Central Lancashire Online Knowledge (CLoK)

Title	Application of Generalized Likelihood Uncertainty Estimation (GLUE) at different temporal scales to reduce the uncertainty level in modelled river flows
Type	Article
URL	https://clok.uclan.ac.uk/id/eprint/33880/
DOI	https://doi.org/10.1080/02626667.2020.1764961
Date	2020
Citation	Ragab, Ragab, Kaelin, Alexandra, Afzal, Muhammad and 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. ISSN 0262-6667
Creators	Ragab, Ragab, Kaelin, Alexandra, Afzal, Muhammad and Panagea, Ioanna

It is advisable to refer to the publisher's version if you intend to cite from the work.
https://doi.org/10.1080/02626667.2020.1764961

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

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

Article

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

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

Ragab, R.[1*]., Kaelin, A.[1], Afzal, M.[1,2], and Panagea[3], I.

Abstract

In this study, the distributed catchment-scale model (DiCaSM) has been applied on six catchments across the UK, with the catchment areas varying from 150 km^2 to over 300 km^2. Given that river flows are of great importance in terms of water supply and for ecosystem services, the river flows were selected to study the uncertainty level in predicting the river flows. The hydrological model was calibrated over a short period of time, and validated over a longer period. For most of the studied catchments, the Nash-Sutcliffe efficiency (NSE) factor, used as indicator of goodness of fit during the model calibration period, was above 0.90, while for the validation test, all the studied catchments showed a NSE of above 0.80 over the entire study period (approx.1961-2012). The Generalized Likelihood Uncertainty Estimation (GLUE) methodology was applied. The uncertainty analysis supported the model efficiency results well. The observed river flows were within the predicted bounds/envelope of 5% and 95% percentiles. These predicted river flows bounds contained above 70% of the observed river flows as expressed by the Containment Ratio (*CR*). In addition to CR ratio, other uncertainty indices, *S, T, B, RB, D, RD* and *R-factor* of uncertainty level in the predicted river flows were also quantified and indicated that the model parameters and predicted river flow have acceptable levels of uncertainty. The GLUE methodology showed lower uncertainty in predicted river flows when increasing the time scale from daily to monthly to seasonal river flows with the lowest uncertainty associated with annual flows. The findings of the study have broader implications for hydrologists, climatologists, and water authorities to study the future impacts of climate and land use changes on water resources availability.

Key words: DiCaSM; GLUE, Model uncertainty, Eden, Don, Ebbw, Frome, Fowey, Pang

---

[1] Centre for Ecology & Hydrology (CEH), Wallingford, Oxfordshire, OX10 8BB, UK
[2] School of Earth and Ocean Sciences Cardiff University, UK CF10 3AT, UK
[3] Division of Soil and Water Management, Katholieke Universiteit Leuven, Belgium
[*] Corresponding author, rag@ceh.ac.uk

1. Introduction

Hydrological models' results 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 inherited 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 a number of hydrological studies by Kan et al. (2019), Xie et al. (2019), Tegegne et al. (2019), Xue et al. (2018) and Aynom et al. (2018).

The methodology assumes that there is an optimal model structure or parameter set that could represent a catchment hydrology. 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 applied the *GLUE* methodology to assess the model performance, little work is carried to study the model performance over different time scales (daily, monthly, seasonal and annual), and over different catchments with different catchment sizes, land use characteristics, geographic location and soil physical characteristics.

The main objective of this study is to assess the uncertainty level in simulated river flow by DiCaSM model (Ragab and Bromley, 2010) using the GLUE methodology on six different catchments and on four different time scales.

The findings of the study are expected to have broader applications which could potentially be used by hydrologists, climatologists, and meteorologists to study the future impacts of climate change on water resources.

2. The DiCASM model, 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 a physically based and considers the 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 stream flow. The model has been developed to estimate the catchment water balance components and to account for the impact of the changes in climate and land use on the catchment water resources, including stream flow and recharge to the groundwater. The model adopts a distributed approach with a variable spatial scale (default in 1km grid square) and requires daily input data of rainfall, temperature, wind speed, vapour pressure, and radiation and runs on a daily time step, however, if hourly rainfall data is available, the model can run with hourly time step. 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), in Italy (D'Agostino et al., 2010) and in Cyprus (Ragab et al., 2010).

2.2 The components of the DiCaSM model

The key model components are rainfall interception, potential evapotranspiration, the catchment water balance, infiltration, surface runoff /overland flow and ground water 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 was calculated according to Raupach (1995) whereas the surface runoff calculation was based on either excess saturation or excess filtration. The infiltration was calculated

3

according to either Philip (1957) or Green and Ampt (Green, 1911) equations. The runoff is a routed between the low points of each grid square along the prevailing slope using the digital terrain model (DTM). The model calculates the soil water balance of the root zone based on the four layers 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 model

The DiCaSM model was applied on six selected catchments from 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 Centre for Ecology and Hydrology (Morris et al., 1990b, Morris and Flavin, 1994) and the National River Flow Archive provided data for the daily river flow for the studied catchments (NRFA, 2014). The river flow data were collected from the Centre for Ecology and Hydrology, 'Digital Rivers 50km GB' Web Map Service (CEH, 2014). The UK Land Cover data were obtained from the Centre for Ecology and Hydrology (Land Cover Map 2007 (25m raster, GB) Web Map Service (Morton et al., 2011).

Figure 1: Case study catchments overview map

2.4 *Key model parameters and model calibration*

To calibrate the model against the observed streamflow several model parameters were used. The river flow in DiCaSM model depends on six parameters. They are the percentage of surface runoff routed to stream, exponent function describing the peak flow, catchment storage/time lag coefficient, stream storage/time lag coefficient, base flow 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 the best periods i.e. those available with the best quality data, were selected for the model calibration. The selected time period was run using a simple iteration algorithm for optimization in which each of the above-selected parameters were assigned a range described by a minimum and a 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 of iterations was assigned to parameters which showed more impact on the streamflow. The model calculates the Nash-Sutcliffe Efficiency factor, *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 Nash-Sutcliffe Efficiency factor, other indices like *ln NSE* (using natural logarithmic values of stream flow) 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 the model validation period varied from several years to the entire available record.

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 Nash-Sutcliffe Efficiency (*NSE*) factor, based on (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* factor of 100% indicates a perfect match.

$$NSE = 100 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \qquad (1)$$

where O_i and S_i *refer* to the observed and simulated flow data, respectively, and $\bar{O}$ is the mean of the observed data. 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):

$$\ln NSE = 100 - \frac{\sum_{i=1}^{n}(ln\ O_i - ln\ S_i)^2}{\sum_{i=1}^{n}(ln\ O_i - ln\ \bar{O})^2} \qquad (2)$$

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

$$R^2 = \left\{ \frac{1}{N} \frac{\sum [(y_0 - \overline{y_0}] \, (\overline{y_s - \overline{y_0}})}{\sigma y_0 - \sigma y_s} \right\} \tag{3}$$

where y_o is the observed value, y_s is the simulated value, N is the total number of observations, $\overline{y}_o$ is the average measured value, $\overline{y}_s$ is the average simulated value, σy_0 is the observed data standard deviation and σy_s is the simulated data standard deviation. The values of this index can range from 1 to 0, with one indicating perfect fit.

3. The 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. The GLUE methodology is somehow related to the Nash–Sutcliffe Efficiency, as the likelihood uncertainty level is calculated as:

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

The likelihood measure, L is the likelihood measure of the *ith model* simulation made with the parameters set θ_i related to the measured stream flow Y, and is a function of the ratio of errors variance, σ_i^2 for the *ith* model simulation (representing the variance of the error between the model prediction and the observed stream flow) and the variance of the observed stream flow, $\sigma_{observed}^2$. 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, together with 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 likelihood measure (i.e. minimum Nash-Sutcliffe Efficiency, NSE value) chosen to evaluate if the selected set of parameters is behavioural or not (Beven, 2006, Beven and Binley, 1992, Viola et al., 2009). Different likelihood measures could be used like the Nash-Sutcliffe Efficiency (NSE) or the sum of squared errors (Beven and Binley, 1992, Freer et al., 1996). The users could define the threshold of efficiency criteria according to their model preference and for each individual catchment. The GLUE methodology rejects the 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 could be used to estimate the quantiles for the predictions at any time step.

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

3.1 Uncertainty indicators

In this study, the uncertainty analysis was carried out on the calibration and validation sub-periods. Different sets of model parameters were used to generate the modelled river flow time series and the Nash-Sutcliffe Efficiency (NSE) factor 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). The *CR* parameter is the *Containment Ratio* which is the percentage of observed river flows that are enveloped by the prediction bounds of 5% and 95% confidence levels ($Q_{5\%}$ - $Q_{95\%}$) likelihood-weighted quantiles. *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 *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 band-width 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 band-width, *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 band-width 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 Annex 1.

3.2 The GLUE methodology application

The DiCaSM model provides for each parameter set combination a single value of simulated stream flow. Performance evaluation was carried out, including rejection of some parameter sets as non-behavioural (NSE < 0.5). This was followed by calculation of the likelihoods of behavioural parameter sets and rescaling to produce a cumulative sum of 1. This was carried out by ranking in ascending order all the simulated stream flows (only of those behavioural parameter set) and the corresponding cumulated efficiencies (NSE). Each cumulated efficiency value, divided by the maximum value,

resulted in a value ranging between 0 and 1. These values are referred to as the 'probability weighted in efficiency' (D'Agostino, et al. 2010, Viola et al., 2009). A cumulative distribution function (CDF) of the simulated stream flow has also been constructed, relating each value of the simulated flow to the corresponding value of the probability weighted in efficiency.

The model uncertainty analysis was carried over daily, monthly, seasonal and annual time scale. This helps to assess the uncertainty level of different time scales.

For calculating an uncertainty level for the simulated streamflow the following steps were followed:

1. Simulate flow with several parameter sets
2. Assess the likelihood of each parameter set (e.g. NSE threshold)
3. Separate simulations into acceptable and non-acceptable
4. Rescale the likelihoods so that the cumulative sum = 1
5. Conduct statistical analysis of the acceptable simulations (5% and 95% quantile)

There are several ways of presenting the results of the uncertainty analysis.

➢ Statistical data about the behavioural and non-behavioural simulations: for example, the percentage of behavioural simulations. If the model was run for 1000 times and achieve 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 are described in "3.1. Uncertainty indicators".
In the description of the results, a combination of the 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 are 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% (Beven & Freer, 2001; Freer et al., 1996; 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 curve of volumes, the upper and lower confidence bands can be identified. Plotting the mean value of the observed volume, 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

In order 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. for two years. This provided more insight into the range of the 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 six catchments (Table 1).

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

Catchment	Total number of iterations	Number of iterations with monthly *NSE* >50%	Range of monthly *NSE* %	Range of monthly *NSE* > 50%
Eden	60	60	50.4 – 84.9	50.4 – 84.9
Frome	490	396	-24.6 – 92.9	50.4 – 92.9
Ebbw	108	103	27.1 - 94.9	50.5 - 95.3
Fowey	112	107	42.0– 96.4	51.5 – 96.4
Pang	648	342	-92.3 – 90.7	50.2 - 90.7
Don	336	225	-35.2 – 89.1	50.0 - 89.1

For all the selected time periods, the threshold for *GLUE* was set at *NSE* = 50%. This means that behavioural simulations must have *NSE* equal or above 50%. All simulations with *NSE* below 50%

were discarded from the *GLUE* analysis. The range of the 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 narrow range (as the model proved the stream flows were less sensitive to those parameters).

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		Base flow 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	1.75E-8– 3.3E-7	3
Ebbw	10 - 90	9	0.02 - 0.4	2	2	1	20	1	2.2E-9 - 2.2E-7	6
Eden	40-95	4	1.0E-5	1	315	1	15-55	3	0.14 - 0.75	5
Fowey	4-95	7	0 - 0.3	2	50	1	23.5-120	2	8.5E-9 - 8.5E-7	4
Don	10-90	8	0 - 0.3	6	143	1	18	1	0.02 - 0.95	7

4.2. Model river flow calibration and validation

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

An example of calibration for the Ebbw catchment is shown Figure 2, which shows a good agreement between the observed the simulated flow. The *NSE* during this period was 91%. Overall the model performed well both during the rainy and dry events and responded according to soil hydrological status, i.e. during the soil moisture deficit period, a small rainfall event did not generate a significant increase in streamflow and during a heavy rainfall event, when the soil was wet especially during the winter months, the model generated stream flow. For all catchments, during the model calibration stages, the model efficiency factor, NSE, on average was around 89% and the maximum percentage error did not exceed 1% (Table 3). The model also performed very well during the well-known 1970s drought events. Generally, the overall model performance over the whole period, 1961-2012 for all catchments was extremely good, the NSE on average was around 85% with maximum error not exceeding 5% for the studied catchments.
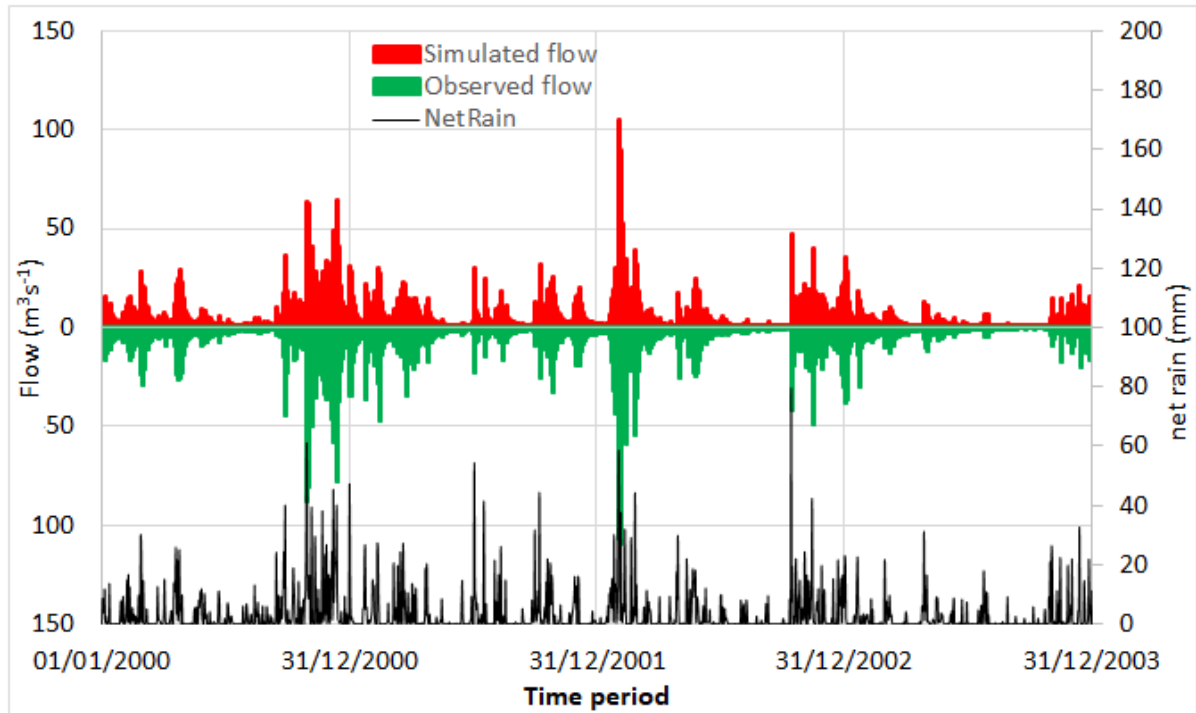
Figure 2: Ebbw catchment model calibration for the period 2000-2003.

Table 3: Model performance during the calibration and validation stages of the 6 catchments

Catchment	Periods	NSE	ln NSE	R^2	Square root of R^2	Modelled flow, m^3s^{-1}	Observed flow, m^3s^{-1}	% error
Ebbw	2000-2003[*]	0.91	0.88	0.92	0.96	7.19	7.23	-0.55
	1971-1980[a]	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
Eden	2000-2001[*]	0.90	0.95	0.89	0.94	5.03	5.04	-0.19
	1971-1980[a]	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
Don	2001-2012[*]	0.91	0.84	0.91	0.95	5.41	5.35	-1.11
	1971- 1980[a]	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
Fowey	1962-1971[*]	0.85	0.87	0.84	0.91	5.10	4.98	2.5
	1971-1980[a]	0.84	0.91	0.85	0.92	5.06	5.11	-0.85
	1962-2012	0.88	0.90	0.86	0.92	5.16	5.38	-2.96
Frome	2000-2001[*]	*0.84*	*0.65*	*0.84*	*0.92*	*2.23*	*2.19*	*-1.82*
	1971-1980 [a]	0.72	0.61	0.73	0.85	1.55	1.62	-5.2
	1962-2012	0.74	0.63	0.76	0.87	1.75	1.74	0.43
Pang	2001- 2003[*]	*0.92*	*0.89*	*0.94*	*0.97*	*0.79*	*0.81*	*-2.14*
	1971-1980[a]	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 and [a] periods with drought events

13

4.3. Model uncertainty analysis

Based on the simulated river flows of the six catchments, during the calibration and validation periods, the envelope of 5% and 95% likelihood-weighted quantiles (the envelope of all behavioural models, i.e. NSE >50%) were 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. An example of monthly river flows of Ebbw catchment is shown in Figure 3. This figure shows the envelope of the 5% and 95% likelihood-weighted quantiles in blue for the calibration period 2000-2004, in brown for the validation period between 1961-2012 and the black line represents the observed river flow. For most of the time, the observed discharge is contained within the calculated uncertainty bounds, the predictions bracketed the observations, given that, for the Ebbw catchment, the calibration had a NSE value of 91% and the validation had a NSE value of 87%, as shown in Table 3.
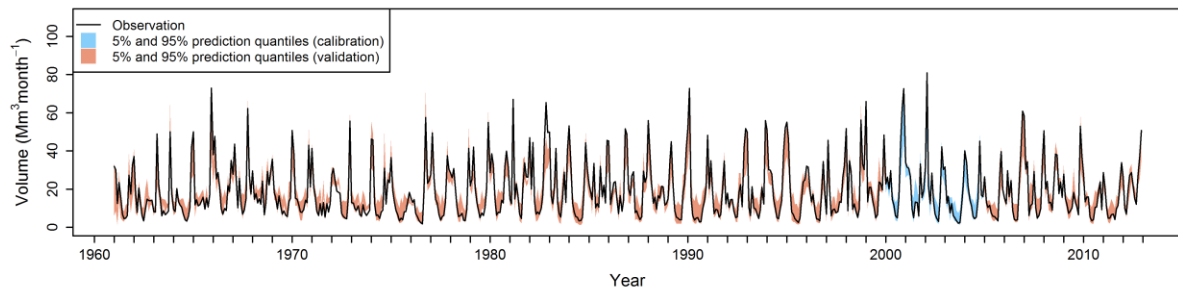

Figure 3: Model output uncertainty boundaries, (the 5th and 95th percentile) performing the GLUE analysis during the model calibration (2000-2004) and validation (1961-2012) periods for the Ebbw catchment river flow (monthly values).

The number of observations contained within the 5% and 95% GLUE uncertainty bounds expressed as Containment Ratio, *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 and show that those sets of parameters used can be considered acceptable in order 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 time scale, 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 over summer. The figure shows the envelope of 5% and 95% likelihood-weighted quantiles together with the observed river seasonal flows. The envelope of the 5% and 95% likelihood-weighted quantiles represents the calibration and the validation periods. For most of the time, the observed discharge is contained within the calculated uncertainty bounds, the predictions bracketed the observations to great extent (CR = 76%) given that the calibration had a NSE value of 91% and the validation had a NSE value of 87% for the Don catchment, as shown in Table 3. Similar results of seasonal flows for other five catchments were obtained, and values are shown in Table 6.
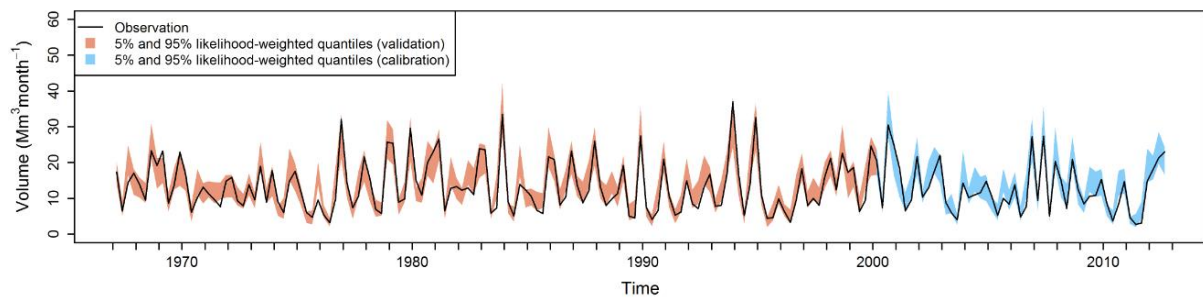

Figure 4: Model output uncertainty boundaries, (the 5th and 95th percentile) performing the GLUE analysis during the model calibration (2001-2012) and validation (1967-2012) periods for the Don catchment river flow (seasonal values).

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 example. The figure shows the envelope of the 5% and 95% likelihood-weighted quantiles for the calibration (period 2001-2012) and for the validation period (1971-2000), where 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 Containment Ratio (*CR*) mean that the model captures the observed flow quite well as more observed values are included in the envelope.
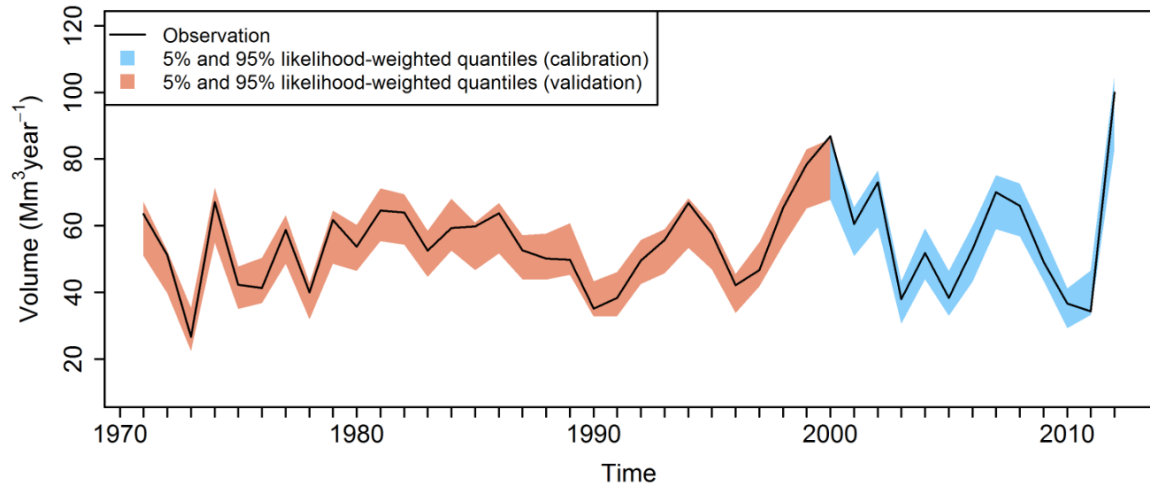
Figure 5 Model output uncertainty boundaries, (the 5th and 95th percentile) performing the GLUE analysis during the model calibration (2001-2012) and validation (1971-2000) periods using the annual observed and simulated data for the Frome catchment.

Based on the methodology given by D'Agostino et al. (2010), Beskow et al. (2011), Hoang et al. (2018), and Viola et al. (2009), the simulated stream flows of each behavioural parameter combination (NSE> 0.5) were ranked in ascending order and the corresponding efficiencies (Nash–Sutcliffe Efficiency, NSE) have been cumulated. Each cumulated efficiency value, divided by their maximum value, resulted in a value ranging between 0 and 1, Figure 6. These values are interpreted as the 'probability weighted in efficiency'. 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, as shown in Figure 6 for the Fowey and Ebbw catchments as examples. Other catchments showed similar results.

In order to compare the behavioural time series with the observed volume, a cumulative distribution function (CDF) is plotted for all behavioural time series. Figure 7 shows the cumulative distribution probability of the Don and Eden catchments as example for the period 1962-2012. It shows that the observed streamflow data fall within the range of the number of simulated values obtained from the iterations.
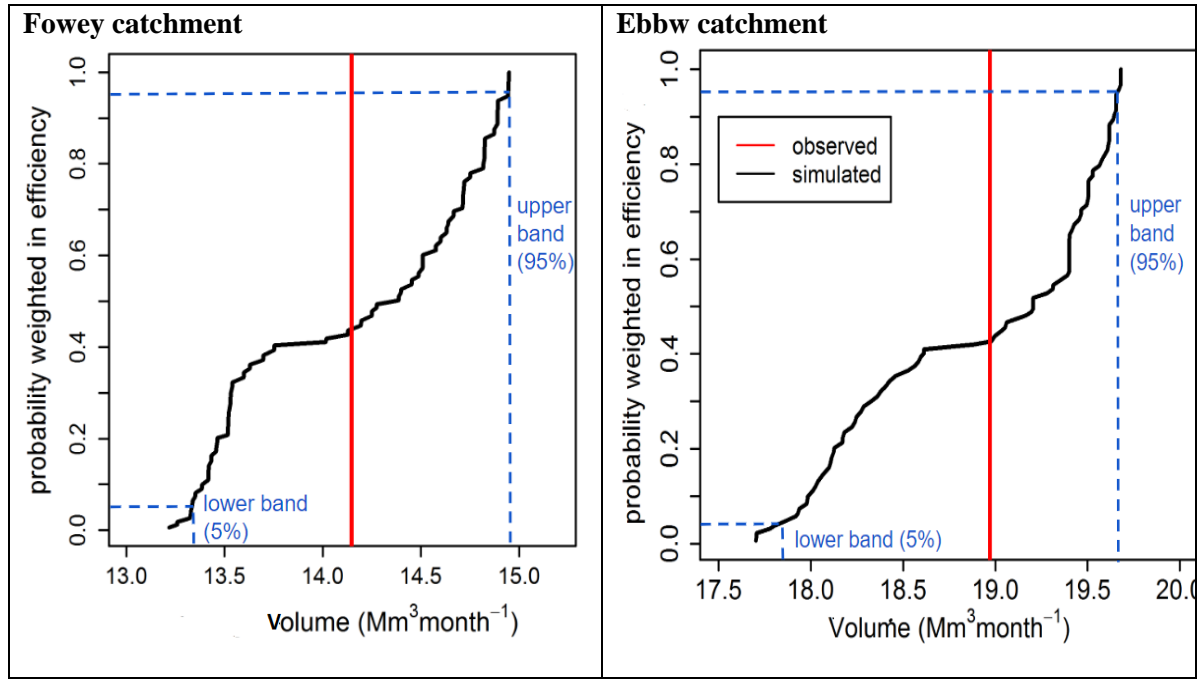
Figure 6: Uncertainty band of the DiCaSM parameters for the period 1962-2012 for the Fowey and Ebbw catchments.
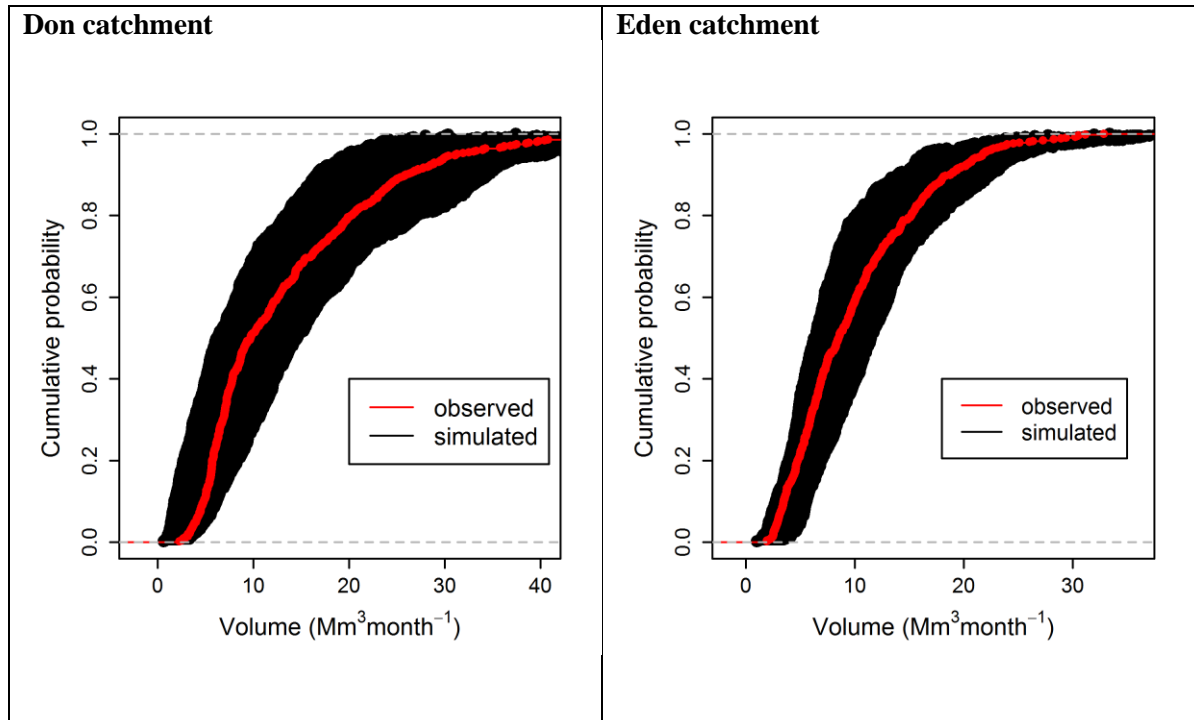
Figure 7: The cumulative probability plot of the flow for the Don and Ebbw catchments (1962-2012)

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, 79% to 86% for the Fowey 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-time record (1961-2012), the *CR* values were 66%, 80%, 57%, 85%, 49% and 36% for the Eden, Ebbw, Don, Fowey, Frome, and Pang catchments, respectively.

The asymmetry degree expressed by *S* and *T* show that the *S* value ranged from 0.39 to 0.50 for the Eden, 0.30 to 0.40 for the Ebbw, 0.53 to 0.61 for the Don, 0.29 to o.35 for the Fowey, 0.75 to 0.93 for the Frome and 0.99 to 1.02 for the Pang catchments. The *S* value for the total period was, 0.43, 0.36, 0.6, 0.3, 0.85, and 1.01 for the Eden, Ebbw, Don, Fowey, Frome and Pang catchments, respectively. The *S* value was within the recommended range $0.0 < S < 0.5$ for the Eden, Ebbw and Fowey catchments. The *T* value ranged from 0.86 to 0.98 for the Eden, 0.76 to 0.87 for the Ebbw, 1.02 to 1.10 for the Don, 0.75 to o.83 for the Fowey, 1.26 to 1.48 for the Frome and 1.54 to 1.57 for the Pang catchment. The *T* value for the whole period was 0.90, 0.82, 1.10, 0.76, 1.38, and 1.56 for the Eden, Ebbw, Don, Fowey, Frome and Pang catchments, respectively. Similar to *S*, the *T* value was within the recommended range $0 < T < 1$ in the Eden, Ebbw and Fowey catchments. One should note here that larger values of *S*, or *T* mean more asymmetrical prediction bounds around the observed flow hydrograph.

The *R-factor* value ranged from 0.72 to 0.90 for the Eden, 0.55 to 0.67 for the Ebbw, 0.64 to 0.70 for the Don, 0.72 to 0.85 for the Fowey, 0.46 to o.52 for the Frome and 0.51 to 0.6 for the Pang catchment. The *R-factor* values for the whole period were 0.86, 0.62, 0.68, 0.74, 0.49, and 0.53 for the Eden, Ebbw, Don, Fowey, Frome and Pang catchments, respectively. A value of *R-factor* less than 1.0 is desirable (Singh et al., 2014). The values obtained for the six catchments largely meet this requirement. The average band width *B* of the prediction bounds ($Q_{0.95}$ - $Q_{0.05}$) and the relative average band width *RB* are also shown in Table 4. The relative band width values for the whole period (1961-2012) were

0.84, 0.81, 0.79, 0.87, 0.93, and 0.37 for the Eden, Ebbw, Don, Fowey, 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, 0.25, 0.48, and 0.25 for the Eden, Ebbw, Don, Fowey, Frome and Pang catchments, respectively.

Both RB and RD values are indicating a small relative band width 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). The authors 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 band-width, 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, showed much improved values of *CR*, *S*, *T*, *RB*, *RD* and *R-factor*. Generally, the annual flows showed better results of low uncertainty than seasonal, seasonal was better than monthly and monthly was better than daily. This will be explained in the next section.

Table 4: Results of daily GLUE prediction bounds for all the studied catchments (Parameter description is provided in Annex 1)

Type	CR	B	RB	S	T	D	RD	R-factor	Time period
Eden	71.31	3.53	0.66	0.41	0.88	1.05	0.22	0.89	2012[*]
	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	86.04	5.29	0.82	0.30	0.76	1.63	0.21	0.55	2000-2004[*]
	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	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
	55.49	4.17	0.83	0.56	1.05	1.76	0.42	0.68	2000-2012[*]
	57.31	4.10	0.79	0.60	1.10	1.76	0.39	0.68	1967-2012
Fowey	79.37	3.51	0.81	0.35	0.83	1.10	0.28	0.85	1962-1971[*]
	86.36	3.75	0.88	0.29	0.75	1.11	0.25	0.72	1972-2012
	84.90	3.70	0.87	0.30	0.76	1.11	0.25	0.74	1962-2012
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
	49.46	1.36	0.97	0.75	1.26	0.61	0.48	0.46	2000-2012[*]
	48.48	1.34	0.93	0.85	1.38	0.64	0.48	0.49	1971-2012
Pang	38.18	0.22	0.37	0.99	1.54	0.15	0.24	0.60	1993-1999
	35.40	0.22	0.37	1.02	1.57	0.16	0.25	0.51	2000-2012
	36.37	0.22	0.37	1.01	1.56	0.15	0.25	0.53	1993-2012

*the time period in grey was used for the model calibration.

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

Type	CR	B	RB	S	T	D	RD	R-factor	Time period
Eden	83.33	6.24	0.47	0.31	0.78	1.51	0.14	1.06	2012*
	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	80.00	9.83	0.86	0.36	0.84	3.69	0.29	0.59	2000-2004*
	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	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
	68.59	7.76	0.70	0.39	0.86	2.83	0.29	0.79	2000-2012
	70.47	7.70	0.66	0.38	0.85	2.76	0.27	0.82	1967-2012
Fowey	76.85	6.50	0.66	0.39	0.87	2.23	0.25	0.79	1962-1971
	77.64	7.05	0.80	0.37	0.85	2.62	0.28	0.68	1972-2012
	77.50	6.95	0.77	0.37	0.85	2.55	0.27	0.69	1962-2012
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
	55.77	2.53	0.72	0.48	0.96	1.12	0.28	0.59	2000-2012
	51.19	2.33	0.73	0.59	1.08	1.21	0.33	0.58	1971-2012
Pang	39.29	0.65	0.44	0.69	1.19	0.37	0.25	0.72	1993-1999
	45.51	0.67	0.43	0.69	1.19	0.38	0.25	0.61	2000-2012
	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.

Type	CR	B	RB	S	T	D	RD	R-factor	Time period
Eden	66.67	4.65	0.35	0.29	0.71	1.06	0.10	1.32	2012*
	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	75.00	7.58	0.59	0.37	0.81	2.76	0.21	0.60	2000-2004*
	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	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
	72.55	6.44	0.59	0.36	0.82	2.30	0.23	0.92	2000-2012*
	75.96	6.43	0.56	0.36	0.83	2.29	0.21	0.90	1967-2012
Fowey	65.00	5.23	0.50	0.39	0.87	2.00	0.20	0.76	1962-1971*
	69.94	5.55	0.59	0.39	0.86	2.14	0.21	0.68	1972-2012
	68.97	5.49	0.57	0.39	0.86	2.11	0.21	0.69	1962-2012
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
	63.83	2.09	0.59	0.38	0.85	0.75	0.20	0.73	2000-2012*
	56.29	2.05	0.61	0.48	0.96	0.95	0.23	0.67	1971-2012
Pang	48.15	0.66	0.46	0.60	1.10	0.35	0.23	0.78	1993-1999*
	52.94	0.67	0.44	0.55	1.04	0.33	0.22	0.64	2000-2012
	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.

Type	CR	B	RB	S	T	D	RD	R-factor	Time period
Eden	100	25.14	0.19	0.54	0.97	9.34	0.09	2.1	2012*
	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	45.72	0.21	0.36	0.82	15.38	0.06	0.61	2000-2004*
	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	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
	76.92	44.27	0.30	0.25	0.70	10.26	0.08	0.75	2000-2012*
	84.78	43.20	0.28	0.26	0.72	10.76	0.08	1.02	1967-2012
Fowey	60.00	26.37	0.17	0.43	0.90	10.28	0.07	0.95	1962-1971*
	75.61	25.39	0.15	0.33	0.80	8.27	0.05	0.73	1972-2012
	72.55	26.55	0.17	0.35	0.82	8.94	0.06	0.78	1962-2012
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
	100.00	15.28	0.29	0.16	0.59	2.46	0.05	0.79	2000-2012*
	97.62	14.55	0.27	0.19	0.65	2.88	0.05	0.99	1971-2012
Pang	100.00	8.74	0.43	0.27	0.76	2.37	0.12	1.38	1993-1999*
	76.92	7.22	0.37	0.32	0.80	2.28	0.12	0.81	2000-2012
	85.00	7.75	0.39	0.30	0.78	2.31	0.12	0.97	1993-2012

4.5 Temporal scale impact on the uncertainty levels

4.5.1. The Containment ratio, CR

The CR values, as shown in Tables 4, 5, 6 and 7, were plotted to illustrate the difference in CR values when calculated for daily, monthly, seasonal and annual river flow/volume. Figure 8 shows an example of four catchments. From the four tables, 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 was associated with the 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 range of $0<S<0.5$ and $0<T<1$ to have a good symmetrical condition, low uncertainty occurs within those ranges. Figure 9 shows the

asymmetry indicators S for four catchments and for four time scales. Generally, the S values were reasonable and showed a better symmetry when considering annual flows (0<S<0.5) then 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 four catchments and four time scales, as shown in Figure 10. Reasonable results were obtained showing a better symmetry when considering annual flows (0<T<1) then seasonal then monthly then daily. Good examples are the cases of Eden, Don, Frome, and Pang catchments.

4.5.3. Average relative band width, RB

Relative band width RB if narrow, indicates lower uncertainty. Tables 4, 5, 6 and 7 show the band width, B and the relative band width, RB. Figure 11 shows the example of RB of four catchments. The RB values look reasonable and indicating a relatively narrower band width when using annual flows then monthly or seasonal or daily flows. The figure shows the significant difference between daily and annual RB values.

4.5.4. Average relative deviation amplitude, RD

Relative deviation amplitude, RD if smaller, indicates lower uncertainty. Tables 4, 5, 6 and 7 show the deviation amplitude D and the relative deviation amplitude, RD. Figure 12 shows the example of RD of four catchments. The RD values look reasonable and indicates a relatively small deviation when using annual flows then monthly or seasonal or daily flows. The figure 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. A value of 1 is ideal. Tables 4, 5, 6 and 7 show reasonable values for R-factor. The example of *R - factor* for four catchments is shown in Figure 13. The figure shows less variations between daily, monthly or seasonal flow with the annual sometimes slightly better.

In contrary to the daily, monthly and seasonal uncertainty indicators such as *CR,* annual river flows had less 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 the Pang catchments were reduced significantly with *CR* of the Pang ranging from 77% to 100% and for the Frome from 90% to 100%. Other uncertainty indicators such as *S* and *T* have also been improved for all the six catchments. Generally, this indicates that the uncertainty level is reduced when considering annual rather than monthly or seasonal river flows.
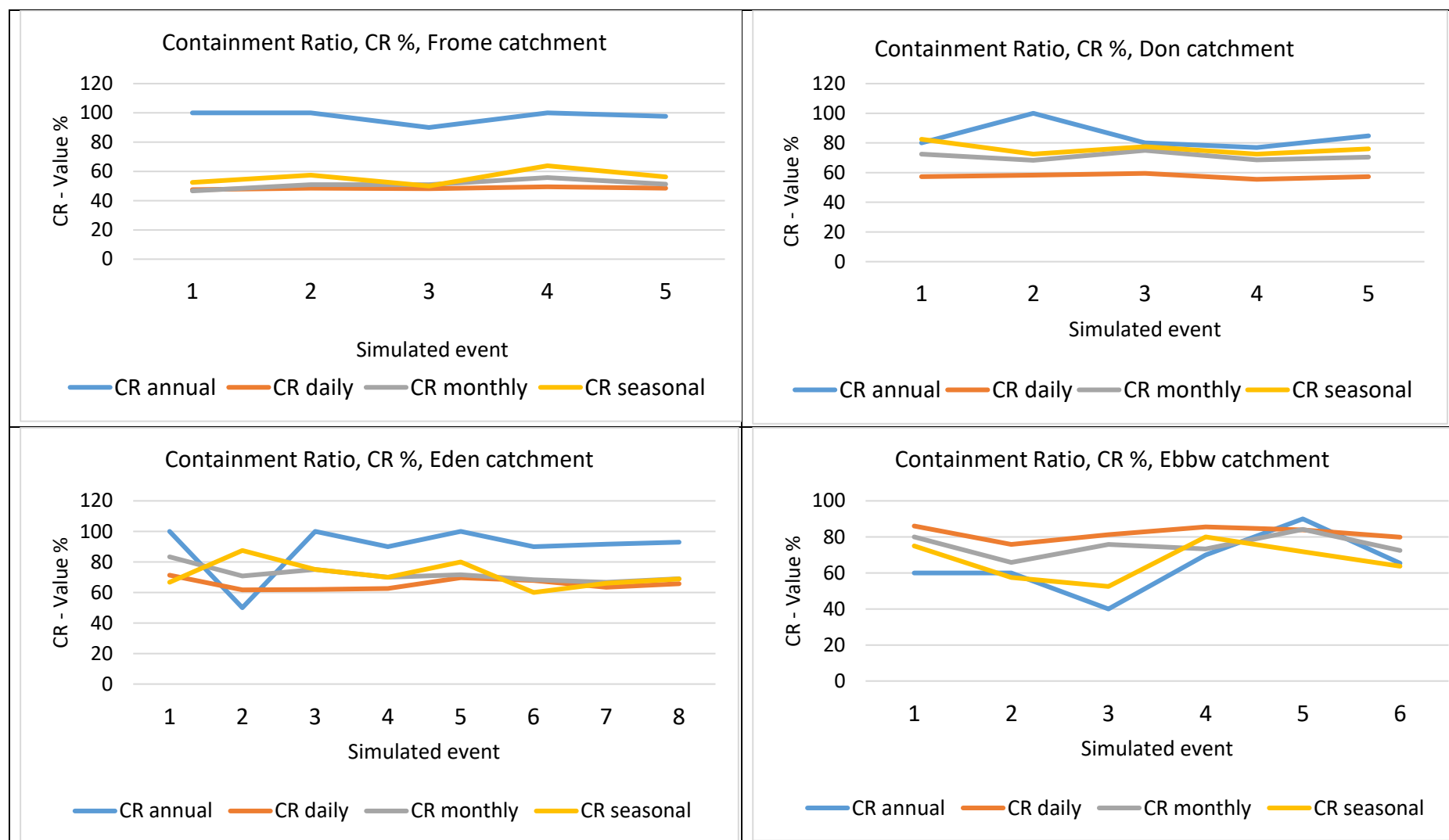
Figure 8. Containment ratio CR at different time scale for the Frome, Eden, Don and Ebbw catchments.
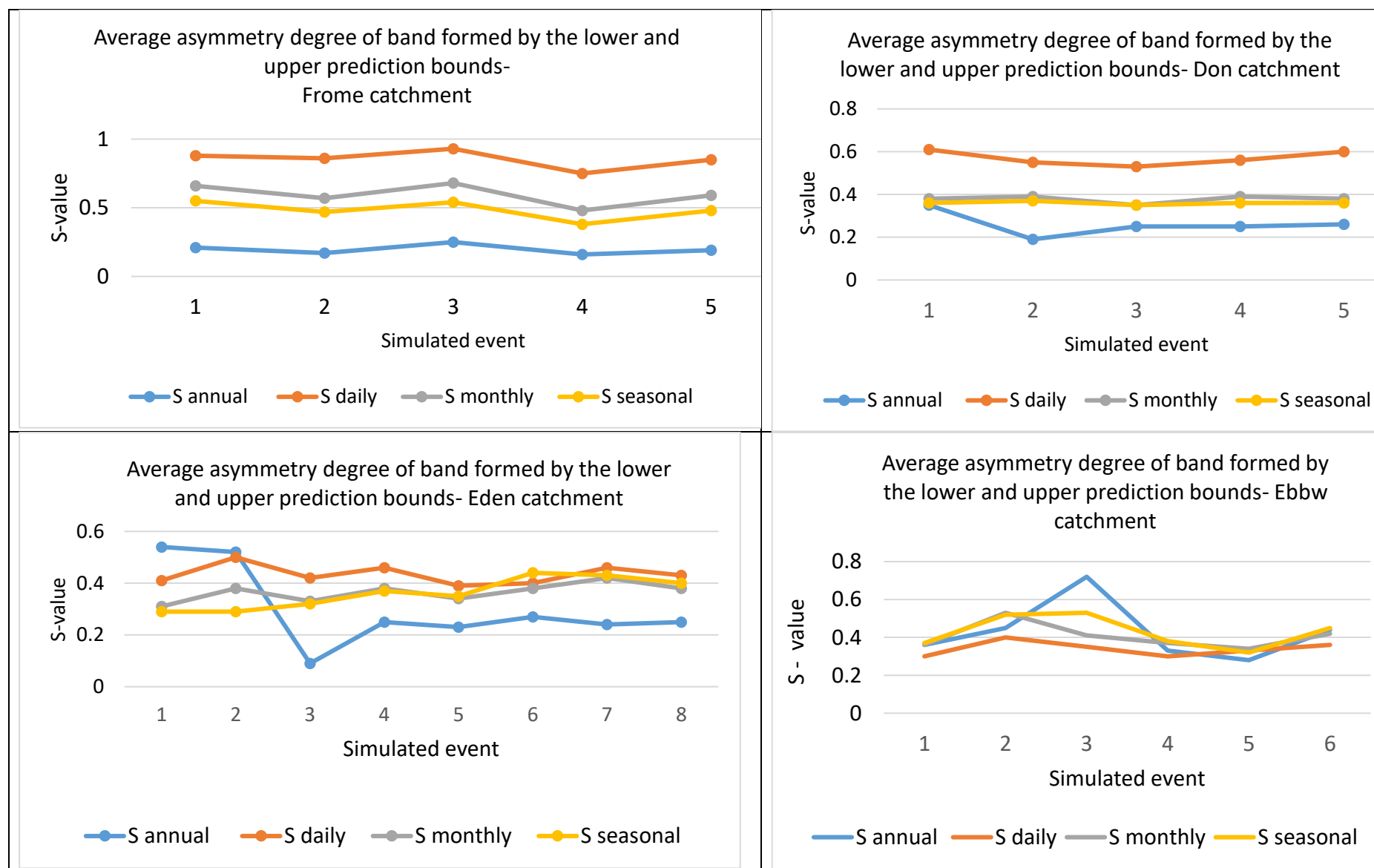
Figure 9. Asymmetric degree S for the Frome, Eden, Don and Ebbw catchments.
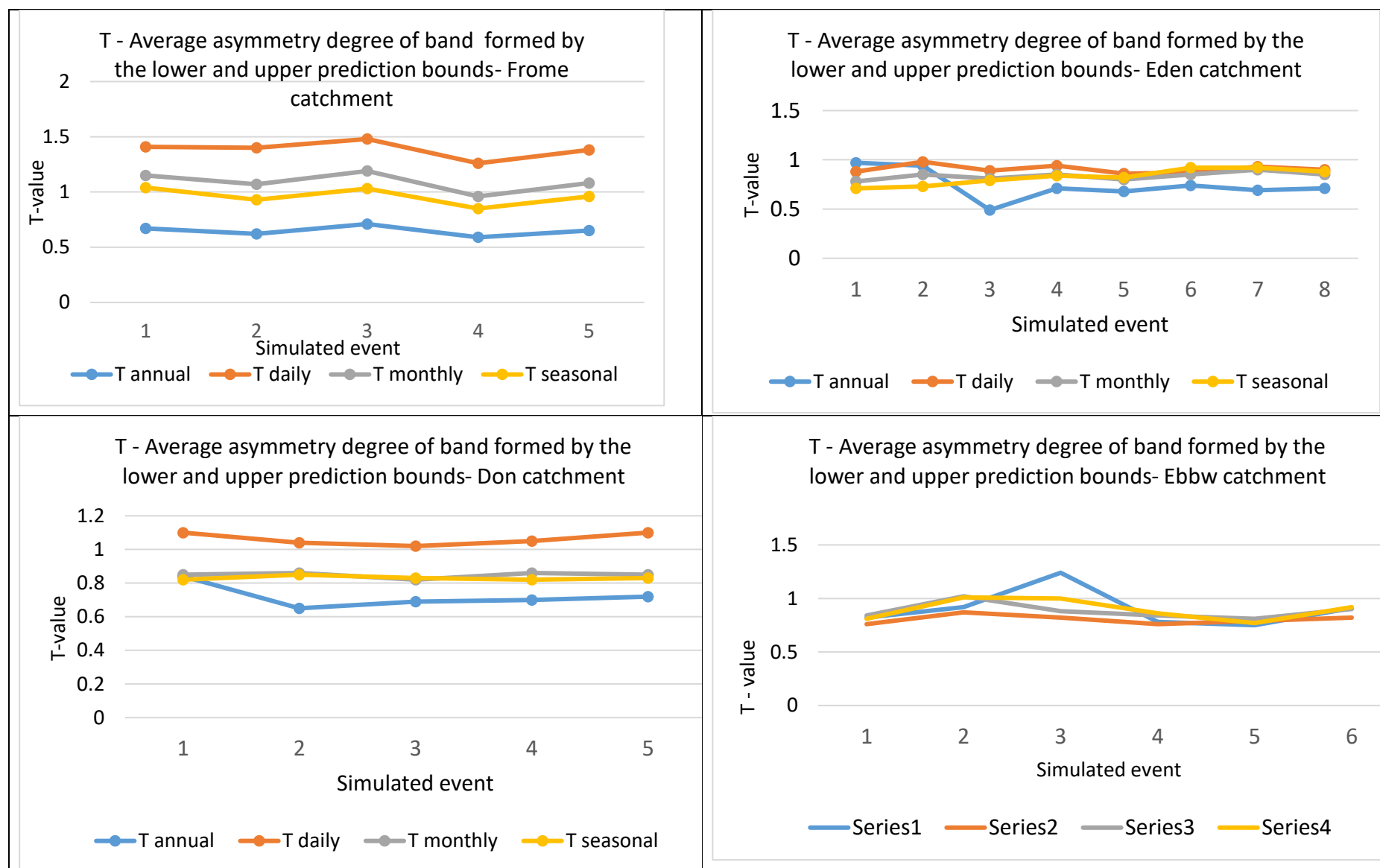
Figure 10 Asymmetric degree T value for the Frome, Eden, Don and Ebbw catchments.
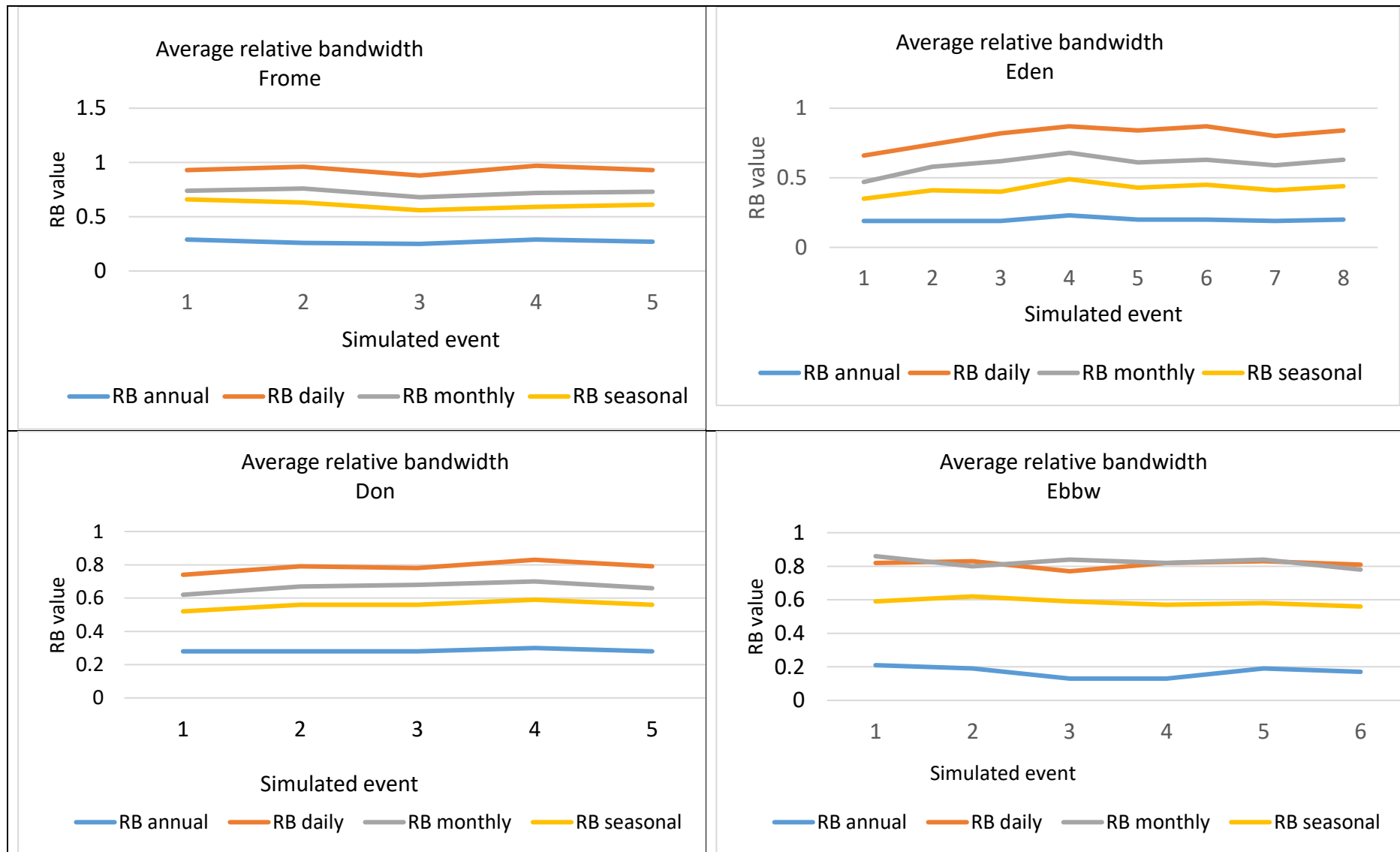
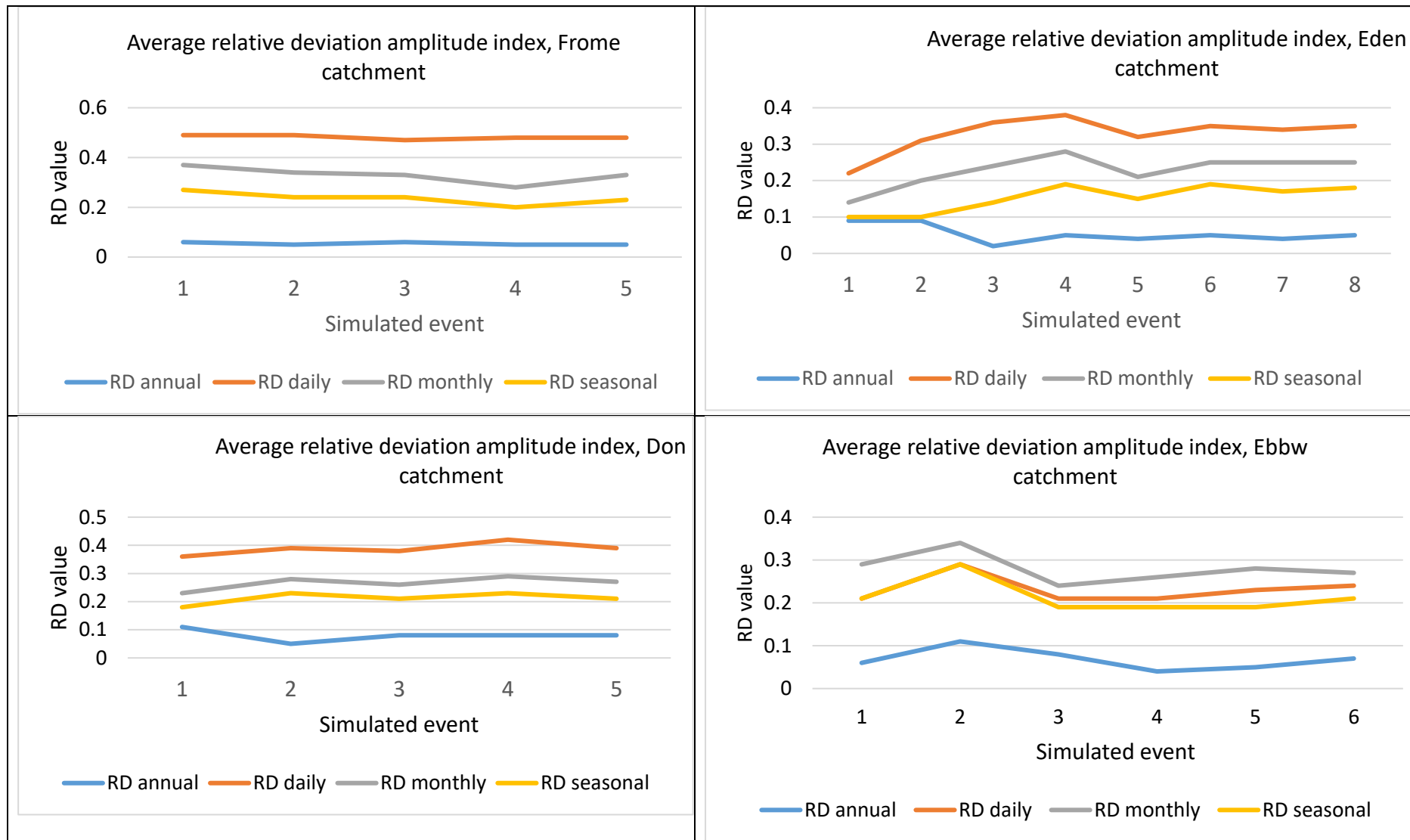Figure 11. Average relative band width, RB value for the Frome, Eden, Don and Ebbw catchments.

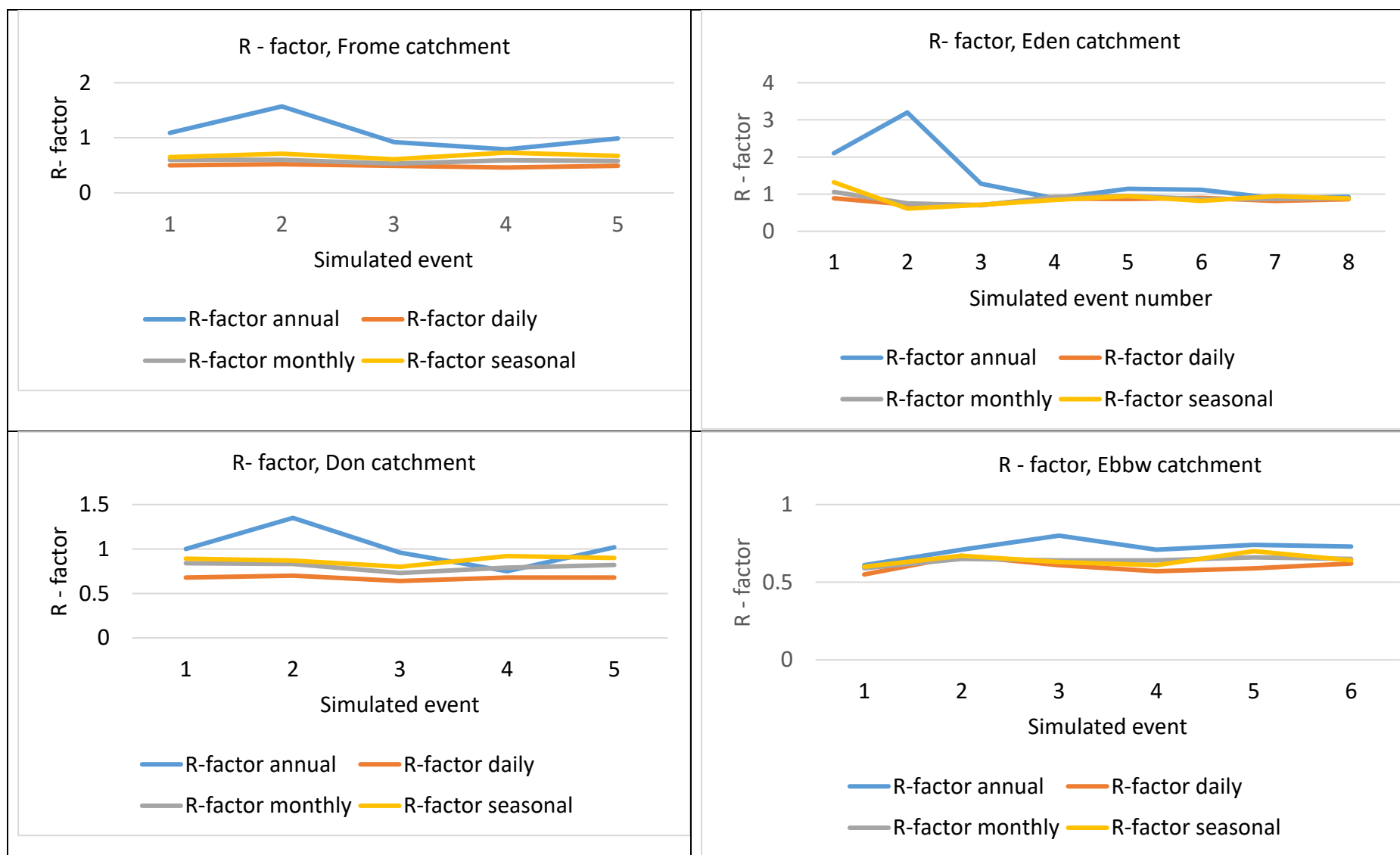Figure 12. Average relative deviation, RD value for the Frome, Eden, Don and Ebbw catchments.

Figure 13 R- factor value for the Frome, Eden, Don and Ebbw catchments.

5. Conclusion

The results of the study using the GLUE methodology were quite important to estimate the uncertainty in the river flow prediction. The analysis is of significant importance, to apply the model parameters for the future climate change scenarios. The application of the GLUE methodology to the Distributed Catchment Scale Model, led to the following conclusions:

> The DiCaSM model is not equally sensitivity to all the model parameters, this sensitivity varied between different catchments. For instance, in the Pang catchment, the river flow was more sensitive to base flow factor in comparison with other catchments.

> The different uncertainty indicators, CR, S, T, B, RB, D, RD and R-factor all gave good values indicating a reasonable low uncertainty level in model prediction.

> The GLUE methodology showed lower uncertainty in predicted river flows when increasing the time scale from daily to monthly to seasonal river flows with the lowest uncertainty associated with annual flows.

In the current study, the reliability of the DiCaSM was assessed when applied to six catchments, the parameters that may cause uncertainty in model output were investigated using a generalized likelihood uncertainty estimation (GLUE) methodology. 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. Finally, the results presented in this paper suggest that DiCaSM river flow or groundwater recharge results could be used by stakeholders and decision makers as a support for planning.

Acknowledgement

50km GB' Web Map Service). Land cover data from Centre for Ecology and Hydrology (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 licensed from NERC Centre for Ecology & Hydrology. Soils data courtesy of Cranfield University (1:250 000 Soilscapes for England and Wales Web Map Service). Hydrogeology data from British Geological Survey (DiGMapGB 1:625 000 scale digital hydrogeological data).

References

AFZAL, M., GAGNON, A. S. & 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.

ASTON, A. 1979. Rainfall interception by eight small trees. *Journal of hydrology,* 42**,** 383-396.

AYNOM T. TEWELDEBRHAN, JOHN F. BURKHART, AND THOMAS V. SCHULER. 2018. Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches. Hydrol. Earth Syst. Sci., 22, 5021–5039, 2018 https://doi.org/10.5194/hess-22-5021-2018

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., & 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., NORTON, L., D. 2011. Development, sensitivity and uncertainty analysis of LASH model. Sci. Agric., 68: 265-274.

CEH. 2014. *CEH digital river network of Great Britain web map service* [Online]. Available: https://data.gov.uk/dataset/3c7ea82e-83e0-45a3-9a3f-8ba653b3211b/ceh-digital-river-network-of-great-britain-web-map-service [Accessed 2014].

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. & 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.

GASH, J. H., LLOYD, C. & 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. *J. Agric. Sci.,* 4**,** 1-24.

GUPTA, H. V., KLING, H., YILMAZ, K. K. & 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. Hydrol. Earth Syst. Sci., 22: 5947–5965

JACKSON, C., WANG, L., PACHOCKA, M., MACKAY, J. & BLOOMFIELD, J. 2016. Reconstruction of multi-decadal groundwater level time-series using a lumped conceptual model. *Hydrological Processes,* 30**,** 3107-3125.

KAN, G., HE, X., DING, L., LI, J., HONG, Y. & 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.

KRAUSE, P., BOYLE, D. P. & 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.

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. & 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. & 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 1990 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.

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

NRFA. 2014. *National River flow Archive* [Online]. Available: http://nrfa.ceh.ac.uk/ [Accessed 2014].

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. & 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. & 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. & 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. & 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. & 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 & 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

VIOLA, F., NOTO, L., CANNAROZZO, M. & 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.

XIE, H., SHEN, Z., CHEN, L., LAI, X., QIU, J., WEI, G., DONG, J., PENG, Y. & 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., WAN, M., WEI, X. & 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. & JENG, Y.-C. 1997. A study on grid based distributed rainfall runoff models. *Water resources management,* 11**,** 83-99.

Annex 1

Containing 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.

Average band width B

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

$$B = \frac{1}{N} \sum_{i=1}^{N} b_i$$

with $b_i = Q_i^u - Q_i^l$

where b_i is the band-width 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.

Average relative band-width 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 band-width of the prediction bounds. This can be done by using a dimensionless index, the average relative band-width of the prediction calculated as:

$$RB = \frac{1}{N} \sum_{i=1}^{N} rb_i$$

$$rb_i = b_i / Q_i$$

where rb_i is the ratio of the band-width of the prediction bounds at time i to the corresponding

observed discharge Q_i .

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:

$$S = \frac{1}{N} \sum_{i=1}^{N} s_i$$

$$s_i = |\quad h_i - 0.5|$$

$$h_i = \frac{Q_i^u - Q_i}{Q_i^u - Q_i^l} = \frac{Q_i^u - Q_i}{b_i}$$

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 band-width, 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 with

respect to the observed flows, is referred to as T, calculated is defined as:

$$T = \frac{1}{N} \sum_{i=1}^{N} t_i$$

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

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.

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:

$$D = \frac{1}{N} \sum_{i=1}^{N} d_i$$

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

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:

$$RD = \frac{1}{N} \sum_{i=1}^{N} rd_i$$

$$rd_i = \frac{\left| \frac{1}{2}(Q_i^u - Q_i^l) - Q_i \right|}{Q_i} = \left| \frac{Q_i^m}{Q_i} - 1 \right|$$

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.

The R – factor

Is calculated as:

$$R - factor = \frac{\bar{d_x}}{\sigma_x}$$

$$d_{\bar{x}} = \frac{1}{n} \sum_{i=1}^{n} (x_u - x_l)$$

Where, σ_x is the standard deviation of the measured stream flow x and $d_{\bar{x}}$ is the average distance between upper and lower boundary ($Q_{0.95}$ and $Q_{0.05}$), n is the number of observations.