Short-term Stream Flow Forecasting at Australian River Sites using Data-driven Regression Techniques

Melise Steyn[1,2], Josefine Wilms[2], Willie Brink[1], and Francois Smit[1]

[1] Applied Mathematics, Stellenbosch University, Stellenbosch, 7600, South Africa,
{melisedt,wbrink,fsmit}@sun.ac.za, http://appliedmaths.sun.ac.za
[2] Council for Scientific and Industrial Research, Stellenbosch, 7600, South Africa,
jwilms@csir.co.za, http://www.csir.co.za

Abstract. This study proposes a computationally efficient solution to stream flow forecasting for river basins where historical time series data are available. Two data-driven modeling techniques are investigated, namely support vector regression and artificial neural network. Each model is trained on historical stream flow and precipitation data to forecast stream flow with a lead time of up to seven days. The Shoalhaven, Herbert and Adelaide rivers in Australia are considered for experimentation. The predictive performance of each model is determined by the Pearson correlation coefficient, the root mean squared error and the Nash-Sutcliffe efficiency. The performance of our data-driven models are compared to that of a physical stream flow forecasting model currently supplied by Australia's Bureau of Meteorology. It is concluded that the data-driven models have the potential to be useful stream flow forecasting tools in river basin modeling.

Keywords: Stream Flow Forecasting, Support Vector Regression, Artificial Neural Networks

1 Introduction

Stream flow is an important component in the hydrological cycle and plays a vital role in many hydraulic and hydrological applications. Research on model-generated stream flow is used by river engineers and scientists for the study of various hydro-environmental aspects, such as the increasing international concern of riverine pollution and the growing flood stages of rivers [5]. The devastating consequences of natural disasters, such as floods, can be lessened or even prevented through accurate stream flow forecasts [15].

Two main types of stream flow forecasting models can be distinguished, based on available information: physical and empirical. A physical model consists of governing partial differential equations that describe the physical laws of a specific system. Empirical or data-driven models are based on observed data that characterize the system [16].

A physical rainfall-runoff model can be used to transform rainfall estimations to runoff by modeling the hydrologic processes within a catchment, such as interception, evaporation, overland and subsurface flow [8]. According to Perrin *et al.* [14], it can be challenging to choose an appropriate model structure and complexity for accurate simulation of hydrological behavior at catchment scale.

During the past few decades, considerable progress has been made in the study of data-driven models to simulate the rainfall-runoff relationship [16]. Various processes within a river basin are characterized by measurable state variables, such as stream flow, precipitation, temperature and humidity. A river basin for which historical time series data are available is therefore a good candidate for the implementation of data-driven models.

In this paper the practicality of data-driven models for stream flow forecasting with a lead time of up to seven days are investigated. In particular, two supervised machine learning models are constructed, namely support vector regression (SVR) and artificial neural network (ANN). Australian river sites are considered, mainly because of a sufficient amount of available historical stream flow and precipitation data.

The Bureau of Meteorology (BOM), Australia's national weather and climate agency, provides a forecasting service that supplies stream flow predictions at more than 100 locations across Australia. These forecasts are determined by a computer based system which uses a rainfall-runoff model known as GR4H as its main component.[3] It determines the total amount of rainfall in a specific catchment, the fraction of rainfall that ends up as runoff, and the accumulation of that runoff in downstream rivers [14]. Forecasts are given for a lead time of up to seven days, and are used for several water management purposes. The predictive capabilities of our data-driven models will be compared to the BOM rainfall-runoff model.

2 Overview of SVR and ANN

We proceed with a cursory theoretical overview of the two data-driven prediction methods considered in this paper.

2.1 Support Vector Regression

Support vector machines were originally developed to solve classification problems, but have been extended to the task of regression and time series prediction in the form of support vector regression (SVR). Many hydrological prediction problems have been addressed using SVR [15].

Consider a training set of n real-valued data pairs $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i$ is an input vector in some space X, with corresponding output value y_i. A generalized continuous-valued target function $f(\mathbf{x})$ is fit to the training set, such that a deviation of at most ϵ is obtained between each true

[3] `http://www.bom.gov.au/water/7daystreamflow/about.shtml`

ouput and its corresponding predicted value, and that $f(\mathbf{x})$ is as flat as possible [6]. Assuming f to be linear, we may write

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \tag{1}$$

where $\mathbf{w} \in X$, $b \in \mathbb{R}$, and $\langle \cdot, \cdot \rangle$ denotes a dot product in X. In order to get f as flat as possible, the orientation parameter (or weight) $\mathbf{w}$ should be minimized. Some of the data pairs might exceed the ϵ margin of error and cause the optimization problem to be infeasible. We introduce slack variables, denoted as ξ and ξ^*, to indicate the vertical distance from each data pair above and below the ϵ margins. The convex optimization problem is solved by minimizing

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*). \tag{2}$$

The positive penalty parameter C determines the tolerated deviations larger than ϵ. The minimization of (2) is a standard constrained optimization problem and can be solved by applying Lagrangian theory [4]. The weight vector is derived as

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathbf{x}_i, \tag{3}$$

where α_i and α_i^* represent Lagrangian multipliers associated with the training points above and below the regression line, respectively. The value of b in equation (1) is computed by exploiting the Karush-Kuhn-Tucker conditions [7, 10], as explained by Granata *et al.* [6].

In many applications the relationship between inputs and outputs in the training data might show complex nonlinear behavior. A kernel function can be introduced to implicitly map the training points from the original input space X to a higher dimensional feature space $\Phi(X)$, such that a linear relationship between the variables exist in $\Phi(X)$. The support vector expansion of the target function for linear regression is then applicable in the feature space. Equation (1) changes to

$$f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}) + b, \tag{4}$$

where

$$k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle. \tag{5}$$

The radial basis function (RBF) is a widely used kernel in hydrological prediction applications [6], and is defined as

$$k(\mathbf{x}, \mathbf{y}) = \exp(-\gamma\|\mathbf{x} - \mathbf{y}\|^2), \tag{6}$$

where $\gamma > 0$ is a kernel-specific hyperparameter. Choosing an optimal value for λ, as well as for ϵ and C, is important when training an SVR model to fit a given dataset [6].

2.2 Artificial Neural Network

Artificial neural networks (ANNs) are especially suitable when the underlying functions that describe complex phenomena are unknown [20], and have been used extensively for hydrological modeling purposes [11].

An ANN contains a set of interconnected nodes that receive, process and send information to one another over weighted connections. These nodes are grouped in different layers. Input values enter the model through the first layer (the input layer). The data is then fed forward through successive hidden layers until it reaches the final layer (the output layer). The hidden layers enable the ANN to learn complex relationships between input and output data [16]. An ANN can be single layered, bilayered or multilayered, depending on the number of hidden layers.

ANNs are further classified as feed-forward or recurrent, based on the direction of information flow and processing between nodes. Feed-forward ANNs allow information to travel only from the input layer to the output layer, while recurrent ANNs allow information to travel in both directions. For each node, an output is determined by calculating the sum of its weighted input nodes and applying a nonlinear activation function. According to Maier and Dandy [11], the sigmoidal-type and logistic sigmoidal-type (such as tanh) activation functions are frequently used in hydrological applications:

$$\text{sigmoidal-type: } g(z) = \frac{2}{1 + \exp(-2z)} - 1, \tag{7}$$

$$\text{tanh: } g(z) = \frac{1}{1 + \exp(-z)}, \tag{8}$$

where z represents the weighted sum of a particular node's inputs. This result is then used as input for the nodes in a succeeding layer. A linear activation function is considered for the final hidden layer of regression models [11].

Training is achieved by finding an optimal set of connection weights that minimize the estimated error between the true output values and the output values that are determined by the network.

3 Methodology

A description of the procedures to construct SVR and ANN models for stream flow forecasting at specific river sites follows.

3.1 Study Area and Data

High quality time series of daily stream flow and precipitation data for the Australian river sites under study were obtained from the Australian Bureau of Meteorology's Hydrologic Reference Stations (HRS) and Climate Data Online (CDO) services, respectively. The HRS network consists of over 200 river sites that are mostly unaffected by water-related systems, such as dam construction

and irrigation services, and located in different hydro-climatic regions across Australia. CDO provides access to precipitation records from the Australian Data Archive for Meteorology.

Three Australian river sites are considered for this study: the Shoalhaven River at Fossikers Flat in New South Wales, the Herbert River at Abergowrie in Queensland, and the Adelaide River at Railway Bridge in Northern Territory. The Shoalhaven River is located in a temperate climate region and has a catchment size of 4660 km^2. The stream flow data for this site were obtained from gauging station 215207 (150.18° E, 34.82° S) and the corresponding precipitation from station 068085, 5.3 km away from station 215207. The Herbert River is in a subtropical climate region and has a catchment size of 7488 km^2. Its stream flow data were obtained from gauging station 116006B (145.92° E, 18.49° S) and the precipitation data from station 032091, 8.7 km away from station 116006B. The Adelaide River is in a tropical climate region and has a catchment size of 638 km^2. Its stream flow data were obtained from gauging station G8170002 (131.11° E, 13.24° S) and the precipitation data from station 014237, 3.3 km away from station G8170002.

Only uninterrupted time series data were used for training: data from 1 January 2000 to 31 December 2014 for training the data-driven models at the Shoalhaven and Herbert rivers, and data from 1 January 2008 to 31 December 2012 for the Adelaide river. For all three river sites, data from 5 February 2017 to 5 May 2017 were used as test data.

3.2 Input Selection, Data Preprocessing and Cross Validation

A moving time window is considered for the generation of input and output data pairs. For each measured stream flow value (which is considered as an output value), a corresponding input vector contains the precipitation and stream flow values of the preceding p-day and q-day time windows, respectively. For this study, p ranges from 0 to 2 and q from 2 to 5. P represents precipitation, Q represents stream flow, t refers to the current day and d refers to the forecasting lead time. An output value Q_{t+d} then has an input vector $\{P_t, P_{t-1}, \cdots, P_{t-p}, Q_t, Q_{t-1}, \cdots, Q_{t-q}\}$. For each model that forecasts with a lead time of d days, an exhaustive search is followed during training to find optimal values for p and q.

Data preprocessing is implemented by normalizing the values in the dataset to a range of $[0, 1]$. This ensures that the influence of large feature values (like stream flow) does not dominate that of smaller feature values (like precipitation) during the training process.

As discussed in Section 3.1, the available datasets are split into a training set and a test set. In order to obtain a model that generalizes well to unseen data, 10-fold cross validation is introduced, i.e. the full training dataset is split into 10 folds of equal size. Each fold is considered as a validation set once, while the remaining 9 folds are combined to form a training set. Ultimately, the model with the lowest average validation error on all 10 trials is used for forecasting purposes, and tested on the test set [16].

3.3 Model Performance Evaluation

Three quantitative indices are considered to evaluate the performance of the SVR and ANN models, and to compare them to the physically based BOM model. These are the Pearson's correlation coefficient (r), the root mean squared error (RMSE) and the Nash-Sutcliffe efficiency (NSE):

$$\mathrm{r} = \frac{\sum_{i=1}^{m}(y_i - \overline{y})(f_i - \overline{f})}{\sqrt{\sum_{i=1}^{m}(y_i - \overline{y})^2}\sqrt{\sum_{i=1}^{m}(f_i - \overline{f})^2}}, \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - f_i)^2}, \tag{10}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{m}(y_i - f_i)^2}{\sum_{i=1}^{m}(y_i - \overline{y})^2}, \tag{11}$$

where y_i and f_i represent each of the m true and forecasted outputs in the test set, respectively. The average of all true outputs is represented by $\overline{y}$ and the average of all forecasted outputs by $\overline{f}$.

Pearson's correlation coefficient gives the extent to which the input and output values are linearly correlated, and ranges between -1 and 1. A value close to -1 or 1 shows a strong linear relationship between the two variables, whereas values close to zero show little to no linear relationship. If the predicted values of the model increase as the input values increase, a positive r-value is obtained. If the predicted values decrease as the input values increase, a negative r-value is obtained.

The RMSE measures the difference between a model's predicted outcomes and the true outcomes from the system that is being modeled. The smaller the RMSE value, the better the performance of the model.

The NSE is used to assess the predictive power of a model and is always less than or equal to 1. A model with an NSE of 1 corresponds to a perfect match of predicted outcomes to true outcomes. An NSE of 0 indicates that the model's predictive capability is the same as considering the mean true outcome value as a predictor. An NSE less than 0 occurs when the mean true outcome value would have been a more reliable predictor than the model [9]. According to Noori and Kalin [13], a model can be considered "good" if the NSE is above 0.5, and "very good" if it is above 0.7.

3.4 SVR Hyperparameters

The SVR model with an RBF kernel is considered for this study. Three parameters have to be selected, namely C, ϵ and γ. We pick possible C values ranging from 1 to 10^4, ϵ values from 10^{-3} to 10^{-1}, and γ values from 10^{-4} to 1. An exhaustive grid search is performed to find the combination of parameters with optimal performance during training and cross validation.

3.5 ANN Architecture

According to Maier and Dandy [11], a one hidden layered feed-forward neural network provides suitable complexity to reproduce the nonlinear behavior of hydrological systems and has been suitable for forecasting hydrological variables in various studies.

It can be challenging to choose an appropriate number of hidden nodes within the hidden layer, as too few might result in a network that cannot capture the complex relationship between input and output, while too many may cause overfitting. This study uses two different methods as bounds for the number of hidden nodes, as proposed by Belayneh and Adamowski [1]. Wanas *et al.* [18] determined that the optimal performance of a neural network is obtained with $\log(n)$ hidden nodes, where n is the number of training samples. Mishra and Desai [12] showed that optimal results are obtained with $2N+1$ hidden nodes, where N is the number of input nodes. Following Belayneh and Adamowski [1], a trial and error approach can be implemented during training to find the optimal number of hidden nodes ranging from $\log(n)$ to $2N+1$.

As discussed, the sigmoidal-type and logistic sigmoidal-type activation functions, given in equations (7) and (8), have been used frequently in hydrological applications. We implement both, and pick the one that achieves the lowest error during training and cross validation.

4 Results and Discussion

Results for the optimal input features, hyperparameter combinations for SVR and architecture for ANN are discussed in the following subsections. The predictive capabilities of our data-driven models are also evaluated, based on the criteria listed in Section 3.3.

4.1 Parameter Selection

Different lead times are considered for stream flow forecasting, ranging from 1 day to 7 days in advance. As stated in Section 3.2, the preceding time windows for stream flow and precipitation that provide an optimal model are found separately during training for each of the different prediction lead times. For SVR, an optimal combination of hyperparameters is also determined, whereas for ANN, an optimal number of hidden nodes and the choice of activation function. Results are listed in Tables 1 and 2.

It can be observed that, when considering different prediction lead times, the preceding time windows for stream flow and precipitation and the combination of model parameters vary. It is also noticeable that only the optimal ANN and SVR models for 7 day lead time forecasting of the Shoalhaven river site do not consider any rainfall values. Apart from this particular case, it appears that rainfall is an important input to the data-driven models for the three considered river sites. Furthermore, each ANN model achieved the lowest error during training and cross validation when considering the tanh activation function.

Table 1. Optimal input features and hyperparameters in the SVR models for the three gauging stations (C, ϵ and γ are SVR parameters; the model uses precipitation data from days $t-p$ to t and stream flow data from days $t-q$ to t to predict stream flow on day $t+d$, with d the lead time).

Lead	**Shoalhaven**					**Herbert**					**Adelaide**				
time (d)	p	q	C	ϵ	γ	p	q	C	ϵ	γ	p	q	C	ϵ	γ
1 day	3	2	100	0.001	0.1	5	2	100	0.001	0.1	5	2	1	0.001	1
2 day	2	1	10	0.001	0.1	5	1	1000	0.001	0.1	4	2	1000	0.001	0.01
3 day	2	2	1	0.001	0.1	4	1	10000	0.01	0.1	5	1	10000	0.01	0.01
4 day	3	2	100	0.001	0.001	4	1	10000	0.01	0.1	5	2	10000	0.01	0.01
5 day	3	2	100	0.001	0.001	2	2	1000	0.001	0.001	5	2	10000	0.01	0.01
6 day	2	2	100	0.001	0.01	2	1	10000	0.01	0.1	5	2	10000	0.01	0.01
7 day	2	0	10	0.001	0.1	2	1	10000	0.01	0.1	2	1	10000	0.01	0.1

Table 2. Optimal input features and architecture (number of nodes in the hidden layer, h) in the ANN models for the three gauging stations.

Lead	**Shoalhaven**			**Herbert**			**Adelaide**		
time (d)	p	q	h	p	q	h	p	q	h
1 day	3	2	9	5	2	12	3	2	4
2 day	3	1	5	3	1	8	4	2	6
3 day	5	1	11	5	2	5	5	2	3
4 day	5	2	10	4	1	4	5	2	3
5 day	5	2	10	4	1	4	3	2	8
6 day	2	2	4	4	1	4	5	1	4
7 day	4	0	3	3	1	3	5	2	3

4.2 Performance Evaluation

The efficiency criteria used in this study are the Pearson correlation coefficient, the root mean squared error and the Nash-Sufcliffe efficiency. Based on these performance indices, the SVR and ANN models that performed optimally on the training and validation sets were applied to the (as yet unused) test sets of the three river sites under study. Results are shown in Tables 3 to 5. For comparison, prediction accuracies made by the Bureau of Meteorology's stream flow forecasting model are also given.

ANN outperforms the SVR and BOM models for stream flow predictions with a lead time of 1 to 2 days at the Shoalhaven river site. The base flow as well as the rising and falling limbs of the hydrographs are well represented by the ANN model. However, the peaks are under- and over-predicted. As the prediction lead time increases, the accuracy of each model decreases. Figures 1a and 1b show how the time lag between observed and forecasted peaks increase. Furthermore, Figure 1c shows that the SVR and ANN models fail to forecast the rising limbs of the hydrograph for predictions with a lead time longer than

Table 3. Performance evaluation for stream flow forecasting at the Shoalhaven river station of our trained SVR and ANN models as well as the physically based model used by the Australian Bureau of Meteorology (BOM).

Lead	**r**			**RMSE**			**NSE**		
time	SVR	ANN	BOM	SVR	ANN	BOM	SVR	ANN	BOM
1 day	0.87	0.90	0.85	541	458	866	0.74	0.81	0.34
2 day	0.74	0.81	0.71	807	629	1357	0.43	0.65	−0.61
3 day	0.71	0.65	0.59	948	826	1601	0.22	0.41	−1.22
4 day	0.66	0.46	0.52	1040	969	1446	0.07	0.19	−0.79
5 day	0.54	0.52	0.28	1118	933	3272	−0.07	0.26	−8.13
6 day	0.39	0.33	0.18	1150	1263	5760	−0.11	−0.35	−27.11
7 day	0.32	0.26	0.18	1195	1094	3957	−0.20	−0.01	−12.17

Table 4. Performance evaluation for stream flow forecasting at the Herbert river station of our trained SVR and ANN models as well as the physically based model used by the Australian Bureau of Meteorology (BOM).

Lead	**r**			**RMSE**			**NSE**		
time	SVR	ANN	BOM	SVR	ANN	BOM	SVR	ANN	BOM
1 day	0.93	0.92	0.95	1627	1728	1748	0.85	0.83	0.83
2 day	0.79	0.82	0.90	2721	2525	2152	0.59	0.64	0.74
3 day	0.70	0.73	0.82	3067	2952	3138	0.48	0.52	0.45
4 day	0.59	0.60	0.74	3514	3691	3707	0.32	0.25	0.25
5 day	0.52	0.53	0.28	5996	3936	12911	−0.95	0.16	−8.04
6 day	0.42	0.48	0.12	4154	3982	23648	0.08	0.15	−28.97
7 day	0.38	0.40	0.09	4116	4391	21508	0.10	−0.02	−23.49

Table 5. Performance evaluation for stream flow forecasting at the Adelaide river station of our trained SVR and ANN models as well as the physically based model used by the Australian Bureau of Meteorology (BOM).

Lead	**r**			**RMSE**			**NSE**		
time	SVR	ANN	BOM	SVR	ANN	BOM	SVR	ANN	BOM
1 day	0.79	0.85	0.84	975	944	925	0.61	0.63	0.65
2 day	0.63	0.67	0.54	1287	1277	1466	0.33	0.34	0.12
3 day	0.53	0.51	0.40	1376	1965	1671	0.21	−0.62	−0.17
4 day	0.41	0.43	0.25	1533	1983	1870	0.00	−0.67	−0.49
5 day	0.38	0.47	0.13	1561	1980	2036	−0.03	−0.66	−0.75
6 day	0.33	0.35	0.16	1567	2188	1856	−0.03	−1.00	−0.44
7 day	0.27	0.26	0.19	1602	2158	1835	−0.06	−0.92	−0.39

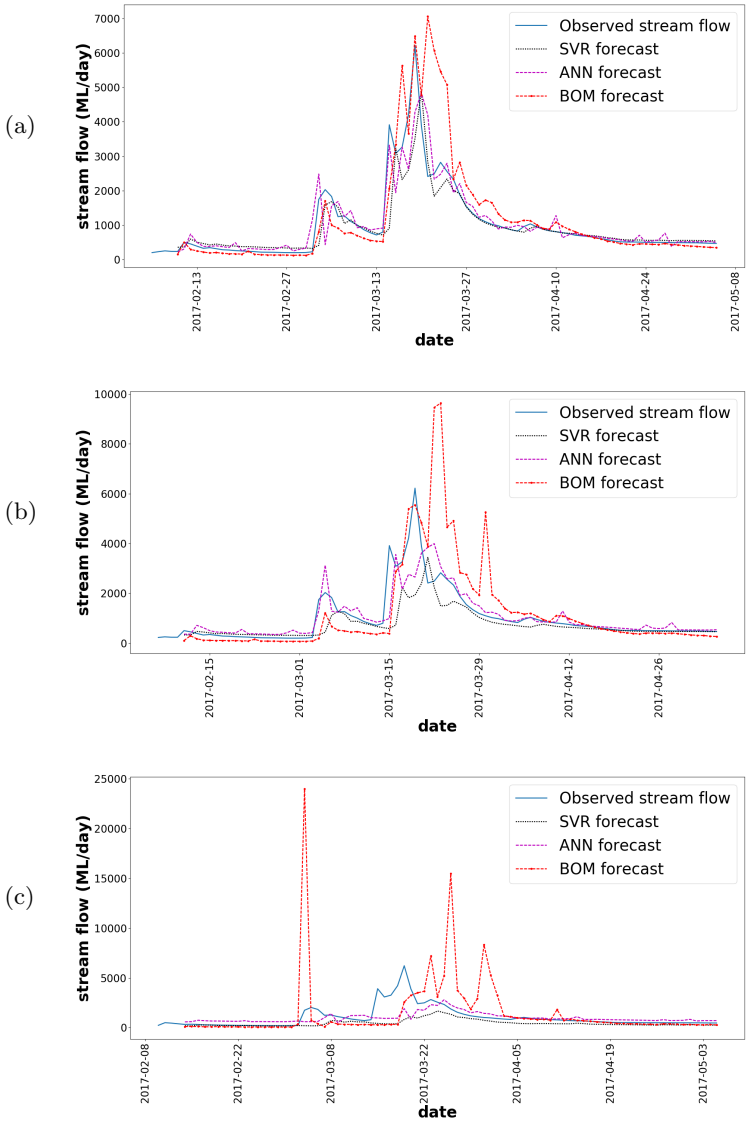

Fig. 1. Daily stream flow predictions for (a) 1 day, (b) 2 day and (c) 5 day lead time forecasts, for the Shoalhaven station.

4 days. This can be attributed to the absence of information (such as rainfall events) when increasing the prediction lead time. For lead times greater than 3 days, the SVR forecasts show the strongest correlation to the observed stream flow, whereas the ANN generally performs better in terms of RMSE and NSE. For 6 and 7 day lead time predictions, the NSE of all three models are negative,

indicating that the mean value of the observed outcomes would have been a more reliable predictor than the forecasting models.

No single model outperforms the rest on the test set of the Herbert river station. For instance, the BOM model obtains the strongest Pearson correlation (0.95) to the observed stream flow when forecasting 1 day in advance, but fails to determine the peaks as accurately as the SVR model. The BOM model does, however, show the better performance in forecasting stream flow with a lead time of 2 days. Similar to the Shoalhaven river models, an increase in prediction lead time causes a decrease in model performance and an increase in lag times between observed peaks and forecasted peaks.

The BOM and ANN models show the better performance on the test set of the Adelaide river station for 1 and 2 day lead time predictions. For instance, as seen in Table 5, comparable r, RMSE and NSE results are obtained for both models. The SVR model shows the better forecasting performance for predictions with a lead time greater than 2 days. Similar to both Herbert and Shoalhaven, the prediction capabilities of all three models worsen with an increase in prediction lead time.

5 Conclusion

This study investigated the ability of data-driven modeling for stream flow forecasting with a lead time of up to 7 days. SVR and ANN models were employed to forecast stream flow at the Shoalhaven, Herbert and Adelaide gauging stations. The predictive capabilities of these data-driven models were compared to that of a physically based rainfall-runoff model. For 1 day lead time forecasts, each data-driven model properly modeled the stream hydrograph shape and the time to peak. However, a noticeable decrease in predictive capabilities with an increase in lead time occurred. The SVR method performed better than the BOM model for the Shoalhaven station, based on the evaluation criteria. For the other stations, no single model outperformed the others.

Based on the results obtained for this study, SVR and ANN models have the potential to be useful tools for short-term stream flow forecasting. They do not require specialized knowledge of physical phenomena, and are therefore especially useful when it is difficult to build a physically based model due to a lack of understanding of the underlying processes. It is also helpful to have modeling alternatives and to validate results obtained from physically based models to that of data-driven models. Furthermore, data-driven models are computationally efficient in the sense that once they are trained, predictions can be made very quickly. Data-driven models could also be combined with physically based models to form even more powerful and accurate hybrid forecasting models.

A limitation of data-driven models are, however, that substantial historical stream flow and precipitation data records should be available. Many of the existing gauging stations have limited available datasets, or a considerable amount of missing data. Developing machine learning techniques to address these problems may be considered in further studies.

References

1. Belayneh, A., Adamowski, J.: Drought Forecasting using New Machine Learning Methods. Journal of Water and Land Development. 18, 3-12 (2013)
2. Bureau of Meteorology Hydrologic Reference Stations, http://www.bom.gov.au/water/hrs/about.shtml
3. Climate Data Online, http://www.bom.gov.au/climate/data/?ref=ftr
4. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2. 121-167 (1998)
5. Falconer, R., Lin, B., Harpin, R.: Environmental Modeling in River Basin Management. International Journal of River Basin Management. 3(1), 169-184 (2005)
6. Granata, F., Gargano, R., De Marinis, G.: Support Vector Regression for Rainfall-runoff Modeling in Urban Drainage: a Comparison with the EPA's Storm Water Management Model. Water. 8(3), 69 (2016)
7. Karush, W.: Minima of Functions of Several Variables with Inequalities as Side Constrains. Master's Thesis, Department of Mathematics, University of Chicago, Chicago, IL, USA (1939)
8. Knapp, H.V., Durgunoglu, A., Ortel, T.W.: A Review of Rainfall-runoff Modeling for Stormwater Management. Illinois State Water Survey, Hydrology Division (1991)
9. Krause, P.; Boyle, D.P.; Bäse, F.: Comparison of Different Efficiency Criteria for Hydrological Model Assessment. Advances in Geosciences. 5, 89-97 (2005)
10. Kuhn, H.W., Tucker, A.W.: Nonlinear Programming. In Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics, pp. 481-492. University of California Press, Oakland, CA, USA (1951)
11. Maier, H.R., Dandy, G.C.: Neural Networks for the Prediction and Forecasting of Water Resources Variables: a Review of Modeling Issues and Applications. Environmental Modelling & Software. 15, 101-124 (2000)
12. Mishra, A.K., Desai, V.R.: Drought Forecasting using Feed-forward Recursive Neural Network. Ecological Modelling. 198(1-2), 127-138 (2006)
13. Noori, N., Kalin, L.: Coupling SWAT and ANN models for enhanced daily streamflow. Journal of Hydrology. 533, 141-151 (2016)
14. Perrin, C., Michel, C., Andréassian, V.: Improvement of a Parsimonious Model for Streamflow Simulation. Journal of Hydrology. 279, 275-289 (2003)
15. Raghavendra, S., Deka, P.C.: Support Vector Machine Applications in the Field of Hydrology: a Review. Applied Soft Computing. 19, 372-286 (2014)
16. Solomatine, D.P., Ostfeld, A. Data-driven Modeling: some Past Experiences and New Approaches. Journal of Hydroinformatics. 10(1), 3-22 (2008)
17. Vapnik, V.: The nature of statistical learning theory. Springer, New York, NY, USA (1995)
18. Wanas, N., Auda, G., Kamel, M.S., Karray, F. On the Optimal Number of Hidden Nodes in a Neural Network. In: Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, pp. 918-921. IEEE Press, Michigan (1998)
19. Wang, Y., Guo, S., Xiong, L., Liu, P., Liu, D.: Daily Runoff Model based on ANN and Data Preprocessing Techniques. Water. 7, 4144-4160 (2015)
20. Zhang, Z.: A Gentle Introduction to Artificial Neural Networks. Annals of Translational Medicine. 4(19), 1-6 (2016)