Hydrological Modelling using Ensemble Satellite Rainfall Estimates in a Sparsely Gauged River Basin: The Need for Whole-Ensemble Calibration

Skinner, Christopher, J.[1]

Bellerby, Timothy, J.[1 a]

Greatrex, Helen.[2b]

Grimes, David, I. F.[2c]

[1]Department of Geography, Environment and Earth Sciences, University of Hull

Cottingham Road,

Hull,

HU6 7RX,

United Kingdom

[2]Department of Meteorology, University of Reading

Earley Gate,

PO Box 243,

Reading,

RG6 6BB,

United Kingdom

Corresponding author:

Dr Christopher Skinner

Department of GEES, University of Hull,

Cottingham Road,

Hull,

HU6 7RX,

United Kingdom

c.skinner@hull.ac.uk

+44 7852 574957

[a] Present details for Dr Timothy Bellerby

t.j.bellerby@hull.ac.uk

Department of GEES, University of Hull,

Cottingham Road,

Hull,

HU6 7RX,

United Kingdom

[b] Present details for Dr Helen Greatrex

greatrex@iri.columbia.edu

61 Route 9W

Monell Building

Palisades, NY 10964-1000

USA

[c] Sadly, Dr David Grimes passed away prior to the completion of this work.

Abstract

The potential for satellite rainfall estimates to drive hydrological models has been long understood, but at the high spatial and temporal resolutions often required by these models the uncertainties in satellite rainfall inputs are both significant in magnitude and spatiotemporally autocorrelated. Conditional stochastic modelling of ensemble observed fields provides one possible approach to representing this uncertainty in a form suitable for hydrological modelling. Previous studies have concentrated on the uncertainty within the satellite rainfall estimates themselves, sometimes applying ensemble inputs to a pre-calibrated hydrological model. This approach does not account for the interaction between input uncertainty and model uncertainty and in particular the impact of input uncertainty on model calibration. Moreover, it may not be appropriate to use deterministic inputs to calibrate a model that is intended to be driven by using an ensemble. A novel whole-ensemble calibration approach has been developed to overcome some of these issues.

This study used ensemble rainfall inputs produced by a conditional satellite-driven stochastic rainfall generator (TAMSIM) to drive a version of the Pitman rainfall-runoff model, calibrated using the whole-ensemble approach. Simulated ensemble discharge outputs were assessed using metrics adapted from ensemble forecast verification, showing that the ensemble outputs produced using the whole-ensemble calibrated Pitman model outperformed equivalent ensemble outputs created using a Pitman model calibrated against either the ensemble mean or a theoretical infinite-ensemble expected value.

Overall, for the verification period the whole-ensemble calibration provided a mean RMSE of 61.7 % of the mean wet season discharge, compared to 83.6 % using a

calibration based on the daily mean of the ensemble estimates. Using a Brier's Skill Score to assess the performance of the ensemble against a climatic estimate, the whole-ensemble calibration provided a positive score for the main range of discharge events. The equivalent score for calibration against the ensemble mean was negative, indicating it showed no skill versus the climatic estimate.

Keywords

Precipitation, Satellite, Catchment, Hydrological, Modelling, Uncertainty

1. Introduction

Satellite rainfall estimates (SRFE) provide a potentially attractive data source for modelling surface hydrology, particularly for large, poorly instrumented catchments. However, the application of satellite techniques in this context is still relatively undeveloped and their potential is yet to be fully explored (Gebremichael and Hossain, 2010). SRFE, particularly when they are produced at the high spatio-temporal resolutions required for many hydrological applications, contain significant uncertainties and a full consideration of these uncertainties, together with an assessment of their propagation to downstream models (and interactions with modelling uncertainties) is essential. Ensemble representations of rainfall uncertainty provide a potentially useful tool in this context. Observational ensembles produce a family of precipitation fields, each consistent with the available input data but containing a stochastic element commensurate with the underlying uncertainty (Bellerby and Sun, 2005; Clark and Slater, 2006; Teo and Grimes, 2007; Wit et al., 2009; Aghakouchak et al., 2010; McMillan et al., 2011; Bellerby, 2013; Greatrex et al., 2014). Each ensemble member can be used in turn as a driver for a downstream model to produce an ensemble set of modelled outputs (Hossain et al., 2004; Nijssen and Lettenmaier, 2004; Hossain and Anagnostou, 2006; Nikolopoulos et al., 2010).

Downstream applications, such as hydrological models, contain their own uncertainties and these will have differing sensitivities to uncertainties in the driving input data. The resulting interactions are potentially complex and non-linear. For hydrological modelling purposes there has been a tendency to separate rainfall uncertainty from the total hydrological method (including model structure selection and calibration). However, there is evidence that input uncertainty and issues such

as model calibration cannot be easily decoupled, particularly for uncertain precipitation data sources such as SRFE. Artan et al. (2007) demonstrated that SRFE are only useful as drivers for a hydrological model when the model has been calibrated against the SRFE itself. Haberlandt and Radtke (2014) suggested that hydrological models require calibration using the same type of precipitation data as will be used for operational model runs. Given that precipitation datasets are characterised by their uncertainties (perfect products would always be the same), there is a clear link between input uncertainty and model calibration that warrants further investigation. Ensemble approaches to SRFE uncertainty representation add further complexity to this issue and consideration needs to be given to the most appropriate model calibration approach when using ensemble datasets. Questions of precipitation uncertainty and uncertainty propagation and their interaction with hydrological model calibration are highly pertinent to operational uses of SRFE in hydrological modelling. Contexts for the operational use of SRFE are likely to include regions covered by extremely sparse, often poorly reporting, rain-gauge networks. Such regions may also display high spatial and temporal heterogeneity in climate and ground conditions.

This study demonstrates the feasibility of quantifying the SRFE-related uncertainty in a poorly instrumented catchment using an ensemble approach based on an established conditional satellite rainfall generator and a well-known lumped hydrological model. In order to successfully link these two components, it proved necessary to develop a novel approach to ensemble hydrological model calibration, based on the simultaneous calibration over the whole ensemble.

2. Methods

2.1. The Process

This study is attempting a step towards a more fully integrated, holistic approach to accounting for uncertainties when using SRFE to drive hydrological models. The focus of the study is the impact of using ensemble approaches for characterising the input uncertainty in SRFE on a downstream hydrological modelling, highlighting the special consideration that needs to be paid to the calibration of that model. This is an aspect often overlooked in analyses of input uncertainties.

The complete ensemble modelling system is constructed as follows -

- Calibrate a conditional satellite-driven stochastic rainfall generator (TAMSIM) against available rain-gauge data.
- Use TAMSIM to produce ensemble rainfall estimates from satellite inputs
- Calibrate a Pitman hydrological model against the complete ensemble satellite input data using a whole-ensemble goodness-of-fit metric.
- Propagate each member of the precipitation ensemble through the Pitman model to yield an ensemble of hydrological model outputs.
- Validate the simulated discharge ensembles using statistical methods adopted from probabilistic forecast verification.

2.2. The Study Area

The study area selected was the Bakoye catchment and the wider Senegal Basin region. The region is semi-arid/arid and is very sparsely covered by rain-gauges, which poses a particular combination of operational challenges. Principally, there is a

lack of rainfall data collected from the ground and this can be overcome by the use of satellite rainfall products.

The coverage of rain recording ground instrumentation in many areas of sub-Saharan Africa has been historically poor, with only sparse rain-gauge networks and very few rainfall radars (Washington et al., 2006). Given this paucity of instrumentation, SRFE have been used to increase the spatial and temporal coverage available (Anagnostou et al., 2010), and these SRFE have been found to be a useful source of data for environmental modellers. Examples of the use of SRFE for downstream applications in sub-Saharan Africa include crop-yield modelling (Teo and Grimes, 2007; Wit et al., 2009; Greatrex, 2012), hydrological modelling (Hardy et al., 1989; Andersen et al., 2002; Diop and Grimes, 2003; Grimes and Diop, 2003; Verdin et al. 2005; Hughes et al., 2006; Stisen et al., 2008) and informing early warning systems (EWS) (Verdin and Klaver, 2002; Verdin et al., 2005).

This study concentrated on the Bakoye catchment which is predominantly in Mali, Western Africa. The catchment is a tributary system of the wider Senegal Basin and lies south of the Sahel climatic zone. Figure 1 shows the location of the Bakoye catchment and highlights the steep north-south rainfall gradient that prevails across the region. The catchment was selected as it has no upstream inputs, and during the study period was not influenced directly by dams, and had a reasonable coverage of rain-gauges that reported regularly compared to other catchments in the wider Basin region. The catchment makes a significant contribution to downstream flows of the River Senegal system, contributing 18 % of the discharge recorded at the Bakel discharge station (the furthest point downstream not influenced by backwash of dams closer to the mouth of the river).

The Bakoye catchment is a large, sparsely gauged, semi-arid catchment with an area of around 86,000 km^2 with a mean rain-gauge density of 1 per 7,000 km^2, and as such it provides an interesting proxy for an operational context. The region is heterogeneous in many aspects, such as displaying a large rainfall gradient (as shown in Figure 1), with mean annual rainfall varying across the catchment with 1,200-1,600 mm in the south and 400-800 mm in the north. Large variations are also seen in the topography, soil, geology and vegetation across the catchment. The south of the catchment has higher elevation (> 500 m), a mix of rocky, poorly developed soils and tropical ferruginous soils, and a savannah form of vegetation, yet the north is lower (< 300 m), has brown/reddish brown soils typical of semi-arid regions and a Sahelian wooded steppe type of vegetation (Jones and Wild, 1979).

The region is poorly served with a sparse network of rain-gauges providing the historic data to measure the rainfall. As well as spatial variability, the region also displays large intra and inter-annual variations in rainfall and discharge. There is a strong seasonality, where at the Bakoye catchment the wet season falls between June and October, and a dry season falls between November and May where there is almost no significant rainfall recorded. In total, 95 % of the rainfall between 1986 and 1996 fell in the wet season (plus a further 4 % in May). Average discharge for the period is 76 $m^3.s^{-1}$, which increases to 182 $m^3.s^{-1}$ for just the wet season, although almost no year in the record displays values close to the average with large inter annual variations. For example, the mean wet season discharges varies from a minimum of 47 $m^3.s^{-1}$ in 1987 to 425 $m^3.s^{-1}$ in 1999.

2.3. Data

The primary data source for this study consisted of 0.5 º spatial resolution,15-minute Cold Cloud Duration (CCD) data from the University of Reading Tropical Applications of Meteorology using Satellites (TAMSAT) archive. This dataset provides daily cold cloud duration totals based on thresholding infra-red imagery from the METEOSAT satellite at a range of temperature thresholds. This dataset has a long history of operational application to rainfall monitoring in Africa. Many satellite rainfall products additionally employ data from passive microwave sensors. However, these sensors must be located on low-earth-orbiting platforms resulting in complex sampling patterns and correspondingly heterogeneous and discontinuous uncertainty structures. Conditional rainfall simulation from microwave or combined microwave/infra-red satellite dataset is still under development (Bellerby, 2013).

Surface rain-gauge data were employed to calibrate and verify the stochastic rainfall generator. The calibration and verification periods were determined by the availability of the data. Rain-gauge data was available from a set of 81 rain-gauges spread across Senegal, Guinea, Mali and Mauritania, for the years 1986-1996 - 13 of the rain-gauges were located within the Bakoye catchment itself. Daily discharge data recorded from the Bakoye discharge stations were available for the years 1986-2005. Thus, the data were split into a calibration period (1986-1996), and a verification period (1997-2005).

2.4. Ensemble Satellite Rainfall Estimates

There are a number of operational SRFE that provide high resolution products at real-time or near real-time, including the Tropical Rainfall Measuring System's (TRMM) Multi-Satellite Precipitation Analysis (TMPA) (Huffmann et al., 2010), the African Rainfall Estimation (RFE 2.0) (NOAA, 2010), the Climate Prediction Centre

Morphing Method (CMORPH) (Joyce et al., 2004), the Precipitation Estimation from Remotely Sensed Information using Neural Networks (PERSIANN; Hsu et al., 1996) and TAMSAT (Milford and Dugdale, 1990). TAMSAT has a long operational history in sub-Saharan Africa (Teo and Grimes, 2007), and produces dekadal estimates of rainfall by assuming a linear relationship between the period of time a cloud top is observed below a calibrated temperature threshold (CCD) and rainfall rate (Milford and Dugdale, 1990; Dugdale et al.1991; Tarnavsky et al., 2014; Maidment et al., 2014). Several studies including Dinku et al. (2007) and Maidment et al. (2013) have shown TAMSAT equalled or outperformed more sophisticated, multi-sensor SRFE, largely attributed to its use of a local calibration derived in conjunction with National meteorological services and using historical in-country rain-gauge datasets. TAMSAT has previously been successfully used on the Senegal Basin region (Andersen et al., 2002; Diop and Grimes, 2003; Grimes and Diop, 2003; Hardy et al., 1989). Teo (2006) and Teo and Grimes (2007) modified the algorithm to operate at a daily timestep and validated it over the Gambia, and the algorithm was further extended to generate an ensemble representation of precipitation uncertainty.

Many of the applications that can use SRFE would require data to be provided at high spatial and temporal resolutions, with the estimates available at real-time, or near real-time, but SRFE products at these resolutions contain large uncertainties which are normally reduced by upscaling. For single inputs, either from satellite data or a SRFE itself, it has been demonstrated that rainfall generators may be used to produce ensemble rainfall estimates (Bellerby and Sun, 2005; Hossain and Anagnostou, 2006; Teo and Grimes, 2007). The methods of Bellerby and Sun (2005) and Teo and Grimes (2007) are similar in that they model full conditional distribution of observed rainfall with regard to the satellite inputs, whereas Hossain and

Anagnostou (2006) model the multiplicative distribution of rainfall errors (using a so-called delta approach) and employ high resolution radar data to simulate satellite data with errors rather than using actual satellite data errors - though their approach could be straightforwardly extended to achieve this. Full conditional simulation based on multivariate satellite inputs, including multi-sensor multi-satellite products remains an ongoing research topic, although some progress has been made for specific types of rainfall-retrieval algorithm (Bellerby, 2007; Bellerby, 2013).

The TAMSAT Simulation (TAMSIM) method was introduced in Teo (2006) and Teo and Grimes (2007) for ensemble representation of SRFE uncertainty, and subsequently extended in Greatrex (2012) and Greatrex et al. (2014). TAMSIM generates an ensemble of rainfall fields, conditioned using the statistical relationship between daily rainfall and observed cold cloud duration (CCD). Each member of the ensemble is generated by combining two random fields (RF) – an indicator field, based on the regional probability of rain at a specified CCD, and a rainfall rate field, based on the regional distribution of positive rainfall rates coincident with a specified CCD. The conditional probability, p_0 of rainfall with respect to cold cloud duration, D_T, is given by:

$$p_0 = \frac{1}{e^{b_0 + b_1 D_T}} \quad (1)$$

where b_0 and b_1 are obtained by regressing empirical rainfall probabilities against discrete coincident values of D_T. The equation extrapolates p_0 to longer cold cloud durations for which insufficient data exist to determine robust empirical conditional probabilities. The conditional probability $p(R|R>0,D_T)$ of rainfall rate R in a raining pixel being associated with cold cloud duration D_T is modelled using a two-parameter gamma distribution with mean μ and shape function θ:

$$p(R \mid R > 0, D_T) = \Gamma(\alpha, \beta) \quad (2)$$

where,

$$\alpha = b_2 + b_3 D_T\,,\ \beta = \frac{1}{b_4 + b_5 \ln(D_T)}$$

Here b_2, ..b_5 are obtained using a two-step process. In the first step, α and β are obtained for each discrete value of D_T for which sufficient raining samples exist. The second step fits a gamma distribution to the data using a maximum likelihood estimator. Teo (2006) and Teo and Grimes (2007) used the method of moments to fit the gamma distribution but this proved unstable for the more challenging dataset available for the Bakoye catchment. Once a separate gamma distribution has been obtained for each feasible discrete value of D_T ,b_2, .. b_5 are derived through regression to yield α and β as continuous functions of cold cloud duration. The choice of a gamma distribution to model positive daily rainfall conditioned on satellite data is compatible with previous studies (Bellerby and Sun, 2005; Teo 2006; Teo and Grimes 2007). Other models have been proposed for positive daily rainfall (e.g. Wilks, 1999) and it is possible that a more extensive examination of distribution models could improve on the results presented here. The theoretical conditional expected value of the TAMSIM distribution at each point is given by:

$$R_{EXP}(D_T) = p_0(D_T)\alpha(D_T)\beta(D_T) \quad (3)$$

The second stage of TAMSIM models the geo-statistical properties of the underlying rainfall field by deriving residual variograms for the probability of rainfall, and non-zero rainfall rate. These variograms facilitate Simple Kriging of both indicator and rain-rate fields and the sequential simulation of individual pixel rainfall occurrences

and rates drawn respectively from the rain/no rain and positive rainfall probability distributions detailed above.

Using the Gambia catchment as a case study, Teo and Grimes (2007) demonstrated that TAMSIM could successfully generate a reliable ensemble of rainfall realisations, each unique but equiprobable and compliant with the underlying rainfall statistics as observed from a set of historic rain-gauges.

For the current study, the TAMSIM algorithm was calibrated against available rain-gauge data. In order to make comparison between the point rain-gauge data and the gridded CCD data, the rain-gauge data were spatially interpolated and aggregated to the grid resolution. Cells containing a rain-gauge were compared to coincident satellite estimates. Gauge data were interpolated using a block double Kriging (DK) methodology based on Barancourt et al. (1992) as implemented by the KrigeRain R-software suite of Greatrex (2009). DK allows the interpolation of a fractional rainfall field, with discrete rain/no-rain areas. TAMSIM was calibrated against the DK rainfall field for the period 1986-1996, and this calibration was used to generate ensemble SRFE for the whole period 1986-2005. The Bakoye catchment proves a challenging environment for which to implement the TAMSIM calibration. The gauge coverage density was just 1 gauge per 7,000 km^2. For comparison, the dataset available to calibrate TAMSIM in Teo and Grimes (2007) provided a network of 1 gauge per 500 km^2.

2.5. Hydrological Modelling

The hydrological model chosen for the study was a version of the Pitman model. This is a lumped conceptual rainfall-runoff model, introduced by Pitman (1973) and widely used across semi-arid regions of Africa (Middleton et al., 1981; Grimes and

Diop, 2003; Hardy et al. 1989; Hughes, 1995; Hughes et al. 2006; Hughes, 2013). The model adopted for this study was based on the one presented in Grimes and Diop (2003), which has been modified to operate at a daily timestep. The version of the Pitman model used has two-buckets and eleven adjustable parameters, and the structure of the model can be seen in Figure 2, and details of the parameters to be calibrated can be seen in Table 1.

Automatic model calibration was performed using the Shuffled Complex Evolution method, developed at the University of Arizona (SCE-UA: Duan et al.,1993). SCE-UA is a global calibration algorithm which uses multiple parameter-set 'complexes' to explore the entire parameter space, identifying areas of local minima and narrowing down the search before refining the final parameter set (Duan et al., 1993). Providing a sufficiently large set of complexes is employed the calibrated, 'optimal', parameter set will be independent of any starting parameter values set, in contrast to local minima optimisation methods (Wang et al., 2010). The SCE-UA algorithm has demonstrated the ability to find optimal parameters from a global set (Duan et al., 1994), and has been shown to be robust in comparison to alternative genetic algorithms in Wang et al. (2010). The original algorithm has been adapted in several studies, including using multi-objective approaches (Yapo et al, 1998; Madsen, 2003), incorporating Markov Chain Monte Carlo (MCMC) methods (Vrugt et al, 2003), or principal component analysis (Chu et al, 2010).

In order to generate hydrological ensembles that reliably represented hydrologic uncertainties, it proved necessary to design a novel whole-ensemble calibration approach. This approach, designated EnsAll in the discussion below, used an extended RMSE error function incorporating all ensemble members:-

$$RMSE_{ALL} = \sqrt{\frac{1}{M.N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(Q_{i,j}^{\mathrm{mod}} - Q_j^{obs}\right)^2}$$

(4)

Where *M* is the number of ensemble members, *N* the number of time steps in the calibration period, Q_j^{obs} is the observed discharge at time step *j* and $Q_{i,j}^{\mathrm{mod}}$ is the modelled discharge for ensemble *i* at time step *j*.

The calibration of the Pitman model using SCE-UA was performed by minimising this extended Root Mean Squared Error ($RMSE_{ALL}$) between modelled and recorded discharges. To provide comparison with more conventional approaches, conventional single-input RMSE-based calibrations were performed using two input datasets: (i) the mean of the ensemble product, designated EnsMean, and (ii) the theoretical expected value for an infinite ensemble given by (3), designated EnsExp. For very large numbers of ensemble members, both EnsExp and EnsMean would be identical, but for realistic ensemble sizes they can differ significantly.

The Pitman model was calibrated using SCE-UA, using a script written for the R environment by Andrews (2012). It was found that provided a high enough number of complexes were used for the calibration the influence of the initial parameter values was negligible – a suitable number in this case was found to be 50. However, due to computational expense the number of complexes for the EnsAll calibration had to be reduced to 10, and to compensate the calibration was run 10 times to ensure suitable convergence. The minimum and maximum values allowed to be sampled by SCE-UA are shown in Table 1 - these were set wide around the parameter values provided from a manual calibration performed on the model in Grimes and Diop (2003).

The operation of the Pitman model requires a potential evapotranspiration (PET) input for each timestep. Monthly climatic derived estimates were used, obtained from the Food and Agriculture Organisation of the United Nations (FAO), available from FAO (2009). Values were taken from the nearest point available to the Bakoye discharge gauging station.

2.6. Statistical Analysis Methods

The Pitman model was calibrated automatically by minimising the RMSE between the modelled and recorded discharge values for the 1986-1996 calibration period. This measure was also used to assess the performance of the Pitman model outputs for the verification period. In order to make a more direct comparison, the RMSE was adjusted by presenting it as a percentage of the mean daily discharge for the wet season of each of the respective periods. This is because the absolute RMSE for a dry period may appear much lower than that for a wetter period, when expressed as a % of the mean discharge for that period it will be greater.

Franz and Hogue (2011) argue that although the hydrological community is more frequently adopting probabilistic methods, the methods used to evaluate the ensemble outputs have largely not moved on from methods used for deterministic evaluations - proposing that methods commonly used for evaluating the performance of ensemble forecasts could also be used to evaluate the performance of probabilistic hydrological model outputs, as the goals are similar. Here, we have adopted three of the probabilistic metrics described in Franz and Hogue (2011) to assess the performance of the ensemble discharges.

The containing ratio (CR) quantifies the number of observations that fall within the ensemble bounds. This was first used to assess the accuracy of ensemble sets

produced by Generalised Likelihood Uncertainty Estimation (GLUE) methods (Beven and Binley, 1992; Montanari, 2005) and was later formalised in Xiong and Connor (2008). The CR calculates the ratio of observations that fall within the minimum and maximum uncertainty bounds defined by ensemble estimates:

$$CR = \frac{1}{n}\sum_{j=1}^{n} I[Q_j^{obs}]$$

(5)

where, $I[]$ is a binary indicator functions, where a value of 1 shows the observation Q_j^{obs} at timestep j is within the ensemble bounds, a value of 0 is not -

$$I[Q_j^{obs}] = \begin{cases} 1, Q_{\min,j}^{\mathrm{mod}} < Q_j^{obs} < Q_{\max,j}^{\mathrm{mod}} \\ 0, otherwise \end{cases}$$

where, $Q_{\min,j}^{\mathrm{mod}}$ and $Q_{\max,j}^{\mathrm{mod}}$ are the minimum and maximum ensemble discharge bounds respectively. The minimum and maximum bounds can vary, for example in Beven and Binley (1992) they were set at the 5th and 95th percentiles, and in Montanari (2005) a 95 % bound was used - this study utilises the full range of the ensemble estimates.

To verify the reliability of an ensemble forecast, it is possible to calculate the ensemble's ability to predict a binary event. A methodology to do so was introduced and developed by Murphy and Winkler (1989; 1992), and it was shown in Franz and Hogue (2011) that this method is suitable for application with ensemble discharge estimates. The methodology used in this study is that presented in Toth et al. (2003). For ensembles of rainfall or discharge estimates this can be achieved by setting a threshold rainfall or discharge value, with the ensemble set said to be reliable when

on days for which a given proportion of ensemble members exceed the threshold value, the same proportion of those days show observed rainfall/discharge over the threshold value. For example, for days where 40 % of ensemble members estimate rainfall over 10 mm, close to 40 % of those same days should be associated with observed rainfall exceeding 10 mm. For each threshold used, the scores at percentage bins are plotted as a reliability curve, with a perfectly reliable 'forecast' being found on a 1:1 line. In addition, the relative probabilities of each bin are plotted as an indicator of ensemble sharpness. The performance of both the TAMSIM SRFE and the modelled discharges from the differently calibrated Pitman models will be assessed using this method.

The Brier's Skill Score (BSS) is an expansion of the original Brier's Score (BS) introduced by Brier (1950), which assesses an ensemble forecast's skill to predict a binary event. BSS compares the BS for both an ensemble forecast and that a climatological forecast. Franz and Hogue (2011) demonstrated how BS can be applied to ensemble discharge estimates, and here this has been expanded to the full BSS, using the methodology described in Toth et al (2003):

$$BSS = 1 - \frac{\langle BS \rangle}{\langle BS_{cl} \rangle} \tag{6}$$

where,

$$BS = \sum_{i-1}^{N} \left(z_i^* - z_i \right)^2$$

$$BS_{cl} = \sum_{i-1}^{N} \left(z_i^{cl} - z_i \right)^2$$

The angle brackets $\langle\ \rangle$ denote the mean of the scores, z_i^* is the probability from the ensemble predictions that the value will exceed the threshold for timestep i, z_i is a binary value based on whether the observed value exceeds the threshold (1) or not (0), and z_i^{cl} is a climatic estimate of whether the threshold will be exceeded for timestep i. The BSS is positively orientated, with a value of 1 indicating a perfect 'forecast', and values of zero or less indicating that the climatological estimate has superior skill over the ensemble estimate.

It should be noted that the BSS is dependent upon the relative skill of the climatic estimate used as a reference, and as such its main value is for comparison between probabilistic estimates and not in assessing the independent skill of a single probabilistic estimate. For example, a BSS score of 1 is only achieved when an ensemble estimate is perfect and the climatic estimate shows no skill at all.

In this instance, a climatic estimate was produced for each day of the year, based on the probability of the discharge exceeding each threshold from recorded discharge for that day based on the eleven year calibration period, 1986-1996. This represents a suitably tough and robust measure to check the skill of the ensemble discharges against. Discharge thresholds were set between 0 $m^3.s^{-1}$ and 700 $m^3.s^{-1}$, with a BSS calculated at 5 $m^3.s^{-1}$ intervals between, encompassing 97.5 % of events. An upper limit was set at 500 $m^3.s^{-1}$ – when mean climatic discharges were calculated for each day from the record, no days exceeded this discharge value and less than 10 % of events exceeded it.

3. Results

3.1. Ensemble Satellite Rainfall Estimates

TAMSIM was used to produce a 200-member ensemble of estimated rainfall fields, each member unique yet equiprobable based on the statistics drawn from the underlying rainfall field. Figure 3a shows the frequency distribution of rainfall rates from the 200 TAMSIM ensemble members for both periods, compared to that from the DK rainfields, showing that the TAMSIM ensembles are consistent with the underlying rainfall field. A similar frequency distribution is observed at individual rain-gauges, as shown in Figure 3b which shows the frequency distribution at the Guene-Gore rain-gauge. This rain-gauge was located in the middle of the catchment and therefore saw limited spatial bias, which was observed at rain-gauges in the drier north and wetter south. The distributions in Figure 3 show that the TAMSIM ensembles are able to reproduce the rainfall distributions at both the catchment level and the gauge-pixel level, with little difference between the calibration and verification periods.

Figure 4 shows the reliability plots for catchment estimates of rainfall produced from the TAMSIM ensemble members, compared to observed rainfall (a catchment estimate from the DK rainfield). TAMSIM shows a good degree of reliability at each of the thresholds, with the reliability curve falling close to the 1:1 line. The charts on the right-hand side show the relative frequency of the events, or sharpness, and it is clear that events that show a lower frequency also show greatest deviation from the 1:1 line.

The annual catchment total rainfall for the Bakoye catchment produced by TAMSIM compared to the catchment average obtained from the DK rainfall field is shown in Figure 5. There are large inter-annual variations in the total seasonal rainfall in the region and this is clearly evident in the DK rainfall field, but much of the variation has been lost in the TAMSIM rainfall.

3.2. Hydrological Modelling

The Pitman model was calibrated using the whole-ensemble approach (EnsAll) and also using two deterministic methods, EnsMean and EnsExp. Figure 6 shows an example of an envelope hydrograph for the 1988 wet season when calibrated by EnsAll and driven by the TAMSIM produced ensemble SRFE - 1988 contained the highest discharge value observed in the calibration period. It is clear from Figure 6 that the majority of the observations are contained within the minimum and maximum bounds of the ensemble discharges, and that these bounds are not symmetrical around the observations which can be a limitation of input error estimation methods that utilise a perturbation of the inputs. The periods where observations are seen to be outside of the ensemble bounds are often limited to periods that coincide with low discharge observations, and in Figure 6 these tend to be at the beginning and end of the wet season.

Figure 7 evaluates hydrological model calibration, showing the spread of the performances from each parameter set. The parameter set produced by calibrating against all of the ensemble members clearly produces the best performances. It could be assumed that calibration against the mean of the ensemble members (either EnsExp or EnsMean) would produce a suitable parameter set for use with ensemble rainfall estimates, but Figure 7 clearly shows this to not be the case. The EnsMean is statistically similar to the EnsExp estimate, which is to be expected, and produces a similar level, and spread, of performances when used with ensemble rainfall inputs.

Calibrations were performed for the period 1986-1996, when data from the rain-gauge network was available. Figure 7 also shows the level and spread of

performances for these calibrations when driven by TAMSIM ensemble members for the verification period 1997-2005. The statistics for this period are very similar to the calibration period, although there has been a slight reduction in the skill of all the calibrations in estimating the discharge. For the EnsAll calibrated Pitman model, the mean RMSE as % of wet season daily discharge was 56.2 % for 1986-1996, and 61.7 % for 1997-2005. EnsAll is still the best performing calibration.

Table 2 shows the CR scores for the bounds of the ensemble sets, for all the observed discharges, and also for discharge observations above 100 $m^3.s^{-1}$. Overall, the EnsAll produced ensemble bounds contained a greater proportion of observations than the other ensemble sets, but shows a greater loss in ratio between the calibration and verification periods. For the higher discharges, each ensemble set shows a similar level of high ratios, with EnsAll being the highest in the calibration period. The EnsAll ensembles again show the greatest drop in ratio between the calibration and verification period, and in the verification period the EnsExp ensemble bounds marginally contain the greatest ratio of observations.

The strength of the EnsAll calibration is also evident when comparing the reliabilities of the discharge ensemble outputs, at various thresholds, against those produced by the EnsExp and EnsMean calibrations (Figure 8), for the verification period. For 75th percentile and mean daily discharge values, the EnsAll calibration clearly produces a more reliable set of discharge ensembles. The reliability of each calibration is poor for the 25^{th} and 50^{th} percentile thresholds, but this is influenced by the highly skewed sharpness in the ensembles, as shown in the charts on the right hand side – for those bins with greater occurrences the reliability is greater.

Figure 9 show plots of the BSS for each ensemble discharge set produced by the three calibrations at 5 $m^3.s^{-1}$ thresholds up to 500 $m^3.s^{-1}$. Events above this threshold are not shown as they account for less than 10 % of all events and their occurrence is too rare to produce a reliable BSS. None of the ensemble discharge estimates show any skill below 100 $m^3.s^{-1}$ for either the calibration or verification period, but the EnsAll ensemble discharge estimates is able to show skill for discharges between 100 and 400 $m^3.s^{-1}$. The EnsExp and EnsMean ensemble discharge estimates only show skill compared to the climatic estimate during the calibration period, and only discharges above 250 $m^3.s^{-1}$.

Table 3 shows the mean BSS values for various sections of the spread of discharges. The EnsAll ensemble discharge estimates overall show no skill compared to the climatic estimate, but this is highly skewed by very high negative BSS values below 100 $m^3.s^{-1}$, and by removing those values from the calculation they show a slightly positive score. This is increased further by removing rare events over 500 $m^3.s^{-1}$. This is contrast with the ensemble discharges produced by the EnsExp and EnsMean calibrations, which showed no skill compared to the climatic estimate at any discharge values.

4. Discussion

This study has demonstrated the necessity of employing a whole-ensemble calibration approach when using observed rainfall ensembles to drive a hydrological model. The ensemble discharges produced by the Pitman model calibrated using the whole-ensemble EnsAll method developed for this study outperformed the results of more conventional single-input model calibrations. They were more closely matched with the recorded discharges (Figure 7), contained a greater ratio of

observations within the ensemble bounds (Table 2), were more reliable (Figure 8) and showed greater skill as compared to a climatic estimate (Figure 9). Ensemble discharges produced by the Pitman model calibrated using the single-input EnsMean and EnsExp approaches did show some good performance statistics, but these were sporadic and largely confined to the calibration period – for example, both showed positive BSS for some discharge thresholds during the calibration period, but both only produced negative BSS for the verification period, showing less skill than a climatic estimate.

Regardless of whichever calibration method was used, the ensemble discharge sets displayed poor performances at different levels of discharge – principally for low discharge levels of < 100 $m^3.s^{-1}$. This is most evident in the BSS plots of Figure 9, where both (a) and (b) show negative BSS for all ensemble sets. This is significant as these levels of discharges represent around 55 % of observed events (although the discharge levels themselves may not represent a level of interest operationally).

The significant cause of this is likely to be propagation of errors in the TAMSIM SRFE through to the hydrological model. It is a known issue with SRFE algorithms that they often show poor performance at estimating trace or zero rainfall levels, and this is also evident in TAMSIM and discussed in more detail below - this is compounded by the fact the study utilised a lumped average of rainfall for the Bakoye catchment which, due to way TAMSIM deals with uncertainty at CCD = 0, results in a very small probability of any input being absolute zero. The BSS during the verification period were all negative above 400 $m^3.s^{-1}$ and this is likely a result of the low observation occurrence of these events in the data, with discharges in excess of this value consisting of around just 10 % of all the observations. It has previously been shown that BSS is negatively biased and contains large errors for

small sample sizes of forecast-observation pairs (Weigel et al, 2007; Bradley et al., 2008), and this was more pronounced for events deemed as rare. It is likely that with a larger data set, and therefore a larger sample of estimate-observation pairs for larger discharges, the BSS for these discharge levels would show a greater level of skill.

The study also demonstrated the implementation and verification of the TAMSIM conditional satellite rainfall algorithm to a large, semi-arid, very sparsely gauged catchment, which represented numerous challenges that serve as a proxy for those likely to be faced in an operational context. The distribution of rainfall values produced by the ensemble SRFE closely resembled those observed in the rain-gauges, both across the whole region and at an individual rain-gauge level. The ensemble SRFE were also shown to be reliable at the catchment scale, compared to a rain-gauge based rainfall field.

The TAMSIM ensemble SRFE displayed some limitations that should be discussed as they will propagate to any downstream application. The TAMSIM produced ensemble set showed less reliability at low rainfall thresholds, especially where the relative occurrences (as indicated by the sharpness of the ensembles) were low. Although the TAMSIM SRFE have been shown to accurately predict the statistical spread of rain/no-rain areas at a pixel level, the uncertainty that exists at the zero CCD level makes it unlikely that zero rainfall will be estimated over a large area – this is significant to this study as the lumped Pitman model required an areal average of rainfall for the Bakoye catchment. SRFE are also known to be particularly poor at estimating trace rainfall levels, and together this has resulted in ensemble rainfall estimates that show almost no days of zero rainfall in the wet season. This was a particular issue highlighted by Teo (2006), where TAMSIM was shown to

overestimate the coverage of rainfall on days which were predominantly zero CCD, resulting higher probability of rainfall at a pixel level - this was shown to be more prominent at the start and end of the rainy season.

The second major limitation with the TAMSIM method applied here to the Senegal Basin region is the poor reproduction of inter-annual variations in the total seasonal rainfall, again another issue highlighted by Teo (2006). TAMSAT, and by extension TAMSIM, rely on calibration against the mean statistics of a set of historic rain-gauges and shows a lack of skill matching the large inter-annual variations observed by the rain-gauges. The only ways to reduce this would be to incorporate real-time or near real-time observations into the process (Teo, 2006), or applying a post-processing to the discharge outputs - such methods could include a modelling of the Eastern African Waves (EAW), or sea surface temperatures (SST) in the Atlantic Ocean, both of which have been shown to be linked to rainfall in Western Africa (Grimes and Diop, 2003; Giannini et al., 2008; Conway, 2009).

The poor performance in representing the full variation in inter-annual rainfall will make the rainfall inputs more consistent year to year, and by using this record to calibrate the Pitman model is less likely to be able to accurately reflect larger variations in rainfall. This is made worse by the known problem of automatic calibration methods that minimise an error score as the objective function, which tend to show bias towards larger discharge values – this will essentially reinforce the errors in the SRFE.

The Pitman model will, of course, produce its own uncertainties with regards to the modelled discharges, regardless of the uncertainties within the SRFE used as an input. Wagener et al. (2003) suggested that a model can be thought of as a sum of

its structure and the calibration of the variable parameters, and therefore the uncertainties are a product of these components. The uncertainty produced by model structure can be assessed by comparing different model structure (as in Butts et al., 2004), yet this study looked at just the Pitman model. There is much debate regarding the appropriate methodologies for the calibration of variable parameters in hydrological models, between optimisation and equifinality approaches (see Beven, 2006 for example). This study has not attempted to address this debate, but adopted an optimisation approach where parameters are altered to minimise a measured value of error (objective function) to produce a single, 'optimal' set of parameter values. This approach was chosen to minimise the uncertainty within the Pitman model as much as possible, but does not attempt to quantify, or account for, the uncertainty produced by the model. The study has also not applied the methodology to different model structures, and for best hydrological ensemble representation an ensemble of model structures and their combined output should be used, as in Seiller et al. (2012).

The natural next step for this research would the application of the methods to a more useful and widely used distributed hydrological model. Such a model would allow full use of the additional spatial rainfall data made available by SRFE, but would increase the complexity of both the modelling process and the nature of the uncertainties themselves. The Senegal Basin region used in this study is large, displaying a strong rainfall gradient across the area and this spatial heterogeneity resulted in a spatial bias in the TAMSIM ensembles - these biases were reduced by using a lumped rainfall product and their influence was virtually eradicated from the downstream model outputs by the automatic calibration (Skinner, 2013). In order to use a method such as TAMSIM in conjunction with a distributed downstream

application it would first be necessary to use a regional calibration method, such as that employed by Greatrex et al. (2014).

5. Conclusions

The study has shown the following –

- It is not appropriate to use deterministic estimates of rainfall to calibrate a hydrological model that is intended to be driven by ensemble rainfall.
- It is possible, and beneficial, to calibrate a hydrological model using an whole-ensemble rainfall input.
- It is feasible to use the TAMSIM algorithm of Teo and Grimes (2007) for a large, semi-arid, very sparsely gauged catchment to produce ensemble satellite rainfall estimates.
- It is feasible to use the TAMSIM ensemble satellite rainfall estimates to drive a hydrological model in an operational context.
- Poor representations of trace and zero rainfall in the satellite rainfall estimates propagate through a hydrological model, a problem which is potentially reinforced via calibration.

The significant outcome of this study is the development of a whole-ensemble method for calibrating the hydrological model. This is a novel development and shows that not only is it possible to calibrate a hydrological model using an ensemble input of rainfall, but that the resulting calibration produces ensemble discharge sets that outperform ensemble discharge sets from a hydrological model calibrated using deterministic rainfall inputs, over a range of statistical measures.

Acknowledgements

This work was completed as part of a University of Hull, UK, 80th Anniversary Studentship award. The data used in this study was provided by the TAMSAT group, University of Reading, unless otherwise stated.

References

Aghakouchak, A., Habib, E., and Bárdossy, A., 2010. A comparison of three remotely sensed rainfall ensemble generators. Atmos. Res., **98(2-4),** 387–399 doi:10.1016/j.atmosres.2010.07.016

Anagnostou, E. N., Maggioni, V., Nikolopoulos, E. I., Meskele, T., Hossain, F., and Papadopoulos, A., 2010. Benchmarking High-Resolution Global Satellite Rainfall Products to Radar and Rain-Gauge Rainfall Estimates. *IEEE Trans. on Geosci. and Remote Sens.,* **48(4),** 1667-1683 doi: 10.1109/TGRS.2009.2034736

Andersen, J., Dybkjaer, G., Jensen, K. H., Refsgaard, J. C., and Rasmussen, K., 2002. Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. *J. of Hydrol.,* **264(1-4),** 34-50 doi:10.1016/S0022-1694(02)00046-X

Andrews, F., 2012. *Shuffled Complex Evolution (SCE) Optimisation*. Available: http://hydromad.catchment.org/man/SCEoptim.html [Accessed 24-03-2013]

Artan, G., Gadain, H., Smith, J. L., Asante, K., Bandaragoda, C. J., and Verdin, J. P., 2007. Adequacy of satellite derived rainfall data for streamflow modeling. *Nat. Hazards,* **43(2)**, 167-185 doi:10.1007/s11069-007-9121-6

Barancourt, C., Creutin, J. D., and Rivoirard, J., 1992. A method for delineating and estimating rainfall fields. *Water Resourc. Res.,* **28(4),** 1133-1144 doi: 10.1029/91WR02896

Bellerby, T. J., and Sun, J., 2005. Probabilistic and ensemble representations of the uncertainty in an IR/Microwave satellite precipitation product. *J. of Hydrometeorol.,* **6(6),** 1032-1044 doi: 10.1175/JHM454.1

Bellerby, T., 2007. Satellite rainfall uncertainty estimation using an artificial neural network. *J. of Hydrometeorol.*, **8(6),** 1397-1412 doi: 10.1175/2007/JHM846.1

Bellerby, T. J., 2013. Ensemble Representation of Uncertainty in Lagrangian Satellite Rainfall Estimates. *J. of Hydrometeorol*., **14(5),** 1483-1499 doi: 10.1175/JHM-D-12-0121.1

Beven, K, 2006. A manifesto for the equifinality thesis. *J. of Hydrol.,* **320(1-2),** 18-36 doi:10.1016/j.hydrol.2005.07.007

Beven, K., and Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process.,* **6(3),** 279-298 doi: 10.1002/hyp.3360060305

Bradley, A. A., Schwartz, S. S., and Hasino, T., 2008. Sampling uncertainty and confidence intervals for the Brier Score and Brier Skill Score. *Weather and Forecast.,* **23(5),** 992-1006 doi:10.1175/2007WAF2007049.1

Brier, G. W., 1950. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.,* **78(1),** 1-3 doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO:2

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. *J. of Hydrol.,* **298(1-4),** 242-266 doi:10.1016/j.jhydrol.2004.03.042

Chu, W., Gao, X., and Sorooshian, S., 2010. Improving the shuffled complex evolution scheme for optimization of complex nonlinear hydrological systems: Application to the calibration of the Sacramento soil-moisture accounting model. *Water Resourc. Res.,* **46(9),** W09530, doi:10.1029/2010WR009224

Clark, M. P., and A. G. Slater, 2006. Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeorol*., **7(1),** 3-22 doi:10.1175/JHM474.1

Conway, G., 2009. *The Science of Climate Change in Africa: Impacts and Adaptation*. *Grantham Institute for Climate Change Discussion Paper.* Imperial College London. Available: https://workspace.imperial.ac.uk/climatechange/public/pdfs/discussion_papers/Grantham_Institue_-_The_science_of_climate_change_in_Africa.pdf [Accessed 24 March 2013]

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S. J., and Ropelewski, C. F., 2007. Validation of satellite rainfall products over East Africa's complex topography. *International J. of Remote Sens.,* **28 (7),** 1503-1526 doi:10.1080/01431160600954688

Diop, M., and Grimes, D. I. F., 2003. Satellite-based rainfall estimation for river flow forecasting in Africa. II: African Easterly Waves, convection and rainfall. *Hydrol. Sci. J.,* **48(4),** 585-599

Duan, Q. Y., Gupta, V. K., and Sorooshian, S., 1993. Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *J. of Optim. Theory and Applications,* **76(3),** 501-521 doi: 10.1007/BF00939380

Duan, Q., Sorooshian, S., and Gupta, V. J., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. of Hydrol.,* **158(3-4),** 265-284 doi: 10.1016/0022-1694(94)90057-4

Dugdale, G., Hardy, S., and Milford, J. R., 1991. V: Daily catchment rainfall estimated from METEOSAT. *Hydrol. Process.,* **5(3),** 261-270 doi: 10.1002/hyp.3360050306

FAO, cited 2014. *AQUASTAT.* Food and Agricultural Organisation of the United Nations. Available: http://www.fao.org/nr/water/aquastat/main/index.stm [accessed 24-02-2014]

Franz, K. J., and Hogue, T. S., 2011. Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community. *Hydrol. Earth Syst. Sci.,* **15,** 3367-3382 doi: 10.5194/hess-15-3367-2011

Gebremichael, M., and Hossain, F., eds. 2010. *Satellite Rainfall Applications for Surface Hydrology.* Springer.

Giannini, A., Biasutti, M., and Verstraete, M. M., 2008. A climate model-based review of drought in the Sahel: Desertification, the re-greening and climate change. *Glob. and Planet. Chang.,* **64(3-4),** 119-128 doi: 10.1016/j.gloplacha.2008.05.004

Greatrex, H., 2009. *KrigeRain v1.2: Geostatistical analysis of precipitation data in R*. University of Reading, UK

Greatrex, H., 2012. *The application of seasonal rainfall forecasts and satellite rainfall estimates to seasonal crop yield forecasting for Africa.* Ph.D. diss., University of Reading

Greatrex, H., Grimes, D., and Wheeler, T., 2014. Advances in the stochastic modeling of satellite-derived rainfall estimates using a sparse calibration dataset. *J. of Hydrometerolo.,* **15(5),** 1810-1831 doi: 10.1175/JHM-D-13-0145.1

Grimes, D. I. F., and Diop, M., 2003. Satellite-based rainfall estimation for river flow forecasting in Africa. I: Rainfall estimates and hydrological forecasts. *Hydrol. Sci. J.,* **48(4),** 567-584 doi: 10.1623/hysj.48.4.567.51410

Haberlandt, U., and Radtke, I., 2014. Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows. *Hydrol. and Earth Syst. Sci.,* **18,** 353-365 doi: 10.5194/hess-18-353-2014

Hardy, S., Dugdale, G., Milford, J. R., and Sutcliffe, J. V., 1989. The use of satellite derived rainfall estimates as inputs to flow prediction in the River Senegal. *New Dir. for Surf. Water Model.,* **181,** 23-30

Hossain, F., Anagnostou, E. N., and Dinku, T., 2004. Sensitivity analyses of satellite rainfall retrieval and sampling error on flood prediction uncertainty. *IEEE Trans. in Geosci. and Remote Sens.,* **42(1),** 130-139 doi: 10.1109/TGRS.2003.818341

Hossain, F. and Anognostou, E. N., 2006. A two-dimensional satellite rainfall error model. *IEEE Trans. on Geosci. and Remote Sens.,* **44,** 1511- 1521 doi: 10.1109/TGRS.2005

Hsu, K., Gupta, H. V., Sorooshian, S., and Gao., X., 1996. An artificial neural network for rainfall estimation from satellite infrared imagery. *Applications of Remote Sensing in Hydrology. (Proceedings of the Third International Workshop), NHRI Symposium no.17,* Greenbelt, Maryland, USA.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E J., 2010. The TRMM Multi-Satellite Precipitation Analysis (TMPA), in: Gebremichael, M., and Hossain, F., (eds), *Satellite Rainfall Applications for Surface Hydrology.* Springer. pp.3-22

Hughes, D. A., 1995. Monthly rainfall-runoff models applied to arid and semiarid catchments for water resource estimation purposes. *Hydrol. Sci. J.,* **40(6),** 751-769 doi: 10.1080/02626669509491463

Hughes, D. A., Andersson, L., Wilk, J., and Savenije, H. H. G., 2006. Regional calibration of the Pitman model for the Okavango River. *J. of Hydrol.,* **331(1-2),** 30-42 doi: 10.1016/j.jhydrol.2006.04.047

Hughes, D. A., 2013. A review of 40 years of hydrological science and practise in southern Africa using the Pitman rainfall-runoff model. *J. of Hydrol.,* **501,** 111-124 doi: 10.1016/j.jhydrol.2013.07.043

Jones, M. J., and Wild, A., 1975. *Soils of the West African Savanna: Technical Communication No. 55*, *Commonwealth Bureau of Soils, Harpenden.* Farnham Royal: Commonwealth Agricultural Bureaux

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. of Hydrometeorol.,* **5(3),** 487-503 doi: 10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO-2

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. in Water Resour.,* **26(2),** 205-216 doi:10.1016/S0309-1708(02)00092-1

Maidment, R. I., Grimes, D. I. F., Allan, R. P., Greatrex, H., Rojas, O., and Leo, O., 2013. Evaluation of satellite-based and model re-analysis rainfall estimates for Uganda. *Meteorological Applications,* **20(3),** 308-317 doi: 10.1002/met.1283

Maidment, R., Grimes, D., Allan, R., Tarnavsky, E., Stringer, M., Hewison, T., Roebeling, R., and Black, E., 2014. The 30-year TAMSAT African Rainfall Climatology and Time-series (TARCAT) Dataset. *Journal of Geophysical Research : Atmospheres,* **119(18),** 10,619-10,644 doi: 10.1002/2014JD021927

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R., 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *J. of Hydrol.,* **400(1-2),** 83-94 doi: 10.1016.j.jhydrol.2011.01.026

Milford, J. R., and Dugdale, G., 1990. Estimations of rainfall using geostationary satellite data. *Applications of Remote Sensing in Agriculture, Butterworth, London. Proceedings of the 48th Easter School in Agricultural Science.* University of Nottingham. April 1989.

Middleton, B. J., Lorentz, S. A., Pitman, W. V., and Midgley, D. C., 1981. *Surface Water Resources of South Africa, Volume V, Drainage Regions MNPQRST, The Eastern Cape, Report no. 12/81*. Hydrological Research Unit, University of Witwatersrand, Johannesburg, South Africa. Available: http://iahs.info/hsj/280/hysj_28_01_0194.pdf [Accessed 24 March 2013]

Montanari, A., 2005. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water Resourc. Res.,* **41,** W08406, doi:10.1029/2004WR003826

Murphy, A. H., and Winkler, R. L., 1987. A general framework for forecast verification. *Monthly Weather Review,* **115(7),** 1330-1338 doi: 10.1175/1520-0493(1987)115<1330AGFFFV>2.0.CO:2

Murphy, A. H., and Winkler, R. L., 1992. Diagnostic verification of probability forecasts. *International J. of Forecasting,* **7(4),** 435-455 doi: 10.1016/0169-2070(92)90028-8

Nijssen, B., and Lettenmaier, D. P., 2004. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *J. of Geophys. Res.: Atmos.,* **109 (D2),** doi: 10.1029/2003JD003497

Nikolopoulos, E. I., Anagnostou, E. N., Hossain, F., Gebremichael, M., and Borga, M., 2010. Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model. *J. of Hydrometeorol.,* **11(2)**, 520–532 doi: 10.1175/2009JHM1169.1

NOAA, 2010.*African Rainfall Estimation Algorithm Version 2.0*. The National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Centre. Available: http://www.cpc.ncep.noaa.gov/products/fews/RFE2.0_tech.pdf [Accessed 24-03-2013]

Pitman, W. V., 1973. *A Mathematical Model for Generating Monthly River Flows from Meteorological Data in South Africa.* A Ph.D Thesis, Hydrological Research Unit, University of Witwatersrand, South Africa

Seiller, G., Anctil, F., and Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrol. and Earth Syst. Sci.,* **16(4),** 1171- 1189 doi: 10.5194/hess-1116-1171-2012

Skinner, C. J., 2013. *Ensemble-characterisation of Satellite Rainfall Uncertainty and its Impacts on the Hydrological Modelling of a Sparsely Gauged Basin in Western Africa.* A Ph.D Thesis, Department of Geography, Environment and Earth Sciences, University of Hull, UK.

Stisen, S., Jensen, K. H., Sandholt, I., and Grimes, D. I. F., 2008. A remote sensing driven distributed hydrological model of the Senegal River basin. *J. of Hydrol.,* **354 (1-4),** 131-148 doi:10.1016/j.jhydrol.2008.03.006

Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R. P., Stringer, M., Chadwick, R., and Kayitakire, F., 2014. Extension of the TAMSAT Satellite-based Rainfall Monitoring over Africa and from 1983 to present. *J. of Applied Meteorol. and Climatology*, **53(12),** 2805-2822 doi: 10.1175/JAMC-D-14-0016.1

Teo, C-K, 2006. *Application of satellite-based rainfall estimates to crop yield forecasting in Africa.* A Ph.D. Thesis, Department of Meteorology, University of Reading, UK

Teo, C-K., and Grimes, D. I. F., 2007. Stochastic modelling of rainfall from satellite data. *J. of Hydrol.,* **346(1-2),** 33-50 doi:10.1016/j.jhydrol.2007.08.014

Toth, Z., Talagrund, O., Candille, G., and Zhu, Y, 2003. Probability and Ensemble Forecasts, in: Jolliffe, I. T., and Stephenson, D. B., (eds), *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* Chichester, UK: John Wiley & Sons. pp.137-164

Verdin, J., and Klaver, R., 2002. Grid-cell-based crop water accounting for the famine early warning system. *Hydrol. Process.,* **16(8),** 1617-1630 doi:10.1002/hyp.1025

Verdin, J., Funk, C., Senay, G., and Choularton, R., 2005. Climate science and famine early warning. *Philos. Trans. of the Royal Soc. B,* **360(1463),** 2155-2168 doi: 10.1098/rstb.2005.1754

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

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrol. Process.,* **17(2),** 455-476 doi:10.1002/hyp.1135

Wang, Y-C., Yu, P-S., and Yang, T-C., 2010. Comparison of genetic algorithms and shuffled complex evolution approach for calibrating distributed rainfall-runoff model. *Hydrol. Process.,* **24(8),** 1015-1026 doi:10.1002/hyp.7543

Washington, R., Kay, G., Harrison, M., Conway, D., Black, E., Challinor, A., Grimes, D., Jones, R., Morse, A., and Todd, M., 2006. African Climate Change: Taking the Shorter Route. In: *Bull. of the Am. Meteorolog. Soc.,* **87(10),** 1355-1366 doi: 10.1175/BAMS-87-10-1355

Weigel, A. P., Liniger, M. A., and Appenzeller, C., 2007. The discrete Brier and ranked probability skill scores. *Mon. Weather Rev.,* **135(1),** 118 -124 doi: 10.1175/MWR3280.1

Wilks, D. S., 1999. Multisite downscaling of daily precipitation with a stochastic weather generator. *Clim Res,* **11,** 125-136

Wit, A. J. W. de., Boogaard, H. L., Diepen, C. A., 2009. Regional crop yield forecasting: a probabilistic approach, EGU General Assembly 2009, Vienna, Austria, 19-24 April 2009. *Geophys. Res. Abstr.,* **11 (2009),** ISSN 1029-7006

Xiong, L., and Connor, K. M., 2008. An empirical method to improve the prediction limits of the GLUE methodology in rainfall-runoff modeling. *J. of Hydrol.,* **349(1-2),** 115-124 doi:10.1016/j.jhydrol.2007.10.029

Yapo, P. O., Gupta, H. V., and Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *J. of Hydrol.,* **204(1-4),** 83-97 doi: 10.1016/S0022-1694(97)00107-8

Figures

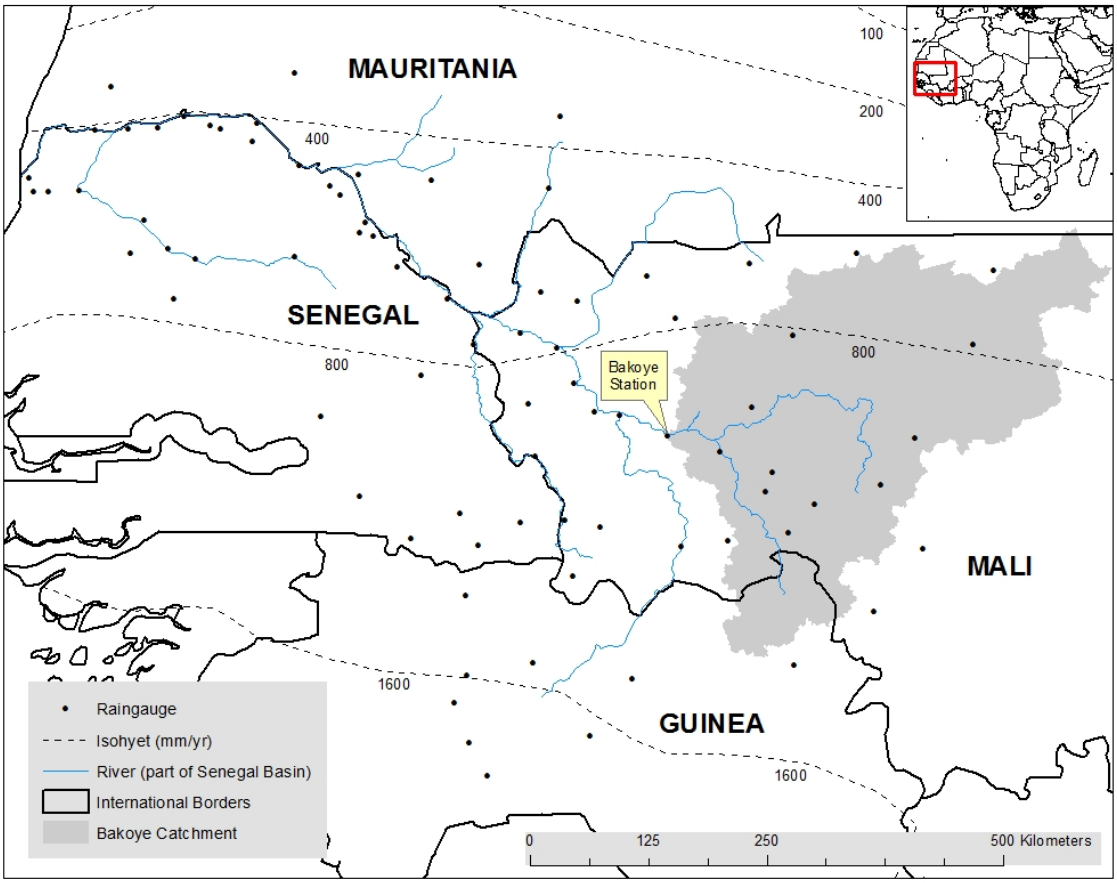

Figure 1 - Map of the wider Senegal Basin, showing the rain-gauge network, the Bakoye catchment and the Bakoye discharge station. Mean annual rainfall isohyets are also shown (after Jones and Wild, 1975).

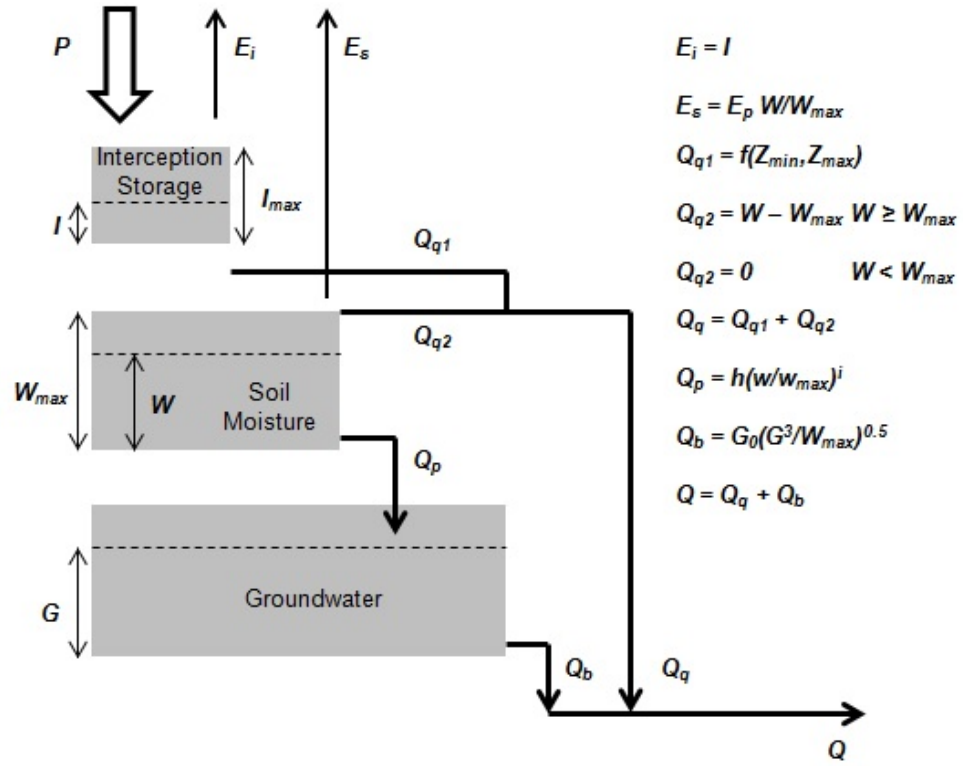

Figure 2 - The Pitman Model (after Grimes and Diop, 2003). Where *P* is rainfall input, E_i is evaporation from the interception storage (*I*), E_s is evaporation from the soil moisture (*W*), E_p is the potential evaporation, I_{max} is the maximum of the interception storage, W_{max} is the maximum of the soil moisture, *G* is groundwater, Q_{q1} is quick flow resulting from rainfall inputs in excess of the maximum infiltration rate (*Z*), Q_{q2} is the quick flow from saturated soil, Q_p is percolation and Q_b is baseflow, and *Q* is the lagged sum of Q_q and Q_b. In the model I_{max}, W_{max}, Z_{min}, Z_{max}, *i*, *h*, and G_0 are calibrated parameters, along with appropriate lags.

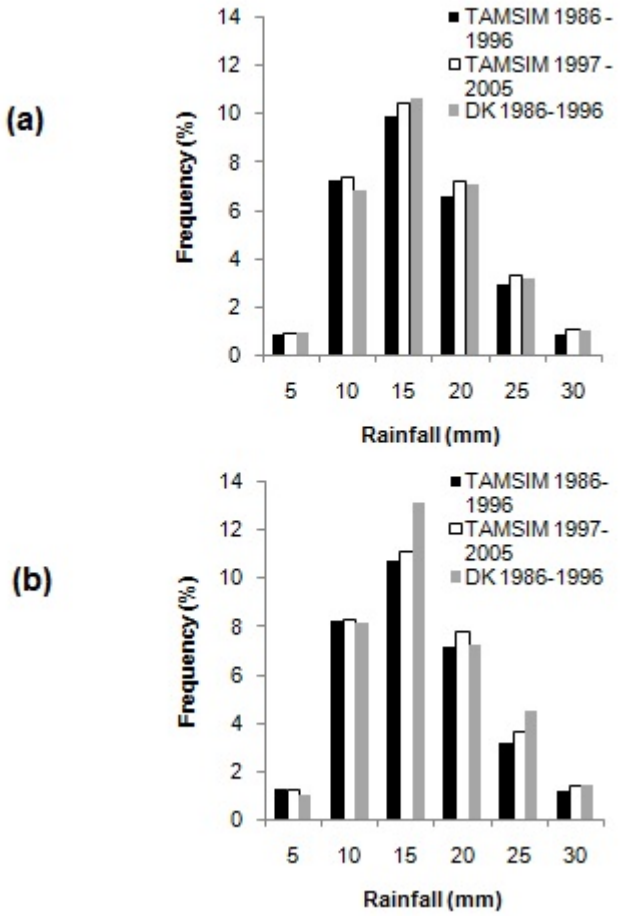

Figure 3 - Comparison of frequency distributions of wet-season daily rainfall for: 1986-1996 DK interpolated estimates; all 200 TAMSIM 1986-1996 ensemble members; all 200 1997-2005 TAMSIM ensemble members. (a) All rain-gauges (b) Guene-Gore rain-gauge. Zero rainfall frequencies are not shown (a. TAMSIM 1986-1996 = 70.8%; TAMSIM 1997-2005 = 69.0%; DK 1986-1996 = 69.4% b. TAMSIM 1986-1996 = 67.7%; TAMSIM 1997-2005 = 65.8%; and DK 1986-1996 = 63.4%).

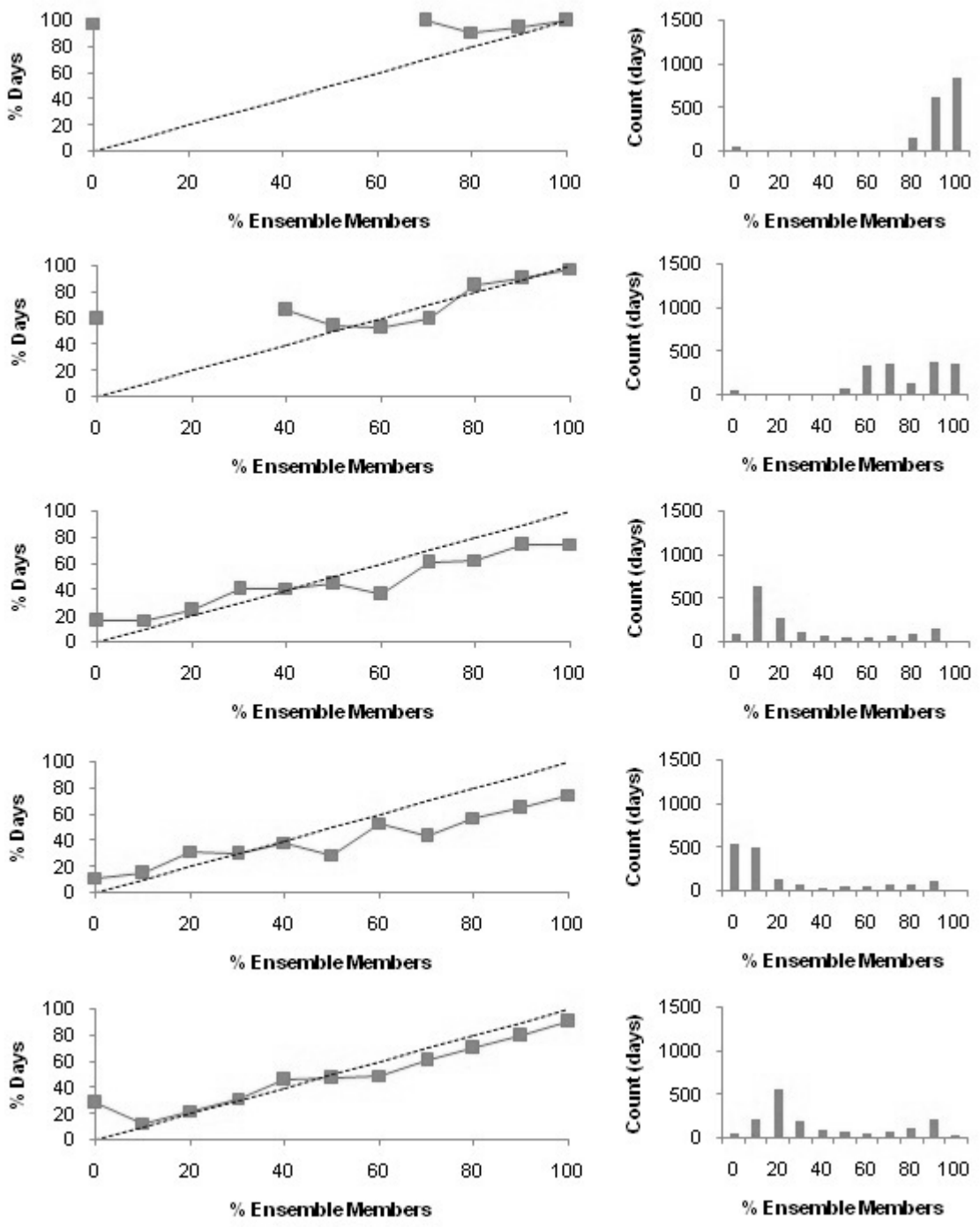

Figure 4 - Forecast reliability plots for TAMSIM rainfall ensembles for the period 1986-1996. From top to bottom the plots show thresholds of zero rainfall, 25th, 50th and 75th percentiles and mean DK rainfall (1986-1996). The blue line shows the TAMSIM rainfall reliability, and the dashed line the ideal 1:1 relationship. The bar charts on the right show the sharpness for the plots.

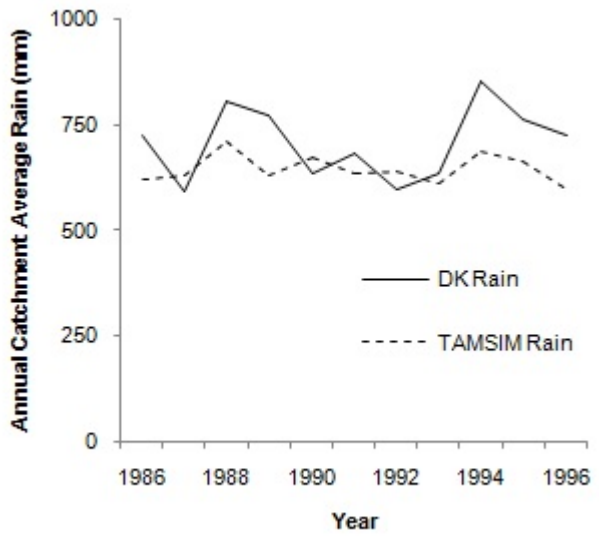

Figure 5 - Mean annual rainfall, as an areal average for the Bakoye catchment, for the years 1986 to 1996. The solid line shows the mean from the DK rainfields and the dashed line from the TAMSIM ensemble members.

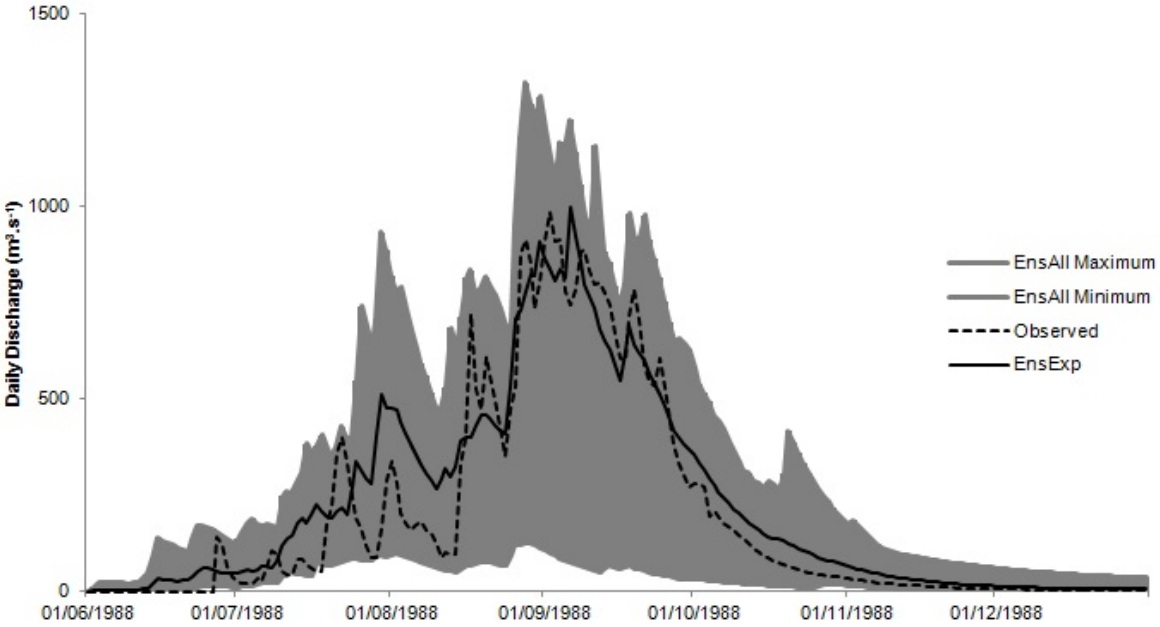

Figure 6 - An example of an ensemble envelope hydrograph produced using the EnsAll calibration of the Pitman model. The hydrograph shows the 1988 wet season, with the shaded area showing the minimum and maximum bounds of the ensemble set, the dashed line showing the observed discharges, and the solid line showing the modelled discharge produced using EnsExp as both the input and calibration method.

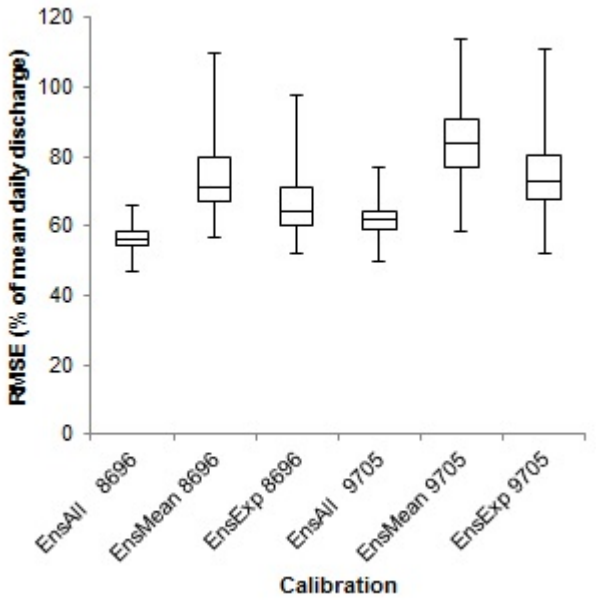

Figure 7 - Box and Whisker plots showing the performance of the TAMSIM ensemble driven Pitman model using different calibrations. The plots show the performance of the Pitman model driven by individual ensemble members for both the calibration period, 1986-1996, and the verification period, 1997-2005.

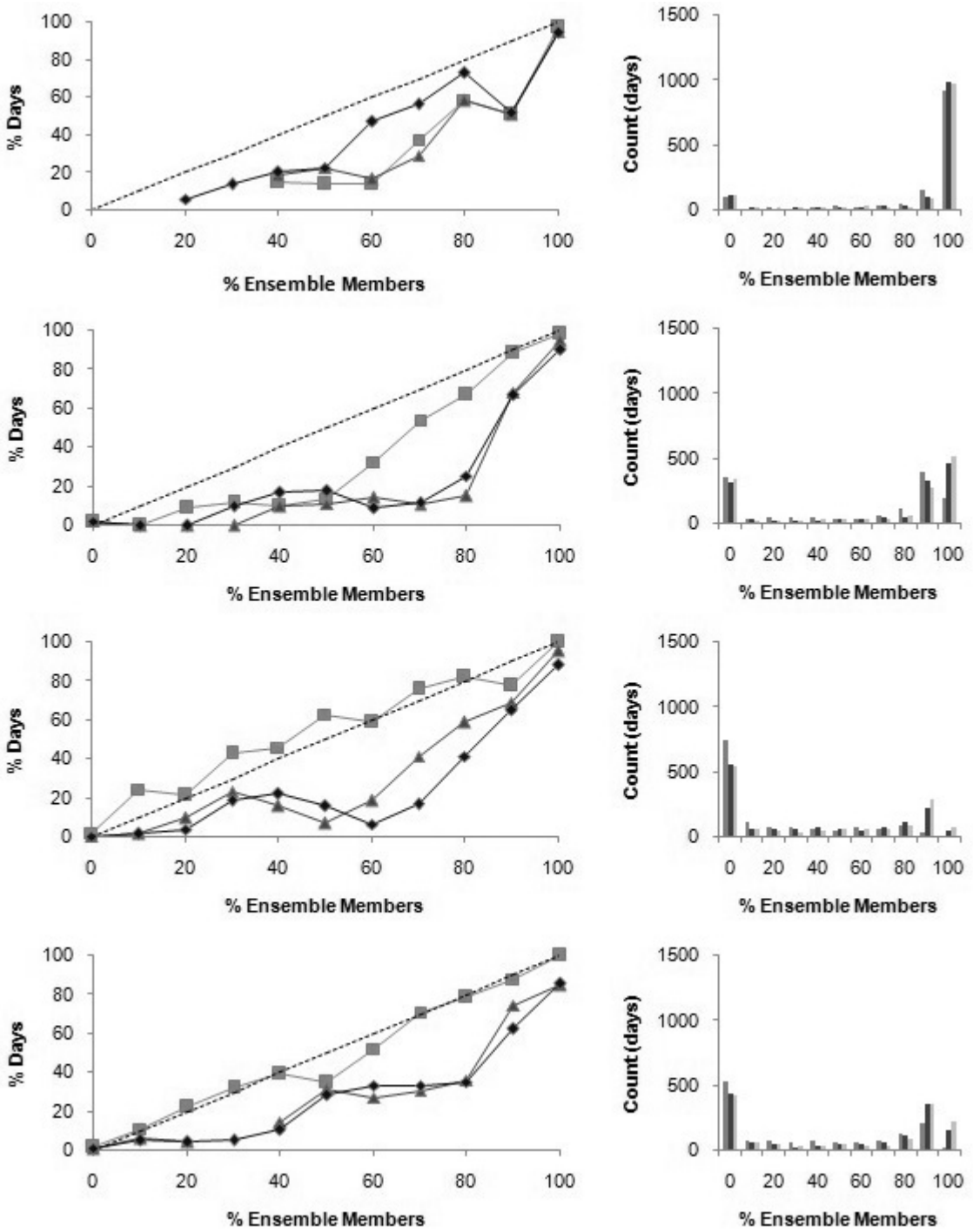

Figure 8 - Reliability plots for Pitman modelled discharges for the verification period 1997-2005. From top to bottom the plots show thresholds of the 25th, 50th and 75th percentiles and mean recorded discharge (1997-2005). The square markers show the discharges from the EnsAll calibrated model, triangle markers from the EnsExp calibrated model and diamond markers from the EnsMean calibrated model. The bar charts on the right show the sharpness for the plots for EnsAll, EnsExp and EnsMean left to right.

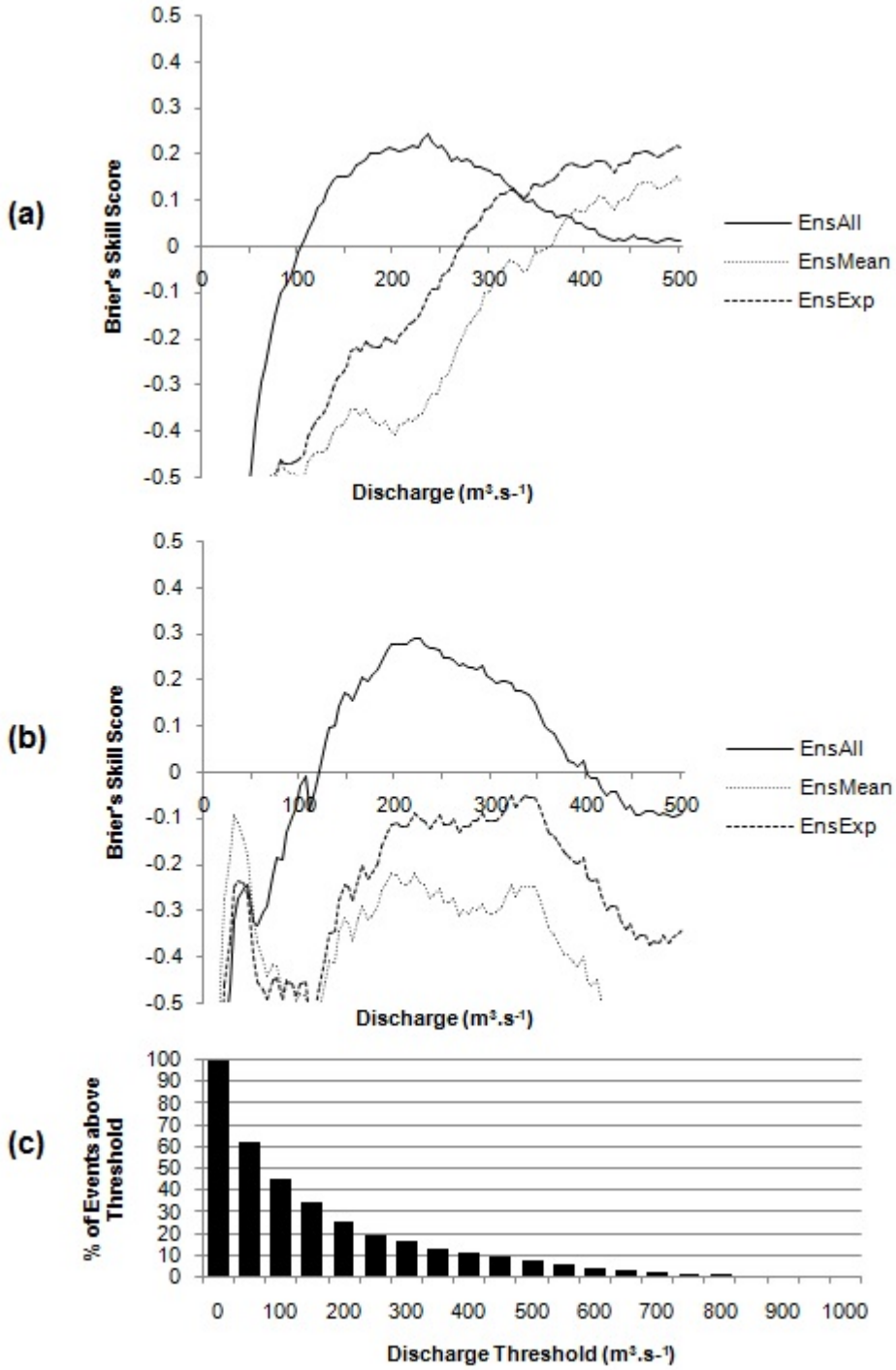

Figure 9 - Brier Skill Score at 5 $m^3.s^{-1}$ thresholds for modelled discharges from each of the calibration methods. Scores above 0 show skill in comparison with a climatic forecast derived from 1986-1996 observed data (a) calibration period 1986-1996 (b) verification period 1997-2005. (c) Relative occurrence of events for the period 1986-1996 as recorded by the Bakoye discharge station for the wet season only.

Tables

Parameter	Description	Minimum	Maximum
Z_{min}	Minimum infiltration rate	0	10
Z_{max}	Maximum infiltration rate	0	100
W_{max}	Storage threshold for Soil Moisture	0.1	1000
W_{min}	Threshold below which no percolation occurs	0	1000
I_{max}	Storage threshold for Interception Storage	0	10
h	Empirical constant used to calculate percolation rate	0	10
i	Empirical constant used to calculate percolation rate	0	20
GL	Recession time constant for baseflow (G_0 = 1/GL)	1	14
TL	Constant used for calculation of the quick flow	0	14
Q_q *Lag*	Lag for quick flow	0	5
Q_b *Lag*	Lag for baseflow	0	5

Table 1 - Table showing the parameters to be calibrated within the Pitman model and the minimum and maximum values used in SCE-UA.

	Containing Ratio - All Discharges		
	EnsAll	**EnsMean**	**EnsExp**
Calibration (1986-1996)	**0.81**	**0.61**	**0.62**
Verification (1997-2005)	**0.70**	**0.62**	**0.64**
	Containing Ratio - Discharge > 100 $m^3.s^{-1}$		
Calibration (1986-1996)	**0.96**	**0.90**	**0.93**
Verification (1997-2005)	**0.89**	**0.84**	**0.90**

Table 2 - Table showing the containing ratio for the ensemble bounds from the ensemble discharge sets produced by the different calibration methods. The table shows the ratios for the full spread of discharges and for only observed discharges above 100 $m^3.s^{-1}$.

	EnsAll	EnsMean	EnsExp
All discharges	-0.05	-0.46	-0.31
> 100 $m^3.s^{-1}$	0.04	-0.44	-0.25
100 - 500 $m^3.s^{-1}$	0.11	-0.38	-0.21

Table 3 - Table showing the mean BSS scores for various sections of the spread of discharges from each calibration method for the verification period 1997-2005. The scores are in comparison to a climatic estimate derived from recorded discharges between 1986-1996.