Accepted manuscript doi:
10.1680/jwama.19.00001

Accepted manuscript
As a service to our authors and readers, we are putting peer-reviewed accepted manuscripts (AM) online, in the Ahead of Print section of each journal web page, shortly after acceptance.

Disclaimer
The AM is yet to be copyedited and formatted in journal house style but can still be read and referenced by quoting its unique reference number, the digital object identifier (DOI). Once the AM has been typeset, an 'uncorrected proof' PDF will replace the 'accepted manuscript' PDF. These formatted articles may still be corrected by the authors. During the Production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal relate to these versions also.

Version of record
The final edited article will be published in PDF and HTML and will contain all author corrections and is considered the version of record. Authors wishing to reference an article published Ahead of Print should quote its DOI. When an issue becomes available, queuing Ahead of Print articles will move to that issue's Table of Contents. When the article is published in a journal issue, the full reference should be cited in addition to the DOI.

Accepted manuscript doi: 10.1680/jwama.19.00001

Submitted: 14 January 2019

Published online in 'accepted manuscript' format: 16 May 2019

Manuscript title: Investigating capabilities of machine learning techniques in forecasting streamflow

Authors: Syed Kabir[1], Sandhya Patidar[1] and Gareth Pender[2]

Affiliations: [1]Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh, UK and [2]Heriot-Watt University, Edinburgh, UK

Corresponding author: Syed Kabir, William Arrol Annexe Building, Gait 4, Heriot-Watt University, EH14 4AS, Edinburgh, UK.

E-mail: Syed.Rezwan.Kabir@gmail.com

Accepted manuscript doi: 10.1680/jwama.19.00001

Abstract

This paper presents a systematic investigation into modelling capacities of three conventional data driven modelling techniques, namely, wavelet based artificial neural network (WANN), support vector regression (SVR), and deep belief network (DBN) for multi-step ahead stream flow forecasting. To evaluate the effectiveness of these modelling techniques, hydro-meteorological hourly datasets from three case-study rivers located in the UK have been used. A heuristic performance analysis of the modelling schemes has been conducted by systematically analysing the key statistics that measure magnitude, scatter and density of model errors. Finally, for each of the modelling technique, performance deterioration rate in time was estimated. The results show that the SVR model can forecast quite accurately up to one to two hours ahead but its performance deteriorates gradually from three hour onwards. Further it has been found that the WANN model performs better when the overall nonlinearity of the system increases whereas the DBN model appeared to show consistent poor predictive capabilities when compared to the other models presented herein. We conclude by stating that for any selected model, it is possible to use identical model structure for up to two step ahead forecasting. Models need to be re-configured beyond that limit.

Keywords: Hydrology and water resource; Computational mechanics; Statistical analysis

Accepted manuscript doi: 10.1680/jwama.19.00001

1. Introduction

Possibly due to considerably high spatial and temporal variability in rainfall distribution attributed to highly nonlinear and inordinately complex rainfall–runoff process, flood forecasting remains one of the most challenging albeit important tasks of operational hydrology (Chang *et al.*, 2007). For decades, a considerable amount of research incorporating physical, conceptual and stochastic models have been carried out to understand the complexity of rainfall–runoff processes. Although physical and conceptual models are important in understanding hydrological processes, such models are difficult to implement and calibrate as they requires detailed information on several parameters (often not readily available), expertise in setting up complicated software tools, time and computational resources (ASCE, 2000, Chang *et al.*, 2007, Raghavendra and Deka, 2014). Due to these various reasons, these modelling procedures often failed to provide any useful information to decision makers during a real time event.

The Flood Forecasting Centre (FFC), UK, uses Grid-to-Grid (G2G) distributed hydrological model (Price *et al.*, 2012) based on probability distributed model (PDM) (Moore, 2007) approach for flood forecasting. The FFC provides six hour to five days ahead flood forecasts (Pilling *et al., 2014)* . The coefficient of determination (*R-squared*) values ranged from 0.64 to 0.87 and mode runtime varied from 12 minutes to 140 minutes. In recent years researchers have successfully applied computationally more efficient numerical models such as the data-based mechanistic (DBM) approach, as an alternative, for adaptive water level forecasting at gauging stations (Romanowicz *et al.*, 2008, Romanowicz *et al.*, 2006). The forecasting performance of DBM approach was tested at several gauging stations along river Severn reach, UK and showed promising results for different lead times. The DBM model was further exploited by Leedal *et al.* (2010) for visualization of inundation information in real time. In their research, a DBM model was coupled with a 2D hydrodynamic model (e.g. LISFLOOD-FP) for 6h ahead flood inundation forecasting. While the 2D model is computationally expensive and not feasible for real time operation, pre-calculated inundation maps stored in a database from the so called ‘offline method’ can be used (Bhola *et al.*, 2018). Forecasted water level at upstream is used to draw the most likely map from the pre-recorded scenario for real time flood inundation. However, such integrated system require multiple platforms to work together and generation of scenarios to build the database using offline method can be expensive. At the same time, storing vast amount of pre-calculated maps would also demand huge storage capacity. To overcome these drawbacks, machine learning (ML) based hydrodynamic models can be applied. Possibly, Liu *et al.* (2009) made the first attempt to introduce ML techniques with an intention to improve performance of fast inundation models. Liu and Pender (2015) applied SVR to predict water depth and velocity at Aldeburgh Marshes, UK. They have concluded that ML techniques can be a potential alternative in fast flood modelling allowing real time simulation of inundation maps. This offers an opportunity for a new computationally efficient and robust data driven end-to-end (e.g. rainfall to inundation) modelling engine on a single platform as an alternative to the previously studied integrated models. A robust ML based rainfall-runoff model can be used to forecast upstream discharge with a suitable lead time and then the forecasted discharge can be fed into the integrated ML based hydraulic model to generate real time flood maps.

A significant amount of research studies have been done to investigate and compare suitability of various data driven modelling techniques for rainfall-runoff modelling. It is beyond the scope of this paper to summarize all of them but within the context of the present paper, findings from some key and representative studies are discussed here. Khan and Coulibaly (2006) applied Bayesian learning approach to train their neural network for daily river flow and reservoir inflow in a cold river basin in Canada. Their results showed that Bayesian neural network (BNN) model outperforms both the conceptual and standard artificial neural network (ANN) model. In addition to that, the BNN has a significant advantage over the standard ANN, is the uncertainty quantification in the form of confidence intervals which are important in operational water resources applications. Makkeasorn *et al.* (2008) used genetic programming (GP) and ANN models for forecasting river discharges in a semi-arid watershed in south Texas. The findings indicated that GP-derived streamflow forecasting models were generally favoured for forecasting over ANNs and the most forward looking GP models had the ability to forecast streamflow 30 day ahead of time with r-square value of 0.84. Chang *et al*. (2007) investigated three types of ANNs, namely, multiple-input multiple-output (MIMO), multi-input single-output (MISO) and serial-propagated structure, for multi-step ahead flood forecasting at two watersheds in Taiwan. The results showed that both MISO and serial-propagated neural networks are capable of providing accurate short-term forecasting for up to one- or two-step-ahead. However, for long term (more than two steps) forecasts, only the serial-propagated neural network was shown to provide satisfactory results, tested in both watersheds.

Wu *et al*. (2008) applied a distributed SVR (D-SVR) with a two-step Genetic Algorithm parameter optimization method to predict river stages. The D-SVR method disaggregates the original training set into few subsets, and then generates a local SVR for each subset independently. The performance of D-SVR against the predictions from linear regression (LR), nearest neighbour (NN) method, and genetic algorithm-based ANN (ANN-GA) methods were compared. The D-SVR model has been shown to predict the water level comparatively better than the other models. Granata *et al.* (2016) developed a SVM based rainfall-runoff model for urban drainage and compared with a storm water management model (SWMM), both the approaches showed comparable results, with SVM slightly underestimating flow while SWMM overestimating it. Han *et al*. (2007) applied SVR for flood forecasting over the Bird Creek catchment, UK, the model performed better when compared with some benchmark models, e.g., linear transfer function model, a trend and a naïve model. Misra *et al*. (2009) used SVR and ANN to simulate daily, weekly and monthly runoff and sediment yield from an Indian watershed, and they concluded that SVM provided significantly improved results as compared to ANN based model estimation. Yu *et al.* (2006) employed SVR for real time multi-step ahead flood stage forecasting and demonstrated for a river in Taiwan that the SVR models can effectively predict the flood stage forecasts one-to-six-hours ahead. Liong and Sivapragasam (2002) compared SVR and ANN for flood stage forecasting and found that forecasting accuracy of SVR was better than the ANN model.

The most comprehensive investigation of predictive capabilities of different data driven modelling techniques could be found in (Elshorbagy *et al.*, 2010b). They compared six modelling techniques (e.g. neural networks, genetic programming, evolutionary polynomial regression, support vector machines, M5 model trees, and K-nearest neighbours) on five different case studies of rainfall-runoff, evapotranspiration, and soil moisture content. Their findings indicate that for highly nonlinear cases ANNs, GP, and K-nearest neighbours (K-NN) could be the better option, whereas, M5 perform very well for linear and semi-linear data. Further, they suggested that Evolutionary Polynomial Regression (EPR) has a potential to perform close to GP given the data sets are linear.

One can observe that in the comparative modelling studies, one of the fundamental means to assess model performance is through evaluating it against other techniques using, in general, correlation coefficient, root mean squared error (RMSE), Nash-Sutcliffe model efficiency coefficient (NSC). However, these comparative studies often considered impaired because of less-than-comprehensive approach being adopted (Elshorbagy *et al.*, 2010a). To highlight the shortcomings of the present machine learning based comparative studies, Abrahart *et al.* (2008) have used the example of neural network applications and suggested solutions or modelling guidelines and frameworks. In general, the key ideology underpinning data driven modelling techniques mainly depend on how well data is represented in the modelling framework. Therefore, exploratory data analysis is identified as an essential step to extract key feature of historical data to ensure greater predictive accuracy (Bai *et al.*, 2016).

For this reason, recently researchers have put much of their efforts on the feature extraction, data pre-processing and data transformation processes (Bai *et al.*, 2016). The wavelet-ANN (WANN) is one of the reliable hybrid model used in time series forecasting problems. Recently, wavelet decomposition has become popular signal processing tool because of its ability to elucidate both spectral and temporal information within the signal (Okkan, 2012, Nourani *et al.*, 2011). The significance of feature learning in forecasting is well understood and therefore, recently deep learning techniques have gained interest. Deep learning techniques are capable of extracting features from the historic data and thus improving the forecasting or predictive performance. Hinton *et al.* (2006) proposed deep belief network (DBN) as one of the deep learning techniques based on probabilistic generative models. There are many hidden layers amalgamate the deep structure of the network for realizing the latent changes by layer wise learning. This technique has already been applied successfully in many different areas, e.g., image processing (Liu *et al.*, 2016), character recognition (Sazal *et al.*, 2014), draught forecasting (Chen *et al.*, 2012), reservoir inflow forecasting (Bai *et al.*, 2016). However, to best of our knowledge very few studies have been reported utilising DBNs for the rainfall-runoff modelling. Bai *et al.* (2016) proposed a DBN for daily reservoir inflow forecasting. The results from the proposed model overwhelmed the standard ANN, least square support vector regression (LSSVR), and WANN.

In rainfall-runoff modelling, one-step-ahead prediction using different data driven techniques has been performed and reported with satisfactory results. However, multi-step-ahead prediction (to make predictions several time steps ahead into the future) using such models is a challenging and complex task (Chang *et al.*, 2007). Considering the lack of comprehensive evaluation of multiple modelling techniques for multi-step ahead runoff forecasting, this paper attempts to test the forecasting abilities of three techniques, namely, a hybrid model- e.g. wavelet decomposed ANN (WANN), SVR, and a deep learning model- e.g. DBN on three UK catchments for up to one to six hours ahead. The forecasting accuracy and uncertainty involved in these models were assessed and compared using multiple error statistics, scatter and density plots of model residuals and the model performance deterioration rate. We also performed gamma tests on the input data to gain some insight into the predictability of the output variable (e.g. river discharge).

2. Methodology

To conduct a comparative investigation of multistep ahead forecasting capabilities of multiple data driven models, firstly, three data driven techniques, e.g. WANN, SVR, and DBN and three case study sites within the

UK have been selected, e.g. the Lune at Killington, the Leven at Newby Bridge, and the Eden at Kirkby Stephen.

Secondly, the identical set of input data, e.g. rainfall and historical streamflow, were used in the each of the three specified modelling scheme. As this study mainly focus on comparing predictive capabilities of widely used data driven modelling techniques for multi hours ahead forecasting, less emphasis has been given on optimising selection of input variables and dataset. Thus, for as long as the set of input variables remains same an unbiased analysis can be carried out to successfully achieve the overall objectives of this research. Preparation of input/output datasets will be explained in more details in a later section of this paper.

Thirdly, for model performance evaluation criteria selection, Elshorbagy *et al.* (2010a) suggested that the forecasting accuracy of the various modelling schemes can be evaluated using four key error related statistics, e.g., the root mean squared error (RMSE), the mean absolute relative error (MARE), the mean bias (MB) and the correlation coefficient (R). The Nash-Sutcliffe model efficiency coefficient (NSC) was also calculated and used for model comparison. The formulae of the error measures specified above are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \quad (1)$$

$$MARE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{O_i - P_i}{O_i}\right| \quad (2)$$

$$MB = \frac{1}{N}\sum_{i=1}^{N}(O_i - P_i) \quad (3)$$

$$R = \frac{\sum_{i=1}^{N}(O_i - \bar{o}_i)(P_i - \bar{p})}{\sqrt{\sum_{i=1}^{N}(O_i - \bar{o})^2 \sum_{i=1}^{N}(P_i - \bar{p})^2}} \quad (4)$$

$$NSC = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{o})^2} \quad (5)$$

Where, N is the sample size, O_i and P_i are the observed and the predicted values, $\bar{o}$ and $\bar{p}$ are the mean of observed and predicted values. However, due to several statistics used for error measurement quite often conflicting results regarding the performance of different modelling techniques could be observed. To avoid such issues, Elshorbagy *et al.* (2010a) proposed a supplemental error measure that combines the effect of four error measurements stated above as a single consistent indicator. The new error measurement statistic, referred as ideal point error (IPE), is based on identifying the ideal point in the four dimensional space that each model aims to reach. The IPE (Eq. 6) defines how far the model is from the ideal point. The value of IPE ranges from 0.0 for the best model to 1.0 for the worst model.

$$IPE_{ij} = \left\{0.25\left[\left(\frac{RMSE_{ij}-0.0}{maxRMSE_{ij}}\right)^2 + \left(\frac{MARE_{ij}-0.0}{maxMARE_{ij}}\right)^2 + \left|\frac{MB_{ij}-0.0}{max|MB_{ij}|}\right|^2 + \left(\frac{R_{ij}-0.0}{1/maxR_{ij}}\right)^2\right]\right\}^{1/2} \quad (6)$$

Where, i denotes model and j denotes technique. Here in this study, $i = 1$ and $j = \{1,2,3\}$.

Fourthly, the uncertainty quantification of these forecasting models was conducted through residual analysis. Residuals for each forecasted timeseries (generated by these three forecasting models) were estimated and presented in the regression plot. A probability density curve was also fitted to the model residuals.

In addition to that, for a robust model validation and comparison of results, gamma test, namely *winGamma* (Jones *et al.*, 2001, Durrant, 2001), is performed on the input variables. The ideology underpinning Gamma test is that for a given input dataset it attempts to estimate the variance of noise associated with each output, which will be an estimate of the best mean squared error that a smooth model can achieve for the corresponding output (Jones *et al.*, 2001, Durrant, 2001). The main aim of conducting gamma test is to gain some insight into the potential of selected input variables in robustly predicting desired output variable.

3. Experimental setup

3.1 Study area and data

To investigate multi-step ahead flow forecasting capabilities of the specified machine learning techniques, hourly catchment average rainfall and flow data from 1st September 2015 to 31st December 2015 was collected for three catchments located around Lake District national park North-Western part of England (Fig. 1).

The Lune at Killington catchment has an area of *219 km²*. This is a wet, natural and completely rural catchment and it drains the eastern Lake District fells and North-western parts of the Yorkshire Dales National Park. Land cover land use is dominated by moorland, grass and arable farming. There are no abstractions and artificial discharges affecting the runoff therefore the gauged flow considered to be within 10% of the natural flow at, or in excess of, the *Q95* flow (Young *et al.*, 2003).

The Leven at Newby Bridge catchment has an area of *247 km²* and drains southern parts of the Lake District massif. The landscape is dominated by mainly grassland and is largely wooded in lower reaches. The catchment contains Windermere, Esthwaite Water, Rydal Water, Grasmere and numerous other small natural water bodies. Catchment runoff is greatly affected by reservoirs, e.g., reduced by water supply abstraction and is increased by effluent returns.

The Eden at Kirkby Stephen is a high relief catchment draining Carboniferous Limestone which forms most of the watershed. This is a small catchment covering only *69.4 km²* and completely natural. There are no known abstractions affecting the runoff for the catchment. Descriptive statistics of the runoff for all three catchments can be found in Table 1. It can be seen that the selected catchments have considerably distinct statistical and hydrological characteristics.

3.2 Model implementation

In order to define the number of input variables (e.g. rainfall, discharge) for the models, the lags associated with the hydro-meteorological variables and the model parameters need to be determined. As mentioned before, for the present work to allow unbiased analysis/comparison of multiple data modelling techniques less significance is given towards identification of optimum number of inputs for individual techniques though same input variables are retained for each modelling technique. However, lags for the input variables has been estimated and assigned systematically (e.g. to select the optimum set of input variables the Gamma test was applied on different input combinations initially, the combination that produced lowest Gamma score was selected. For further validation, the WANN model was simulated using the selected input variables and other input candidates. Then results were compared in terms of NSE and R). Following the procedure lag for rainfall (R) variable has been set to five and the lag for discharge (Q) variable has been set to three.

Thus the input vector for each model contains eight elements. The model structure for performing one to six hour ahead forecasting can be defined as:

$$Q(t+n) = f_{(t)}[Q(t+1-Q_l), R(t+1-R_l)] \qquad (7)$$

Where, $Q_l = 1,2,3$; $R_l = 1,2,3,4,5$; $f_{(t)}$ indicates the model function, e.g. WANN, SVR and DBN; t is the time index and $n = 1,2,3,4,5,6$.

In addition to suitable input variables selection, special care was given while preparing the training and validation set. It was insured that the training set included the global maxima (e.g. the peak) in all three catchments. The Lune at Killington training set had seven peak over threshold (POT) - e.g.127.364 m^3/s- events and validation set included four POT events. The Leven at Newby Bridge had three and three POT (34.999 m^3/s) events in training and validation sets respectively. Training set for the Eden at Kirkby Stephen catchment had five POT (46.092 m^3/s) events while validation set had four POT events. Figure 2 shows the flow series used during the training process for the catchments.

3.3 Wavelet- Artificial Neural Network (WANN)

The wavelet transformation has been carried out on the specified set of input variables using the *Daubechies* wavelets in MATLAB. The decomposed timeseries was then fed in to a fully connected feed forward back propagation neural network with one hidden layer. More than one hidden layer can also be used. However, large

Accepted manuscript doi:
10.1680/jwama.19.00001

number of previous studies, for example- Vijayashanthar *et al.* (2018), Kourgialas *et al.* (2015), Elshorbagy *et al.* (2010b), Wu *et al.* (2009), Jain and Kumar (2007), Khan and Coulibaly (2006) etc., suggest that for rainfall-runoff modelling one hidden layer with sufficient number of hidden units is enough for achieving desired output. Thus, the general structure of the ANN model consisted one input layer, one hidden layer and one output layer. The input layer had eight nodes for each of the input variables and flow at T+1h, T+2h,….,T+6h was used as target data for output layer. All input and output data sets were normalized (0 to 1).

Tuning hyperparameters, e.g. selecting hidden units, learning rate, number of training iteration, activation function, optimization algorithm, of a given general ANN structure in order to optimize model performance is a labour intensive process. These are the parameters that are not learned by the network during training process (Vijayashanthar *et al.*, 2018). A range of hidden units were tested during training process (e.g. 10 to 60). Since there were only eight input variables at each input-output instance, this was a reasonable range of hidden units to build the model without overfitting. Multiple activation functions were also considered. The candidate activation functions were *sigmoid*, *tanh*, and *linear*. The WANN training was done by the so called trial and error approach and the network performance was monitored. The number of nodes in the hidden layer was systematically increased from 10 to 60 and stopped at the number where lowest SSE and highest R values were found. The final WANN architecture consisted 50 neurons in the hidden layer. In this case the best results were found when *sigmoid* activation function was applied to both hidden and output layer. The widely used *Levenberg-Marquardt* algorithm was selected as the optimizer and hundred training iterations were carried out during the training process.

This optimal network architecture was derived using T+1h flow as the target output. As one of our research objectives was to observe rate of performance degradation going ahead in time, hence, network architecture was kept identical for training all 18 (one to six hours for all three catchments) WANN models.

3.4 Support Vector Regression (SVR)

The *e1071* package (Dimitriadou *et al*., 2011) in *R* software was used to implement the SVR to all normalized datasets. In this experiment, the nu-SVR (ν-SVR), a variant of SVR was selected as the preferred technique. Nu-SVR has one additional parameter (nu) that allows to control the proportion of the number of support vectors in the model and automatically provides an optimal value of parameter ε (Abe, 2010). This approach makes it possible to overcome the problem of estimating optimal value of ε, by definition, which can be any positive real number. SVR is sensitive to the kernel choice; therefore kernel function needs to be selected carefully.

Accepted manuscript doi:
10.1680/jwama.19.00001

Numerous studies have suggested that for modelling nonlinear regression process 'radial basis function (RBF)' are more effective than other kernel functions (Elshorbagy *et al.*, 2010b, Han *et al.*, 2007, Lin *et al.*, 2006).

A RBF kernel was used for training the model. An exhaustive grid search method, a common practice to find SVR parameters, was applied to find the optimal values. An interesting finding was that the optimal values found from the grid search method had produced lower accuracy than a set of values found from labour intensive trial and error approach. In both cases 'cost', 'gamma' and 'nu' were searched from 0.001 to 1000, 0.001 to 1 and 0.001 to 1 respectively for the SVR model. We selected the values found using trial and error approach for our predictive analysis. The selected parameter values were: $C = 10$, $nu = 0.5$, $gamma = 0.5$. Then 18 SVR models were developed and tested on the corresponding testing set.

3.5 Deep Belief Network (DBN)

The network was constructed in *Python* using the *dbn* package (Albertbup, 2017). This implementation utilizes *Numpy* and *Tensorflow* libraries. Similar to the WANN and SVR technique, the network parameters, e.g. learning rate, batch size, number of hidden layer nodes, and number of hidden layers were perturbed manually to optimize the model performance using training datasets. Particular care was taken while selecting optimum "batch size - the number of training samples present in a single batch" as for example; DBNs technique is susceptible to overfitting if larger batch size is selected whereas smaller batch size could increase the computational cost. The best set of values for the network parameters are shown below (Table 2).

18 separate DBN models were trained using the same parameters then tested on the corresponding testing datasets.

4. Results and analysis

4.1 Lune at Killington

The performance analysis of the WANN, SVR and DBN techniques for multi hour ahead stream flow forecasting in lune at Killington case study is provided in Table 3. It can be observed that the different forecasting techniques have different performance responses measured across different error statistics and along different hours in future. Font colour corresponding to the forecasting techniques for which error statistics is measure lowest has been highlighted in 'bold'. In summary it can be conclude that, for first two hours SVR performed significantly better than WANN and DBN. From third to sixth hours onwards, performance of WANN appears to gradually improve in comparison to SVR and DBN techniques. Interestingly, various error statistics (except MB) corresponding to DBN seems to be considerably large in comparison to other forecasting techniques. This shows that DBN is less biased than the other two methods compared here. The IPE indicator (presented in Table 5) provides a considerably good representation of overall performance of each technique. It can be seen that SVR models had lowest IPE values for one to fifth hour ahead forecasting, whereas WANN had comparatively lower IPE at sixth hour. However, averaged IPE value corresponding to six hours ahead forecasting outcomes conclude that SVR method has the most efficient forecasting abilities than WANN and DBN methods.

For visual comparison, predicted flow using three different techniques against validation data with different lead times is illustrated in Figure 3. At T+3h and further ahead the SVR models overestimates the peaks with slight delays. This can be said that for over T+3h ahead predictions one has to change the SVR model configuration for better estimations. Figure 4 shows the residual scatter plots of observed versus forecasted streamflow data for one, three and six hour ahead forecasted values generated by the WANN, SVR and DBN methods. It can be noticed visually that for one hour ahead forecasting the SVR residuals scatter fit closely around the regression line. Looking at three hour ahead residual plots, it is apparent that the SVR performance is on a par with the WANN, however, on closely inspecting the residual distances from the regression line it appears that the SVR performed slightly better. Nevertheless, it should be noted that the performance of the WANN is comparatively better than the SVR and DBN methods specifically for the comparison of six hour ahead forecasted values. Thus the residual plots support the conclusion made earlier regarding the predictive capabilities of the techniques using IPE indicator.

4.2 Leven at Newby Bridge

The second case study location is river Leven at Newby Bridge. This case study has considerably different rainfall-runoff process compared to the first case study as discussed in section 4.1. The runoff process for the Leven case study is greatly affected by the abstraction and response reduced by water supply abstraction and is increased by effluent returns. Table 4 presents the analysis of error statistics for one- to six-hour-ahead forecasts for three forecasting techniques compared in this paper for the Leven. It can be observed that SVR model have consistently performed better than the WANN and DBN models along 1 to 5 hours ahead forecasted values and across all the error measuring criteria. For sixth hour ahead forecast, SVR model occasionally appears to performance moderately less in comparison to WANN model. The DBN models seem to generate satisfactory results although the error statistics measurement are slightly high than the other models. The analysis of IPE indicator providing a comprehensive representation of overall performance of each technique is presented in table 5.

The visual comparison between predicted flows against validation data (Fig. 5) clearly shows that while SVR models are more stable and out performs other two models at T+1h and T+3h, the WANN model more accurately predicts the peaks at T+6h. Further, Figure 6 (following the analogy of Figure 4) displays the residual scatter plots of observed versus forecasted value of streamflow data for one, three and six hour ahead in future. On close inspection of plots of residuals of forecasted model it can be seen that, the models are somewhat analogous with regard to the R squared value for one hour ahead forecasting, although the SVR marginally outperforms the WANN and DBN. The SVR appears to perform distinctively better than the WANN and DBN models along two to five hours ahead forecast. While, the R-squared values for both the WANN and SVR are same at six hours ahead forecasting, the WANN has slightly better forecasted model residual spread compared to the SVR. These observations are further supported by the analysis of outcomes of the IPE indicators (presented in Table 5).

It can be observed that unlike the SVR and DBN model, the predictive performance of WANN model does not decreases gradually with every time step ahead in future. This means that when nonlinearity increases in the model, WANN can provide considerably more reliable forecast than the SVR and DBN models.

4.3 Eden at Kirkby Stephen

The third study area river Eden at Kirkby Stephen has considerably similar geomorphological characteristic to that of the Lune at Killington. The comparative performance analysis using five different error statistics for the three modelling techniques is given in Table 6. The SVR models appear to perform considerably better than other two competitors for forecasting at one and two hours ahead in the future. It is interesting to note that with increase in overall complexity of system WANN models seem to demonstrate good response, as mentioned in previous section. Further, it is noted that after two hour ahead forecasting, overall performance of the SVR models appear to decline significantly and the WANN models seem to provide more reliable outcomes (with lowest RMSE and highest coefficient of correlation, and also lowest MB for in four out of six cases). The IPE indicator supports the findings of the error statistics analysis presented in Table 6 for one and two hour ahead forecasting, where it can be seen that the SVR models are performing better than the other two models. Additionally, it should be noted that comparative measures for IPE seem to differ from the general error statistics analysis (presented in Table 6) for three to six hours ahead forecasting. Since except MARE, for all other error statistics the WANN models were consistently providing lowest error measures, therefore, one would expect the WANN to have lower IPE values, given their dominance over other models. Although the DBN model performances were considerably lower than their counterparts, the IPE indicator shows that for five and six hours ahead forecasting DBN models are performing better. This is counter intuitive and requires further investigation.

Figure 7 shows the observed and predicted flows from different models. The visual representation again confirms that the SVR models capture the peaks better than other models when lead time is short. The residual regression plot (Fig. 8) also confirms that the WANN models are performing better when compared to the SVR and DBN models for three to six hours ahead forecasting. These findings are not in agreement with the IPE indicators.

To quantify model uncertainty analysis, probability density distribution (pdd) of residuals of the various techniques has been developed. Figure 9 presents the pdd plots of all the models for three study areas. The plots show the spread of the distribution of residuals for each method. For a good fit model, pdd plots of the residuals, in general, follow a normal distribution with mean close to zero. Further, model with wider spread (e.g. large standard deviation) represent large uncertainty in predictive capabilities. The pdd plots of the SVR models for one hour ahead prediction are close to a normally distributed curve for all three catchments. The models deviate from the zero mean with each time step ahead. This indicates that model predictions become biased and more uncertain with added nonlinearity.

The pdd plots of Eden at Kirkby Stephen for three and six hour ahead forecasting show disparities between the models when compared against IPE measures, residual plot and five error statistics. The pdd plots of residual for three hours ahead forecast demonstrates that the SVR model is performing better, whereas, the residual regression plot shows that the WANN performed better. This is also noticed that the residual regression and density plot both indicate that the WANN model outperformed other models at six hour ahead forecasting, while IPE value shows that the DBN had a better performance.

5. Discussion

The fundamental objective of this study was to investigate whether there are real differences between the predictive capabilities of three major and widely applied machine learning techniques in projecting multi-step ahead stream flow forecasts without having to change model parameters. This is particularly important in order to develop a fully machine learning based rainfall-inundation model on a single platform. Because for real time flood forecasting using a data driven model one would also want to take scalability of the models into account. Constructing multiple models for multiple lead times for each catchment may not be a viable option. In other words, it could be more efficient and realistic to develop a single robust model which would forecast multi step ahead flows with minimum added uncertainty.

The investigation conducted the performance evaluation of the techniques using multiple error measurements and through a thorough analysis of error distribution. It has been found that, the SVR technique based models are more reliable for forecasting one to three hours ahead stream flow, and this has been confirmed by multiple error measure statistics, the IPE values, residual regression plots and pdd. This has also been shown that the WANN based models generally tend to perform better when the system nonlinearity increases that is forecasting four to six hours ahead stream flow. Moreover, for short term forecasting (e.g. T+1h to T+3h) the SVR might be the preferred modelling option while the WANN could be chosen for T+4h to T+6h head forecasting. Alternatively, new SVR model configuration is needed for >T+3h ahead forecasting. However, four to six hours ahead forecasting results for Eden at Kirkby Stephen case study in light of five distinct error measurements, through residual regression analysis, the IPE, and the pdd seemed to contradict each other. While error statistics and residual regression have showed the WANN had performed better, the IPE values illustrated that the DBN has had the better outcome. This shows that there is a need for further analysis with different datasets to draw a conclusion on the usefulness of IPE statistics.

During model implementation, the optimum model parameters were chosen using input data sets, e.g., rainfall and river discharge data with their lags, from Lune at Killington catchment to forecast one hour ahead stream flow. The model was then applied to other case studies for one hour ahead forecasting. The results were satisfactory; therefore the model parameters for all techniques were kept constant to further evaluate the rate of performance deterioration. For a generalized overview of deterioration rate we looked into correlation coefficient (R) and Nash- Sutcliffe efficiency (NSC) measures. The SVR and DBN model performance seem to deteriorate drastically for each hour ahead of forecasting. For example, R and NSC values of the SVR technique deteriorated by 34.4% and 73.5%, whereas the WANN deteriorated by 4.9% and 21.6%, and the DBN deteriorated by 31.5% and 64.3% (Lune at Killington). However, for comprehensive analysis of model deterioration rate the percent deterioration should be calculated for each technique by dividing the difference between training and testing performance by the training performance. Less deterioration may indicate excellent generalization ability, e.g. a higher level of reliability and less uncertainty of a particular technique when testing against unseen data (Elshorbagy *et al.*, 2010b).

The best structure of the WANN and DBN model was found through iterating training process using various parameters or in other words the best fit WANN and DBN model was selected from multitude of model candidates at T+1h. There is no universal method of determining the best set of parameter values for SVR. Researchers have applied exhaustive grid search algorithms, evolutionary algorithms or trial and error process to find the optimum parameter values. Here we have used the exhaustive grid search method and intensive trial and error process for SVR model parameter identification. It was discovered that a different set of parameter values found though trial and error process produced slightly better results. Therefore, later was chosen more model configuration. We also recommend using at least two approaches to select SVR model parameters for any future studies.

In this study we used 70% of the data for training the models and 30% for testing predictive capabilities of the three driven techniques. It is understood that an ensemble prediction analysis could provide robust comparison between the various forecasting techniques and this can be done by randomly sampling complete set of observed dataset into multiple subsets without replacement, or observed data from multiple years can serve the purpose. Earlier studies have reported that the data driven models often fail to simulate extreme events effectively. These issues can be tackled by incorporating and effectively fitting suitable extreme value distribution for sampling rare events. In the realm of present study, pdd plots show that most of the models analysed here are not effective in simulating high flow projections for more than two hours ahead and specially if the catchment is sensitive.

The non-parametric Gamma test, using winGamma software (Jones *et al.*, 2001, Durrant, 2001) was applied to every dataset to explore the predictability and complexities of the modelling procedure. Table 7 provides the summary statistics of the key components.

The V-ratio, gradient and gamma statistic are calculated using the scaled data. Lower gradient value means that a noncomplex smooth function can be used for modelling the process and V-ratio closer to zero indicates the higher predictive abilities of model. The standard error in the gamma statistics shows the reliability of the test. On analysing the V-ratio and gradient values from the table above, it is evident that the model complexity increases as we move from one to six hours ahead for prediction and overall predictability decreases concurrently. The gamma test can also be applied to select the best model among the potential candidate models and this is particularly useful when multiple techniques perform equally consistent. Further, M-test can be used to define minimum number of samples required to develop a predictive model. However, in this study M-test is not used.

6. Conclusion

This study compares the predictive capabilities of three popular data driven modelling techniques for multi hour ahead streamflow forecasting. The results indicate that the SVR models are considerably good in predicting one to two hour ahead forecasting but their overall performance drops significantly from three hours onward projections. This behaviour indicates that with increase in nonlinearity of the system the reliability of SVR models reduces. Comparatively, the WANN models appear to perform satisfactory under highly nonlinear conditions. Thus, depending on the application type for forecasting over five to six hours ahead in future WANN model could be the most preferred option. The DBN technique has gained a great success in image processing and pattern recognition; however, in nonlinear regression modelling this technique appears to be a less effective.

There are few limitations of this study that should be noted. Firstly, the effect of the model inputs variables: to allow unbiased comparative analysis the inputs dataset remain unchanged across all the different model variants. Adding more inputs for one technique or removing some of them for another may affect the input-output relationship, consistency and thus overall model performance. Secondly, due to limited availability of dataset at fine resolution (hourly) this work is conducted by using three months of hourly dataset. For more comprehensive comparative analysis multiyear data should be modelled, or multiple groups (realizations) of each dataset should be randomly generated by sampling without replacement for ensemble prediction analysis. Future comprehensive studies can be developed based on the outcomes of the work presented in this paper, which could investigate various other modelling techniques applied in different watersheds, and should consider overcoming some of the limitations mentioned here. A k-fold cross validation technique can also be considered for future model robustness analysis. Hydrographs generated by the best performing models can be directly used for further hydraulic modelling in order to manage flood inundation and other water resources.

Acknowledgement

We acknowledge Dr. Jeffrey Neal, School of Geographical Sciences, University of Bristol, for providing the hydrological datasets required for this work.

Accepted manuscript doi: 10.1680/jwama.19.00001

List of notations

N	is the number of instances
O_i	is the observed values
P_i	is the predicted values
$\bar{o}$	is the mean of observed values
$\bar{p}$	is the mean of predicted values
C	is the cost parameter of Support Vector Regression model
γ	is the Radial Basis Function kernel parameter
ν	is the nu-Support Vector Regression model parameter

References

Abe, S. (2010) Function Approximation. In *Support Vector Machines for Pattern Classification*.) Springer London, London, pp. 395-442.

Abrahart, R. J., See, L. M. & Dawson, C. W. (2008) Neural Network Hydroinformatics: Maintaining Scientific Rigour. In *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications*. (Abrahart, R. J., See, L. M., andSolomatine, D. P. (eds)) Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 33-47.

Albertbup (2017) *A Python implementation of Deep Belief Networks built upon NumPy and TensorFlow with scikit-learn compatibility*, https://github.com/albertbup/deep-belief-network.

Asce (2000) Artificial Neural Networks in Hydrology. I: Preliminary Concepts. *Journal of Hydrologic Engineering* **5(2)**:115-123.

Bai, Y., Chen, Z., Xie, J. & Li, C. (2016) Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology* **532(Supplement C)**:193-206.

Accepted manuscript doi: 10.1680/jwama.19.00001

Bhola, P. K., Leandro, J. & Disse, M. (2018) Framework for Offline Flood Inundation Forecasts for Two-Dimensional Hydrodynamic Models. *Geosciences* **8(9)**:346.

Chang, F.-J., Chiang, Y.-M. & Chang, L.-C. (2007) Multi-step-ahead neural networks for flood forecasting. *Hydrological Sciences Journal* **52(1)**:114-130.

Chen, J., Jin, Q. & Chao, J. (2012) Design of Deep Belief Networks for Short-Term Prediction of Drought Index Using Data in the Huaihe River Basin. *Mathematical Problems in Engineering* **2012**:16.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D. & Weingessel, A. (2011) e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.5-25.), Vienna, Austria.

Durrant, P. (2001) winGamma TM : a non-linear data analysis and modelling tool with applications to flood prediction. In *Department of Computer Science*.) Cardiff University.

Elshorbagy, A., Corzo, G., Srinivasulu, S. & Solomatine, D. P. (2010a) Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. *Hydrol. Earth Syst. Sci.* **14(10)**:1931-1941.

Elshorbagy, A., Corzo, G., Srinivasulu, S. & Solomatine, D. P. (2010b) Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. *Hydrol. Earth Syst. Sci.* **14(10)**:1943-1961.

Granata, F., Gargano, R. & De Marinis, G. (2016) Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with the EPA's Storm Water Management Model. *Water* **8(3)**:69.

Han, D., Chan, L. & Zhu, N. (2007) Flood forecasting using support vector machines. *Journal of Hydroinformatics* **9(4)**:267-276.

Hinton, G. E., Osindero, S. & Teh, Y.-W. (2006) A fast learning algorithm for deep belief nets. *Neural Computation* **18**:1527-1554.

Jain, A. & Kumar, A. M. (2007) Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing* **7(2)**:585-592.

Jones, A. J., Margetts, S. & Durrant, P. (2001) *The winGamma User Guide*.

Khan, M. S. & Coulibaly, P. (2006) Bayesian neural network for rainfall-runoff modeling. *Water Resources Research* **42(7)**:W07409.

Kourgialas, N. N., Dokou, Z. & Karatzas, G. P. (2015) Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed. *Journal of Environmental Management* **154**:86-101.

Leedal, D., Neal, J., Beven, K., Young, P. & Bates, P. (2010) Visualization approaches for communicating real-time flood forecasting level and inundation information. *Journal of Flood Risk Management* **3(2)**:140-150.

Lin, J.-Y., Cheng, C.-T. & Chau, K.-W. (2006) Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal* **51(4)**:599-612.

Liong, S.-Y. & Sivapragasam, C. (2002) Flood Stage Forecasting With Support Vector Machines. *JAWRA Journal of the American Water Resources Association* **38(1)**:173-186.

Liu, J., Gong, M., Zhao, J., Li, H. & Jiao, L. (2016) Difference representation learning using stacked restricted Boltzmann machines for change detection in SAR images. *Soft Computing* **20(12)**:4645-4657.

Liu, Y. & Pender, G. (2015) A flood inundation modelling using v-support vector machine regression model. *Engineering Applications of Artificial Intelligence* **46, Part A**:223-231.

Liu, Y., Pender, G. & Neélz, S. (2009) Improving the Performance of Fast Inundation Models Using v-Support Vector Regression and Particle Swarm Optimisation. *In the 33rd IAHR 2009 Congress*:1436–1443.

Makkeasorn, A., Chang, N. B. & Zhou, X. (2008) Short-term streamflow forecasting with global climate change implications – A comparative study between genetic programming and neural network models. *Journal of Hydrology* **352(3)**:336-354.

Misra, D., Oommen, T., Agarwal, A., Mishra, S. K. & Thompson, A. M. (2009) Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosystems Engineering* **103(4)**:527-535.

Moore, R. J. (2007) The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci.* **11(1)**:483-499.

Nourani, V., Kisi, Ö. & Komasi, M. (2011) Two hybrid Artificial Intelligence approaches for modeling rainfall–runoff process. *Journal of Hydrology* **402(1–2)**:41-59.

Okkan, U. (2012) Wavelet neural network model for reservoir inflow prediction. *Scientia Iranica* **19(6)**:1445-1455.

Pilling, C., Price, D., Wynn, A., Lane, A., Cole, S. J., Moore, R. J. & Aldridge, T. (2014) From drought to floods in 2012: operations and early warning services in the UK. In *Hydrology in a Changing World: Environmental and Human Dimensions Proceedings of FRIEND-Water*) IAHS Publ., Montpellier, France, vol. 363.

Price, D., Hudson, K., Boyce, G., Schellekens, J., Moore, R. J., Clark, P., Harrison, T., Connolly, E. & Pilling, C. (2012) Operational use of a grid-based model for flood forecasting. *Proceedings of the Institution of Civil Engineers - Water Management* **165(2)**:65-77.

Raghavendra, N., Sujay & Deka, P. C. (2014) Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing* **19**:372-386.

Romanowicz, R. J., Young, P. C. & Beven, K. J. (2006) Data assimilation and adaptive forecasting of water levels in the river Severn catchment, United Kingdom. *Water Resources Research* **42(6)**.

Romanowicz, R. J., Young, P. C., Beven, K. J. & Pappenberger, F. (2008) A data based mechanistic approach to nonlinear flood routing and adaptive flood level forecasting. *Advances in Water Resources* **31(8)**:1048-1056.

Sazal, M. M. R., Biswas, S. K., Amin, M. F. & Murase, K. (2014) Bangla handwritten character recognition using deep belief network. In *2013 International Conference on Electrical Information and Communication Technology (EICT)*.), pp. 1-5.

Vijayashanthar, V., Qiao, J., Zhu, Z., Entwistle, P. & Yu, G. (2018) Modeling Fecal Indicator Bacteria in Urban Waterways Using Artificial Neural Networks. *Journal of Environmental Engineering* **144(6)**:05018003.

Wu, C. L., Chau, K. W. & Li, Y. S. (2008) River stage prediction based on a distributed support vector regression. *Journal of Hydrology* **358(1)**:96-111.

Wu, C. L., Chau, K. W. & Li, Y. S. (2009) Methods to improve neural network performance in daily flows prediction. *Journal of Hydrology* **372(1–4)**:80-93.

Young, A. R., Grew, R. & Holmes, M. G. R. (2003) Low Flows 2000: a national water resources assessment and decision support tool. *Water Science and Technology* **48(10)**:119-126.

Accepted manuscript doi: 10.1680/jwama.19.00001

Yu, P.-S., Chen, S.-T. & Chang, I. F. (2006) Support vector regression for real-time flood stage forecasting. *Journal of Hydrology* **328(3–4)**:704-716.

Table 1. Descriptive statistics for all three catchments

	Lune at Killington (m^3/s)	**Leven at Newby Bridge (m^3/s)**	**Eden at Kirkby Stephen (m^3/s)**
Minimum	1.21	1.35	0.19
Maximum	621	220.75	139
Q5	1.32	2.21	0.21
Q95	103.77	90.88	34.51
Mean	26.69	30.51	8.06
Std. deviation	56.34	36.45	17.53
CV	211.06	119.47	217.51

Table 2. The DBN parameters

Parameters	**Values**
Number of hidden layer nodes	100
Number of hidden layers	5
Learning rate for the RBM	0.01
Learning rate for the DBN	0.01
Number of epochs for RBM	10
Number of back propagation iterations	200
Batch size	10
Activation function	'relu' (Rectified Linear Unit)

Table 3. Comparing five different error statistics (RMSE, MARE, MB, COR and NSC) for three forecasting techniques (WANN, SVR and DBN) along one to six hours ahead forecast in future for the river Lune at Killington data.

Hour	**RMSE_WANN**	**RMSE_SVR**	**RMSE_DBN**	**MARE_WANN**	**MARE_SVR**	**MARE_DBN**	**MB_WANN**	**MB_SVR**	**MB_DBN**	**COR_WANN**	**COR_SVR**
One	12.045	**5.841**	16.115	0.106	**0.039**	0.224	1.007	**0.544**	0.409	0.974	**0.994**
Two	21.496	**13.122**	21.064	0.142	**0.080**	0.270	1.304	1.015	**0.608**	0.937	**0.968**
Three	**21.141**	21.377	27.097	0.159	**0.117**	0.306	3.188	1.382	**0.380**	0.914	**0.918**
Four	**21.421**	30.070	32.973	0.175	**0.162**	0.340	3.431	1.220	**0.798**	**0.920**	0.844
Five	**26.087**	38.563	39.117	0.239	**0.206**	0.367	-5.248	2.016	**-0.634**	**0.913**	0.747
Six	**26.663**	44.767	44.074	**0.233**	0.243	0.408	1.788	3.325	**0.800**	**0.925**	0.650

Accepted manuscript doi: 10.1680/jwama.19.00001

Table 4. Comparing five different error statistics (RMSE, MARE, MB, COR and NSC) for three forecasting techniques (WANN, SVR and DBN) along one to six hours ahead forecast in future for the Leven at Newby Bridge data.

Hour	RMSE_WANN	RMSE_SVR	RMSE_DBN	MARE_WANN	MARE_SVR	MARE_DBN	MB_WANN	MB_SVR	MB_DBN	COR_WANN	COR_SVR	COR_DBN
One	1.004	**0.547**	1.298	0.012	**0.007**	0.017	-0.204	**0.006**	-0.593	0.998	**0.999**	0.997
Two	1.444	**0.738**	1.814	0.014	**0.010**	0.023	0.179	**0.014**	-0.553	0.995	**0.999**	0.993
Three	1.541	**0.986**	2.180	0.016	**0.013**	0.027	0.059	**-0.010**	-0.557	0.995	**0.998**	0.990
Four	1.615	**1.272**	2.751	0.018	**0.016**	0.035	0.291	**-0.033**	-0.281	0.994	**0.996**	0.983
Five	1.863	**1.518**	3.293	**0.019**	**0.019**	0.041	0.089	**-0.059**	-0.495	0.992	**0.995**	0.976
Six	1.863	**1.857**	3.808	**0.020**	0.023	0.047	**-0.009**	-0.110	-0.512	**0.992**	**0.992**	0.969

Table 5. Comparing ideal point error statistics for three forecasting techniques (WANN, SVR and DBN) along one to six hours ahead forecast in future for all three case studies.

	Lune at Killington			**Leven at Newby Bridge**			**Eden at Kirkby Stephan**		
Hour	**IPE_WANN**	**IPE_SVR**	**IPE_DBN**	**IPE_WANN**	**IPE_SVR**	**IPE_DBN**	**IPE_WANN**	**IPE_SVR**	**IPE_DBN**
One	0.332	**0.132**	0.418	0.538	**0.219**	0.559	0.341	**0.099**	0.410
Two	0.518	**0.268**	0.559	0.611	**0.299**	0.577	0.369	**0.230**	0.427
Three	0.602	**0.400**	0.543	0.590	**0.387**	0.624	0.651	**0.396**	0.476
Four	0.635	**0.513**	0.756	0.807	**0.507**	0.570	**0.570**	0.583	0.797
Five	0.861	**0.687**	0.758	0.713	**0.639**	0.746	0.820	0.724	**0.633**
Six	**0.720**	0.883	0.883	**0.707**	0.866	0.829	0.835	0.878	**0.806**
Avg	0.611	**0.481**	0.653	0.661	**0.486**	0.651	0.598	**0.485**	0.592

Table 6. Comparing five different error statistics (RMSE, MARE, MB, COR and NSC) for three forecasting techniques (WANN, SVR and DBN) along one to six hours ahead forecast in future for the Eden at Kirkby Stephen data.

Hour	RMSE_WANN	RMSE_SVR	RMSE_DBN	MARE_WANN	MARE_SVR	MARE_DBN	MB_WANN	MB_SVR	MB_DBN	COR_WANN	COR_SVR	COR_DBN	NW
One	4.165	**1.818**	4.835	0.073	**0.029**	0.184	-0.304	**0.210**	0.514	0.980	**0.996**	0.970	0
Two	4.462	**3.817**	6.351	0.113	**0.067**	0.197	**-0.108**	0.560	0.410	0.975	**0.982**	0.947	0
Three	**4.715**	6.246	8.178	0.184	**0.111**	0.220	**0.772**	1.073	0.319	**0.973**	0.950	0.912	**0.**
Four	**7.133**	9.205	10.247	0.164	**0.156**	0.318	**-0.198**	1.637	1.091	**0.938**	0.888	0.860	**0.**
Five	**6.918**	11.776	12.319	0.229	**0.195**	0.275	0.904	1.940	**0.256**	**0.938**	0.809	0.791	**0.**
Six	**8.334**	14.309	14.190	**0.228**	0.232	0.302	**0.802**	2.360	0.856	**0.919**	0.708	0.713	**0.**

Table 7. Gamma test results

	Lune at Killington			**Leven at Newby Bridge**			**Eden at Kirkby Stephen**		
***Γ*- Stat**	1 Hour	3 Hour	6 Hour	1 Hour	3 Hour	6 Hour	1 Hour	3 Hour	6 Hour
Gamma	0.0009	0.0028	0.0219	-0.0015	-0.0019	-0.0023	0.0005	0.0096	0.0499
Gradient	0.0404	0.0857	0.1557	0.0451	0.0520	0.0612	0.0459	0.0682	0.0841
Std. Error	0.0005	0.0009	0.0028	0.0002	0.0003	0.0005	0.0006	0.001	0.0034
V-ratio	0.0035	0.0112	0.0877	-0.0059	-0.0077	-0.0094	0.0018	0.0383	0.1996

Figure captions

Figure 1. Selected catchments for the study.

Figure 2. Observed flow data used to train the models. The Y axes is normalized flow data.

Figure 3. Comparing the predicted flow against validation data at different lead times for Lune at Killington catchment.

Figure 4. Comparing residuals of three forecasting techniques (WANN, SVR and DBN) for one (row 1), three (row 2), and six hours (row 3) ahead forecast for the river Lune at Killington data.

Figure 5. Comparing the predicted flow against validation data at different lead times for Leven at Newby Bridge catchment.

Figure 6. Comparing residuals of three forecasting techniques (WANN, SVR and DBN) for one (row 1), three (row 2), and six hours (row 3) ahead forecast for the Leven at Newby Bridge data.

Figure 7. Comparing the predicted flow against validation data at different lead times for Eden at Kirkby Stephen catchment.

Figure 8. Comparing residuals of three forecasting techniques (WANN, SVR and DBN) for one (row 1), three (row 2), and six hours (row 3) ahead forecast for the Eden at Kirkby Stephen data.

Accepted manuscript doi:
10.1680/jwama.19.00001

Figure 9. Pdd plots of model residuals. Row 1, 2 and 3 represent Lune at Killington, Leven at Newby Bridge and Eden at Kirkby Stephen respectively. Column 1, 2 and 3 compares densities at one, three and six hours ahead forecast.

Accepted manuscript doi:
10.1680/jwama.19.00001

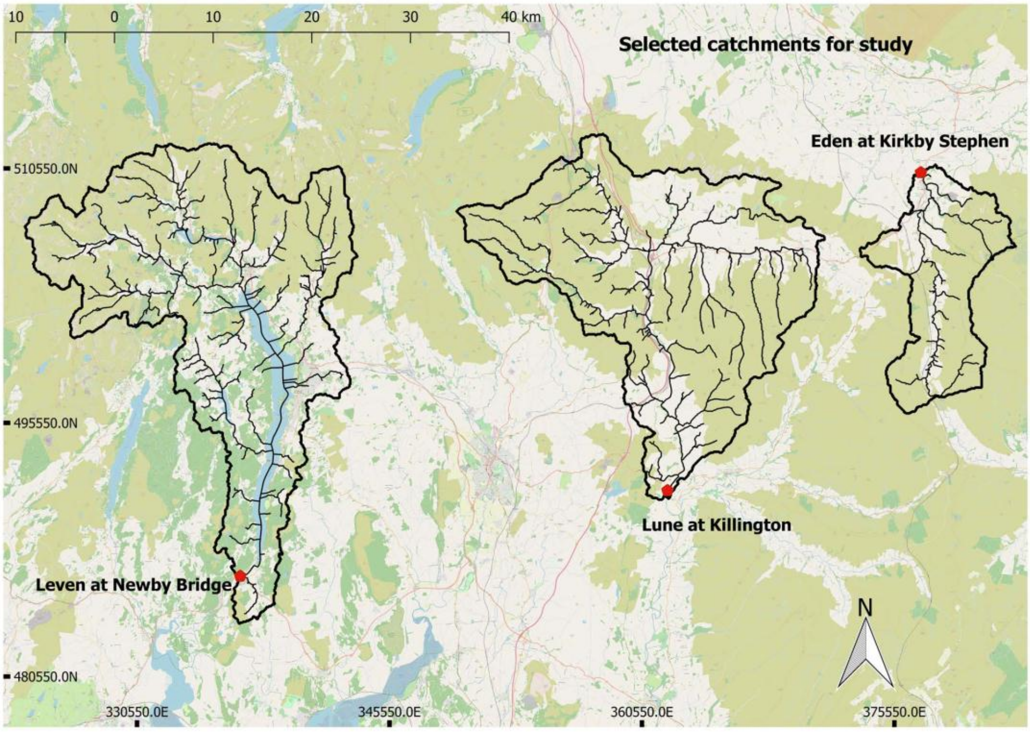

Figure 1.jpg

Accepted manuscript doi: 10.1680/jwama.19.00001

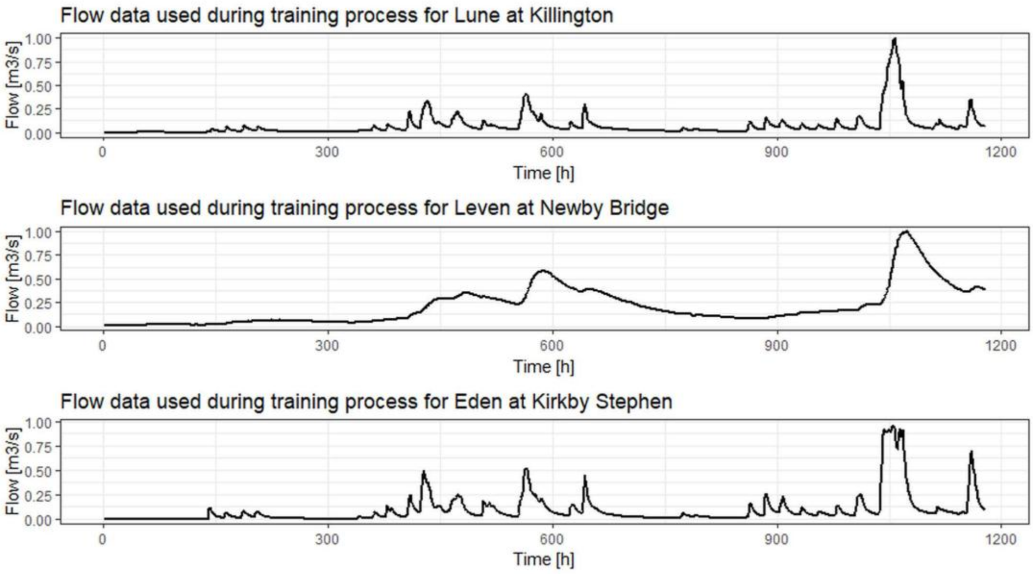

Figure 2.jpg

Accepted manuscript doi:
10.1680/jwama.19.00001

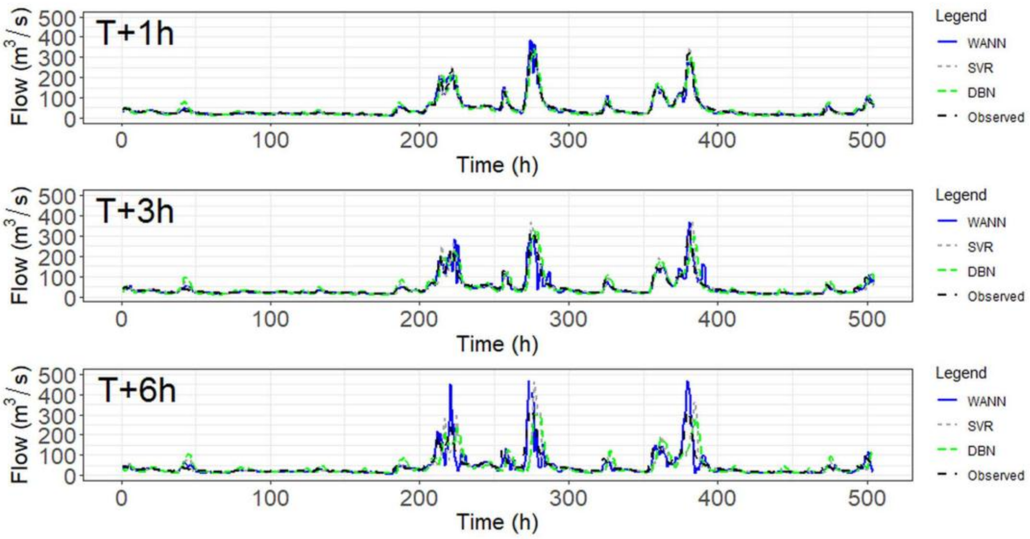

Figure 3.tif

Accepted manuscript doi:
10.1680/jwama.19.00001

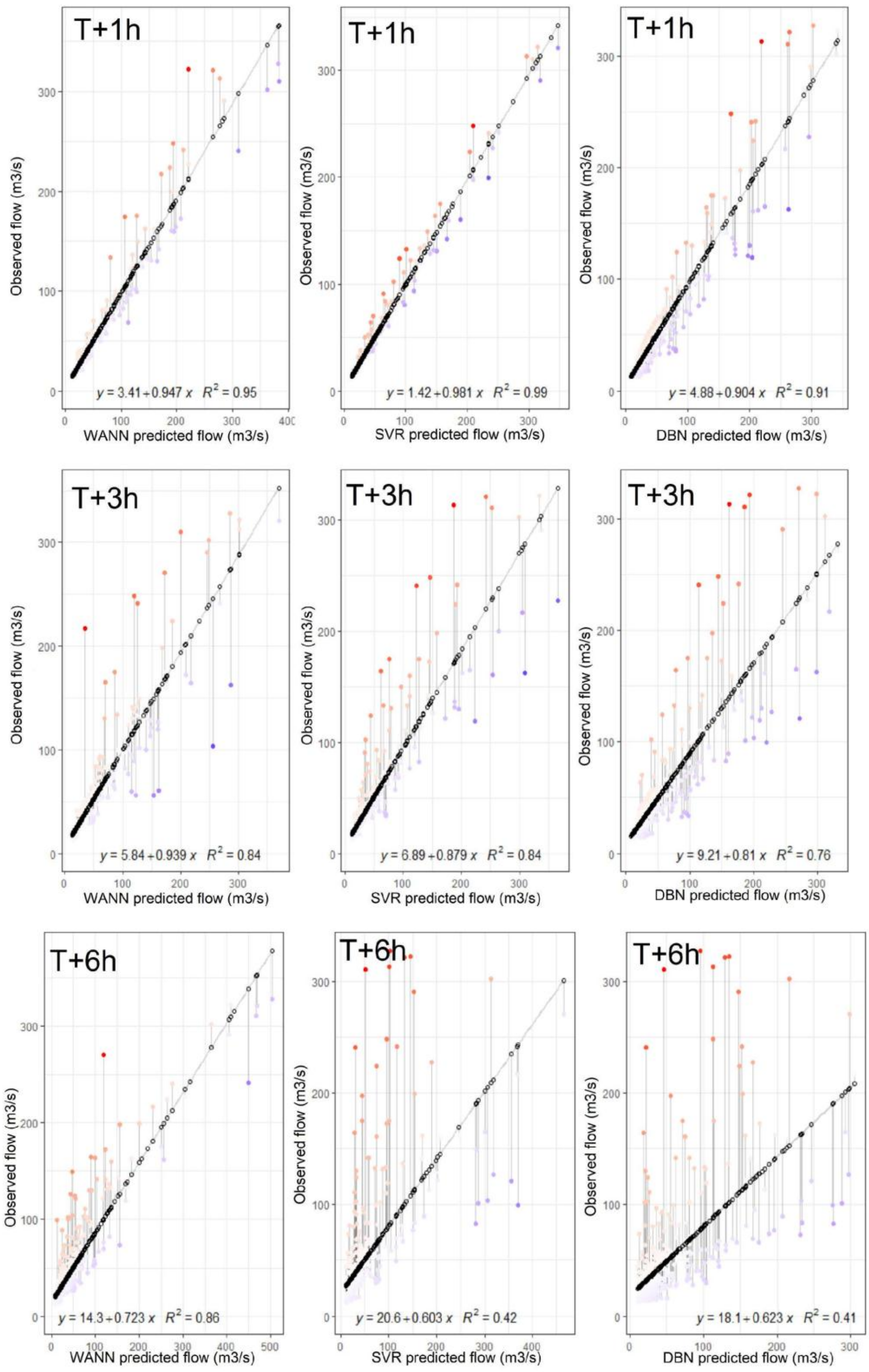

Figure 4.jpeg

Accepted manuscript doi:
10.1680/jwama.19.00001

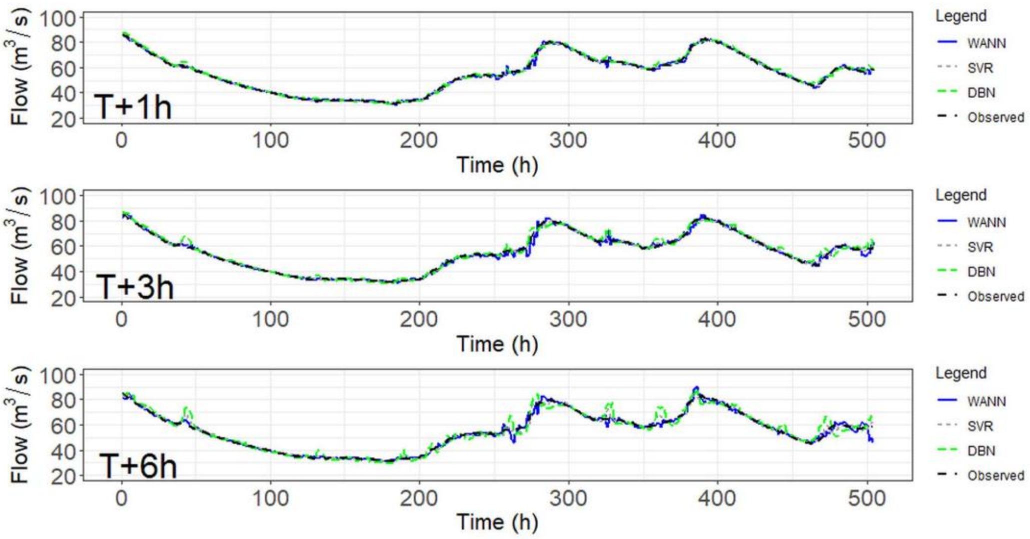

Figure 5.tif

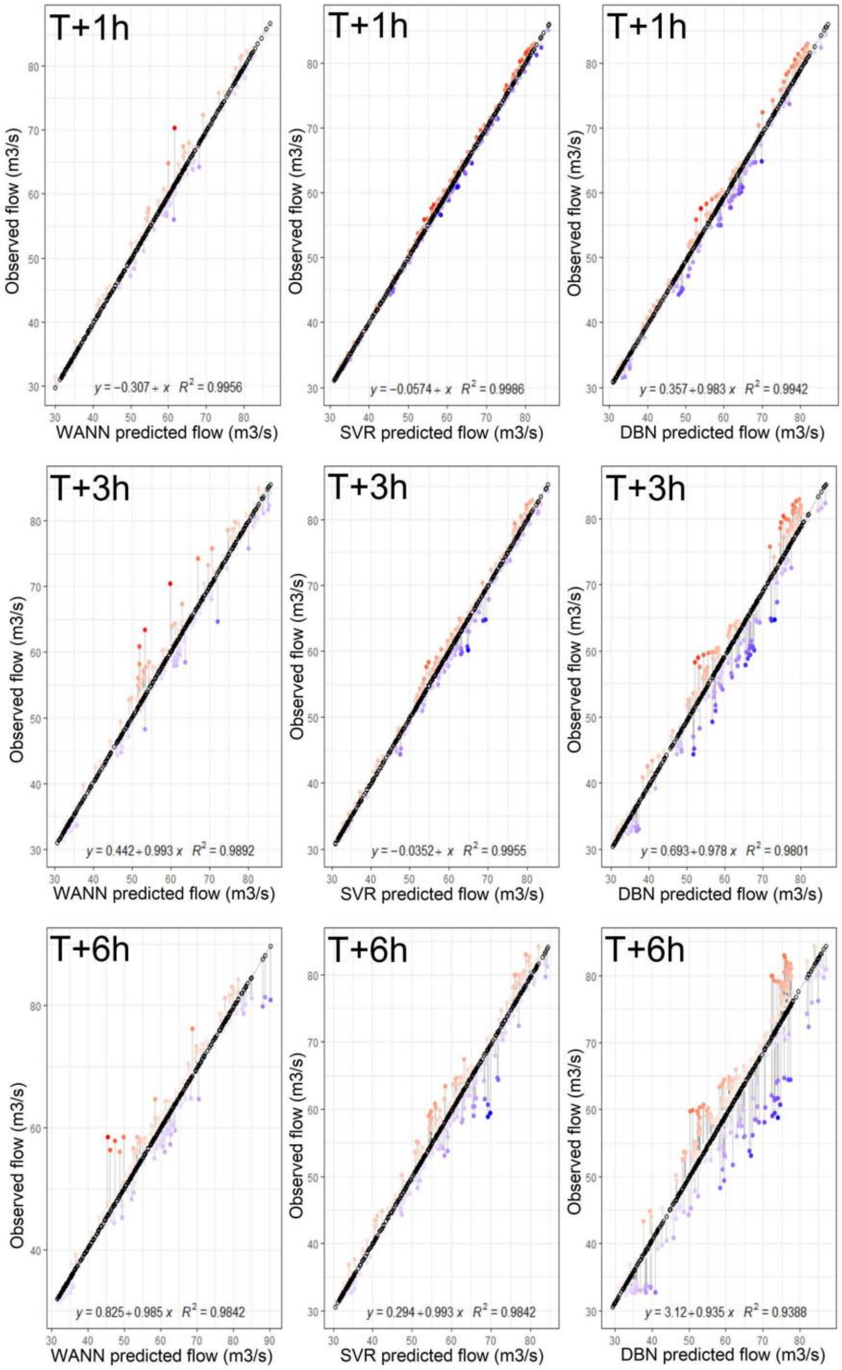
T+1h
T+1h
T+1h
y = −0.307 + x   R² = 0.9956
y = −0.0574 + x   R² = 0.9986
y = 0.357 + 0.983 x   R² = 0.9942
WANN predicted flow (m3/s)
SVR predicted flow (m3/s)
DBN predicted flow (m3/s)
Observed flow (m3/s)
T+3h
T+3h
T+3h
y = 0.442 + 0.993 x   R² = 0.9892
y = −0.0352 + x   R² = 0.9955
y = 0.693 + 0.978 x   R² = 0.9801
T+6h
T+6h
T+6h
y = 0.825 + 0.985 x   R² = 0.9842
y = 0.294 + 0.993 x   R² = 0.9842
y = 3.12 + 0.935 x   R² = 0.9388

Figure 6.jpg

Accepted manuscript doi: 10.1680/jwama.19.00001

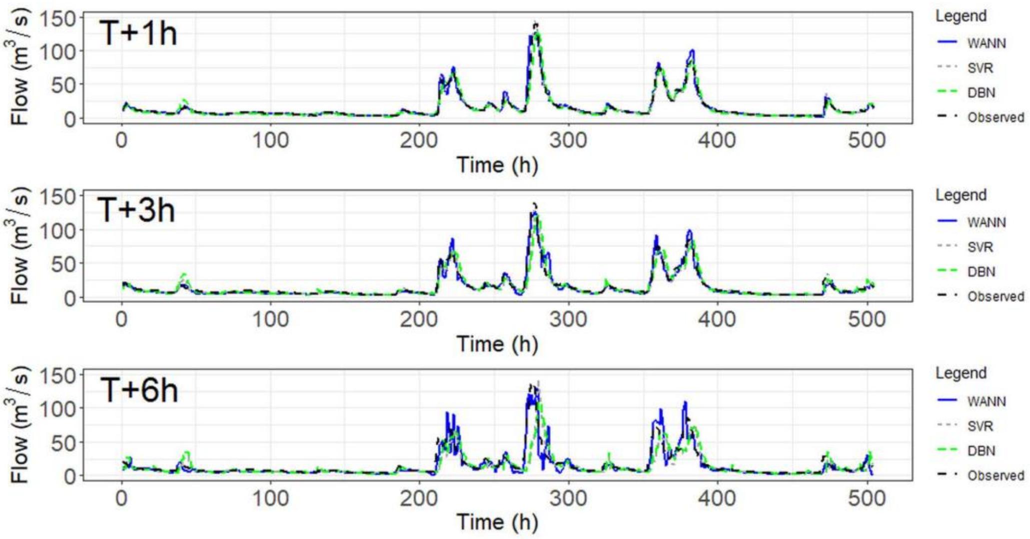

Figure 7.tif

Accepted manuscript doi:
10.1680/jwama.19.00001

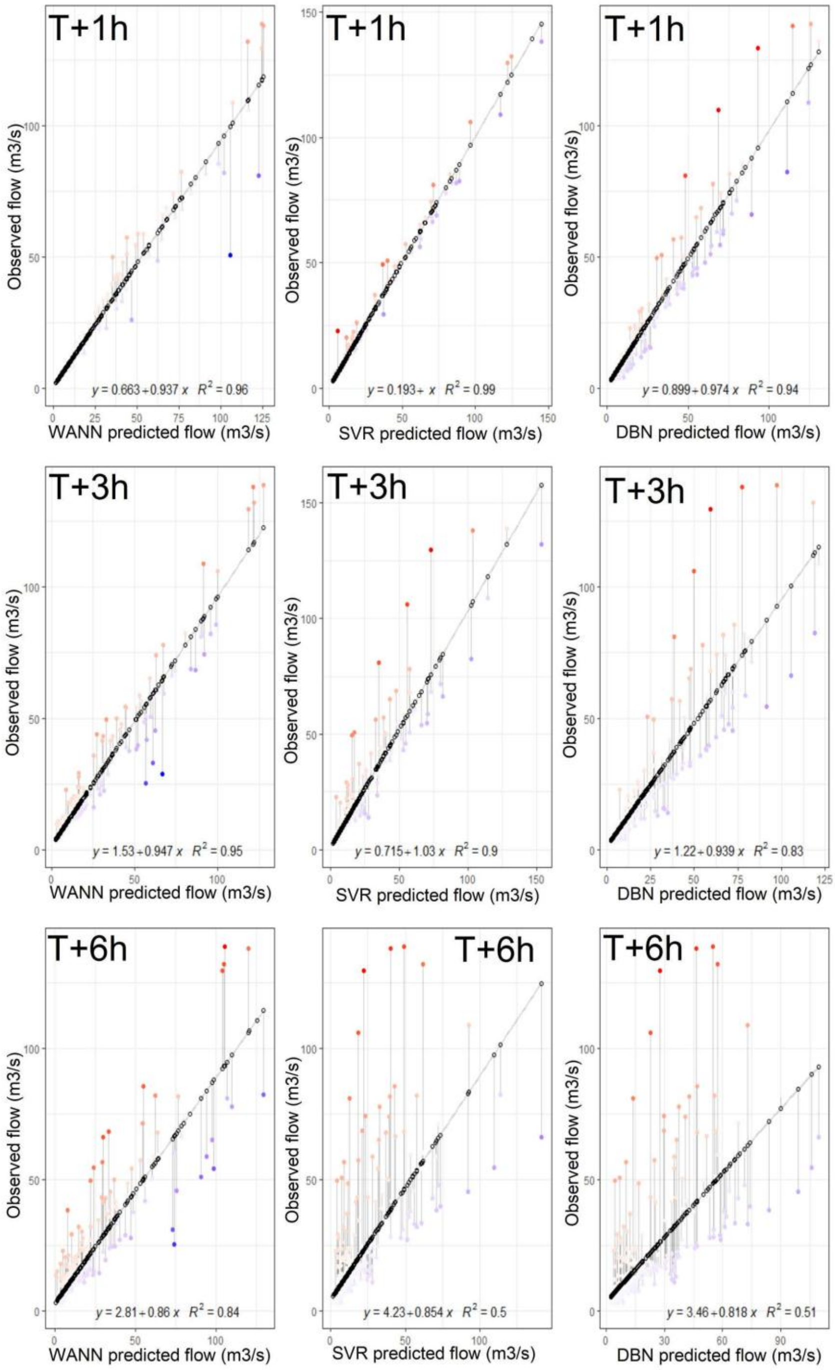

Figure 8.jpg

Accepted manuscript doi:
10.1680/jwama.19.00001

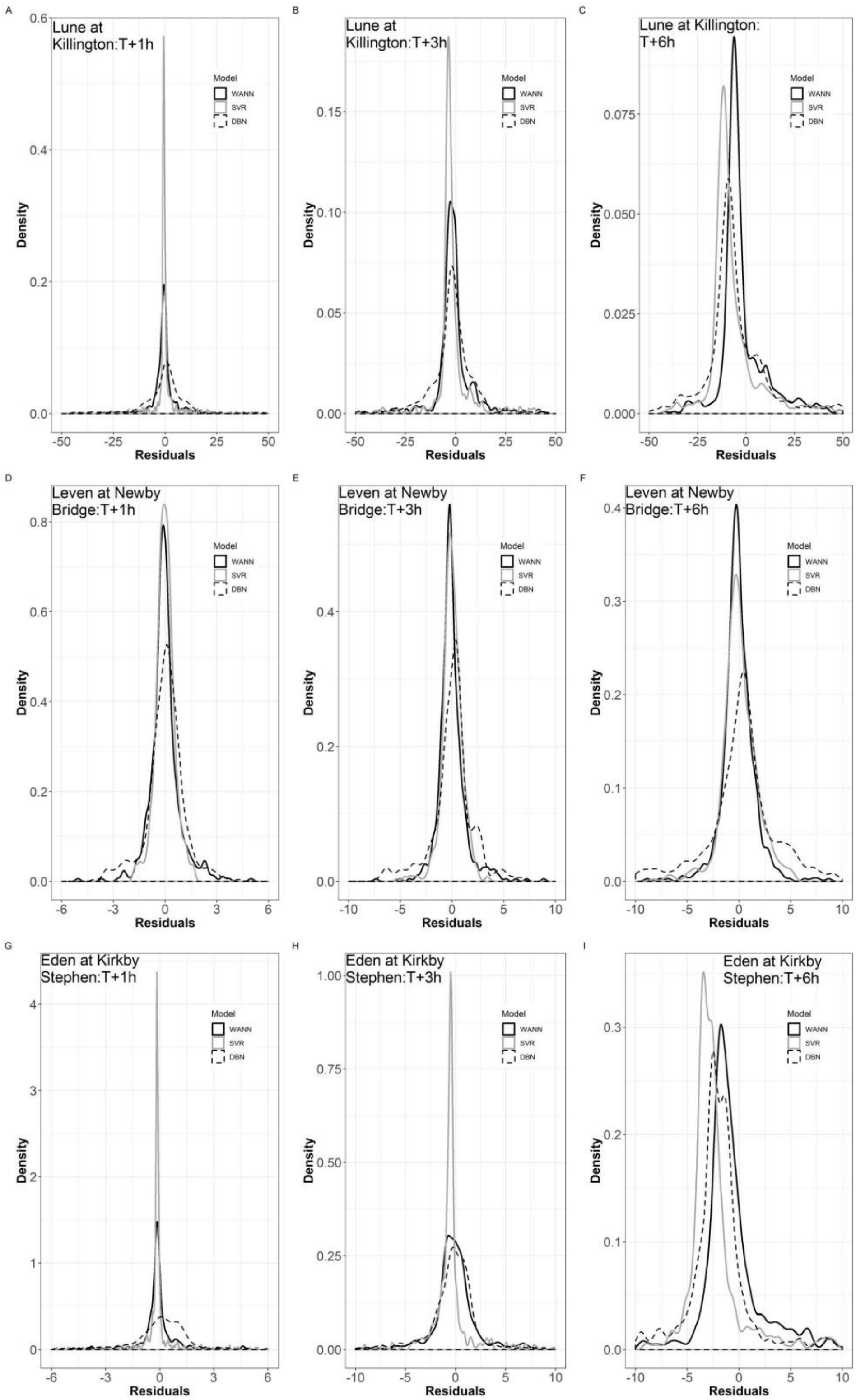

Figure 9.tiff