Journal of Hydrology (NZ) 58 (2): 81-103 2019

Improving hydrological multi-model prediction by elimination of double counting in the ensemble

Shailesh Kumar Singh,[1] Markus Pahlow,[2] Qingyun Duan[3] and George Griffiths[1]

[1] *National Institute of Water and Atmospheric Research, Christchurch, New Zealand*
[2] *Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand*
[3] *State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University,Nanjing, Jiangsu, China*
Corresponding author: shailesh.singh@niwa.co.nz

Abstract

Hydrological models are a mathematical representation of heterogeneous and non-linear hydrological processes. In the past, a great many simple to complex hydrological models have been developed, but none of these models is superior to the others for all types of practical applications. These hydrological model alternatives have different strengths in representing and capturing complex natural hydrological processes. Yet generally a single hydrological model is used in practice, which may represent certain processes of the catchment well and may be less adequate for others. Moreover, the use of a single hydrological model is restrictive, as the conceptual uncertainty associated with the model structure cannot be identified and quantified. To overcome these issues, the multi-model ensemble approach has recently been applied more commonly to take advantage of the diverse skills of different models. In this study, the multi-model ensemble approach to account for model structure uncertainty is employed to improve hydrological model prediction. While a certain hydrological model may represent particular processes or (extreme) events better than another, two distinct models may represent these processes or events with comparable accuracy. If members of a hydrological ensemble model capture the same process and if they are similar in process representation, then these members will not supply any additional information for prediction and therefore will not improve the accuracy. Hence, by identifying similar models, there is potential to increase the reliability of hydrological ensemble predictions and to reduce computing costs without reducing accuracy. In this study a methodology is presented to identify similar models. The methodology is applied and tested for the Tuapiro catchment in New Zealand. A range of verification statistics are computed to ascertain the validity of the approach. Overall, the multi-model ensemble-based hydrological prediction where non-informative members have been removed is shown to not compromise prediction accuracy. For the case study streamflow prediction an increased flatness of the rank histogram, insignificant changes in the continuous rank probability score, and improved accuracy in terms of Nash-Sutcliffe

coefficient, Kling-Gupta efficiency and Root Mean Square Error were found, at lower computing costs.

Keywords

multi-model ensemble, data depth, hydrological prediction

Introduction

Various simple to complex hydrological models have been developed over time. A general practice for hydrological prediction is to employ a single hydrological model (Duan *et al.*, 2007). It is well known that hydrological modelling accuracy is challenged by various types of uncertainty; for example, measurement uncertainty, input uncertainty, model structure uncertainty and parameterisation uncertainty, amongst others (see, for instance, Liu and Gupta, 2007; McMillan *et al.*, 2011; Tyralla and Schumann, 2016; Zhang and Zhao, 2012). It remains difficult to handle the different uncertainties simultaneously. A practical way forward is to use multi-model ensemble hydrological prediction to sample the respective share of the total uncertainty (Bastola *et al.*, 2011; Duan *et al.*, 2007; Georgakakos *et al.*, 2004; Roudier *et al.*, 2016; Velázquez *et al.*, 2011, 2013). As demonstrated with atmospheric ensemble forecasts, the estimates of predictive uncertainty provide forecasters and users with objective guidance on the level of confidence that they may place in the forecasts. The end users can decide to take action based on their risk tolerance (Demargne *et al.*, 2014).

However, the computational cost is frequently a limiting factor in both uncertainty-based and optimisation-based calibration of computationally intensive environmental simulation models such as hydrological models (Razavi *et al.*, 2010). The situation is aggravated when moving from a single model to an ensemble of models. To this end, model pre-emption has been suggested as a sensible solution, whereby a simulation model is terminated early if it is recognised through intermediate simulation model results that a given solution (model parameter set) is so poor that it will not benefit the search strategy, thereby reducing unnecessary simulation time steps (Razavi *et al.*, 2010; Sharii *et al.*, 2015).

A further challenge is to employ appropriate mathematical or statistical methods, or both, to post-process the ensemble information for optimum results. For example, Arsenault *et al.* (2015) concluded in their comparative analysis of nine multi-model averaging approaches in hydrological continuous streamflow simulation that multi-model averaging increases prediction skill better than any single model. In more recent work Arsenault and Brissette (2016) postulated that the averaging aspect is quite well understood and promising for use on a single basin; for example, to estimate streamflow during calibration using multi-model averaging and then applying that to validation is common and has been shown to be efficient. Yet the authors found that multi-model averaging techniques may be less well suited in regionalisation applications and, furthermore, models selected for such a purpose must undergo a careful selection process to be as robust as possible at a given study site. Arsenault *et al.* (2015) found that Bayesian Model Averaging (BMA) showed good performance but is not as robust as others. They concluded that the unconstrained Granger-Ramanathan averaging variant (Granger and Ramanathan, 1984) is comparable to BMA in terms of performance, but excels in ease of implementation and robustness, following the findings of Diks and Vrugt (2010). Various adjustments to hydrological ensemble modelling using BMA have been proposed, such as integration of BMA and data assimilation (Parrish *et al.*, 2012), BMA using particle filtering and Gaussian mixture

modelling (Rings *et al.*, 2012), BMA-based Bayesian Neural Networks combined with genetic algorithms, a BMA-based ensemble multi-wavelet Volterra nonlinear model by Rathinasamy *et al.* (2013), an approach to consider BMA model-weighting uncertainty (Schöninger *et al.*, 2015) and application of BMA to post-process raw grand ensemble runoff forecasts statistically as exemplified by Qu *et al.* (2017).

It must also be acknowledged that alternative hydrological models have different strengths in representing and capturing complex hydrological processes. A single hydrological model may represent a certain process within the catchment exceptionally well, yet at the same time is not suited to capture others. Furthermore, as noted above, using a single model can be restrictive, as the conceptual uncertainty associated with the model structure cannot be captured. To overcome these challenges, one can take advantage of the diverse skills of different models (Duan *et al.*, 2007; Perrin *et al* 2003). It is acknowledged that a single model is associated with uncertainty but so is the combination of different models (Dong *et al.*, 2013; Velázquez *et al.*, 2013). While a particular model can represent certain processes or events with higher accuracy than another, at the same time two models may represent the process equally well. If members of an ensemble represent the same process comparably and if those members are similar in process representation, then these will not add information for accurate prediction. Hence, by elimination of similar models, ensemble prediction may be more robust, reliable and cost-effective, without any decrease in performance of the ensemble. The research question pursued here is: how can similar models be detected and removed from an ensemble to increase reliability and to eliminate double counting, thereby increasing cost-effectiveness? The research objectives to answer this question are (i) to develop a methodology that allows for identification of similar models, and (ii) to test and verify the methodology for a case study catchment in New Zealand using a ten-member ensemble.

We show that by removing double counting in the ensemble, performance of multi-model-based hydrological ensemble prediction is not decreased and can even be improved in terms of accuracy and reliability of stream flow prediction in a catchment, while reducing computational cost.

Study area, data and model ensemble

Study area and data

The proposed methodology of multi-model-based hydrological prediction was tested in the Tuapiro (Station No. 13310) catchment located in the Bay of Plenty region (Fig. 1), in the North Island of New Zealand. Tables 1 and 2 provide details of catchment flow statistics and their properties. The mean annual precipitation in the catchment is 2,096 mm. The mean annual flow is 1.2 m^3/s, whereas the mean annual low flow is 0.2 m^3/s. Elevation of the catchment ranges from 39 to 738 m above mean sea level and the main land use is agriculture (dairy and sheep farming). Hourly runoff and precipitation data from the period 1990 to 1997 were obtained from the National Institute of Water and Atmospheric Research and Bay of Plenty Regional Council. The catchment physical and hydrological properties were derived from a variety of sources including a digital river network (Snelder and Biggs, 2002), 30 m Digital Elevation Model, and land cover and soils databases (Newsome *et al.*, 2000).

Model Ensemble

HYDROlogical Modelling Assessment and Development (HydroMAD) (Andrews *et al.*, 2011) is a modelling framework with a range

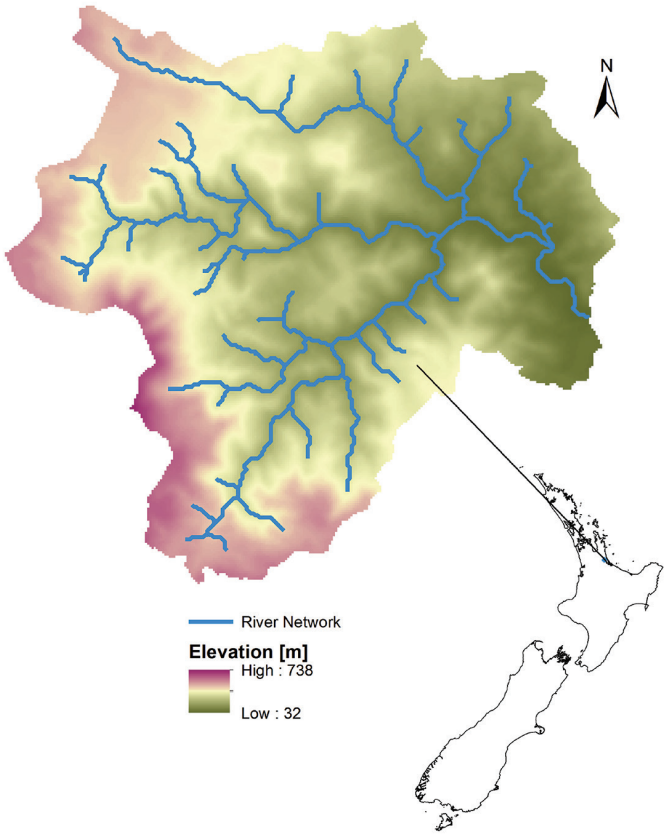

Figure 1 – Study area Tuapiro catchment (Station No. 13310), located in the Bay of Plenty Region, in the North Island of New Zealand.

Table 1 – Catchment characteristics of the study area

Station Nr.	Station reach Nr.	Area (km^2)	Length of stream reach (m)	Width of stream reach (m)	Average ln (a/tan(β))	Minimum elevation of stream reach (m amsl)	Maximum elevation of stream reach (m amsl)	Average slope of stream reach (degree)
13310	4000199	37.9	1641	9.3	6.8	39.1	56.1	0.010

Table 2 – Flow statistics in (mm/d) for the study area

Station Nr.	Station reach Nr.	Minimum discharge (Q_{min})	Median discharge (Q_{median})	Mean discharge (Q_{mean})	Maximum discharge (Q_{max})	5th percentile discharge (Q5)	95th percentile discharge (Q95)
13310	4000199	2.73E-05	6.39E-05	4.47E-04	4.32E-02	4.10E-05	1.66E-03

of soil moisture accounting modules and routing functions. The R implementation of HydroMAD was utilised in this study (Andrews and Guillaume, 2018). The class of hydrological models considered are spatially aggregated conceptual models. There are two main components, a Soil Moisture Accounting (SMA) module and a routing or unit hydrograph module. The SMA module converts rainfall and temperature into effective rainfall, which is converted into streamflow by the routing module.

The nine SMA models presented below were chosen for the analysis, along with the exponential components transfer function as the routing model (Jakeman *et al.*, 1990). In addition a physically-based model was selected (TopNet), so that in total ten models were used:

M1. CMD (Catchment Moisture Deficit)
M2. CWI (Catchment Wetness Index)
M3. Sacramento (Soil Moisture Accounting model)
M4. AWBM (Australian Water Balance Model)
M5. Bucket (Single-bucket Soil Moisture Accounting models)
M6. BDM (Data-Based Mechanistic modelling)
M7. Snow (Simple degree day factor)
M8. Intensity (Runoff as rainfall to a power)
M9. Runoff ratio
M10. TopNet

A description of each model, together with the model parameters, is provided in the Appendix.

Methodology

In this section, we first define the mathematical framework of the double counting of ensemble members, followed by a brief description of the data depth function. It is shown how this approach can be used to identify similar models. Lastly, the approaches used to calibrate and validate the hydrological models, plus to verify the results, are presented.

Double counting in the multi-model ensemble approach for prediction

Let us consider a hydrological flow prediction of N hydrological models for time step t. The weighted prediction P_t from N models is given by the equation:

$$P_t = \sum_{i=1}^{N} M_{i,t} w_i \quad (1)$$

where $M_{i,t}$ is the hydrological prediction from model i for time step t and w_i is the weight. The observed flow at time t is Q_t, and the error in the prediction E_t can be defined as:

$$E_t = P_t - Q_t \quad (2)$$

Let us assume $N = 5$ hydrological models, whereby the prediction from the models for time t is 3, 3, 4, 5, 5 m^3/s, respectively. Let us further assume equal weights for all of these models. So, P_t, the ensemble prediction from $N = 5$ models, will be 4 m^3/s. The models 1 and 2 provide similar or exactly the same prediction, as do models 4 and 5. Hence, for this example, if we remove models 2 and 5 and recalculate P_t, we still obtain P_t equal to 4 m^3/s. Similarly, in ensemble prediction, if there are models that yield similar predictions then those do not add any information towards the ensemble prediction. Hence, their elimination will not negatively affect the results, yet at the same time will speed up the computation.

Data depth function

Data depth is a quantitative measurement of how central (or deep) a point is with respect to a data set or a distribution that can be utilised for central outward ordering of multivariate data points and provides a means to quantify the many complex multivariate features of the underlying multivariate distribution

(Li *et al.*, 2012; Liu *et al.*, 1999). A depth function was first introduced by Tukey (1975) to identify the centre (a generalised median) of a multivariate dataset. Several generalisations of this concept have been defined (Barnett, 1976; Liu *et al.*, 2006; Rousseeuw and Struyf, 1998; Zuo and Serfling, 2000). Data depth functions have been applied in several fields of non-parametric multivariate analysis (Cheng *et al.*, 2000; Hamurkaroglu *et al.*, 2006; Liu, 1995; Liu and Singh, 1993; Messaoud *et al.*, 2004; Serfling, 2002; Stoumbos *et al.*, 2001). Application of the data depth function is relatively new in the field of water resources. It has been used in the field of regional flood frequency analysis (Chebana and Ouarda, 2008; Wazneh *et al.*, 2013a, 2013b), depth-based multivariate descriptive statistics in hydrology (Chebana and Ouarda, 2011), regionalisation of hydrological model parameters (Bárdossy and Singh, 2011), robust estimation of hydrological model parameters (Bárdossy and Singh, 2008), defining predictive uncertainty of a model (Singh *et al.*, 2013), and in selection of critical events for model calibration (Singh and Bárdossy, 2012). Several types of data depth functions have been developed, e.g., half-space, L1 and Mahalanobis depth functions. The half-space data depth function was used in this study because it satisfies all the properties of the data depth function and it is robust in calculation (Dutta and Ghosh, 2012).

Formally, the half-space depth of a point p with respect to the finite set X in the d-dimensional space is defined as the minimum number of points of the set X lying on one side of a hyperplane through the point p. The minimum is calculated over all possible hyperplanes. The half-space depth of the point p with respect to set X is:

Here $\langle x,y \rangle$ is the scalar product of the d dimensional vectors, and n_h is an arbitrary unit vector in the d dimensional space representing the normal vector of a selected hyperplane. If the point p is outside the convex hull of X then its depth is 0. The convex hull of a set of points, S, is the smallest convex set (e.g., a convex polygon in two dimensions) which encloses S. More specifically, the convex hull of a set of M points in two dimensions is the smallest polygon area that encloses all M points. Mathematically, the convex hull of a set of M points in d dimensions is the intersection of all convex sets containing all M points. Conceptually, data points that are close to the center of the dataset have high depths, while those that are near the boundary of the dataset have low depths. An example of a convex hull is given in Figure 2. Points on and near the boundary have low depth while central (i.e., deep) points have high depth. One advantage of this depth function is that it is invariant to affine transformations

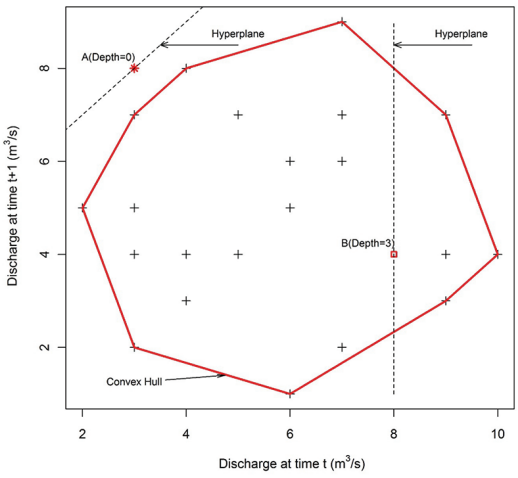

Figure 2 – Example of a convex hull using the hydrological variable, discharge.

$$D_X(p) = \min_{n_h}\left(\min(|\{x \in X | \langle n_h, x-p\rangle > 0\}|), (|\{x \in X | \langle n_h, x-p\rangle < 0\}|)\right) \quad (3)$$

of the space. This means that the different ranges of the variables have no influence on the calculated depth. In this study, we normalised the data depth between 0 and 1 by dividing the depth with half the number of total points in the convex hull.

Model discharge can be used to define model similarity. We denote the selected time series (discharge) as $\mathbf{X_d}$. Considering d consecutive time steps, the d-dimensional set is defined by the following equation (Singh and Bardossy, 2012):

$$\mathbf{X_d} = \{(X(t-d+1), X(t-d+2), ..., X(t)) \quad t = d, ..., T\}, (4)$$

where T is the total number of observation time steps. Higher values of d result in a higher dimensional problem, which in turn is computationally more expensive. However, lower values may not capture the response dynamics and the memory of the catchment. Several studies (Singh and Bardossy, 2012; Singh *et al.*, 2013; Tiejun *et al.*, 2018) suggest setting $d = 4$. After testing different values of d and analysing the autocorrelation, $d = 4$ has been identified to reflect the dynamics and memory of the catchment, thereby suggesting any autocorrelation beyond is negligible. Hence, considering the dynamics and the memory of the catchment, it is a four-dimensional problem. From the discharge, the boundary of the convex hull from the d-dimensional dataset is prepared for the model. The following section describes how the data depth function can be used to define the model similarity-based model discharge.

Application of data depth function to define a similar model

If a model is similar to another, then it does not bring any additional information to the ensemble prediction. An example is given in Figure 3, where a ten-model (M1 to M10) ensemble is presented. From Figure 3 we can see that models M4, M6, M8, and M9 have different flow magnitudes and dynamics, whereas models M7 and M5 have very similar magnitudes and dynamics, as do models M1, M2 and M3. Our hypothesis is that keeping both M7 and M5 does not contribute additional information to the ensemble prediction. The same reasoning applies for M1, M2 and M3. Hence, out of M7 and M5 only one model is enough for representing both models, similarly for M1, M2 and M3. Figure 4 shows the model ensemble hydrographs, observed data, the mean of the ensemble and the mean after eliminating the non-informative members of the ensemble. We can see that after removing the non-informative members the mean of the ensemble changes only slightly, as those members do not contribute additional information.

To find similar models the data depth

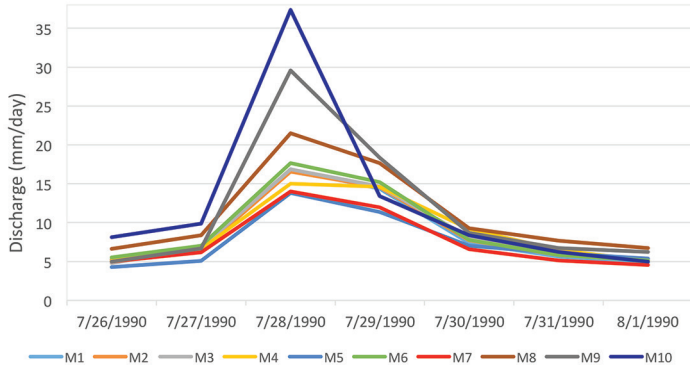

Figure 3 – Example of ensemble of models. Here 10 models are used.

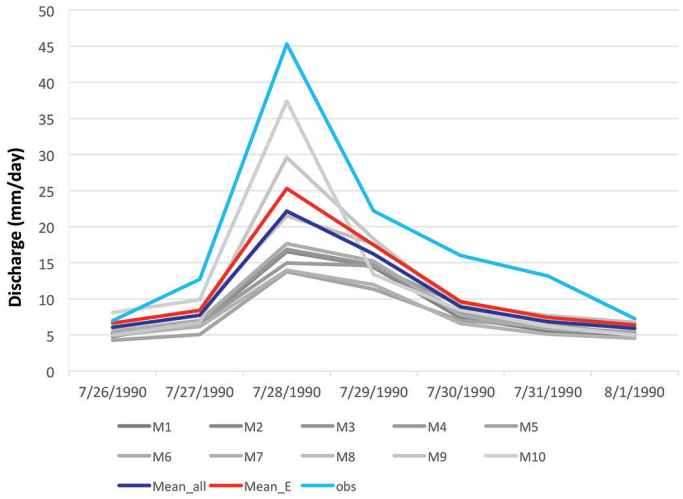

Figure 4 – Ensemble of model hydrographs (M1 to M10), along with the observed discharge (obs), the mean of the ensemble (Mean_all), and the mean discharge after eliminating the non-informative members from the ensemble (Mean_E).

function is used herein, where the depth of one model response was calculated against the other. If the convex hull of a model discharge is within the convex hull of another model discharge, then one model is a subset of the other. A two-dimensional example of a convex hull of two hydrological model outputs is given in Figure 5. It can be seen that the convex hull of model M2 discharge lies within the convex hull of the model M3 discharge. This suggests that M2 and M3 are very similar in nature.

The concept can be used to eliminate non-informative models from the ensemble. If a model A is similar to model B then the convex hull of responses from models A and B will lie within each other, whereas if they are different one convex hull will not enclose the other convex hull (Fig. 6).

Once the non-informative ensemble members have been eliminated, the ensemble needs to be combined using an adequate strategy. To illustrate the methodology, the Simple Model Average (SMA) technique

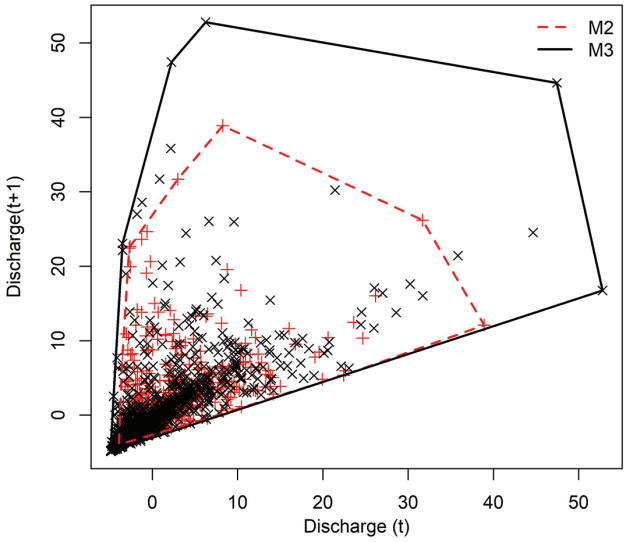

Figure 5 – Example of the convex hull of two models, M2 and M3.

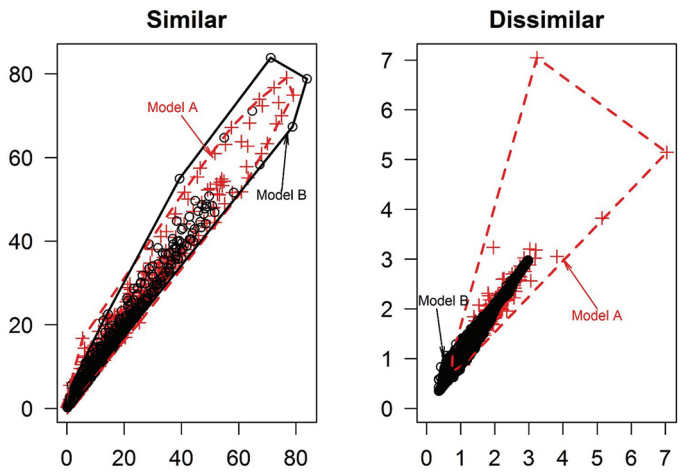

Figure 6 – Example of similar (left) and dissimilar (right) model response based on the convex hull.

by Georgakakos *et al.* (2004) was used as a benchmark for evaluating the benefit of eliminating non-informative members from the ensemble. SMA is expressed by the following equation:

$$(Q_{SMA})_t = \overline{Q}_{obs} + \sum_{i=1}^{N} \frac{(\overline{Q}_{sim})_{i,t} - (\overline{Q}_{sim})_i}{N} \quad (5)$$

where $(Q_{SMA})_t$ is the multi-model streamflow simulation obtained through SMA at time t, $(Q_{sim})_{i,t}$ is the i^{th} model streamflow simulation at time t, $(\overline{Q}_{sim})_i$ is the time average of the i^{th} model streamflow simulation, $(\overline{Q}_{obs})$ is the corresponding observed average streamflow, and N is the number of models under consideration. Note that SMA has equal model weights and removes the bias by replacing the simulation mean with the observed mean.

Q_{SMA} was calculated by considering all ensemble members (termed as Q_{SMA_all}) and Q_{SMA} is also calculated with eliminated non-informative members from the ensemble (termed Q_{SMA_elim}). A comparison was carried out utilising Q_{SMA_all}, Q_{SMA_elim} and the observed discharge.

Model calibration and validation

All models except TopNet were calibrated using the fitByOptim function provided in the HydroMAD R software package using the Nash-Sutcliffe efficiency as objective function (Nash and Sutcliffe, 1970). This objective function gives the proportion of the variance of the data explained by the model (Pechlivanidis *et al.*, 2011):

$$NSE = 1 - \frac{\sum_{t=1}^{n} \left(Q_{obs}(t) - Q_{sim}(t)\right)^2}{\sum_{t=1}^{n} \left(Q_{obs}(t) - \overline{Q_{obs}(t)}\right)^2} \quad (6)$$

where $Q_{obs}(t)$ and $Q_{sim}(t)$ are the observed and simulated discharge at time step t, respectively, and $\overline{Q_{obs}(t)}$ is the mean observed discharge over the entire simulation period of length n. NSE ranges from -∞ to 1, whereby 1 indicates perfect agreement between simulations and observations and NSE = 0 is interpreted that the simulation results are as accurate as the mean of the observed data. TopNet was calibrated using the Robust Parameter Estimation (ROPE) algorithm (Bardossy and Singh, 2008) using log-Nash-Sutcliffe efficiency as objective function. The parameters of the models which require calibration, along with their description, are

provided in the Appendix. The models were calibrated at daily time step for the period 1990–1993, with a validation period of 1994–1997.

The second performance indicator used here was the Kling–Gupta efficiency (KGE) (Gupta *et al.*, 2009). KGE is based on a decomposition of NSE into its constitutive components, i.e., correlation, variability bias and mean bias:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (7)$$

whereby r represents the linear correlation between simulations and observations, α is a measure of the flow variability error, and β is a bias term. Note that, analogous to NSE, KGE = 1 indicates perfect agreement between simulations and observations.

The root mean squared error (RMSE) is the third performance indicator used here:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2} \quad (8)$$

An RMSE value of zero points to a perfect agreement between simulation and observation, whereas the agreement decreases for increasing RMSE values.

Verification

Two verification measures were used in this study: the rank histogram and the continuous rank probability score. Rank histograms are a tool for evaluating ensemble forecasts. They are useful for determining the reliability of ensemble forecasts and for diagnosing errors in its mean and spread. Rank histograms are generated by repeatedly tallying the rank of the verification (usually an observation) relative to values from an ensemble sorted from lowest to highest (Hamill, 2001). Let $\mathbf{R} = (r_1, \ldots, r_{n+1})$ represent a rank histogram with n+1 possible ranks. The population of a rank histogram element is determined from (Hamill, 2001):

$$r_j = \overline{P\left(x_{j-1} \le V < x_j\right)} \quad (9)$$

where x_i are the i sorted ensemble members, V is the true state (described with a probability distribution) and $\overline{(\cdot)}$ is the average over a large sample of statistically independent points. Hence the population of rank j is the fraction of times when the true state, when pooled with the sorted ensemble, is between sorted members j-1 and j.

The continuous ranked probability score (CRPS) is a much-used measure of performance for probabilistic forecasts of a scalar observation (see, for example, Zamo and Naveau, 2018). It is a quadratic measure of the difference between the forecast cumulative distribution function (CDF) and the empirical CDF of the observation. Following Hersbach (2000):

$$CRPS = CRPS(P, x_a) = \int_{-\infty}^{\infty}\left[P(x) - P_a(x)\right]^2 dx \quad (10)$$

where $P(x) = \int_{-\infty}^{x}\rho(y)dy$ and $P_a(x) = H(x - x_a)$, with the parameters of interest being x (here the flow), the forecast ϱ, the observation x_a and the Heaviside function H. CRPS values can range from zero to infinity, whereby zero is achieved for $P = P_a$, i.e., in the case of a perfect deterministic forecast.

Results and discussion

An ensemble of model discharges was generated using the ten models. The ensemble members, along with the observed discharge and the mean of the ensemble, are shown in Figure 7 for an example time period. The spread of the ensemble members indicates the uncertainty. It can be seen from Figure 7 that the spread in the hydrograph varies from one event to another. Some of the ensemble members are similar in both magnitude and dynamics. These members do not contribute additional information to the ensemble discharge prediction and hence can be eliminated; the M2, M3 and M7 models were eliminated using the method described above. More specifically, M7 was eliminated

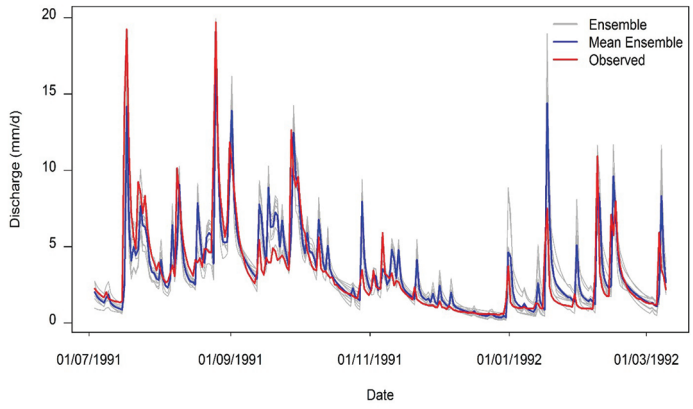

Figure 7 – Ensemble discharge based on 10 hydrological models, along with the mean and the observed discharge for an example time period.

since it is similar to M5, whereas M2 and M3 are similar to M1. It must be noted that further quantification of uncertainty and estimation of confidence intervals of the developed multi-model ensemble system is desirable (e.g., Abbaspour *et al.*, 2007; Aqil *et al.*, 2007; Noori *et al.*, 2010). Such an assessment adds further information towards the selection of the optimal ensemble and will be carried out in future work.

The quality of the ensembles was evaluated using several measures. The RMSE with respect to observed data for the mean of all ensemble members and the RMSE for the ensemble where non-informative members have been removed are 2.97 and 2.83 mm/day, respectively (Table 3). Thus, model accuracy improved by eliminating the non-informative members from the ensemble. Figure 8 shows the rank histogram for the calibration time period, using all ensemble members (top) and after eliminating the non-informative members (bottom). The rank histogram is more strongly skewed for the all-member case compared to the case with the non-informative members removed.

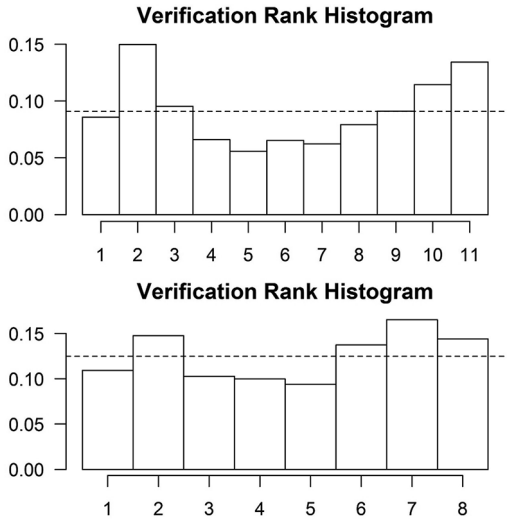

Figure 8 – Rank histogram using all ensemble members (top) and using the ensemble after non-informative models have been eliminated (bottom), for the calibration time period.

Table 3 – Performance statistics during calibration and validation

	Calibration		Validation	
	All	**Eliminated**	**All**	**Eliminated**
NSE	0.61	0.64	0.57	0.61
KGE	0.67	0.70	0.57	0.59
RMSE (mm/d)	2.97	2.83	4.93	4.70

The increased flatness of the rank histogram after removing the non-informative members suggests that the ensemble spread is satisfactorily covering the observed data; i.e., according to Talagrand *et al.* (1997) and Hersbach (2000) the reliability has increased. The results are similar for the validation time period (see Fig. 9). Hence, removing the non-informative ensemble members improved the ensemble performance, or at least did not decrease their performance.

The continuous rank probability score (CRPS) for the case where all ensemble members have been kept (Q_{SMA_all}) and for the case after elimination of non-informative ensemble members (Q_{SMA_elim}) are shown in Figure 10. The CRPS values show that low flows are, in general, better captured than peak flows for this particular example period. The difference between the CRPS of the two cases is rather small, with an absolute deviation of up to 1.5. This again confirms that the elimination of non-informative ensemble members did not negatively impact the ensemble performance.

The flow duration curves (FDCs) for Q_{SMA_all} and Q_{SMA_elim}, along with the observed FDC, are shown in Figure 11. The FDCs for both cases follow each other closely for all flow ranges. This suggests that a flow simulation result that deviates very little can be obtained with fewer ensemble members, i.e., a computationally less expensive solution. Q_{SMA_all} and Q_{SMA_elim} hydrographs are plotted along with the observed discharge in Figure 12 for an example time period. The Nash-Sutcliffe efficiency NSE for Q_{MSA_all} and Q_{SMA_elim} are 0.61 and 0.64, respectively. The KGEs are 0.67 and 0.70 and the RMSEs are 2.97 and 2.83 mm/day, respectively, for Q_{MSA_all} and Q_{SMA_elim}. The results were confirmed for the validation time period (see Table 3). After elimination of non-informative ensemble members performance did not worsen; in fact, it improved slightly when compared

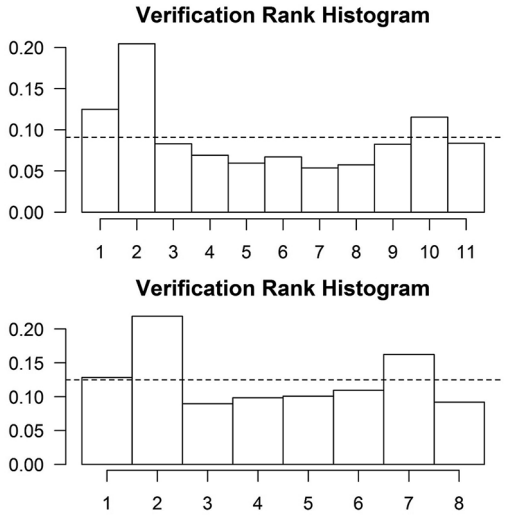

Figure 9 – Rank histogram using all ensemble members (top) and using the ensemble after non-informative models have been eliminated (bottom), for the validation time period.

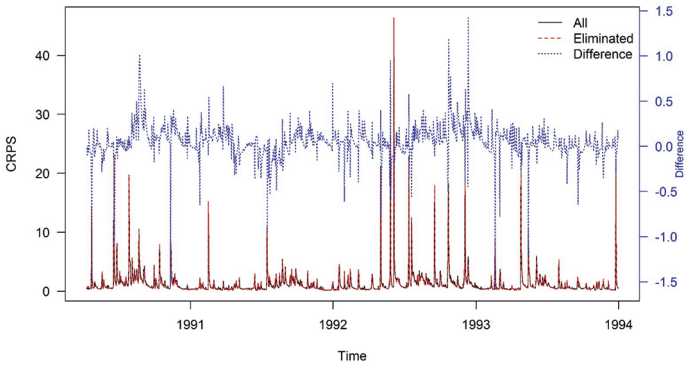

Figure 10 – CRPS using all ensemble members and CRPS using ensemble after non-informative models have been eliminated (for an example time period). The secondary axis (on the right) shows the difference between the CRPS using all ensemble members and the CRPS using the ensemble where non-informative models have been removed.

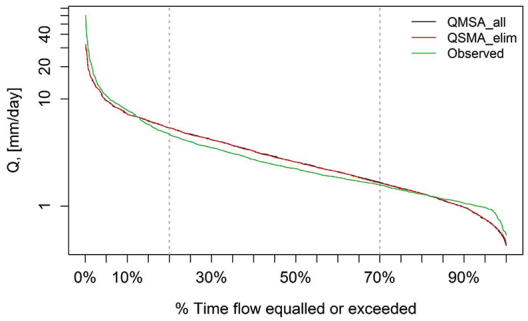

Figure 11 – FDC for all ensemble members (Q_{SMA_all}) and the case where non-informative members have been eliminated from the ensemble (Q_{SMA_elim}), along with that of the observed discharge.

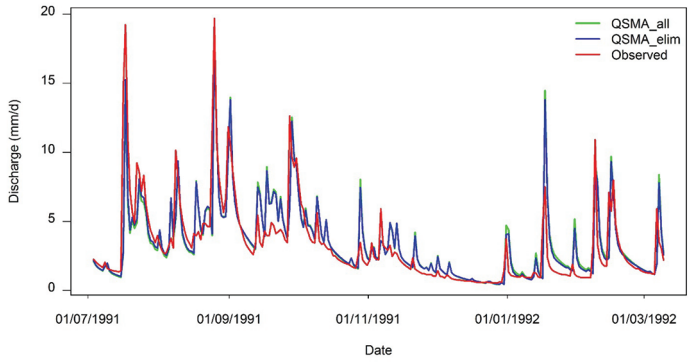

Figure 12 – Hydrographs for all ensemble members (Q_{SMA_all}) and the case where non-informative members have been eliminated from the ensemble (Q_{SMA_elim}), along with the observed discharge.

to the case when all ensemble members are considered. This suggests that the presence of non-informative members in the ensemble does not result in an advantage with respect to accuracy, whereas it rather increases the computation cost. It can also potentially lead to double counting in weight calculation for ensemble prediction.

Summary and conclusions

A particular model may represent certain hydrological processes or (extreme) events more adequately than another, yet at the same time two models may represent these processes or events equally well. Hydrological model ensembles may comprise model types with similarities, which may in turn lead to the situation of no information gain for prediction and so will not improve accuracy. Therefore, by identifying such similar models, there is potential to increase the reliability of hydrological ensemble predictions and to reduce computing costs without compromising accuracy. We presented a method based on the data depth function to eliminate non-informative members from the ensemble. In the practical application to a catchment in New Zealand it was shown that after eliminating non-informative members, the accuracy of the ensemble prediction of discharge was not diminished. On the contrary, it was improved in terms of the Nash-Sutcliffe coefficient, the Kling-Gupta efficiency and the Root Mean Square Error – for both the calibration and validation time periods. Various verification statistics were shown to further support the proposed elimination strategy. The flatness of the rank histogram increased once non-informative ensemble members had been removed, which points to an increase in reliability. Modellers may use the gained knowledge on similar ensemble members to reduce computation cost or to utilise freed-up computation time for additional complementary model runs that may lead to a true information gain

that would not otherwise be feasible. Razavi *et al.* (2010) or Shafii *et al.* (2015) show that computational savings may be substantial when applying a model pre-emption strategy and hence savings can be expected for a strategy where similar models are eliminated from an ensemble.

While one single case study was the focus of this paper to introduce the concept and to show a first application, the concept presented is not restricted to a certain catchment type. It should, furthermore, be noted that the ensemble members can be obtained from any type of model (e.g., rainfall-runoff model, data driven model, etc.). However, the selection of adequate models needs to be based on the right criteria for a given task (Höge *et al.*, 2018, 2019).

One limitation of the current study is that conceptual hydrological models prevailed in this analysis. Different combinations of model types may improve the ensemble-based prediction. This is left for future work, together with defining the weights based on the data depth function and comparing results with outcomes based on other available weighting strategies. Lastly, we note that quantification of uncertainty is required for any kind of prediction. In this paper, our purpose was to eliminate non-informative ensemble members. Hence, we leave further quantification of uncertainty for future work.

Acknowledgments

The authors wish to thank the editor and two anonymous reviewers for their constructive comments, which helped us to improve the paper.

References

Abbaspour, K.C.; Yang, J.; Maximov, I.; Siber, R.; Bogner, K.; Mieleitner, J.; Zobrist, J.; Srinivasan, R. 2007: Modeling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology 333*: 413-430.

Andrews, F.; Croke, B.; Jakeman, A. 2011: An open software environment for hydrological model assessment and development. *Environmental Modelling & Software 26(10)*: 1171-1185.

Andrews, F.; Guillaume, J. 2018: hydromad: Hydrological Model Assessment and Development. Available at: http://hydromad.catchment.org, R package version 0.9-26.

Aqil, M.; Kita, I.; Yano, A.; Nishiyama, S. 2007. Analysis and prediction of flow from local source in a river basin using a neuro-fuzzy modeling tool. *Journal of Environmental Management 85*: 215-223.

Arsenault, R.; Brissette, F. 2016: Multi-model averaging for continuous streamflow prediction in ungauged basins. *Hydrological Sciences Journal 61(13)*: 2443-2454. DOI: 10.1080/02626667.2015.1117088

Arsenault, R.; Gatien, P.; Renaud, B.; Brissette, F.; Martel, J.-L. 2015: A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology 529 (3)*: 754–767. DOI: 10.1016/j.jhydrol.2015.09.001

Bandaragoda, C.; Tarboton, D.G.; Woods, R. 2004: Application of TOPNET in the distributed model intercomparison project. *Journal of Hydrology 298(1)*: 178-201.

Bárdossy, A.; Singh, S.K. 2008: Robust estimation of hydrological model parameters. *Hydrology and Earth System Sciences 12*: 1273-1283.

Bardossy, A.; Singh, S.K. 2011: Regionalization of hydrological model parameters using data depth. *Hydrology Research 42(5)*: 356-371.

Barnett, V. 1976: The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society Series A 139(3)*: 318-354.

Bastola, S.; Murphy, C.; Sweeney, J. 2011: The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. *Advances in Water Resources 34*: 562–576.

Beven, K.; Kirkby, M. 1979: A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal 24(1)*: 43-69.

Boughton, W. 2004: The Australian water balance model. *Environmental Modelling & Software 19(10)*: 943-956.

Chebana, F.; Ouarda, T. 2008. Depth and homogeneity in regional flood frequency analysis. *Water Resources Research 44(11)*: W11422.

Chebana, F.; Ouarda, T.B.M.J. 2011. Multivariate extreme value identification using depth functions. *Environmetrics 22(3)*: 441-455.

Cheng, A.Y.; Liu, R.Y.; Luxhoj, J.T. 2000: Monitoring multivariate aviation safety data by data depth: control charts and threshold systems. *IIE Transactions 32(9)*: 861-872.

Clark, M.P.; Rupp, D.E.; Woods, R.A.; Zheng, X.; Ibbitt, R.P.; Slater, A.G.; Schmidt, J.; Uddstrom, M.J. 2008: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources 31(10)*: 1309-1324.

Croke, B.F.; Jakeman, A.J. 2004: A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environmental Modelling & Software 19(1)*: 1-5.

Demargne, J.; Wu, L.; Regonda, S.K.; Brown, J.D.; Lee, H.; He, M.; Seo, D.; Hartman, R.; Herr, H.D.; Fresch, M.; Schaake, J.; Zhu, Y. 2014: The science of NOAA's operational hydrologic ensemble forecast service. *Bulletin of the American Meteorological Society 95*: 79–98. https://doi.org/10.1175/BAMS-D-12-00081.1

Diks, C.G.H.; Vrugt, J.A. 2010: Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment 24(6)*: 809–820. DOI: 10.1007/s00477-010-0378-z

Dong, L.; Xiong, L.; Yu, K.-X. 2013: Uncertainty analysis of multiple hydrologic models using the Bayesian Model Averaging method. *Journal of Applied Mathematics 2013(2)*: 1-11.

Duan, Q.; Ajami, N.K.; Gao, X.; Sorooshian, S. 2007: Multi-model ensemble hydrologic prediction using Bayesian Model Averaging. *Advances in Water Resources 30(5)*: 1371-1386.

Dutta, S.; Ghosh, A.K. 2012: On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics 64(3)*: 657-676.

Georgakakos, K.P.; Seo, D.-J.; Gupta, H.; Schaake, J.; Butts, M.B. 2004: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology 298(1)*: 222-241.

Granger, C.W.J.; Ramanathan, R. 1984. Improved methods of combining forecasts. *Journal of Forecasting 3(2)*: 197–204.

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(1-2)*: 80-91. DOI: 10.1016/j.jhydrol.2009.08.003.

Hamurkaroglu, C.; Mert, M.; Saykan, Y. 2006: Nonparametric control charts based on Mahalanobis depth. *Quality Contol and Applied Statistics 51(1)*: 21-26.

Hamill, T.M. 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review 129*: 550-560.

Hersbach, H., 2000. Decomposition of the continous ranked probability score for ensemble prediction systems. *Weather and Forecasting 15*: 559-570.

Höge, M.; Wöhling, T.; Nowak, W. 2018. A primer for model selection: The decisive role of model complexity. *Water Resources Research 54*: 1688-1715. https://doi.org/10.1002/2017WR021902

Höge, M.; Guthke, A.; Nowak, W. 2019: The hydrologist's guide to Bayesian model selection, averaging and combination. *Journal of Hydrology 572*: 96-107. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.01.072

Jakeman, A.J.; Littlewood, I.G.; Whitehead, P.G. 1990: Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology 117*: 275-300.

Kokkonen, T.; Koivusalo, H.; Jakeman, A.; Norton, J. 2006: Construction of a degree-day snow model in the light of the 'Ten iterative steps in model development'. *In*: Voinov, A.; Jakeman, A.J.; Rizzoli, A.E. (*Eds.*) *Proceedings of the iEMSs Third Biennial Meeting: 'Summit on Environmental Modelling and Software'*. International Environmental Modelling and Software Society, Burlington, USA.

Li, J.; Cuesta-Albertos, J.A.; Liu, R.Y. 2012: DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association 107(498)*: 737-753.

Liu, R.Y. 1995: Control charts for multivariate processes. *Journal of the American Statistical Association 90(432)*: 1380-1387.

Liu, R.Y.; Parelius, J.M.; Singh, K. 1999. Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion and a rejoinder by Liu and Singh). *Annals of Statistics 27(3)*: 783-858.

Liu, R.Y.; Serfling, R.J.; Souvaine, D.L. 2006: Data depth: robust multivariate analysis, computational geometry and applications. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science 72*. American Mathematical Society.

Liu, R.Y.; Singh, K. 1993: A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association 88(421)*: 252-260.

Liu, Y.; Gupta, H.V. 2007: Uncertainty in hydrologic modeling: *Toward an integrated data assimilation framework. Water Resources Research 43*: W07401, doi:10.1029/2006WR005756.

McMillan, H.; Jackson, B.; Clark, M.; Kavetski, D.; Woods, R, 2011: Input uncertainty in hydrological models: An evaluation of error models for rainfall. *Journal of Hydrology 400(1-2)*: 83-94.

Messaoud, A.; Weihs, C.; Hering, F. 2004: A Nonparametric Multivariate Control Chart Based on Data Depth. Technical Report/Universität Dortmund, SFB 475 Komplexitätsreduktion in Multivariaten Datenstrukturen.

Nash, J.; Sutcliffe, J. 1970: River flow forecasting through conceptual models part I – A discussion of principles. *Journal of Hydrology 10(3)*: 282-290.

Newsome, P.F.J.; Wilde, R.H.; Willoughby, E.J. 2000: Land resource information system spatial data layers. Landcare Research Technical Report No.84

Noori, R.; Hoshiyaripour, G.A.; Ashrafi, K.; Araabi, B.N. 2010: Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. *Atmospheric Environment 44*: 476-482.

Parrish, M.A.; Moradkhani, H.; DeChant, C.M. 2012: Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. *Water Resources Research 48*: 1–18.

Pechlivanidis, I.; Jackson, B.; McIntyre, N.; Wheater, H. 2011: Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal 13(3)*: 193-214

Peck, E.L. 1976: Catchment modeling and initial parameter estimation for the National Weather Service River Forecast System. Office of Hydrology, National Weather Service.

Perrin, C.; Michel, C.; Andréassian, V. 2003: Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology 279(1)*: 275-289.

Qu, B.; Zhang, X.; Pappenberger, F.; Zhang, T.; Fang, Y. 2017: Multi-model grand ensemble hydrologic forecasting in the Fu river basin using Bayesian Model Averaging. *Water 9(2)*: 74.

Rathinasamy, M.; Adamowski, J.; Khosa, R. 2013: Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. *Journal of Hydrology 507*: 186-200.

Razavi, S.; Tolson, B.A.; Matott, L.S.; Thomson, N.R.; MacLean, A.; Seglenieks, F.R. 2010: Reducing the computational cost of automatic calibration through model preemption. *Water Resources Research 46*. W11523. doi:10.1029/2009WR008957.

Rings, J.; Vrugt, J.A.; Schoups, G.; Huisman, J.A.; Vereecken, H. 2012: Bayesian model averaging using particle filtering and Gaussian mixture modeling: theory, concepts, and simulation experiments. *Water Resources Research 48*. W05520.

Roudier, P.; Andersson, J.C.M.; Donnelly, C.; Feyen, L.; Greuell, W.; Ludwig, F. 2016: Projections of future floods and hydrological droughts in Europe under a +2 °C global warming. *Climate Change 135*: 341–355.

Rousseeuw, P.J.; Struyf, A. 1998: Computing location depth and regression depth in higher dimensions. *Statistics and Computing 8(3)*: 193-203.

Schöniger, A.; Wöhling, T.; Nowak, W. 2015: A statistical concept to assess the uncertainty in Bayesian model weights and its impact on model ranking. *Water Resources Research 51*: 7524– 7546. doi:10.1002/2015WR016918.

Serfling, R. 2002: Generalized quantile processes based on multivariate depth functions, with applications in nonparametric multivariate analysis. *Journal of Multivariate Analysis 83(1)*: 232-247.

Shafii, M.H.; Tolson, B.A.; Matott, L.S. 2015: Improving the efficiency of Monte Carlo Bayesian calibration of hydrologic models via model pre-emption. *Journal of Hydroinfomatics 17(5): 763-770*. doi:10.2166/hydro.2015.043

Singh, S.K.; Bárdossy, A. 2012: Calibration of hydrological models on hydrologically unusual events. *Advances in Water Resources 38*: 81-91.

Singh, S.K.; McMillan, H.; Bárdossy, A. 2013: Use of the data depth function to differentiate between case of interpolation and extrapolation in hydrological model prediction. *Journal of Hydrology 477*: 213-228.

Snelder, T.H.; Biggs, B.J.F. 2002: Multiscale river environment classification for water resources managements. *Journal of the American Water Resources Association 38(5)*: 1225-1239.

Stoumbos, Z.G.; Jones, L.A.; Woodall, W.H.; Reynolds, M. 2001: On nonparametric multivariate control charts based on data depth. *Frontiers in Statistical Quality Control 6*: 207-227.

Talagrand, O.; Vautard, R.; Strauss, B. 1997: Evaluation of probabilistic prediction systems. Proceedings, ECMWF Workshop on Predictability, 20-22 October 1997, ECMWF, pp. 1–25. Available at https://www.ecmwf.int/en/elibrary/12555-evaluation-probabilistic-prediction-systems

Tukey, J. 1975: Mathematics and the picturing of data. Proceedings of the 1974 International Congress of Mathematics, Vancouver, Canada, vol. 2, pp. 523–531.

Tyralla, C.; Schumann, A.H. 2016: Incorporating structural uncertainty of hydrological models in likelihood functions via an ensemble range approach. *Hydrological Sciences Journal 61(9)*: 1679-1690. DOI: 10.1080/02626667.2016.1164314

Velázquez, J.A.; Anctil, F.; Ramos, M.H.; Perrin, C. 2011: Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geoscience 29: 33-42*. DOI:10.5194/adgeo-29-33-2011.

Velázquez, J.A.; Schmid, J.; Ricard, S.; Muerth, M.J.; Gauvin St-Denis, B.; Minville, M.; Chaumont, D.; Caya, D.; Ludwig, R.; Turcotte, R. 2013: An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrology and Earth System Sciences 17*: 565-578. DOI:10.5194/hess-17-565-2013.

Wazneh, H.; Chebana, F.; Ouarda, T. 2013a: Depth-based regional index-flood model. *Water Resources Research 49(12)*: 7957-7972.

Wazneh, H.; Chebana, F.; Ouarda, T.B.M.J. 2013b: Optimal depth-based regional frequency analysis. *Hydrology and Earth System Sciences 17(6)*: 2281-2296.

Zamo, M.; Naveau, P. 2018: Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts. *Mathematical Geosciences 50(2)*: 209-234. https://doi.org/10.1007/s11004-017-9709-7

Zhang, X.; Zhao, K. 2012: Bayesian Neural Networks for Uncertainty Analysis of Hydrologic Modeling: A Comparison of Two Schemes. *Water Resources Management 26(8)*: 2365–2382.

Zuo, Y.; Serfling, R. 2000: General notions of statistical depth function. *Annals of Statistics 28(2)*: 461-482.

Appendix

Hydrological models used in this study

CMD

The IHACRES Catchment Moisture Deficit (CMD) model calculates the effective rainfall by taking account of evapotranspiration and the catchment moisture deficit. The effective rainfall is calculated from input rainfall by using a mass balance approach (Croke and Jakeman, 2004).

CWI

The IHACRES Catchment Wetness Index (CWI) model uses a temperature-dependent drying rate to estimate a wetness index which defines the runoff ratio.

Sacramento model

The Sacramento model was developed by the US National Weather Service. It is the most complex model of the package. Two soil zones, upper and lower, are defined. The upper zone contains interception storage while the lower zone represents the bulk of the soil moisture and groundwater storage. In each soil zone, two storage moistures are represented, tension water and free water. Another notable aspect of the model is the representation process of the percolation from the upper zone to the lower zone. Evapotranspiration is taken account of from each part in the model according to a hierarchy of priorities. A mass balance approach is used to calculate the effective rainfall from lateral drainage that occurs from each of the soil zones (Peck, 1976).

AWBM

The Australian Water Balance Model (AWBM) is a conceptual model. It is developed from concepts of saturation over-land flow (excess rainfall after reaching the surface storage capacity of the catchment) and generation of runoff. Watersheds are divided into three different areas with different storages. The effective rainfall is the sum of excess water in each area (Boughton, 2004).

BUCKET model

The single-bucket (BUCKET) model takes into account interception, saturation excess runoff and subsurface flow (Andrews *et al.*, 2011).

BDM model

The BDM model is the typical initial model used in Data-Based Mechanistic modelling. The observed streamflow raised to a power defines an index of antecedent wetness. Rainfall is scaled by using this index (Andrews *et al.*, 2011). BDM and runoff ratio model cannot be used for prediction as the Soil Moisture Account (SMA) uses streamflow data. The routing discharge is an exponential component transfer function model. It is a linear transfer function which translate an input time series U into an output series X. The unit hydrograph is described by exponential decaying components. Each component is defined by its recession rate α and peak response β. For SMA models use, two components (slow and quick) are used and these components are in a parallel configuration. The total simulated flow is the sum of slow and quick components.

SNOW

A degree-day snow (SNOW) model is essential in cold regions to estimate the snowmelt input to be used in streamflow forecasting. A daily snowmelt discharge series and an estimate of the water stored in the snow pack are produced in this model (Kokkonen *et al.*, 2006). This model is coupled with the CMD soil moisture model.

Intensity model

The intensity model uses a runoff ratio generated by raising rainfall to a power. It increases up to a full runoff level.

Runoff ratio model

The runoff ratio model scales the rainfall to a runoff coefficient, estimated by a moving average through the data.

TopNet

TopNet is a physically-based and semi-distributed model based on TOPMODEL (Beven and Kirkby, 1979) concepts. It has two main components, namely water balance over each sub-catchment and routing streamflow for each sub-catchment. The water balance model represents storages and fluxes of water in canopy, snowpack, unsaturated and saturated soil zone. The catchment is divided into sub-catchments linked by a river network. Then a kinematic wave is used to model the output discharge through the digital stream network. The routing component has three sources of runoff from each sub-catchment: saturation excess runoff from excess precipitation, infiltration excess runoff and base flow within the saturated zone. Saturation excess runoff occurs in the saturated portion of the catchment when soil water storage reaches its capacity. Infiltration runoff excess occurs in the uninfluenced and influenced portions of the catchment when the effective rainfall exceeds the infiltration rate.

TopNet uses TOPMODEL concepts for the representation of the soil moisture deficit using a topographic index to model the dynamics of variable source areas contributing to saturation excess runoff (Bandaragoda *et al.*, 2004; Clark *et al.*, 2008).

The input data are precipitation, minimal and maximal temperature at hourly time steps for each sub basin, relative humidity, shortwave radiation, wind speed and mean sea level pressure.

Parameters of the hydrological models used in this study

Table A1 – Summary of CMD parameters

Parameter	Description	Units
f	Stress threshold, which defines the minimum amount of daily rainfall needed during dry condition to generate effective precipitation	–
e	Evapotranspiration coefficient	–
d	Drainage threshold, which defines the catchment deficit above which some rainfall becomes effective rainfall	mm
shape	Rainfall effectiveness (i.e. drainage proportion) power in drainage equation. Value less than 1 selects the linear form; a value of 1 selects the trigonometric form.	–
ts	Time constant for slow flow	days
tq	Time constant for quick flow	days
vs	Relative volume of slow flow	–

Table A2 – Summary of CWI parameters

Parameter	Description	Units
tw	Rate at which the catchment wetness declines in the absence of rainfall	days
f	Temperature modulation factor dependence of drying rate	$°C^{-1}$
scale	Mass balance term (c in the literature)	
l	Moisture threshold for producing flow	s
p	Power on soil moisture (above the threshold l).	–
tref	Reference temperature in units of E	°C
ts	Time constant for slow flow	days
tq	Time constant for quick flow	days
vs	Relative volume of slow flow	–

Table A3 – Summary of SACRAMENTO parameters

Parameter	Description	Units
uztwm	Maximum capacity upper zone tension water	mm
uzfwm	Maximum capacity upper zone free water	mm
uzk	Lateral drainage rate of upper zone free water expressed as a fraction of contents per day	–
pctim	The fraction of the catchment which produces impervious runoff during low flow conditions	*Decimal fraction*
adimp	The fraction of the catchment which becomes impervious as all tension water requirements are met	*Decimal fraction*
zperc	Factor which defines the proportional increase in percolation from saturated to dry lower zone soil moisture conditions (indicates the maximum percolation rate possible when upper zone storages are full and the lower zone soil moisture us 100% deficient)	–
rexp	An exponent determining the rate of change of the percolation rate as the lower zone deficiency ratio varies from 1 to 0 (1 = completely dry, 0 = lower zone storage completely full)	–
lztwm	Maximum capacity of lower zone tension water	mm
lzfsm	Maximum capacity of lower zone supplemental free water storage	mm
lzfpm	Maximum capacity of lower zone primary free water storage	mm
lzsk	Lateral drainage rate of lower zone supplemental free water expressed as a fraction of contents per day	–
lzpk	Lateral drainage rate of lower zone primary free water expressed as a fraction of contents per day	–
pfree	Percentage of percolation water which directly enters the lower zone free water without a prior claim by lower zone tension water	–
t_s	Time constant for slow flow	days
t_q	Time constant for quick flow	days
v_s	Relative volume of slow flow	–

Table A4 – Summary of AWBM parameters

Parameter	Description	Units
cap.ave	Average soil water storage capacity	mm
etmult	Multiplier for the E input data	–
t_s	Time constant for slow flow	days
t_q	Time constant for quick flow	days
v_s	Relative volume of slow flow	–

Table A5 – Summary of BUCKET parameters

Parameter	Description	Units
Sb	Maximum soil water storage	mm
fc	Field capacity	–
α_{ei}	Interception coefficient	–
M	Fraction of catchment area covered by deep rooted vegetation	–
α_{ss}	Recession coefficients for subsurface flow from saturated zone	–-
t_s	Time constant for slow flow	days
t_q	Time constant for quick flow	days
v_s	Relative volume of slow flow	–

Table A6 – Summary of BDM parameters

Parameter	Description	Units
power	Power to apply to streamflow values	–
qlag	Number of time steps to lag the streamflow before multiplication	–
scale	Constant multiplier of the result for mass balance	–
t_s	Time constant for slow flow	days
t_q	Time constant for quick flow	days
v_s	Relative volume of slow flow	–

Table A7 – Summary of SNOW parameters

Parameter	Description	Units
T_{max}	Temperature threshold for rain, all rain is liquid above this threshold	°C
T_{min}	Temperature threshold for rain, all rain is snow below this threshold	°C
C_r	Correction factor for rainfall	–
C_s	Correction factor for snowfall	–
k_d	Degree-day factor for snowmelt	$mm°C^{-1}\ day^{-1}$
k_f	Degree-day factor for freezing	$mm°C^{-1}\ day^{-1}$
rcap	Retention parameter for liquid water retention capacity of snowpack	–
f	Stress threshold / wilting point (as fraction of d)	–
e	Evapotranspiration coefficient	–
d	CMD threshold for producing flow	mm
t_s	Time constant for slow flow	days
t_q	Time constant for quick flow	days
v_s	Relative volume of slow flow	–

Table A8 – Summary of INTENSITY parameters

Parameter	Description	Units
power	Power on rainfall used to estimate effective rainfall	–
maxP	Level of rainfall at which full runoff occurs	–
scale	Constant multiplier of the result for mass balance	–
t_s	Time constant for slow flow	days
t_q	Time constant for quick flow	days
v_s	Relative volume of slow flow	–

Table A9 – Summary of RUNOFF RATIO parameters

Parameter	Description	Units
rrthresh	Threshold value of the runoff ratio, below which there is no effective rainfall	–
scale	Constant multiplier of the result for mass balance	–
t_s	Time constant for slow flow	days
t_q	Time constant for quick flow	days
v_s	Relative volume of slow flow	–

Table A10 – Summary of TopNet parameters

Parameter	Description	Units
topmodf	TOPMODEL f parameter	m^{-1}
hydcon0	Saturated hydraulic conductivity	$m.s^{-1}$
swater1	Drainable water	m
swater2	Plant-available water	m
dthetat	Soil water content	m
overvel	Overland flow velocity	$m.s^{-1}$
gucatch	Gauge under-catch for snowfall	-
th_accm	Threshold for snow accumulation	K
th_melt	Threshold for snow melt	K
snowddf	Mean degree-day factor for snow melt	$mm.K^{-1}.day^{-1}$
minddfd	Minimum degree-day-factor day	day
maxddfd	Maximum degree-day-factor day	day
snowamp	Seasonal amplitude of degree-day factor for snow melt	$mm.K^{-1}.day^{-1}$
cv_snow	Coefficient of variation in sub-grid SWE	-
r_man_n	Manning's n	-

Manuscript received 12 December 2019; accepted for publication 7 January 2020