5C1	1958-2000

Table 2 Summary of the gauged sub-basins used in the study and some comments on the sources of observational data uncertainties.

Sub-basin	Area (km ²)	Assume quality of forcing climate data	Other key observational data uncertainties
KAF4	21 582	Good	Effects of water use. Very high in stream flow data for some sub-basins & high in recharge estimates related to underlying sandy aquifer.
KAF11	8 310	Good	
BAR3	69 933	Good	
BAR4	33 284	Good	
BAR7	78 253	Good	Effects of small farms dams & water use (including direct river & groundwater abstractions). MAP3 and MAP4 are also impacted by upstream major dams & their operation.
CHB2	35 002	Good	
GWA3	17 965	Moderate (the key uncertainties are in the expected high level of spatial variation in these semi-arid areas).	
GWA4	19 925		
MAZ2	13 224		Moderate to high in stream flow data.
MAP2	5 243		
MAP3	2 407		
MAP4	4 333		
RUK1	3 905	Poor	Very high in stream flow data. High in stream flow data.
RUK2	12 007	Moderate	
RUK3	1 706	Poor	
RUH1	14 331	Good	
RUH2	1 944	Moderate to poor	Moderate to high in stream flow data. Moderate to high in stream flow data.
NAM1	10 532	Good	

Table 3 Summary of results, including the maximum statistic value (CS), number of behavioural ensembles and range of process proportions simulated by the model.

Sub-basin	Max. CS	No. behavioural	Processes range as % of total runoff	
			Sat. Surface	Int. excess surface
KAF4	3.563	98	8.5 – 27.6	2.6 – 21.4
KAF11	3.469	34	0.0 – 14.5	1.9 – 22.1
BAR3	3.507	56	6.7 – 30.4	3.6 – 16.8
BAR4	3.305	20	0.0 - 13.7	0.6 - 15.8
BAR7	3.501	12	0.0 - 1.7	0.5 - 6.1
CHB2	2.669	6	0.0 - 0.8	0.0 - 3.0
GWA3	1.347	11	20.6 - 32.0	28.3 - 58.8
GWA4	1.405	17	0.0 - 9.0	26.8 - 63.4
MAZ2	1.454	59	9.6 - 24.7	20.6 - 50.0
MAP2	1.460	8	13.9 - 41.9	10.0 - 64.3
MAP3	1.500	552	0.0 - 40.8	17.2 - 99.7
MAP4	1.465	121	0.0 – 29.1	14.6 – 91.5
RUK1 (CRU)	3.191	19	0.0 – 8.4	11.7 – 46.1
RUK1 (UNI)	3.217	207	0.0 – 31.4	7.3 – 66.2
RUK2	3.238	84	21.0 – 37.6	4.9 – 27.0
RUK3	3.112	96	0.0 – 17.2	27.0 – 60.6
RUH1	3.658	986	4.1 – 30.5	5.8 – 50.0
RUH2	3.283	75	5.6 – 21.2	7.2 – 23.7
NAM1	3.358	34	21.3 – 44.4	5.9 – 45.0

Note: Max. CS for the Zimbabwe sub-basins (GWA3 to MAP4) do not include the log transformed statistics (i.e. max. value = 2) and the behavioural ensembles are further constrained by the number of months of zero flows. The behavioural ensemble members for sub-basins with maximum CS values of >3.2 are chosen as those with CS values of >0.05*MaxCS, the others are chosen as those with CS values of >0.1*MaxCS (to obtain enough behavioural ensemble members to analyse the range of different process proportions).

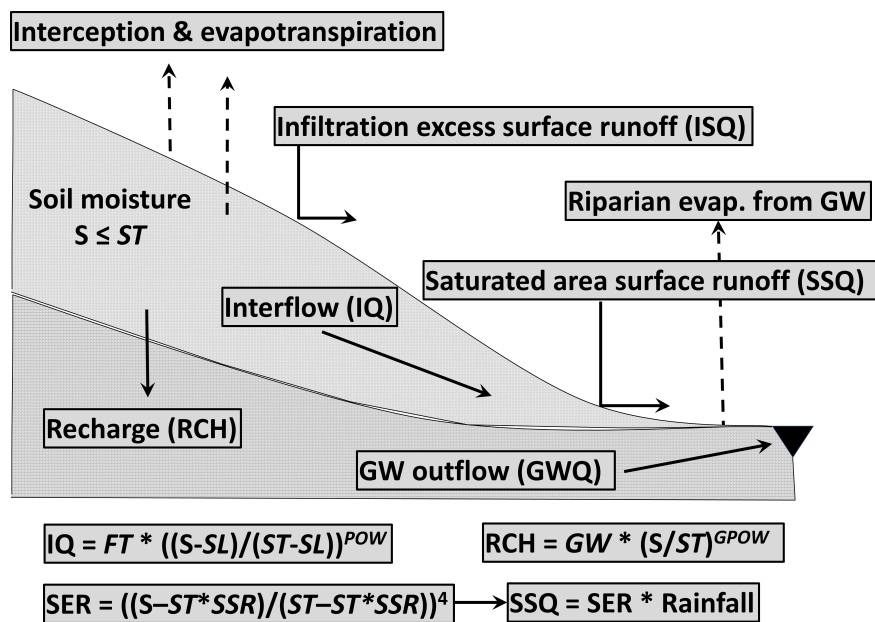
Table 4 Assessment of combined observational uncertainties for the Lake Malawi/Nyasa sub-basins.

WB component	Sub-area	Sub-area	Sub-area	Sub-area	Sub-area	Sub-area
	RUK1	RUK2	RUK3	RUH1	RUH2	NAM1
Stream flow (mm y ⁻¹)	621.6	111.7	901.2	310.0	540.0	104.6
MODIS AET (mm y ⁻¹)	759.6	855.4	877.2	876.0	879.6	826.4
Water balance rainfall (mm y ⁻¹)	1381.2	967.1	1778.4	1186.0	1419.6	931.0
CRU rainfall (mm y ⁻¹)	964.8	966.7	1399.2	1172.4	1174.8	911.0
UNIDEL rainfall (mm y ⁻¹)	1628.4	844.8	2214.0	1204.8	1381.2	912.8

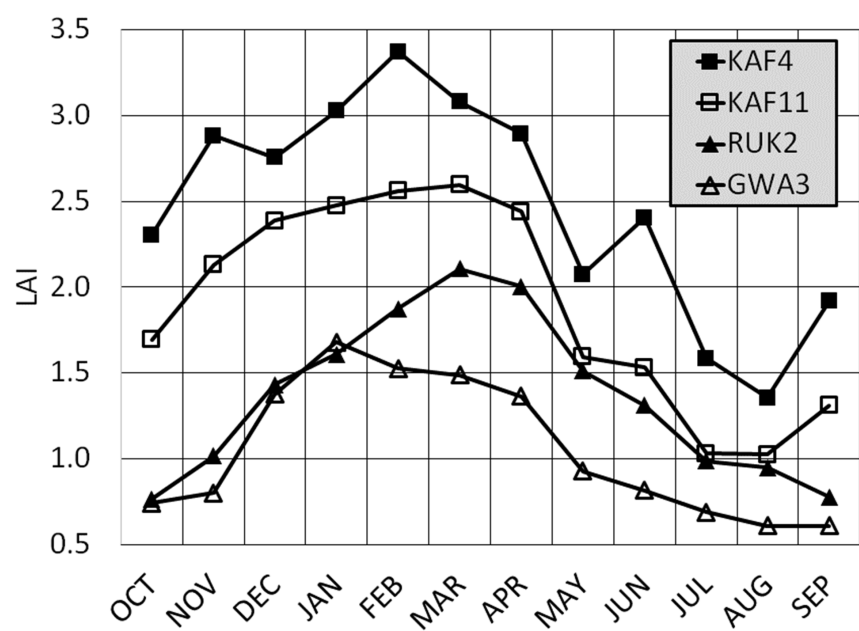
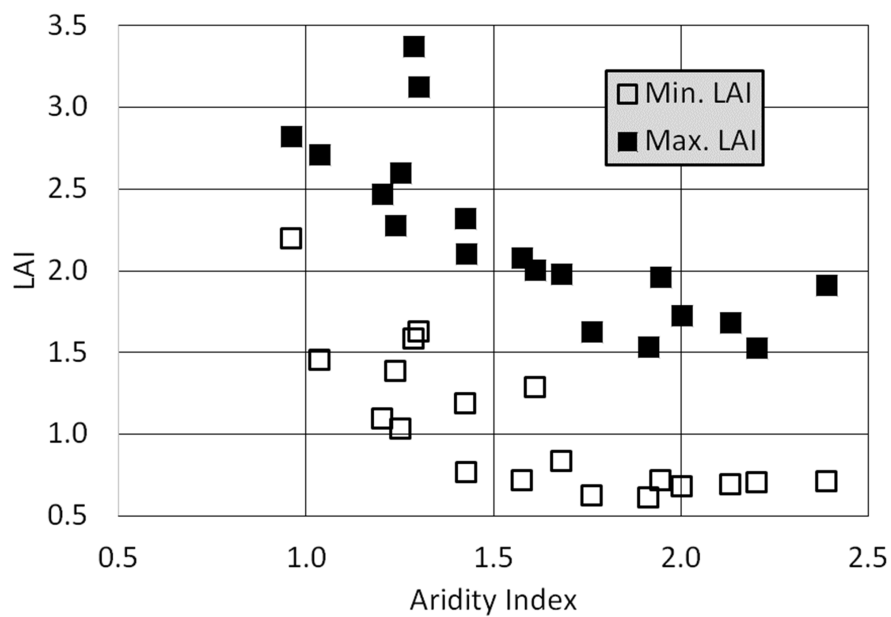
WB component	Sub-area	Sub-area	Sub-area	Sub-area	Sub-area	Sub-area
Comments	Observed stream flow data after 2000 is higher than 1985 to 2000.	Observed flows appear consistent throughout the record (1986-2008).	Observed stream flow data before 1972 is much higher.	Observed data after 2008 appears to be very wrong.	Observed stream flow data before 2000 is ~10% higher.	Observed flows after 2000 appear to be too high.
Decision	Use scaled UNIDEL	Use CRU	Use scaled UNIDEL	Use CRU or UNIDEL	Use UNIDEL	Use CRU or UNIDEL

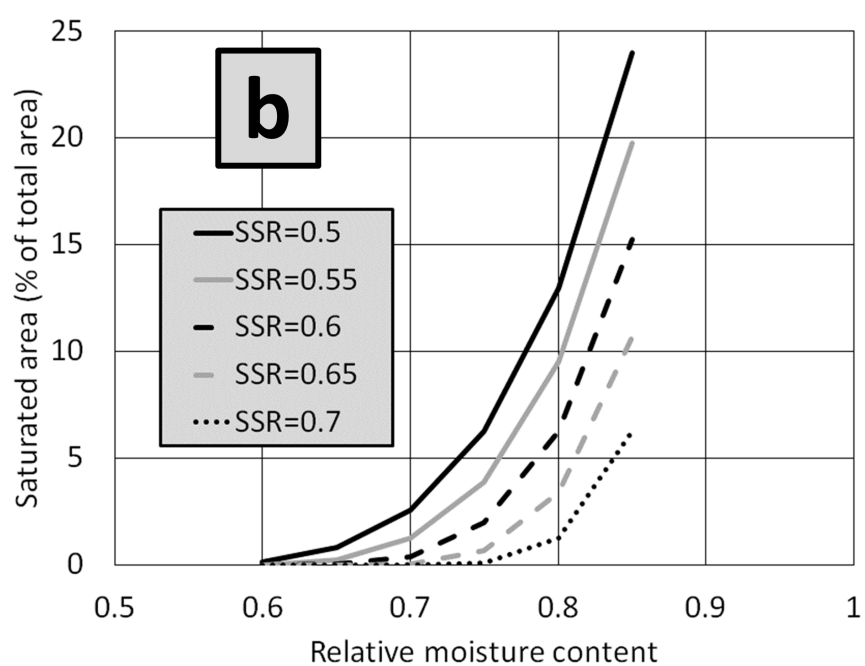
Table 5 Proportions of different processes represented in the four ensemble members shown in Figure 8 for sub-basin RUH1.

Process	Ensemble member description	Ensemble member description	Ensemble member description	Ensemble member description
	Lowest Sat. Area	Highest Sat. Area	Lowest Low Flows	Highest Low Flows
Sat. surface	4.1%	30.6%	26.1%	11.5%
Int. excess surface	45.7%	9.5%	24.2%	20.8%
Interflow	22.2%	4.9%	7.6%	34.6%
GW Outflow	23.9%	24.3%	15.8%	21.5%

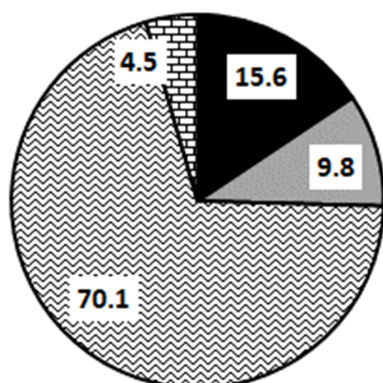








Low recharge (17 mm y⁻¹)



High recharge (51 mm y⁻¹)

