Ragab, Ragab; Kaelin, Alexandra; Afzal, Muhammad; Panagea, Ioanna. 2020. **Application of generalized likelihood uncertainty estimation (GLUE) at different temporal scales to reduce the uncertainty level in modelled river flows**. *Hydrological Sciences Journal*, 65 (11). 1856-1871.

Publisher: Taylor & Francis & IAHS

Journal: *Hydrological Sciences Journal*

DOI: 10.1080/02626667.2020.1764961

Application of Generalized Likelihood Uncertainty Estimation (GLUE) at different temporal scales to reduce the uncertainty level in modelled river flows

Ragab Ragab[a], Alexandra Kaelin[a], Muhammad Afzal[b], and Ioanna Panagea[c]
[a]*UK Centre for Ecology & Hydrology, Wallingford, Oxfordshire, UK;*
[b]*School of Earth and Ocean Sciences Cardiff University, Cardiff, UK;*
[c]*Division of Soil and Water Management, Katholieke Universiteit Leuven, Leuven, Belgium*

Corresponding author, rag@ceh.ac.uk

Abstract

In this study, the distributed catchment-scale model, DiCaSM, was applied on five catchments across the UK. Given its importance, river flow was selected to study the uncertainty in streamflow prediction using the Generalized Likelihood Uncertainty Estimation (GLUE) methodology at different timescales (daily, monthly, seasonal and annual). The uncertainty analysis showed that the observed river flows were within the predicted bounds/envelope of 5% and 95% percentiles. These predicted river flow bounds contained most of the observed river flows, as expressed by the high containment ratio, CR. In addition to CR, other uncertainty indices – bandwidth *B*, relative bandwidth RB, degrees of asymmetry *S* and *T*, deviation amplitude *D*, relative deviation amplitude RD and the *R* factor – also indicated that the predicted river flows have acceptable uncertainty levels. The results show lower uncertainty in predicted river flows when increasing the timescale from daily to monthly to seasonal, with the lowest uncertainty associated with annual flows.

Keywords distributed catchment-scale model, DiCaSM; GLUE, model uncertainty, River Eden, River Don, River Ebbw, River Frome, River Pang, UK

1 Introduction

The results of hydrological models are judged by their reliability, accuracy and level of uncertainty. There are a number of factors that affect the model results. They include the model structure, the accuracy in describing the hydrological processes, the parameter values and the inherent errors in input and observation data. These factors, if not closely representing the natural system, could lead to imprecision and uncertainty in model results (Loucks and Van Beek, 2017). One of the most applied methods for assessing the uncertainty is the Generalized Likelihood Uncertainty Estimation (GLUE), proposed by Beven and Binley (1992). This methodology has been used in numerous hydrological studies (e.g. Xue et al. 2018, Teweldebrhan et al. 2018, Kan et al. 2019, Xie et al. 2019, Tegegne et al. 2019).

The GLUE methodology assumes that there is no optimal parameter set that could describe the catchment hydrology. As stated by Beven and Binley (1992), "*there may be many sets of parameter values that are equally likely as simulator of the hydrological system*." The GLUE methodology states that the performance of simulation is not decided by one specific parameter, but by the combination of parameters in a parameter set. The methodology recognises the possible equifinality and the multiple behavioural (Beven and Binley, 1992) of the different parameter sets and assesses the likelihood of a set being acceptable when compared with the observed datasets.

Although several studies have applied the GLUE methodology to assess model performance, little work has been carried to study model performance over different timescales (daily, monthly, seasonal and annual), and over different catchments of different sizes, land-use characteristics, geographical location and soil physical characteristics.

This work is part of a large project, DRY[1], aimed at investigating the impact of climate and land use on water resources, with special emphasis on the drought aspect. The aim of this paper is to assess the level of uncertainty in predicted river flows at different timescales, since its quantification at different timescales will be valuable. The project involves national and local stakeholders, each of which has different interest. For example, the National Farmers Union (UK) has an interest in the seasonal water requirement for the farming community, which is more associated with seasonal predictions, while local councils have an interest in

[1] http://dryproject.co.uk/

water availability for the new urban developments, which is more associated with annual prediction of water resources availability.

2 DiCaSM, the data and the studied catchments

2.1 The DiCaSM model

This study applied the Distributed Catchment Scale Model, DiCaSM (Ragab and Bromley, 2010, Ragab et al., 2010). The model is physically based and considers commonly known hydrological processes, such as rainfall interception, infiltration, evapotranspiration, surface runoff to streams, recharge to groundwater, water uptake by plants, soil moisture dynamics and streamflow. The model has been developed to estimate the catchment water balance components and to account for the impact of changes in climate and land use on the catchment water resources, including streamflow and recharge to the groundwater. The model adopts a distributed approach with a variable spatial scale (default is a 1 km × 1 km grid square). It requires daily input data of rainfall, temperature, wind speed, vapour pressure and radiation. The model runs on a daily timestep; however, if hourly rainfall data is available, the model can run on an hourly timestep. The model also addresses the heterogeneity of input parameters of soil and land cover within the grid square using three different algorithms (Ragab and Bromley, 2010). The model has been successfully applied on a catchment in Brazil (Montenegro and Ragab 2010, 2012), Italy (D'Agostino et al. 2010) and Cyprus (Ragab et al. 2010).

2.2 Components of DiCaSM

The key model components of DiCaSM are rainfall interception, potential evapotranspiration, the catchment water balance, infiltration, surface runoff /overland flow and groundwater recharge. The processes include rainfall interception by grass surface (calculated according to Aston 1979), by crops (according to Von Hoyningen-Huene 1981) and by trees (according to Gash et al. 1995). Potential evapotranspiration of mixed vegetation is calculated according to Raupach (1995), whereas the surface runoff calculation is based on either excess saturation or excess filtration. The infiltration is calculated according to the equations of either Philip (1957) or Green and Ampt (Green 1911). The runoff is routed between the low points of each grid square along the prevailing slope using a digital terrain model (DTM). The model calculates the soil water balance of the root zone based on the four-layer model of Ragab et al. (1997) and calculates overland and

channel flow according to Yu and Jeng (1997). Further details about the model are given in Ragab et al. (2010) and Ragab and Bromley (2010).

2.3 Input data for DiCaSM

The DiCaSM model was applied on five selected catchments located in different parts of the UK (Fig. 1), as part of a project to study the climate change impact on water resources. The data required to run the model are: climate data, elevation map data, land cover map data, soil cover map data, soil hydraulic properties and land cover properties. The climate data required are temperature, wind speed, vapour pressure and radiation, as well as the rainfall. The climatic data were obtained from the Climate, Hydrology and Ecology research Support System (CHESS) (Robinson et al., 2015, Tanguy et al., 2016). The catchment boundary and gauging station location data were collected from the UK Centre for Ecology and Hydrology (Morris et al., 1990b, Morris and Flavin, 1994) and the National River Flow Archive (NRFA[2]) provided data for the daily river flow for the studied catchments. The river flow data were collected from the UK Centre for Ecology and Hydrology: "Digital Rivers 50km GB"' Web Map Service[3] and the land cover data were obtained from the "Land Cover Map 2007 (25m raster, GB)" Web Map Service (Morton et al., 2011).

2.4 ***Key model parameters and model simulations***

The observed streamflow data were obtained from the UK Environment Agency (one of the stakeholders in the project). The river flow in the DiCaSM model depends on six parameters: the percentage of surface runoff routed to stream, an exponent function describing the peak flow, the catchment storage/time lag coefficient, the stream storage/time lag coefficient, the baseflow index and the streambed infiltration//leakage. In addition, there are other parameters that affect the calibration, such as the soil hydraulic parameters. For the model calibration, the model was run with a range of the above-mentioned model parameters using first the best periods, i.e. those available with the best quality data (no gaps), for the model simulations. The selected time period was chosen using a simple iteration algorithm for optimization, in which each of the above-selected parameters was assigned a range described by a minimum and a

[2] http://nrfa.ceh.ac.uk/ [Accessed 2014]
[3] https://data.gov.uk/dataset/3c7ea82e-83e0-45a3-9a3f-8ba653b3211b/ceh-digital-river-network-of-great-britain-web-map-service [Accessed 2014]

maximum value. Each range was divided into several steps and the number of total iterations is the product of multiplication of the steps of the six key parameters. The number of iterations for each parameter was assigned according to the parameter sensitivity, i.e. a higher number was assigned to parameters that showed more impact on the streamflow. The model calculates the Nash-Sutcliffe efficiency, NSE, for each iteration. The model optimisation process helps in finding a good set of parameters that produces a good model efficiency factor. In addition to the NSE, other indices such as lnNSE (using natural logarithmic values of streamflow) and R^2 were also used to compare the simulated and the observed data.

Generally, the model calibration was carried out over a shorter period and then the model was validated for several years to the entire available record to ensure the consistency and coherence of the parameter ranges. For distributed models such as DiCaSM (grid square area of 1 km^2 and using 52 years of data), this process helps in identifying the range of each parameter to reduce the number of iterations required by the GLUE methodology. Conceptual non-distributed models such as rainfall–runoff models have less computing power requirement and can be run with a large number of iterations within significantly a shorter time period than distributed models.

2.5 Assessing model efficiency/performance

To determine the model efficiency/goodness of fit, the modelled and observed river flow data were compared using a number of indices, including the NSE criterion (Nash and Sutcliffe, 1970). The NSE is the most widely used factor to assess the performance of hydrological models (Gupta et al., 2009). An NSE of 1 indicates a perfect match.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (2)$$

where O_i and S_i refer to the observed and simulated flow data, respectively, $\bar{O}$ is the mean of the observed data and n is the number of observations. The calibration procedure consisted of adjusting the model parameters to achieve the best model fit, with the latter assessed using the NSE values. Krause et al. (2005) indicated that extreme values in a time series can result in a low NSE coefficient because hydrological

models tend to underestimate river flow during peak flows. For this reason, they suggested calculating the NSE coefficient with natural logarithmic values of the flow, as used in Afzal et al. (2015):

$$\ln \mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(\ln O_i - \ln S_i)^2}{\sum_{i=1}^{n}(\ln O_i - \ln \bar{O})^2} \quad (2)$$

In addition, the model performance was also evaluated using the statistical indicators, namely the coefficient of determination, R^2 as follows:

$$R^2 = \left\{\frac{1}{N}\frac{\sum_{i=1}^{n}[(O_i - \bar{O}]\,(\bar{S} - \bar{O}}{\sigma O_i - \sigma S_i}\right\} \quad (3)$$

where N is the total number of observations, $\bar{y}_o$ is the average measured (observed) value, $\overline{S_i}$ is the average simulated value, σO_i is the observed data standard deviation and σS_i is the simulated data standard deviation. The values of this index can range from 1 to 0, with one indicating perfect fit.

3 Generalized Likelihood Uncertainty Estimation (GLUE) methodology

Although there are a number of ways to evaluate the uncertainty, the GLUE methodology has the advantage of using only a small number of assumptions and of being simple in its application. It is based on the estimation of the weights, or probabilities, associated with different parameter sets. The set that produces the least errors (good fit) is usually associated with the highest likelihood function, and the highest probability. In the GLUE methodology, the likelihood uncertainty level is calculated as:

$$L\left(\frac{\theta_i}{Y}\right) = \left(1 - \frac{\sigma_i^2}{\sigma_{\mathrm{obs}}^2}\right) \quad (4)$$

where L is the likelihood measure of the ith model simulation made with the parameter set θ_i related to the measured streamflow Y, and is a function of the ratio of errors variance, $\sigma^2{}_i$ for the ith model simulation (representing the variance of the error between the model prediction and the observed streamflow) and the variance of the observed streamflow, $\sigma^2{}_{\mathrm{obs}}$. A distribution function is obtained by rescaling of the likelihood measures such that the sum of all the likelihood values would equal 1. The cumulative distribution and the prediction quantiles (95% and 5% confidence levels) are used to assess the uncertainty level.

The application and results of the GLUE approach vary, based on the threshold assigned for acceptable goodness-of-fit indicator and the likelihood measure (i.e. minimum NSE value) chosen to evaluate whether the selected set of parameters is behavioural or not (Beven and Binley 1992, Beven 2006, Viola et al., 2009). Different likelihood measures could be used, such as the NSE or the sum of squared errors (Beven and Binley 1992, Freer et al. 1996). Users could define the threshold of efficiency criteria according to their model preference and for each individual catchment. The GLUE methodology rejects non-behavioural parameter sets when the likelihood measure selected takes lower values than the designated threshold. The behavioural sets are retained together with the likelihood values, which are used for the weight calculation. The cumulative likelihood weighted distribution of predictions can be used to estimate the quantiles for the predictions at any timestep.

To apply the GLUE methodology, one needs to define the threshold value of the likelihood measure that differentiates between behavioural and non-behavioural models.

3.1 Uncertainty indicators

In this study, the uncertainty analysis was carried out on calibration and validation sub-periods of the dataset. Different sets of model parameters were used to generate the modelled river flow time series and the NSE criterion was chosen as a likelihood measure indicator. Based on previous studies (Jackson et al., 2016), the NSE threshold was set to 0.5 (50%), which implies that all parameter sets with NSE below 0.5 are considered non-behavioural and not included in the GLUE analysis. The uncertainty levels are evaluated with a number of indicators: CR, *B,* RB, *S*, *T, R* factor*, D*, and RD as reported by Xiong et al. (2009) and defined in the Appendix. The CR parameter is the containment ratio, which is the percentage of observed river flows that are enveloped by the prediction bounds of the 5% and 95% confidence levels, i.e. the $Q_{5\%}$–$Q_{95\%}$ likelihood-weighted quantiles. The CR is probably the most basic requirement for the prediction bounds. A high CR for the estimated prediction bounds is always the aim.

The indices *S* and *T* are used for assessing the geometric structure/average asymmetry degree of the band formed by the lower and upper prediction bounds. An average asymmetry degree index value of $S < 0.5$

indicates that, on average, the river hydrograph lies within the prediction bounds. In the completely symmetrical case, the value of S is zero. Desirable bounds should have values of $0 < S < 0.5$ and $0 < T < 1$.

The bandwidth of the prediction bounds $Q_{5\%}$–$Q_{95\%}$, B, should be as narrow as possible, so as to capture the most important information about the modelling uncertainty. The average relative bandwidth, RB is used to facilitate the comparison of results of the prediction bounds on different catchments, it is necessary to eliminate the impact of discharge magnitude on the bandwidth of the prediction bounds.

The average deviation amplitude, D, quantifies the discrepancy between the trajectory consisting of the middle points of the prediction bounds and the observed discharge hydrograph. The average relative deviation amplitude, RD, eliminates the impact of discharge magnitude on the value of the index of average deviation amplitude.

The uncertainty parameter R factor is the average thickness of the band divided by the standard deviation of the observed data. A value of less than 1 is a desirable measure for the R factor (Singh et al., 2014). More details about the indicators are given in the Appendix.

3.2 GLUE methodology application

The DiCaSM model provides, for each parameter set combination, a single value of simulated streamflow. Performance evaluation is carried out, including rejection of some parameter sets as non-behavioural (NSE < 0.5). This is followed by calculation of the likelihoods of behavioural parameter sets and rescaling to produce a cumulative sum of 1. This is carried out by ranking in ascending order all the simulated streamflow data (only of those behavioural parameter sets) and the corresponding cumulated efficiencies (NSE). Each cumulated efficiency value, divided by the maximum value, results in a value ranging between 0 and 1. These values are referred to as the 'probability weighted in efficiency' (Viola et al., 2009, D'Agostino et al. 2010). A cumulative distribution function (cdf) of the simulated streamflow is also constructed, relating each value of the simulated flow to the corresponding value of the probability weighted in efficiency.

The model uncertainty analysis is carried out over daily, monthly, seasonal and annual timescales. This helps to assess the uncertainty level of different timescales. To calculate the uncertainty level for the simulated streamflow, the following steps are followed (see also Fig. A1 in the Appendix):

Step 1. Select a likelihood measure; the NSE was selected as it is widely used.

Step 2. Assign parameter ranges (minimum, maximum, number of steps).

Step 3. Assign a threshold value for the likelihood measure to differentiate between behavioural and non-behavioural parameter sets; a value of NSE = 0.5 (Jackson et al., 2016) was selected.

Step 4. Simulate flow with several parameter sets and record the likelihood value of each parameter set.

Step 5. Retain the behavioural parameter sets together with the likelihood values which are considered the simulation weights.

Step 6. Rescale the weights obtained in Step 6 so that the sum equals one, to produce a cumulative distribution.

Step 7. From the cumulative likelihood distribution function of streamflow prediction, derive the 5% and 95% quantiles of uncertainty in the streamflow prediction at any timestep.

The procedure followed here is in accordance with Beven and Binley (1992) and similar to the procedure followed by a number of scientists, such as Freer et al. (1996), Blasome et al. (2008), Freni et al. (2008), Xiong and O'Connor (2008), Viola et al. (2009), Xiong, et al. (2009), Jin et al. (2010), Beskow et al. (2011), Chen et al. (2013), Khoi and Thom (2015), Jackson et al. (2016) and Teweldebrhan, et al. (2018).

There are several ways of presenting the results of the uncertainty analysis, as given below.

- Statistical data about the behavioural and non-behavioural simulations: for example, the percentage of behavioural simulations. If the model is run 1000 times and achieves 100 behavioural simulations, then the percentage of behavioural simulations = 100/1000 = 10%. Table 1 shows that this ratio varies from >50% to >90% for the studied catchments.
- Different indicators (Section 3.1). In the description of the results, a combination of indicators can give valuable information about the uncertainty results. For example, a low average bandwidth

combined with high CR denotes that the uncertainty bounds are low (low average bandwidth) and that a large part of the observed values is included in these bounds (high CR). This would show that the model and the parameter sets are reliable and could be used for further analysis and decision making.

- Simulated time series of the observed, calibration and validation periods together with the envelopes of percentiles 5%–95% (Freer et al., 1996; Beven and Freer, 2001; Jackson et al., 2016). Average volumes are plotted against their rescaled likelihoods, which results in a plot of variation interval of the average volume. By projecting the probability weighted in efficiency values of 0.05 and 0.95 onto the volume curve, the upper and lower confidence bands can be identified. The mean values of the observed volume, plotted as a vertical line, should lie within the confidence region of the model. In order to compare the behavioural time series with the observed volume, a cumulative distribution function (cdf) can be plotted for all behavioural time series.

4 Results and discussion

4.1 Model river flow simulations for uncertainty analysis

To reduce the number of simulations, the sensitivity of each parameter for each catchment was tested by running the model for a short time period, e.g. 2 years. This provided more insight into the range of parameters and number of iteration steps that could be considered in order to get the best results (e.g. to get iterations with an NSE between 50% and 100%). The benefit of the sensitivity analysis prior to application of the GLUE was to reduce the number of iterations, which led to a reasonable number of parameter sets selected for the five catchments (Table 1).

For all the selected time periods, the threshold for GLUE was set at NSE = 50%. This means that behavioural simulations must have an NSE value equal to or above 50%. All simulations with NSE below 50% were discarded from the GLUE analysis. The range of parameters and the number of iterations used in the study for all studied catchments are shown in Table 2. Some parameters were sampled over a very broad

range, while others were sampled over a narrow range (as the model proved the streamflow data were less sensitive to those parameters).

4.2 Model river flow calibration and validation

The five study catchments were calibrated and validated using the observed naturalized river flows for different periods from 1961 to 2012. Detailed goodness-of-fit indices as model performance indicators for the calibration and validation periods of the five studied catchments are shown in Table 3.

An example of calibration for the Ebbw catchment is shown in Figure 2, which shows a good agreement between the observed simulated flows. The NSE for this period is 91%. Overall, the model performed well for both the rainy and dry events and responded according to the soil hydrological status, i.e., for the soil moisture deficit period, a small rainfall event did not generate a significant increase in streamflow and for a heavy rainfall event when the soil was wet, especially in winter months, the model generated streamflow. For all catchments, in the model calibration stages, the NSE was, on average, around 89% and the maximum percentage error did not exceed 1% (Table 3). The model also performed very well for the well-known 1970s drought events. Generally, the overall model performance for the whole period (1961–2012) for all catchments was extremely good, with an NSE, on average. of around 85% and a maximum error not exceeding 5%.

4.3 Model uncertainty analysis

Based on the simulated river flows of the five catchments, for the calibration and validation periods, the envelope of 5% and 95% likelihood-weighted quantiles (the envelope of all behavioural models, i.e. NSE >50%) was plotted against the observed time series. Tables 4, 5, 6 and 7 show the uncertainty indicator values for daily, monthly, seasonal and annual river flows, respectively. An example of monthly river flows of the Ebbw catchment is shown in Figure 3, which shows the envelope of the 5% and 95% likelihood-weighted quantiles for the calibration period 2000–2004 (blue) and the validation period 1961–2012 (brown), and the observed river flow (solid line). For most of the time, the observed discharge is contained within the calculated uncertainty bounds, and the predictions bracketed the observations, given that, for the

Ebbw catchment, the calibration had an NSE value of 91% and the validation NSE of 87%, as shown in Table 3.

The number of observations contained within the 5% and 95% GLUE uncertainty bounds, expressed as CR, ranged from 72% to 84%. Such high values of CR mean that the model captured the observed flow quite well, as more observed values are included in the envelope, showing that those sets of parameters used can be considered acceptable to be used for future projections, such as climate change scenarios. Similar results with good CR values were also obtained for monthly flows of the other catchments, as shown in Table 5.

The uncertainty level could differ according to the timescale. Figure 4 shows the seasonal river flow for the Don catchment as an example. The results reveal that the model performed well over different seasons, including summer. The envelope of the 5% and 95% likelihood-weighted quantiles is shown for the calibration and validation periods. For most of the time, the observed discharge is contained within the calculated uncertainty bounds, i.e. the predictions bracketed the observations to a great extent (CR = 76%), given that the calibration NSE value was 91% and the validation NSE value was 87% for the Don catchment, as shown in Table 3. Similar results of seasonal flows for other catchments were obtained, and values are shown in Table 6.

Figure 5 shows the simulated annual flows and the envelope of 5% and 95% likelihood-weighted quantiles compared against the observed annual time series for the Frome catchment as an example. In Figure 5, the envelope of the 5% and 95% likelihood-weighted quantiles are given for the calibration (2001–2012) and the validation period (1971–2000); the black line represents the observed flow. All the time the observed discharge is contained within the calculated uncertainty bounds, the predictions bracketed the observations, where CR ranged from 90% to 100%. In most of the studied catchments the annual CR was above 80, and the lowest was 73%. The higher values of CR mean that the model captures the observed flow quite well as more observed values are included in the envelope.

Figure 6 shows the 'probability weighted in efficiency' based on the methodology given by Viola et al. (2009), D'Agostino et al. (2010), Beskow et al. (2011) and Hoang et al. (2018). The solid (red) vertical line represents the average value of the measured river flow for the simulated period. It should be noted that this

value falls within the confidence region of the model for the Don and Ebbw catchments as examples (Fig. 6). Other catchments showed similar results.

To compare the behavioural time series with the observed volume, the cdf is plotted for all behavioural time series. Figure 7 shows, as an example, the cumulative distribution probability of the Don and Eden catchments for the period 1962–2012. It is shown in Figure 7 that the observed streamflow data fall within the range of the number of simulated values obtained from the iterations.

4.4 Statistic indices

The uncertainty indicators were calculated for daily (Table 4), monthly (Table 5), seasonal (Table 6) and annual (Table 7) river flows for different periods. Table 4 shows the uncertainty indicators for daily river flows. The containment ratio CR ranged from 62% to 71%, for the Eden catchment, 76% to 86% for the Ebbw catchment, 56% to 60% for the Don catchment, 48% to 50% for the Frome catchment and 35% to 38% for the Pang catchment. The CR values differ from one period to another, as the validation and calibration goodness of fit are also different for different periods. However, for the full period of record (1961–2012), the CR values were 66%, 80%, 57%, 49% and 36% for the Eden, Ebbw, Don, Frome, and Pang catchments, respectively.

The asymmetry degree expressed by S and T shows S value ranges of 0.39–0.50, 0.30–0.40, 0.53–0.61, 0.75–0.93 and 0.99–1.02 for the Eden, Ebbw, Don, Frome, and Pang catchments, respectively, with S for the total period of 0.43, 0.36, 0.6, 0.85 and 1.01, respectively. The S value was within the recommended range $0.0 < S < 0.5$ for the Eden and Ebbw catchments. The T value was in the ranges 0.86–0.98, 0.76–0.87, 1.02–1.10, 1.26–1.48 and 1.54–1.57 for the Eden, Ebbw, Don, Frome, and Pang catchments, respectively, with T for the whole period of 0.90, 0.82, 1.10, 1.38 and 1.56, respectively. Similar to S, the T value was within the recommended range $0 < T < 1$ in the Eden and Ebbw catchments. One should note here that larger values of S or T represent more asymmetrical prediction bounds around the observed flow hydrograph.

The R factor value was in the ranges 0.72–0.90, 0.55–0.67, 0.64–0.70, 0.46–0.52 and 0.51–0.6 for the Eden, Ebbw, Don, Frome, and Pang catchments, respectively, with R for the whole period of 0.86, 0.62, 0.68, 0.49,

and 0.53, respectively. A value of *R* less than 1.0 is desirable (Singh et al., 2014) and the values obtained for the five catchments largely meet this requirement.

The average bandwidth *B* of the prediction bounds $Q_{0.95}$–$Q_{0.05}$ and the relative average bandwidth RB are also shown in Table 4. The relative bandwidth values for the whole period (1961–2012) were 0.84, 0.81, 0.79, 0.93, and 0.37 for the Eden, Ebbw, Don, Frome and Pang catchments, respectively.

Table 4 shows the average deviation amplitude of the middle points of the predicted bounds from the observed flow hydrograph, *D* and the relative average RD. The relative average deviation amplitude values for the whole record were 0.35, 0.24, 0.39, 0.48 and 0.25 for the Eden, Ebbw, Don, Frome and Pang catchments, respectively.

Both RB and RD values indicate a small relative bandwidth and deviation amplitude relative to observed flow values, respectively. The *S, T, B,* RB*, D* and RD values are comparable with the results of Xiong et al. (2009), who found that higher CR values are associated with lower values of *S* and *T* and higher values of *B*, RB and *D*. They stated that it is very difficult to achieve a desirable level of the CR, *T* and *B*, e.g. a high CR associated with a narrow bandwidth, *B*, and a low average asymmetry *S* and *T* with respect to the observed flows.

The monthly, seasonal and annual flow uncertainty indicators, as presented in Tables 5, 6 and 7, respectively, show much improved values of CR, *S*, *T*, RB, RD and *R*. Generally, the annual flows show better results of low uncertainty than seasonal; seasonal results are better than monthly and monthly better than daily. This is explained in the next section.

4.5 Temporal scale impact on the uncertainty levels

4.5.1 Containment ratio (CR)

The CR values (Tables 4–7) were plotted to illustrate the difference in CR when calculated for daily, monthly, seasonal and annual river flow/volume. Figure 8 shows an example of four catchments. From Tables 4–7, in most of the catchments, the CR value was improving (getting larger} when the temporal scale increased from daily to monthly to seasonal to annual. In most cases, the highest CR was associated with

annual flows and the lowest with daily flows. The CR annual value increased to almost 100% from lower values at daily, monthly or seasonal CR for the Eden, Don, Frome and Pang catchments.

4.5.2 Asymmetry degree, S and T

Ideally the asymmetry indicators such as S and T should be in the ranges $0 < S < 0.5$ and $0 < T < 1$ to have a good symmetrical condition; i.e. low uncertainty occurs within those ranges. Figure 9 shows the asymmetry indicator S for four catchments and at four timescales. Generally, the S values are reasonable and show a better symmetry when considering annual flows ($0 < S < 0.5$), followed by seasonal, then monthly, then daily flows. Clear examples are the cases of the Eden, Don, Frome and Pang catchments. Similar results obtained for the T indicator for the four catchments and at four timescales are shown in Fig. 10. Reasonable results were obtained showing a better symmetry when considering annual flows ($0 < T < 1$), followed by seasonal, then monthly, then daily. Good examples are the cases of the Eden, Don, Frome and Pang catchments.

4.5.3 Average relative bandwidth, RB

Relative bandwidth (RB), if narrow, indicates lower uncertainty. Tables 4–7 show both the bandwidth (B) and the RB. Figure 11 shows an example of RB for the four study catchments. The RB values look reasonable and indicate a relatively narrower bandwidth when using annual flows, followed by monthly, then seasonal, then daily flows. Figure 11 shows the significant difference between daily and annual RB values.

4.5.4 Average relative deviation amplitude, RD

The relative deviation amplitude (RD), if smaller, indicates lower uncertainty. Tables 4–7 show the deviation amplitude (D) and the RD. Figure 12 shows an example of RD of the four study catchments. The RD values look reasonable and indicate a relatively small deviation when using annual flows, followed by monthly, then seasonal, then daily flows. Figure 12 shows the significant difference between daily and annual RB values.

4.5.5 The R factor

The *R* factor gives the average thickness of the band $Q_{0.95}$–$Q_{0.05}$ relative to the standard deviation of the observed data, where a value of 1 is ideal. Tables 4–7 show reasonable values for the *R* factor. An example of *R* factor for the four study catchments is shown in Figure 13; it can be seen that there are fewer variations between daily, monthly and seasonal flow, with the annual flows sometimes slightly better.

In contrast to the daily, monthly and seasonal uncertainty indicators, such as CR, annual river flows had lower uncertainty levels. For example, the annual CR for some periods was as high as 100% for some catchments. Moreover, in comparison to the monthly and seasonal flows, the uncertainty levels of annual river flows of both the Frome and Pang catchments were reduced significantly, with the CR of the Pang ranging from 77% to 100% and that for the Frome from 90% to 100%. Other uncertainty indicators such as *S* and *T* were also been improved for all five catchments. Generally, this indicates that the uncertainty level is reduced when considering annual rather than monthly or seasonal river flows.

5 Conclusion

The uncertainty indicators of the five catchments revealed the following results:

- The average containment ratio (CR) value that gives the percentage of observed river flows enveloped by confidence levels $Q_{5\%}$–$Q_{95\%}$ likelihood-weighted quantiles for daily, monthly, seasonal and annual flow values, was 60%, 64%, 66% and 84%, respectively (a higher value is desirable).
- The average relative bandwidth (RB) values for daily, monthly, seasonal and annual flows were 0.77, 0.67, 0.52 and 0.25, respectively (a narrow value is desirable).
- The average asymmetry indicator (*S*) values for daily, monthly, seasonal and annual flows were 0.59, 0.46, 0.43 and 0.30, respectively, and the average *T* values were 0.09. 0.94, 0.91 and 0.76, respectively (desirable bounds should have values of $0 < S < 0.5$ and $0 < T < 1$).
- The average relative deviation amplitude (RD) for daily, monthly, seasonal and annual flows were 0.34, 0.27, 0.20 and 0.07, respectively (a smaller value is desirable).
- The average *R* factor values for daily, monthly, seasonal and annual flows were 0.66, 0.73, 0.76 and 1.08, respectively (a value closer to 1 is desirable).

Generally, the different uncertainty indicators (CR, *S*, *T*, *B*, RB, *D*, RD and *R* factor) all gave desirable values indicating a reasonable low uncertainty level in model prediction.

The GLUE methodology showed lower uncertainty in predicted river flows when increasing the timescale from daily to monthly to seasonal river flows with the lowest uncertainty associated with annual flows. The results showed that DiCaSM provided a small level of uncertainty in the predicted river flows and subsequently, a higher confidence level in the results. The results presented in this paper, for different timescales, could be useful for various stakeholders, water resources planners and decision makers.

Funding

The authors acknowledge the Natural Environment Research Council (NERC) funding for this 4-year project "Drought Risk and You, DRY" [grant reference NE/L010292/1].

Acknowledgements

The authors are very thankful to our colleagues at UK Centre for Ecology and Hydrology (UKCEH), especially Yan Weigang, Egon Dumont, Virginie Keller, James Blake and Nikolaos Vavlas, who helped us in preparing the model input data and model application. The authors would like to acknowledge the data sources: Ordnance Survey (1:250 000 Scale Colour Raster) for background mapping; and UKCEH for catchment boundary and gauging station location data (Morris et al., 1990a, Morris and Flavin, 1994), river and waterbody data ("Digital Rivers 50km GB" Web Map Service) and land cover data ("Land Cover Map 2007 (25m raster, GB)" Web Map Service; Morton et al., 2011). Standardized precipitation index time series for IHU groups (1961–2012) (SPI_IHU_groups) data were licensed from Cranfield University ("1:250 000 Soilscapes for England and Wales" Web Map Service); and hydrogeology data from the British Geological Survey (DiGMapGB 1:625 000-scale digital hydrogeological data[4]).

References

Afzal, M., Gagnon, A. S. and Mansell, M. G. 2015. The impact of projected changes in climate variability on the reliability of surface water supply in Scotland. *Water Science and Technology: Water Supply,* 15, 736-745.

[4] https://www.bgs.ac.uk/research/groundwater/datainfo/hydromaps/hydro_map_625.html

Aston, A. 1979. Rainfall interception by eight small trees. *Journal of hydrology,* 42, 383-396.
Beven, K. 2006. A manifesto for the equifinality thesis. *Journal of hydrology,* 320, 18-36.
Beven, K. & Binley, A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes,* 6, 279-29
Beven, K., and Freer, J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1–4), 11–29. https://doi.org/10.1016/S0022-1694 (01)00421-8
Beskow, S., Rogerio del mello, C., and Norton, L., D. 2011. Development, sensitivity and uncertainty analysis of LASH model. S*cientia Agricola*., 68: 265-274.
Blasone, R.-S., Vurgt, J. A., Masden H., Rosbjerg, D., Robinson, B. A. and Zyvoloski, G. A. 2008. Generalised likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Advance is Water Resources*, 31, 630-648.
Chen, X., Yang, T., Wang, X., XU, C.-Y.& Yu, Z. 2013. Uncertainty intercomparison of different hydrological models in simulating extreme flows. *Water Resources Management*, 27 1393-1409.
D'Agostino, D. R., Trisorio, L. G., LamaddalenA, N. & Ragab, R. 2010. Assessing the results of scenarios of climate and land use changes on the hydrology of an Italian catchment: modelling study. *Hydrological Processes,* 24, 2693-2704.
Freer, J., Beven, K. and Ambroise, B. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resources Research,* 32, 2161-2173.
Freni, G., Mannina, G.and Viviani, G. 2008. Uncertainty in urban stormwater quality modelling: The effect of acceptability threshold in the GLUE methodology. *Water Research*, 42, 2061-2072.
Gash, J. H., Lloyd, C. and Lachaud, G. 1995. Estimating sparse forest rainfall interception with an analytical model. *Journal of Hydrology,* 170, 79-86.
Green, W. H. 1911. Studies on soil physics, part I, the flow of air and water through soils. *Journal of Agricultural Sciences,* 4, 1-24.
Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology,* 377, 80-91.
Hoang, L., Mukundan, R., Moore, K. E. B., Owens, E. M., and Steenhuis, T. S. 2018. The effect of input data complexity on the uncertainty in simulated streamflow in humid, mountinous watershed. *Hydrology and Earth System. Sciences*, 22: 5947–5965
Jackson, C., Wang, L., Pachocka, M., Mackay, J. and BloomfielD, J. 2016. Reconstruction of multi□decadal groundwater level time□series using a lumped conceptual model. *Hydrological Processes,* 30, 3107-3125.
Jin, X., XU, C.-Y., Zhang, Q. and Singh, V.P. 2010. Parameter and modelling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology*, 383, 147-155.
Kan, G., HE, X., Ding, L., LI, J., Hong, Y. and Liang, K. 2019. Heterogeneous parallel computing accelerated generalized likelihood uncertainty estimation (GLUE) method for fast hydrological model uncertainty analysis purpose. *Engineering with Computers*, 1-22.
Khoi, D. N. & Thom, V. T. 2015. Parameter uncertainty analysis for simulating streamflow in a river catchment of Vietnam. *Global Ecology and Conservation*, 4, 538-548.
Krause, P., Boyle, D. P. and Bäse, F. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences,* 5, 89-97.
Loucks, D. P. & Van Beek, E. 2017. *Water resource systems planning and management: An introduction to methods, models, and applications*, Springer. Cham, Switzerland
Montenegro S, Ragab R. 2012. Impact of possible climate and land use changes in the semi-arid regions: a case study from North Eastern Brazil. *Journal of Hydrology*. 434-435, 55-68.
Montenegro A, Ragab R. 2010. Hydrological response of a Brazilian semi-arid catchment to different land use and climate change scenarios: a modelling study. *Hydrological Processes*. 24(19): 2705-2723.
Morris, D. and Flavin, R. 1994. Sub-set of the UK 50 m by 50 m hydrological digital terrain model grids. *NERC, Institute of Hydrology, Wallingford*.
Morris, D., Flavin, R. and MOORE, R. 1990a. *A digital terrain model for hydrology*.
Morris, D., Flavin, R. & Moore, R. A digital terrain model for hydrology. 1990b. 4th International Symposium on Spatial Data Handling, 23-27 July 1990b Zürich. 250-262

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R. & Simpson, I. 2011. Final Report for LCM2007-the new UK land cover map. Countryside Survey Technical Report No 11/07. Wallingford, UK: UK Centre for Ecology and Hydrology (report available at https://www.ceh.ac.uk/sites/default/files/LCM2007%20Final%20Report.pdf

Nash, J. E. and Sutcliffe, J. V. 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology,* 10, 282-290.

Philip, J. 1957. The theory of infiltration: 1. The infiltration equation and its solution. *Soil science,* 83, 345-358.

Ragab, R. & Bromley, J. 2010. IHMS—Integrated Hydrological Modelling System. Part 1. Hydrological processes and general structure. *Hydrological processes,* 24, 2663-2680.

Ragab, R., Bromley, J., Dörflinger, G. and Katsikides, S. 2010. IHMS—Integrated Hydrological Modelling System. Part 2. Application of linked unsaturated, DiCaSM and saturated zone, MODFLOW models on Kouris and Akrotiri catchments in Cyprus. *Hydrological Processes,* 24, 2681-2692.

Ragab, R., Finch, J. and Harding, R. 1997. Estimation of groundwater recharge to chalk and sandstone aquifers using simple soil models. *Journal of Hydrology,* 190, 19-41.

Raupach, M. 1995. Vegetation-atmosphere interaction and surface conductance at leaf, canopy and regional scales. *Agricultural and Forest Meteorology,* 73, 151-179.

Robinson, E., Blyth, E., Clark, D., Comyn-Platt, E., Finch, J. and Rudd, A. 2015. Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961-2015)[CHESS-PE].

Singh, A., Imtiyaz, M., Isaac, R. and Denis, D. 2014. Assessing the performance and uncertainty analysis of the SWAT and RBNN models for simulation of sediment yield in the Nagwa watershed, India. *Hydrological Sciences Journal,* 59, 351-364.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. and Keller, V. 2016. Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890–2015)[CEH-GEAR]. *NERC Environmental Information Data Centre, doi,* 10.

Tegegne, Getachew, Young-Oh KIM, Seung Beom Seo and YoungiL Kim. 2019. Hydrological modelling uncertainty analysis for different flow quantiles: a case study in two hydro-geographically different watersheds, *Hydrological Sciences Journal*, 64:4, 473-489, DOI: 10.1080/02626667.2019.1587562

Teweldebrhan, A. T., Burkhart, J. F., and Schuler, T.V. 2018. Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches. *Hydrology and Earth System. Sciences*, 22, 5021–5039 https://doi.org/10.5194/hess-22-5021-2018

Viola, F., Noto, L., Cannarozzo, M. and LA Loggia, G. 2009. Daily streamflow prediction with uncertainty in ephemeral catchments using the GLUE methodology. *Physics and Chemistry of the Earth, Parts A/B/C,* 34, 701-706.

Von Hoyningen-Huene, J. 1981. *Die Interzeption des Niederschlags in landwirtschaftlichen Pflanzenbeständen*, Arbeitsbericht Deutscher Verband für Wasserwirtschaft und Kulturbau, DVWK.An

Xie, H., Shen, Z., Chen, L., Lai, X., Qiu, J., Wei, G., Dong, J., Peng, Y. and Chen, X. 2019. Parameter Estimation and Uncertainty Analysis: A Comparison between Continuous and Event-Based Modeling of Streamflow Based on the Hydrological Simulation Program–Fortran (HSPF) Model. *Water,* 11, 171.

Xiong, L. and O'Connor, K. M. 2008. An empirical method to improve the prediction limit of the GLUE methodology in rainfall-runoff modelling. *Journal of Hydrology*, 349, 115-124.

Xiong, L., Wan, M., Wei, X. and O'Connor, K. M. 2009. Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation/Indices pour évaluer les bornes de prévision de modèles hydrologiques et mise en œuvre pour une estimation d'incertitude par vraisemblance généralisée. *Hydrological Sciences Journal,* 54, 852-871.

Xue, Lianqing, Fan Yang, Changbing Yang, Guanghui Wei, Wenqian Li and Xinlin He. 2018. Hydrological simulation and uncertainty analysis using the improved TOPMODEL in the arid Manas River basin, China. *Scientific Reports* volume 8, Article number: 452.

Yu, P.-S. and Jeng, Y.-C. 1997. A study on grid based distributed rainfall runoff models. *Water Resources Management,* 11**,** 83-99.

Figure captions

Figure 1. Overview map of the UK showing the case study catchments.

Figure 2. Ebbw River catchment – model calibration for the period 2000–2003.

Figure 3. Model output uncertainty boundaries (5th and 95th percentiles) when performing the GLUE analysis on model calibration (2000–2004) and validation (1961–2012) periods for the Ebbw catchment river flow (monthly values).

Figure 4. Model output uncertainty boundaries (5th and 95th percentiles) when performing the GLUE analysis on model calibration (2001–2012) and validation (1967–2012) periods for the Don catchment river flow (seasonal values).

Figure 5. Model output uncertainty boundaries (5th and 95th percentiles) when performing the GLUE analysis on model calibration (2001–2012) and validation (1971–2000) periods for the Frome catchment using annual observed and simulated data.

Figure 6. Uncertainty band of the DiCaSM parameters for the period 1962–2012 for the Don and Ebbw catchments. The solid (red) vertical line represents the average value of the measured river flow for the simulated period.

Figure 7. Cumulative probability plot of flows for the Don and Ebbw catchments (1962–2012).
*the time period in grey was used for the model calibration.

Figure 8. Containment ratio (CR) at different timescales for the Frome, Eden, Don and Ebbw catchments.

Figure 9. Asymmetric degree *S* for the Frome, Eden, Don and Ebbw catchments.

Figure 10. Asymmetric degree *T* for the Frome, Eden, Don and Ebbw catchments.

Figure 11. Average relative band width, RB, for the Frome, Eden, Don and Ebbw catchments.

Figure 12. Average relative deviation, RD, for the Frome, Eden, Don and Ebbw catchments.

Figure 13 *R* factor values for the Frome, Eden, Don and Ebbw catchments.

Table 1. Number of iterations and ranges of NSE for monthly percentiles.

Catchment	Total no. of iterations	No. of iterations with monthly NSE > 0.50	Range of monthly NSE	Range of monthly NSE > 0.50
Eden	60	60	0.50–0.85	0.50–0.85
Frome	490	396	-0.25–0.93	0.50–0.93
Ebbw	108	103	0.27–0.95	0.51–0.95
Pang	648	342	-0.92–0.91	0.50–0.91
Don	336	225	-0.35–0.89	0.50–0.89

Table 2. Key model parameter ranges and number of iterations.

Catchment	Model parameters									
	Base percentage of flow routed to stream		Exponent function of flow routed to stream		Catchment storage/ time lag		Stream storage/ time lag		Baseflow factor	
	Range/ Iterations		Range/ Iterations		Range/Iterations		Range/Iterations		Range/ Iterations	
Frome	9–90	10	0.02–0.04	7	45	1	15	1	0.2–0.8	7
Pang	0.2–1.8	9	0 to 0.004	3	0.1–1	8	0.015	1	$1.75\ 10^{-8}$–3.3×10^{-7}	3
Ebbw	10 - 90	9	0.02–0.4	2	2	1	20	1	2.2×10^{-9}–2.2×10^{-7}	6
Eden	40-95	4	1.0×10^{-5}	1	315	1	15–55	3	0.14–0.75	5
Don	10-90	8	0–0.3	6	143	1	18	1	0.02–0.95	7

Table 3. Model performance for the calibration and validation stages of the six catchments studied.

Catchment	Period	NSE	lnNSE	R^2	Square root of R^2	Modelled flow ($m^3 s^{-1}$)	Observed flow ($m^3 s^{-1}$)	% Error
Eden	2012†	0.90	0.95	0.89	0.94	5.03	5.04	–0.19
	1971–1980	0.79	0.89	0.79	0.89	3.60	3.54	1.69
	1971–2012	0.79	0.90	0.80	0.89	4.11	4.13	–0.48
Ebbw	2000–2003†	0.91	0.88	0.92	0.96	7.19	7.23	–0.55
	1971–1980	0.87	0.82	0.88	0.94	6.70	6.53	2.56
	1961–2012	0.87	0.82	0.88	0.93	6.98	7.21	–3.17
Don	2011–2012†	0.92	0.86	0.91	0.95	5.41	5.32	1.81
	2001–2012	0.87	0.73	0.87	0.93	4.86	4.73	2.61
	1971–1980	0.82	0.66	0.83	0.91	4.68	4.90	– 4.63
	1966–2012	0.83	0.73	0.84	0.91	5.06	5.08	–0.60
Frome	2001–2012†	0.86	0.83	0.86	0.93	1.79	1.78	0.94
	1971–1980	0.80	0.78	0.81	0.90	1.56	1.48	4.93
	1962–2012	0.82	0.80	0.83	0.91	1.71	1.74	–2.25
Pang	2001–2003†	0.92	0.89	0.94	0.97	0.79	0.81	–2.14
	2000–2012	0.90	0.83	0.90	0.95	0.66	0.64	2.96
	1971–1980	0.78	0.79	0.78	0.88	0.62	0.61	1.76
	1971–2012	0.81	0.80	0.83	0.91	0.66	0.64	3.47

†calibration period.

Table 4. Results of daily GLUE prediction bounds for all the studied catchments (See the Appendix for description of the parameters).

Catchment	Daily prediction bounds								
	CR	*B*	RB	*S*	*T*	*D*	RD	*R* factor	Period
Eden	61.70	2.29	0.74	0.50	0.98	0.88	0.31	0.72	1975–1976
	61.97	2.82	0.82	0.42	0.89	0.97	0.36	0.71	1976–1977
	62.61	2.51	0.87	0.46	0.94	0.96	0.38	0.88	1971–1980
	69.69	2.94	0.84	0.39	0.86	1.06	0.32	0.87	1981–1990
	67.78	2.91	0.87	0.40	0.87	1.05	0.35	0.90	1991–2000
	63.43	2.99	0.80	0.46	0.93	1.15	0.34	0.82	2001–2012
	65.76	2.85	0.84	0.43	0.90	1.06	0.35	0.86	1971–2012
Ebbw	75.83	4.54	0.83	0.40	0.87	1.50	0.29	0.67	1971–1980
	81.22	5.25	0.77	0.35	0.82	1.74	0.21	0.61	1981–1990
	85.60	5.34	0.82	0.30	0.76	1.64	0.21	0.57	1991–2000
	83.87	4.96	0.83	0.33	0.79	1.57	0.23	0.59	2001–2010
	79.87	4.97	0.81	0.36	0.82	1.63	0.24	0.62	1961–2012
Don	55.49	4.17	0.83	0.56	1.05	1.76	0.42	0.68	2000–2012
	57.35	3.88	0.74	0.61	1.10	1.67	0.36	0.68	1971–1980
	58.38	4.24	0.79	0.55	1.04	1.82	0.39	0.70	1981–1990
	59.51	4.16	0.78	0.53	1.02	1.72	0.38	0.64	1991–2000
	57.31	4.10	0.79	0.60	1.10	1.76	0.39	0.68	1967–2012
Frome	47.50	1.23	0.93	0.88	1.41	0.60	0.49	0.50	1971–1980
	48.60	1.37	0.96	0.86	1.40	0.63	0.49	0.52	1981–1990
	48.15	1.38	0.88	0.93	1.48	0.71	0.47	0.49	1991–2000
	48.48	1.34	0.93	0.85	1.38	0.64	0.48	0.49	1971–2012
Pang	35.40	0.22	0.37	1.02	1.57	0.16	0.25	0.51	2000–2012
	38.18	0.22	0.37	0.99	1.54	0.15	0.24	0.60	1993–1999
	36.37	0.22	0.37	1.01	1.56	0.15	0.25	0.53	1993–2012

Table 5. Results of monthly GLUE prediction bounds for all the studied catchments.

Catch-ment	Monthly prediction bounds								
	CR	*B*	RB	*S*	*T*	*D*	RD	*R* factor	Period
Eden	70.83	4.72	0.58	0.38	0.85	1.37	0.20	0.75	1975–1976
	75.00	5.23	0.62	0.33	0.81	1.48	0.24	0.70	1976–1977
	70.00	5.09	0.68	0.38	0.85	1.62	0.28	0.93	1971–1980
	71.67	5.63	0.61	0.34	0.80	1.77	0.21	0.95	1981–1990
	68.33	5.40	0.63	0.38	0.85	1.89	0.25	0.87	1991–2000
	66.67	5.57	0.59	0.42	0.90	2.13	0.25	0.87	2001–2012
	69.05	5.43	0.63	0.38	0.85	1.87	0.25	0.89	1971–2012
Ebbw	65.83	8.39	0.80	0.53	1.02	3.34	0.34	0.65	1971–1980
	75.83	10.11	0.84	0.41	0.88	3.86	0.24	0.64	1981–1990
	73.33	10.17	0.82	0.37	0.84	3.75	0.26	0.64	1991–2000
	84.17	9.32	0.84	0.34	0.81	3.11	0.28	0.66	2001–2010
	72.44	9.27	0.78	0.42	0.90	3.51	0.27	0.65	1961–2012
Don	68.59	7.76	0.70	0.39	0.86	2.83	0.29	0.79	2000–2012
	72.50	7.28	0.62	0.38	0.85	2.54	0.23	0.84	1971–1980
	68.33	7.99	0.67	0.39	0.86	2.83	0.28	0.83	1981–1990
	75.00	7.91	0.68	0.35	0.82	2.62	0.26	0.73	1991–2000
	70.47	7.70	0.66	0.38	0.85	2.76	0.27	0.82	1967–2012
Frome	46.67	2.18	0.74	0.66	1.15	1.20	0.37	0.60	1971–1980
	50.83	2.36	0.76	0.57	1.07	1.21	0.34	0.60	1981–1990
	50.83	2.36	0.68	0.68	1.19	1.40	0.33	0.53	1991–2000
	51.19	2.33	0.73	0.59	1.08	1.21	0.33	0.58	1971–2012
Pang	45.51	0.67	0.43	0.69	1.19	0.38	0.25	0.61	2000–2012
	39.29	0.65	0.44	0.69	1.19	0.37	0.25	0.72	1993–1999
	43.33	0.66	0.44	0.69	1.19	0.38	0.25	0.64	1993–2012

Table 6. Results of seasonal GLUE prediction bounds for all the studied catchments.

Catch-ment	Seasonal prediction bounds								
	CR	*B*	RB	*S*	*T*	*D*	RD	*R* factor	Period
Eden	87.50	3.64	0.41	0.29	0.73	0.67	0.10	0.61	1975–1976
	75.00	4.19	0.40	0.32	0.79	1.23	0.14	0.71	1976–1977
	70.00	3.77	0.49	0.37	0.84	1.25	0.19	0.84	1971–1980
	80.00	4.12	0.43	0.35	0.82	1.37	0.15	0.95	1981–1990
	60.00	3.95	0.45	0.44	0.92	1.68	0.19	0.82	1991–2000
	65.96	4.00	0.41	0.43	0.92	1.67	0.17	0.95	2001–2012
	68.86	3.96	0.44	0.40	0.88	1.50	0.18	0.88	1971–2012
Ebbw	57.50	6.73	0.62	0.52	1.01	2.85	0.29	0.67	1971–1980
	52.50	7.50	0.59	0.53	1.00	3.32	0.19	0.63	1981–1990
	80.00	7.52	0.57	0.38	0.86	2.87	0.19	0.61	1991–2000
	71.79	7.27	0.58	0.32	0.77	2.29	0.19	0.70	2001–2010
	63.77	7.04	0.56	0.45	0.92	2.85	0.21	0.64	1961–2012
Don	72.55	6.44	0.59	0.36	0.82	2.30	0.23	0.92	2000–2012
	82.50	6.24	0.52	0.36	0.82	2.16	0.18	0.89	1971–1980
	72.50	6.58	0.56	0.37	0.85	2.36	0.23	0.87	1981–1990
	77.50	6.56	0.56	0.35	0.83	2.32	0.21	0.80	1991–2000
	75.96	6.43	0.56	0.36	0.83	2.29	0.21	0.90	1967–2012
Frome	52.50	1.90	0.66	0.55	1.04	0.97	0.27	0.65	1971–1980
	57.50	2.10	0.63	0.47	0.93	0.94	0.24	0.71	1981–1990
	50.00	2.08	0.56	0.54	1.03	1.15	0.24	0.61	1991–2000
	56.29	2.05	0.61	0.48	0.96	0.95	0.23	0.67	1971–2012
Pang	52.94	0.67	0.44	0.55	1.04	0.33	0.22	0.64	2000–2012
	48.15	0.66	0.46	0.60	1.10	0.35	0.23	0.78	1993–1999
	50.63	0.66	0.44	0.57	1.06	0.34	0.22	0.68	1993–2012

Table 7. Results of annual GLUE prediction bounds for all the studied catchments.

Catchment	Annual prediction bounds								
	CR	*B*	RB	*S*	*T*	*D*	RD	*R* factor	Period
Eden	50.00	20.32	0.19	0.52	0.94	9.02	0.09	3.20	1975–1976
	100.00	23.46	0.19	0.09	0.49	2.12	0.02	1.28	1976–1977
	90.00	22.42	0.23	0.25	0.71	5.01	0.05	0.89	1971–1980
	100.00	24.11	0.20	0.23	0.68	5.65	0.04	1.14	1981–1990
	90.00	23.64	0.20	0.27	0.74	6.11	0.05	1.12	1991–2000
	91.67	24.13	0.19	0.24	0.69	5.48	0.04	0.89	2001–2012
	92.86	23.60	0.20	0.25	0.71	5.56	0.05	0.93	1971–2012
Ebbw	60.00	34.62	0.19	0.45	0.92	17.05	0.11	0.71	1971–1980
	40.00	30.73	0.13	0.72	1.24	21.50	0.08	0.80	1981–1990
	70.00	32.29	0.13	0.33	0.78	10.38	0.04	0.71	1991–2000
	90.00	39.24	0.19	0.28	0.75	10.37	0.05	0.74	2001–2010
	65.38	36.10	0.17	0.44	0.91	15.15	0.07	0.73	1961–2012
Don	76.92	44.27	0.30	0.25	0.70	10.26	0.08	0.75	2000–2012
	80.00	41.60	0.28	0.35	0.84	14.43	0.11	1.00	1971–1980
	100.00	44.18	0.28	0.19	0.65	8.04	0.05	1.35	1981–1990
	80.00	43.91	0.28	0.25	0.69	10.34	0.08	0.96	1991–2000
	84.78	43.20	0.28	0.26	0.72	10.76	0.08	1.02	1967–2012
Frome	100.00	13.99	0.29	0.21	0.67	2.97	0.06	1.09	1971–1980
	100.00	14.34	0.26	0.17	0.62	2.44	0.05	1.57	1981–1990
	90.00	14.42	0.25	0.25	0.71	3.74	0.06	0.92	1991–2000
	97.62	14.55	0.27	0.19	0.65	2.88	0.05	0.99	1971–2012
Pang	76.92	7.22	0.37	0.32	0.80	2.28	0.12	0.81	2000–2012
	100.00	8.74	0.43	0.27	0.76	2.37	0.12	1.38	1993–1999
	85.00	7.75	0.39	0.30	0.78	2.31	0.12	0.97	1993–2012

Appendix

A.1.1 Containment ratio, CR

The containment ratio, is the percentage of the number of observed flows enveloped by its prediction bounds to the total number of the observed flows. This index is commonly used for measuring the goodness of the prediction bounds. The larger the value of CR, the greater is the proportion of the observed flow points that fall within the interval defined by the prediction bounds. A high CR for the estimated prediction bounds is always the aim.

A.1.2 Average band width, B

The average bandwidth, *B*, of the prediction bounds for the whole simulated period is calculated as:

$$B = \frac{1}{N}\sum_{i=1}^{N} b_i \quad \text{with } b_i = Q_i^u - Q_i^l \qquad \text{(A1)}$$

where b_i is the bandwidth of the prediction bounds for the flow at time *i*. For a given confidence level. Q_i^u and Q_i^l represent the upper and lower prediction bounds of flows, respectively and are associated with a particular confidence level (5% and 95% selected for this study). Narrow bandwidth is considered better than wide band width.

A.1.3 Average relative bandwidth, RB

In order to compare the results of the prediction bounds of different catchments, it is necessary to eliminate the impact of flows magnitude on the bandwidth of the prediction bounds. This can be done by using a dimensionless index, the average relative bandwidth of the prediction calculated as:

$$\text{RB} = \frac{1}{N}\sum_{i=1}^{N} \text{rb}_i \;;\quad \text{rb}_i = b_i/Q_i \quad \text{(A2)}$$

where rb_i is the ratio of the bandwidth of the prediction bounds at time *i* to the corresponding observed discharge Q_i.

A.1.4 Asymmetry degree indices, S and T

There are two indices for assessing the average asymmetry degree of the prediction bounds with respect to the observed flows. These two indices are referred to as *S* and *T*. The index *S* is calculated as:

$$S = \frac{1}{N}\sum_{i=1}^{N} s_i \qquad \text{(A3a)}$$

$$s_i = |h_i - 0.5| \quad \text{(A3b)}$$

$$h_i = \frac{Q_i^u - Q_i}{Q_i^u - Q_i^l} = \frac{Q_i^u - Q_i}{b_i} \quad \text{(A3c)}$$

where s_i represents the asymmetry degree of the prediction bounds with respect to the corresponding observed discharge, Q_i. s_i is a function of h_i, which is the ratio of the difference between the upper limit, Q_i^u and the observed discharge, Q_i to the actual bandwidth, b_i. An average asymmetry value of $S < 0.5$ would mean that, on average, the river flows lie within the prediction bounds. In a 100% completely symmetrical case the value of S would be zero. The larger the value of S, the greater asymmetrical the prediction bounds are around the observed flows.

The second index for assessing the average asymmetry degree of the prediction bounds withrespect to the observed flows, is referred to as T, calculated is defined as:

$$T = \frac{1}{N} \sum_{i=1}^{N} t_i \quad \text{(A4a)}$$

$$t_i = \left(\frac{\left(Q_i^u - Q_i\right)^3 + \left(Q_i^l - Q_i\right)^3}{[Q_i^u - Q_i^l]^3} \right)^{1/3} \quad \text{(A4b)}$$

The variations of the t_i values depend on the location of the observed flows with respect to the prediction bounds. It is expected that $0 \leq t < 1$, with $t = 0$ when the value of Q_i is equal to the lower and upper prediction bounds. The larger the value of T, the more asymmetrical the prediction bounds are around the observed flows.

A.1.5 Average deviation amplitude

In some cases, where the estimated prediction bounds are asymmetric with respect to the observed flows, the middle point of the prediction bounds Q^m deviates from the corresponding observed flow Q. To quantify the actual discrepancy between the trajectory consisting of the middle points the prediction bounds and the observed flows, another index, D, the average deviation amplitude of the prediction bounds from the observed flow is calculated as:

$$D = \frac{1}{N} \sum_{i=1}^{N} d_i \quad \text{(A5a)}$$

$$d_i = |Q_i^m - Q_i| = \left| \frac{1}{2}(Q_i^u - Q_i^l) - Q_i \right| \quad \text{(A5b)}$$

A.1.6 Average relative deviation amplitude

To eliminate the impact of flow magnitude on the value of the *D* index, the dimensionless relative average deviation amplitude RD, would be a better option. It is calculated as:

$$\mathrm{RD} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{rd}_i \quad \text{(A6a)}$$

$$\mathrm{rd}_i = \frac{\left|\frac{1}{2}(Q_i^u - Q_i^l) - Q_i\right|}{Q_i} = \left|\frac{Q_i^m}{Q_i} - 1\right| \quad \text{(A6b)}$$

where rd_i is the relative deviation of the mid-point of the prediction bounds Q_i^m from the corresponding observed flow, Q_i at time *i*.

A.1.7 The R factor

The *R* factor is calculated as:

$$R = \frac{\bar{d_x}}{\sigma_x} \quad \text{(A7a)}$$

$$\bar{d_x} = \frac{1}{n}\sum_{i=1}^{n}(x_u - x_l) \quad \text{(A7b)}$$

where σ_x is the standard deviation of the measured streamflow *x*; $\bar{d_x}$ is the average distance between the upper and lower boundaries ($Q_{0.95}$ and $Q_{0.05}$); and *n* is the number of observations.

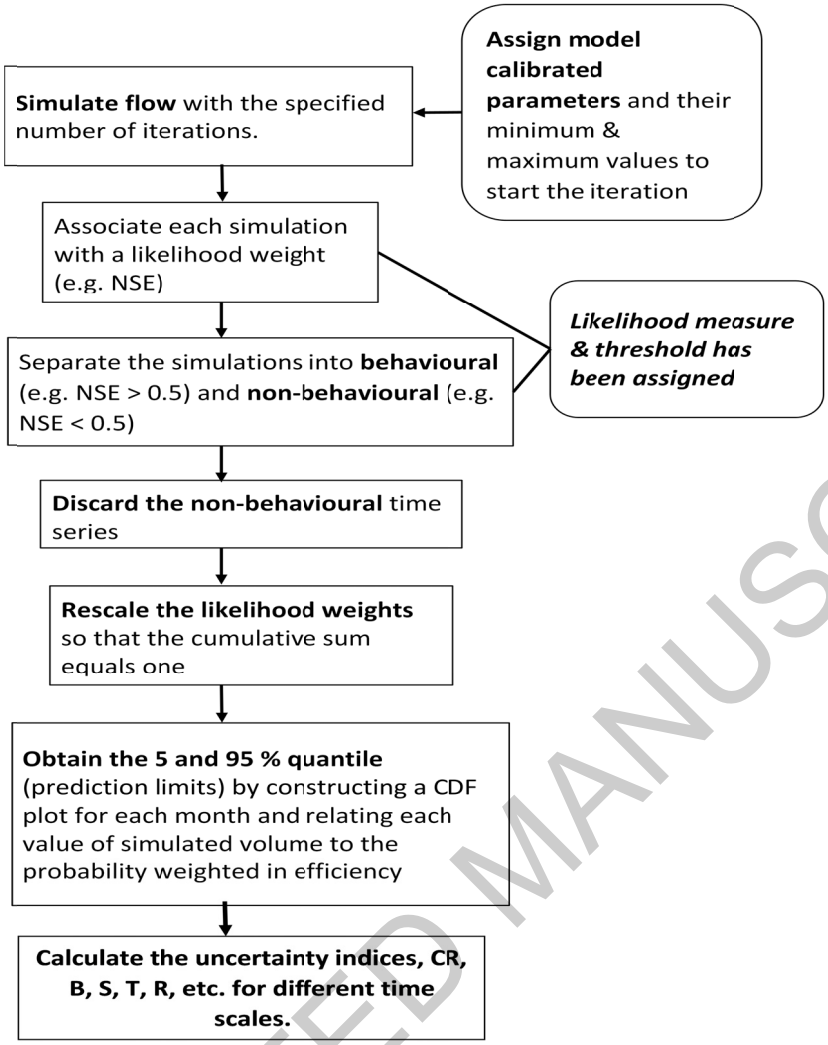

Figure A1. Flowchart for the uncertainty analysis procedure.

Figure 1

Figure 2

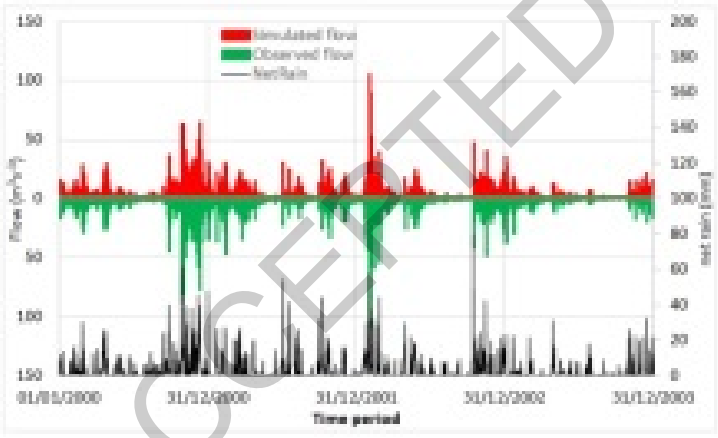

Figure 3

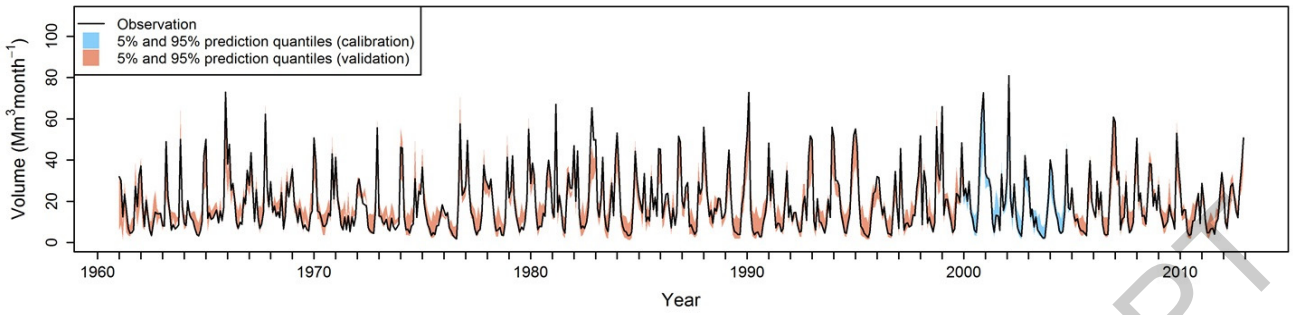

Figure 4

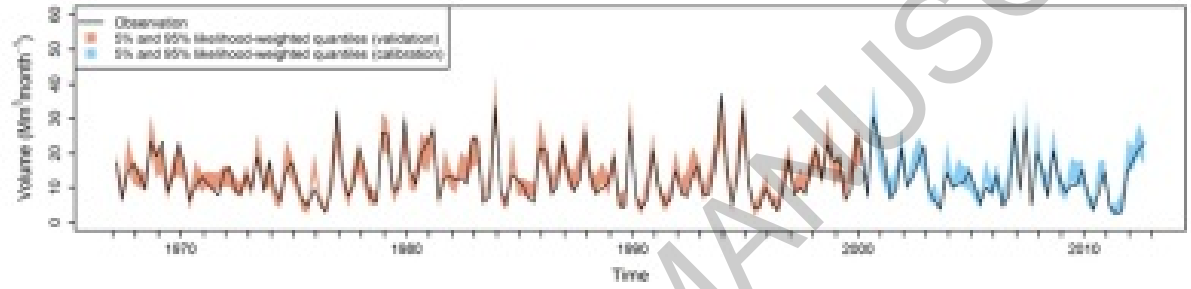

Figure 5

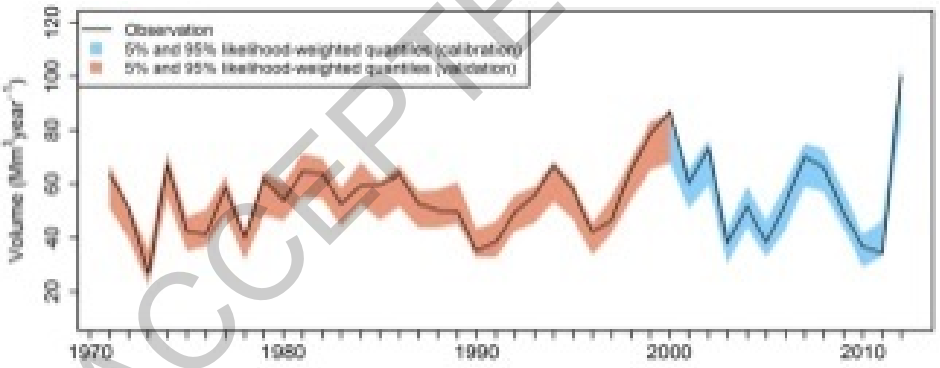

Figure 6

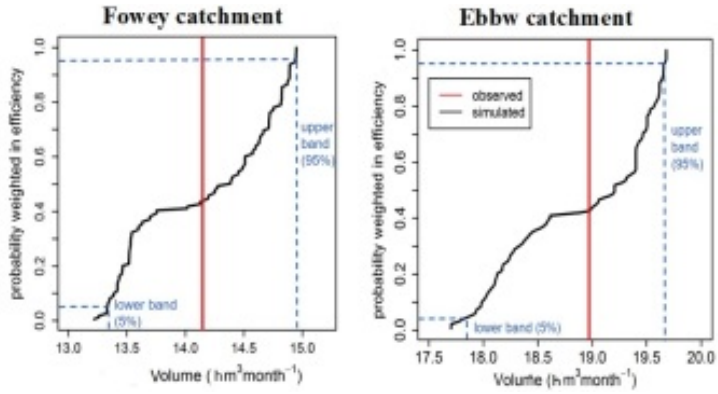

Figure 7

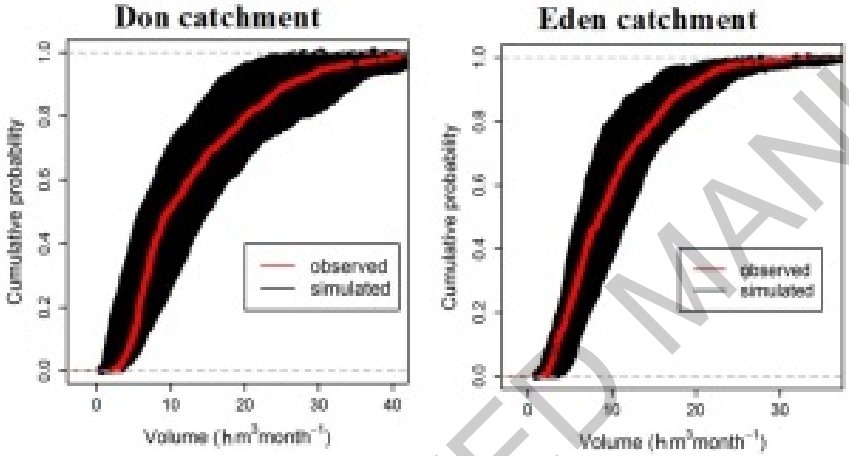

Figure 8

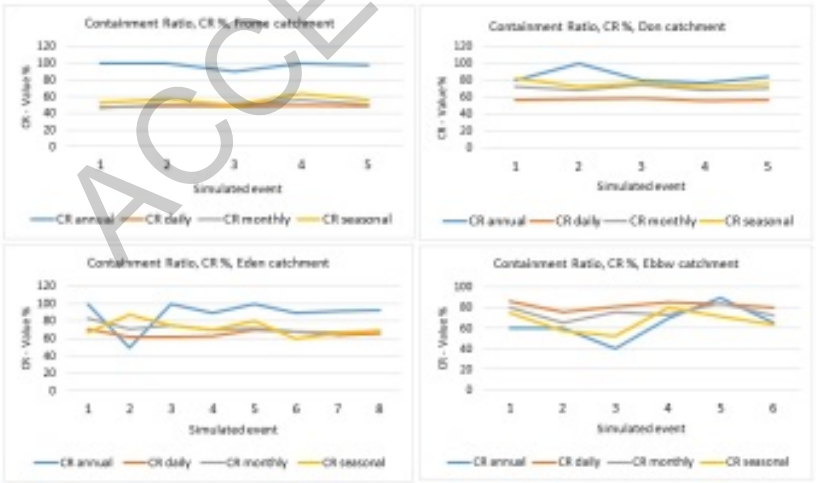

Figure 9

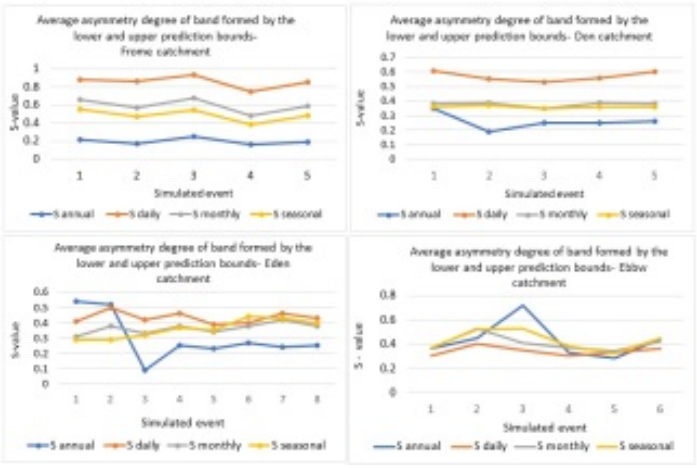

Figure 10

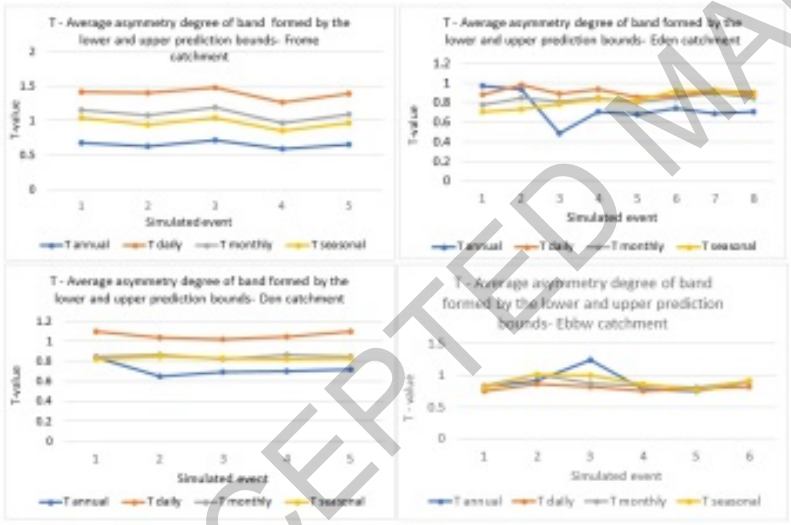

Figure 11

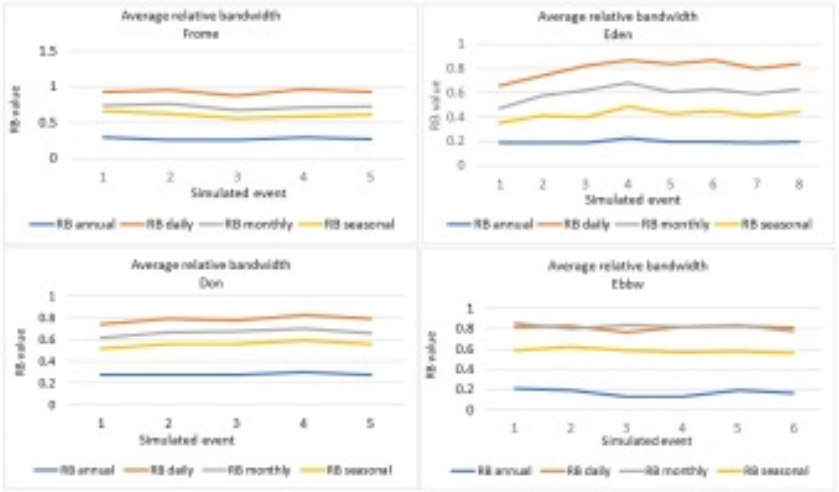

Figure 12