Neural Computing and Applications, Volume 32, Issue 7, April 2020, pp. 2457 - 2470
DOI:10.1007/s00521-018-3745-1

River Discharge Simulation Using Variable Parameter McCarthy–Muskingum and Wavelet- Support Vector Machine Methods

Basant Yadav[*1], Shashi Mathur[2]

[1]Cranfield Water Science Institute, Cranfield University,

Cranfield MK43 0AL, UK

[2]Department of Civil Engineering, Indian Institute of Technology Delhi, India

[1*]Corresponding author email- basant.yadav@cranfield.ac.uk ; basant1488@gmail.com

ABSTRACT

In this study an extended version of variable parameter McCarthy–Muskingum (VPMM) method originally proposed by Perumal and Price (2013) was compared with the widely used data based model namely support vector machine (SVM) and hybrid wavelet-support vector machine (WA-SVM) to simulate the hourly discharge in Neckar River wherein significant lateral flow contribution by intermediate catchment rainfall prevails during flood wave movement. The discharge data from the year 1999 to 2002 has been used in this study. The extended VPMM method has been used to simulate nine flood events of the year 2002 and later the results were compared with SVM and WA-SVM models. The analysis of statistical and graphical results suggest that the extended VPMM method was able to predict the flood wave movement better than the SVM and WA-SVM models. A model complexity analysis was also conducted which suggest that the two parameter based extended VPMM method has less complexity than the three parameters based SVM and WA-SVM model. Further, the model selection criteria also gives the highest values for VPMM in 7 out of 9 flood events. The simulation of flood events suggested that both the approaches were able to capture the underlying physics and reproduced the target value close to the observed hydrograph. However, the VPMM models is slightly more efficient and accurate, than the SVM and WA-SVM model which are based only on the antecedent discharge data. The study captures the current trend in the flood forecasting studies and showed the importance of both the approaches (Physical and data based modeling). The analysis of the study suggested that these approaches complements each other and can be used in accurate yet less computational intensive flood forecasting.

Keywords- Flood forecasting, VPMM, SVM, Wavelet Transform

1. Introduction

Accurate forecasting of discharge is extremely important in flood management, reservoir management and hydropower design. The accuracy in forecasting discharge depends on the type of simulation model adopted and a review of literature shows that long term and short term discharge forecasting models are being used extensively in various water management problems such as flood control, drought management, water supply utilities operations, irrigation supply management and sustainable development of water resources. In the last few decades, researchers have proposed many models to improve the accuracy of discharge forecasting. These models can be broadly classified as physically based, conceptual and data driven models. A physically based model include as much of small-scale physics and natural heterogeneity as is computationally possible by considering variables such as groundwater, precipitation, evapotranspiration, initial soil moisture content and temperature (Loague and Vander Kwaak, 2004). These can be further classified as hydraulic and hydrologic routing methods. The hydrologic routing methods are widely used in the field practices since early thirties and they have been developed essentially to overcome the tedious computations involved in the hydraulic routing methods (Perumal et al., 2017). Among the many lumped hydrological routing methods, the Muskingum method introduced by McCarthy (1938) is well known in the literature (Chow et al., 1988). The Muskingum method was studied by Ponce and Yevjevich (1978) resulting in the development of Variable Parameter Muskingum-

Cunge (VPMC) method. However, the VPMC method was criticised for the mass conservation problem (Perumal and Sahoo, 2008). To overcome this problem, Todini (2007) revisited the original Muskingum-Cunge (MC) flood routing approach and suggested that the error in mass conservation occurs due to the use of time variant parameters. Later, Price (2009) proposed a nonlinear Muskingum method as an approximation of the one-dimensional Saint-Venant equations and suggested a way out to include any uniformly distributed time-dependent lateral inflow along the river. Recently Perumal and Price (2013) proposed a fully mass conservative approach to study the flood wave propagation in channels (without lateral flow) named variable parameter Muskingum method based on the Saint-Venant equations. Although, these methods successfully captured the flood wave movements and also tackled the problem of mass conservation, the consideration of lateral flow along the river reach still was the cause for erroneous river discharge prediction. A separate approach was suggested by O'Donnell (1985) to include lateral flow in the Muskingum method assuming that the lateral flow has the same form as the inflow hydrograph as pointed out by Perumal et al. (2001). This concept was further studied by Karahan et al., (2014) using the approach of O'Donnell (1985) to incorporate lateral flow and proposed a nonlinear Muskingum flood routing model. This three parameter based semi-empirical Muskingum method has limitation about its applicability to only those events which were similar to the observed past events. To overcome this problem, Yadav et al., (2015) proposed an extended VPMM method considering uniformly distributed lateral flow along the river reach. This study extended the approach of Perumal and Price (2013) and successfully captured the significant amount of lateral flow due to intervening catchment rainfall. Recently, Swain and Sahoo (2015) also studied the fully mass conservative VPMM model and extended it to exclusively incorporate the spatially and temporally distributed non-uniform lateral flow while routing the flood events for compound river channel flows.

Although a physical method provide reasonable accuracy, their implementation and calibration typically present various difficulties (Nayak et. al., 2007). Moreover, in situations particularly in developing countries where the data about the processes to be modelled is limited, physically based model cannot be built, or they are inadequate. A well-calibrated conceptual model can also provide reasonable simulation accuracy, however, their uses are limited, because entire physical process in the hydrologic cycle is mathematically formulated in the conceptual models. Thus, they are composed of a large number of parameters making the model very complicated and slow. This in turn leads to problems of over parameterization (Beven, 2006) which may manifest itself in large prediction uncertainty (Uhlenbrook et al., 1999). In the last few decades, data driven techniques capable of handling large data sets have been adopted while dealing with water resources problems. In forecasting of river discharge, data-based hydrological methods are gaining popularity because they can be developed very rapidly with requirement of minimal information (Yadav et al., 2016b). Though they may lack the ability to provide a physical interpretation and insight into the catchment processes, they are nevertheless able to forecast relatively accurate discharge values (Adamowski and Sun, 2010). The lack of extensive data and cost of collection coupled with inaccessibility of sites compels one to select models based on past recorded flow data while simulating river flow variability (Kisi, 2008, Shiri and Kisi, 2010). Further, data-driven models that operate on an interrelationship between input-output data only without capturing the complete dynamics of the system, may therefore be preferred in certain cases (e.g., in contexts of limited data).

With the advent of computers and the availability of high computational facilities, many researchers have employed data driven techniques while forecasting discharge (e.g., Dawson and Wilby 1998; Sudheer et al. 2002; Shiri et al., 2012; Ghalkhani et al., 2013; Badrzadeh et al., 2013; Rezaeianzadeh et al., 2014; Kasiviswanathan et al, 2016). Much research has been carried out in the recent past on the use of artificial neural networks (ANN) for discharge forecasting since it is reliable and promising and plethora of literature is available with its applications. Study of hydrological processes using data based models mainly depends on the time series of the considered process. The length of the time series is also important as it captures the short term and long term trend of the process, which can also help in accurate simulation and prediction of the future events. The neural network based models were also used successfully for the trend analysis of time series (Maier and Dandy, 2000; Rafael et al., 2011; Lin et al., 2017). Similarly, Genetic programing (Koza, 1992) is another data based approach which has been successfully applied to many studies in water resources engineering problems. However, the most notable one was the support vector machine (SVM), a kernel based technique based on the Vapnik–Chervonenkis (VC) theory (Vapnik, 1995). The main advantage of this relatively new machine learning method is that it not only possesses the strengths of ANN but is able to overcome the problems associated with local minimum and network over fitting (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000). Further, despite the flexibility and usefulness of data driven methods in modeling hydrological processes, they have some drawbacks with highly non-stationary responses or seasonality (Cannas et al., 2006, Tiwari and Chatterjee, 2010, Adamowski and Chan, 2011, Nourani et al., 2014). To handle such problems a method called wavelet analysis (WA) has been used in various hydrological studies. Sang (2013a) highlighted that the understanding of hydrologic series can be improved from wavelet analysis. Recent application of wavelet analysis in hydrological modeling (Kalteh, 2013, Suryanarayana et al., 2014, Agarwal et al., 2016, Yadav et al., 2017) suggest that the WA approach provides a superior alternative to the data driven models and can enhance the accuracy by developing the more detailed input–output combinations. In light of the above facts, an attempt has been made herein to assess the abilities of the wavelet based support vector machine to predict the discharge in a river reach where the lateral flow is very significant. Further, we also intend to compare the two distinctively discharge prediction approaches to suggest an accurate yet less complex discharge prediction method for such a catchment conditions. The techniques were experimented on a 24.2 km stretch of Neckar River between Rottweil and Oberndorf.

2. Methodology

2.1 Variable parameter McCarthy–Muskingum (VPMM)

The fully mass conservative VPMM was developed by Perumal and Price (2013). After a decade of research, VPMM is capable to conserve volume absolutely and also follow the heuristic assumption of the prism and wedge storage established by McCarthy (1938) in the development of the classical Muskingum method. The method fundamentally makes use of a parallel approach followed by Perumal (1994a, 1994b) in the development of the VPM routing method. The VPMM method is developed from an approximation of the momentum equation of the Saint-Venant equations. This approximation is applied directly to the one-dimensional continuity equation of the Saint-Venant equations, leading to a fully conservative routing method which has the same routing equation as the classical Muskingum method proposed by McCarthy in (1938). The use of hydraulic principle in the development of the VPMM method allow the characterization of the considered channel reach storage into prism and wedge storage which complies with the heuristic assumption of McCarthy (1938) who developed the Muskingum method. The equation derived in the VPMM method for the travel time and weighting parameter is same as the classical Muskingum method and based on the flow and channel characteristics. The equations governing the one-dimensional unsteady flow in channels and rivers are given below (Perumal and Price (2013) as

$$\frac{\partial Q}{\partial x}+\frac{\partial A}{\partial t}=0 \tag{1}$$

$$S_f = S_o - \frac{\partial y}{\partial x} - \frac{v}{g}\frac{\partial v}{\partial x} - \frac{1}{g}\frac{\partial y}{\partial t} \tag{2}$$

The Eq. (1) and (2) represents the continuity and momentum equation, respectively. The discharge at any section of the routing reach using the VPMM method can be obtained using the equation as (Perumal and Price (2013) as

$$Q_M = Q_{o,M}\left\{1-\frac{1}{S_o}\frac{\partial y}{\partial x}\left[1-\frac{4}{9}F_M^2\left(\frac{P}{B}\frac{dR}{dy}\right)_M^2\right]\right\}^{1/2} \tag{3}$$

where, t is the time; x is the distance along the channel; y is the flow depth; v is the average cross-sectional velocity; A is the cross-sectional area; Q represents the discharge; g is the acceleration due to gravity; S_f is the frictional slope; S_o is the bed slope; $(\partial y/\partial x)$ is the longitudinal gradient of water profile; $(v/g)(\partial v/\partial x)$ is the convective acceleration slope and $(1/g)(\partial v/\partial t)$ is the local acceleration slope; P_M, B_M and R_M, respectively, represents the wetted perimeter, top width and hydraulic radius corresponding to flow depth y_m. The notation Q_M is the average discharge at the mid-section of the reach at any time and $Q_{o,M}$ is the normal discharge at the midsection corresponding to flow depth y_m and F_M is the Froude number.

The developed VPMM method was further modified to account lateral flow in flood routing study using the similar approach suggested by O'Donnell (1985). Though, the fundamental principle remains same, lateral flow was incorporated in a distributed form throughout the river stretch (Fig. 1). For the detailed explanation on the lateral flow estimation approach readers can refer to Yadav et al., (2015). Accordingly, the lateral flow hydrograph q_L is assumed to have the similar shape as the inflow hydrograph and it is supplied uniformly along the river stretch at each time interval. Hence, the original continuity equation in the VPMM method is modified as

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q_L \qquad (4)$$

where, q_L is the lateral flow per unit length of the channel.

The contribution of lateral flow in the river stretch is assumed to be perpendicular to the channel reach, hence the channel flow receives no or very negligible momentum. Accordingly, in the modified VPMM method the momentum equation (Eq. (2)) remains unaltered. The modified continuity equation and the original momentum equation were further solved to account the uniformly distributed lateral flow and the approach arrived at following (Yadav et al., 2015) as

$$Q_{i+1}^{j+1} = C_1 Q_i^{j+1} + C_2 Q_i^j + C_3 Q_{i+1}^j + C_4 q_{Lavg} \qquad (5)$$

The coefficients C_1, C_2, C_3, C_4 and q_{Lavg} are expressed as

$$C_1 = \frac{\Delta t - 2K^{j+1}\theta^{j+1}}{\Delta t + 2K^{j+1}\left(1-\theta^{j+1}\right)}$$

$$C_2 = \frac{\Delta t + 2K^{j}\theta^{j}}{\Delta t + 2K^{j+1}\left(1-\theta^{j+1}\right)}$$

$$C_3 = \frac{-\Delta t + 2K^{j}\left(1-\theta^{j+1}\right)}{\Delta t + 2K^{j+1}\left(1-\theta^{j+1}\right)}$$

$$C_4 = \frac{2K\Delta t\Delta x}{\Delta t + 2K^{j+1}\left(1-\theta^{j+1}\right)}$$

$$q_{Lavg} = \frac{q_{L,j+1} + q_{L,j}}{2}$$

Considering the shape of lateral flow hydrograph as same as the inflow hydrograph, the lateral flow rate q_L joining to the river stretch (discharge per unit length of the channel) is obtained (Yadav et al., 2015) as

$$q_L = \frac{I}{\sum_{i=1}^{N} I\Delta t} \times \frac{V_L}{L} \tag{6}$$

where, I is the inflow discharge at any time; L is the length of the river reach in meter; V_L is the volume of lateral flow .

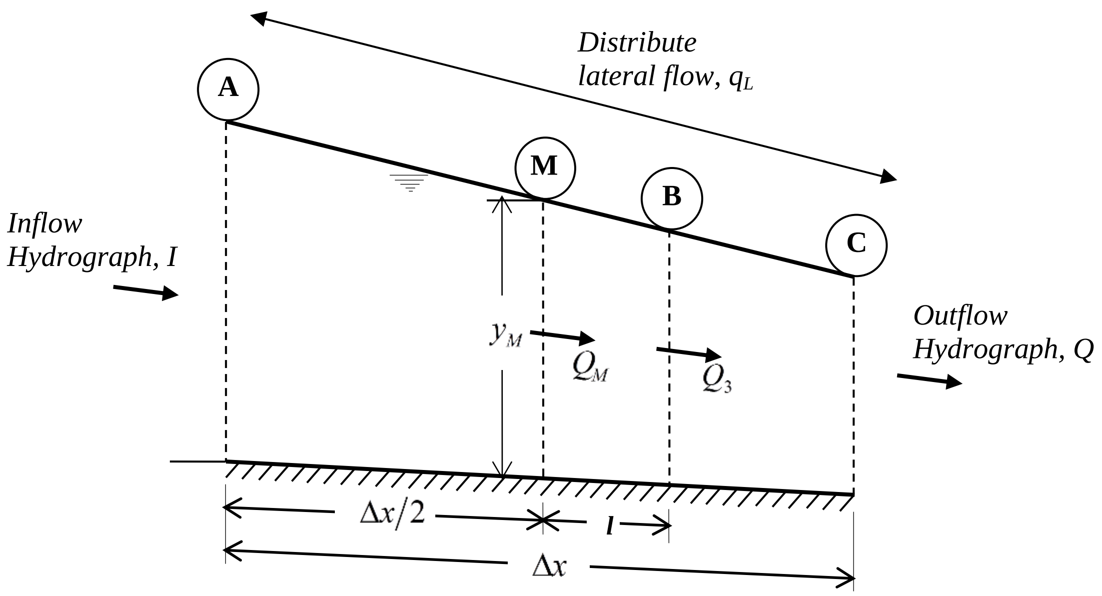

Fig: 1. Concept diagram of VPMM method considering distributed lateral flow in river reach (Yadav et al., 2015)

To calculate the discharge values at the downtream location, the VPMM method requires the following data- Manning's roughness value, Bed slope, River width (meter), Side slope and Cross-sectional shape. The method also requires river dischrage data of the upstream gauging station and rainfall data to calculate the lateral flow of the intervening catchment. As the

VPMM method is a fully mass conservative, physically based method, it does not require any calibration. The precipitation and discharge data of year 2002 was used in simulation of the discharge at the downstream location.

2.2 Support vector machine

Vapnik et. al. (1995) proposed a kernel-based algorithm as support vector machine (SVM) which has a function form like physical models, however, the level of complexity is to be decided by the data used to train the model. The method was developed using the similar principle like ANN, however by using a novel way to approximate various functions (i.e. linear (LN), polynomial (PL), radial basis function (RBF), and sigmoid (SIG)) using the method of structural risk minimization (opposite to the empirical risk minimization). A kernel function is used to transform the data into higher dimensional feature space. The SRM principle allows the method to have a good generalization ability for the unseen data. Let $\{(x_1, y_1)...,...,(x_n, y_n)\}$ be assumed to be the given training data sets, where $x_i \subset R^n$ represents the input sample space and $y_i \subset R^n$ for $i = 1,...,l$ denotes respective target output, elements in the training data set represented by l. Error tolerance level is fixed by a value of ε (errors < ε). The linear regression in SVM is estimated by solving the equation (7) as

$$Minimize\,\frac{1}{2}\|w\|^2 + C\sum_{i=0}^{n}\left(\xi + \xi^*\right) \tag{7}$$

$$\text{Subject to} \begin{cases} y_i - (w, x_i) - b \le \varepsilon_i + \xi \\ (w, x_i) + b - y_i \le \varepsilon_i + \xi^* \\ \xi_i \xi_i^* \ge 0, i = 1,...,l \end{cases}$$

w denotes the normal vector, b is a bias, C represents a regularization constant, ε is the error tolerance level of the function, and the ξ , ξ^* are slack variables.

The support vector machine have variety of kernel function (mathematical function) and its selection based on the problem at hand, which in turn has a direct impact on the accuracy of the model (Yao et al., 2008). Various studies suggest that the RBF has higher generalization ability and produce more accurate results than the other kernel types (Harpham and Dawson, 2006; Yang et al., 2009; Tehrany et al., 2015, Yadav et al., 2017). A study by Tehrany et al., (2014) suggested that RBF may produce less accurate results in case of longer range extrapolation. However, RBF as a kernel function for SVM used by many researchers in the past (Yu et al., 2004; Choy and Chan, 2003; Suryanarayana et al., 2014; Yadav et al., 2016a, Yadav et al., 2017, Yadav et al., 2018) and has been found to be suitable for simulation and prediction studies. RBF is defined as

$$K\left(X_i, X_j\right) = \exp\left(-\gamma\left\|X_i - X_j\right\|^2\right) \tag{8}$$

where X_i and X_j are vectors in the input space, such as the vectors of features computed from training and testing. γ is defined by, $\gamma = -\frac{1}{2\sigma^2}$ for which σ is the Gaussian noise level of standard deviation.

The output of the SVM is critically dependent on the parameters such as regularization constant(C)insensitive loss function(ε), and parameter of radial basis function(γ). Trial and error procedure was used in the present study to optimize these parameters based on the RMSE value. The trial continues by using different combinations of all three parameters till the value of RMSE was minimized. Once the optimal parameters are obtained, the methods requires time series of upstream and downstream gauging locations to simulate the discharge values at the downstream location. The effect of lateral flow on the downstream discharge values is automatically considered in the method as the lateral flow calculation is based on the input and output discharge data. The time series data from the year 1999 to 2001 was used for the training while the data from year 2002 was used for the testing.

2.3 Wavelet analysis

A wavelet analysis is based on Fourier analysis and was developed to analyze stationary and non-stationary data. Wavelet decomposition is a technique used in case of non-periodic and transient signals to extract the relevant time-frequency information by disintegrating the data into low frequency and high frequency components. Wavelet decomposition breaks the signal into low and high frequency components and utilizes the information hidden in the original signal. The lower frequency components (approximation) are obtained using low pass filter and captures the rapidly changing details of the signal. The higher frequency components (details) are obtained using high pass filter to encompass the slowly changing features of the signals. In this study, discrete wavelet transform (DWT) was used and the discharge time series was decomposed into four resolution interval. Thus, some features of the subseries can be seen more clearly than the original signal series. Though, DWT is able to decompose the time series in many interval, it is important to note that higher number of resolution may also slow down the computational speed. For each component a separate SVM model need to be developed and the decomposed component may be given as the input for SVM. Later, the output of the all the developed SVM (i.e. four in this case) will be summed to get the final output in the form of recomposed time series.

There are two basic form of wavelet analysis, continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The continuous wavelet transform (CWT) of a signal $x(t)$ is defined as follows (Kalteh, 2013):

$$CWT_x^{\psi}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-\tau}{s} \right) dt \tag{9}$$

where $\psi(t)$ is the mother wavelet function; s represents the scale parameter, τ is the translation parameter.

The discrete wavelet transform (DWT) is defined as follows:

$$\psi_{m,n}(t) = a^{-m/2}\psi\left(\frac{t - n\tau_0 a^m}{a^m}\right) \tag{10}$$

m and n is the resolution level and position which controls the scale and time; t is the time; a is a specified fixed dilation step greater than 1; τ_0 is the location parameter that must be greater than zero. The term $a^{-m/2}$ in the above equation normalizes the functions.

The two form of wavelet has been used in many studies, however it was observed that the CWT is computationally costly and requires large number of data. On the other hand the development and application of DWT is much simpler and easy to use (Adamowski and Chan, 2011; Kalteh, 2013). Therefore, DWT has been used in this study where a father wavelet function used for the extraction of low frequency components while the high frequency component is extracted by using a complementary of the father wavelet, a mother wavelet function. The decomposition of the data series is represented by the approximation series A_m and the detail series D_m. Later, the both the approximation and detail series were recomposed to get the final output of the model.

2.4. Evaluation criteria

The VPMM method was originally developed by Perumal and Price (2013) and further the extended version considering the lateral flow was evaluated by Yadav et al., (2015). In the flood forecasting study value of flood peak and its time of arrival is very important, hence in this study three important evaluation criteria which is error in peak discharge (Q_{er}), error in time to peak (t_{Qe}) and error in volume ($EVOL$) are adopted. The criteria for error in volume has different definition than the one proposed by Perumal and Price (2013) as in their method the objective was to assess the error in mass conservation. However, in this study the lateral inflow from the intervening catchment is very significant hence the mass reproduction at the downstream location is bound to have higher value than the upstream location. Therefore, this study evaluated the volume reproduction ability of the selected methods based on the observed discharge at the downstream location. Further, the performance of VPMM, SVM and WA-SVM was also evaluated using the statistical indicators like root mean square error (RMSE), normalized mean square error (NMSE) and coefficient of determination (R^2). The aforementioned statistical indicator gives the interpretation about the overall reproduction ability of the selected models, and may not provide the information that how the model behaved throughout the flood event. Therefore, another evaluation criteria called absolute average relative error (AARE) was adopted to assess the model performance at each discharge ordinate. Furthermore, the performance of the selected methods was also evaluated using graphical analysis where the closeness with which the proposed method reproduces the benchmark solution, including the closeness of shape and size of the hydrograph, can be measured using the Nash–Sutcliffe (NSE) efficiency criterion. The definition of RMSE, NMSE, NSE and R^2 can be found easily in the literature, however the definition for some of the specific performance measures are given as follows:

Error in peak discharge (Q_{er})

$$Q_{er} = \left(\frac{Q_s}{Q_o} - 1 \right) \times 100 \tag{11}$$

Relative error in time to peak (t_{Qe})

$$t_{Qe} = t_{Qs} - t_{Qo} \tag{12}$$

Error in volume ($EVOL$)

$$EVOL = \left[\left\{ \sum_{i=1}^{N} Q_{si} \Big/ \sum_{i=1}^{N} Q_{oi} \right\} - 1 \right] \times 100 \quad (13)$$

Absolute average relative error (AARE)

$$AARE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Q_{oi} - Q_{si}}{Q_{si}} \right| \times 100 \quad (14)$$

where Q_{er} represents the percentage error in simulated peak discharge; Q_s is the simulated peak discharge of the flood event at the downstream location (m^3/s); Q_o is the observed peak discharge of the flood event at the downstream location (m^3/s); t_{Qe} is the relative error in time to peak of the simulated flood event (hr); t_{Qs} time to peak of the simulated flood event (hr); t_{Qo} time to peak of the observed flood event (hr); $EVOL$ is the error in volume is simulated flood event (%); Q_{si} is the *i*th ordinate of the simulated flood event (m^3/s); Q_{oi} is the *i*th ordinate of the observed flood event (m^3/s) and N is the total number of ordinates in the flood event.

2.5 Evaluation of model complexity

The level of complexity of a specific model is tested using Akaike information criterion (AIC) and model selection criteria (MSC). The most appropriate model based on the model complexities is the one with the smallest values of the AIC and largest value of MSC. The performance measures are also defined as;

$$AIC = N\ln\left[\sum_{i=1}^{N}(Q_{oi}-Q_{si})^2\right]+2N_p \quad (15)$$

$$MSC = \ln\left[\frac{\sum_{i=1}^{N}(Q_{oi}-\overline{Q_s})}{\sum_{i=1}^{N}(Q_{oi}-Q_{si})}\right]-\frac{2N_p}{N} \quad (16)$$

where $\overline{Q_s}$ represents the average simulated discharge and N_p represents the number of model parameters.

3. Study area and data

The research work as a part of this study was mainly performed on a part of the Neckar River basin (Fig.2). This region is situated in the South- Western part of Germany in the state of Baden Württemberg. The river in the catchment is unaffected by large hydropower generation plants and other such water management structures or navigational reasons, which are the most common reasons influencing the runoff characteristics of the catchment area. The study area of this research is characterised by strong differences in altitude between the foothills of the Black Forest in the west, the valley of the Neckar in the centre and once again the steep ascent to the Schwäbische Alb in the east. The catchment consists of lots of narrow valleys. There is a wide variety of vegetation in the study catchment. In the western part of the catchment the soil is acidic and poor in minerals which supports only Spruce, fir and beech trees. The same forest is also found in the sandy soil of Keuper. The pasture, meadows, fruits, vines, ash tress, elm and lime trees are also found in the smaller pockets. The study was conducted between the two initial observation stations Rottweil and Oberndorf on the Neckar River. Distance between two stations is 24.2 km. The intermediate drainage area between two stations is 235 km^2 which is around 34 % of the total drainage area of Oberndorf gauging station. The hourly amounts of precipitation used in the VPMM for the period from 1999 to 2002 are obtained from 3 precipitation stations which are distributed in and around the catchment area. The data based modeling (SVM and WA-SVM) is based only on the discharge time series from 1999 to 2002 (Fig. 3) which was provided by the University of Stuttgart, Germany. The description of the study area is partly based on the description of Das (2006) and CCHYDRO (1999). The discharge time series from 1999-2001 was used for the training and, nine flood events from the year 2002 are selected for the comparative analysis of the selected methods. The event selection was completely random but keeping in mind that the peak discharge value should be high and lateral flow contribution must be more than 10% in all events. The parameters for the application of extended VPMM were taken from the study of Yadav et al., (2015).

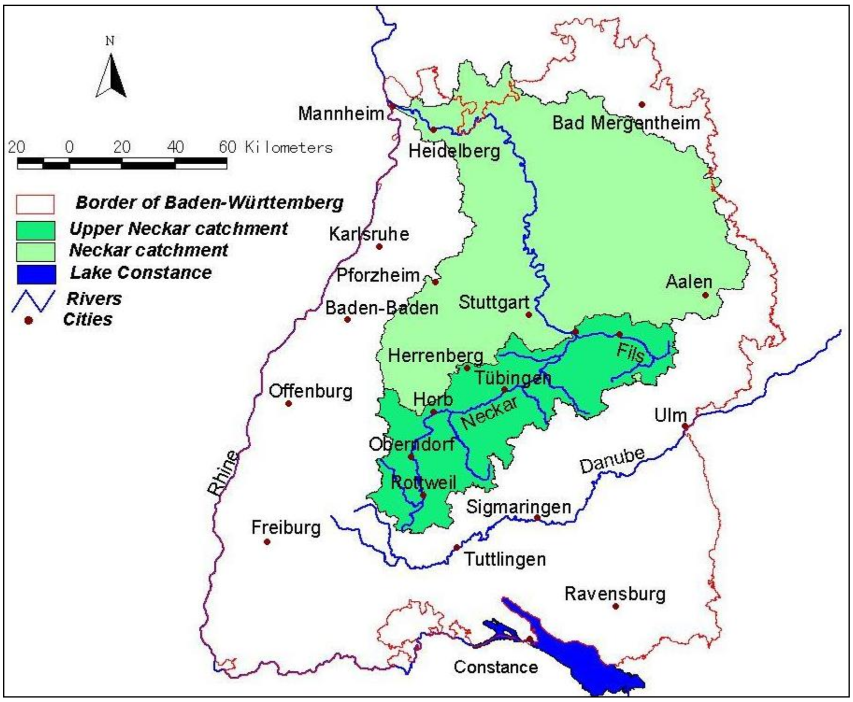

Fig.2 Neckar catchment (IWS, Stuttgart)

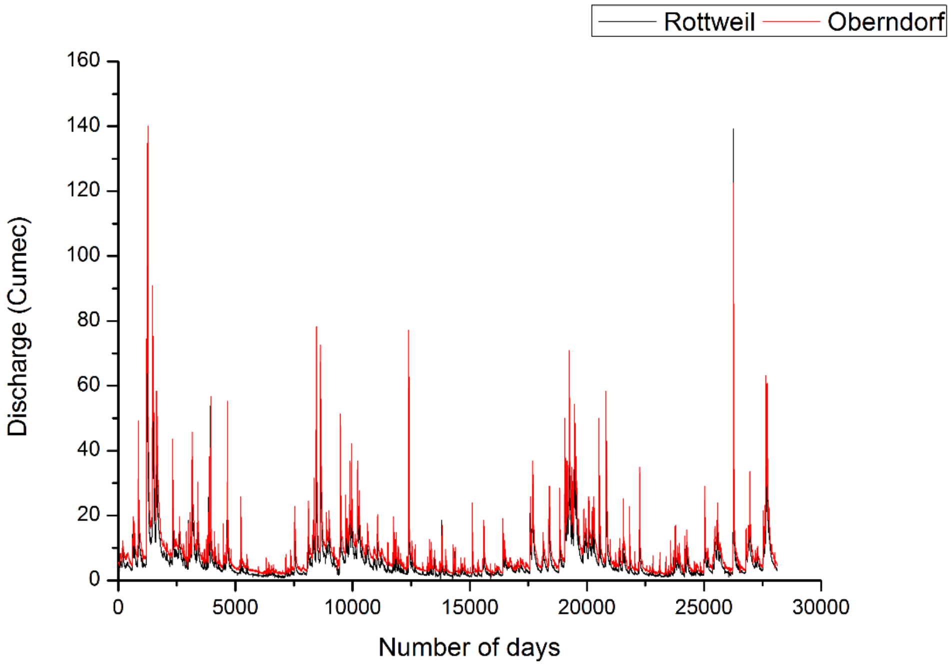

Fig. 3 Time series at Rottweil (upstream) and Oberndorf (downstream) gauging stations

4. Results and discussions

4.1 Flood Routing Using VPMM, SVM and WA-SVM

The extended VPMM method and its parameters were obtained from the study conducted by Yadav et al., (2015) for the same river stretch. The VPMM method under consideration has only two parameters K and θ which depends on the cross sectional information and the flow characteristics. The routing reach information such as bed slope and Manning's roughness value were obtained from the study reports of the Neckar river catchment, But the bed width and side slope of the cross section (Table 1) of the channel reach is optimized by the ROPE algorithm (Singh, 2008, Yadav et al., 2015). To avoid the influence of lateral inflow on the parameter optimization process the flood event with a minimum lateral flow among the 9 events is considered for the analysis. The data based model namely SVM and WA-SVM were developed using LIBSVM toolbox (Chang et al. 2011) to predict the discharge at the downstream location. The Radial Basis Function was adopted as kernel function for SVM and its parameters C and γ were obtained using a trial and error procedure where the trial continues till the value of RMSE was minimized. Later, the SVM model was suitability coupled with wavelet analysis (WA) which decomposes the input discharge time series using DWT into approximation and detailed time series (Fig. 4). The parameters for SVM and WA-SVM has been presented in Table 2. After the model calibration (VPMM) or training (SVM, WA-SVM) they were used to predict the discharge hydrograph of 9 flood events of the year 2002.

Table 1. Parameters for the development of VPMM method

Parameter	Value
Manning’s roughness	0.035
Bed slope	0.0034
River width (meter)	8.417
Side slope	1.035
Cross-sectional shape	Trapezoidal

Table 2. Optimal SVM and WA-SVM parameters for various decomposition series

Model	Decomposed series	best C	best γ
SVM		3.104	0.0412
WA-SVM	Approximation series	42	0.0611
	D1 series	3	0.0412
	D2 series	3	0.0712
	D3 series	7	0.0912
	D4 series	9	0.0812

Table 3 presents the statistical analysis of the simulated hydrograph obtained by VPMM, SVM and WA-SVM. The VPMM reproduced 7 out of 9 flood events with highest accuracy, where the error measures like NMSE, RMSE values ranges between 0.018 to 0.083 and 1.471

(m^3/s) to 4.301 (m^3/s), respectively. Similarly the values for R^2 and NSE ranges between 0.968 to 0.997 and 0.872 to 0.982, respectively. In case of SVM, the values obtained for NMSE and RMSE were significantly high for most of the flood events and ranges between 0.046 to 0.176 and 2.932 (m^3/s) to 5.918 (m^3/s), respectively. The fitness criteria (R^2 and NSE) also follows the similar trend like error measures and ranges between 0.831 to 0.966 and 0.822 to 0.954, respectively. The inclusion of wavelet analysis has definitely improved the accuracy of SVM and outperforms it in all flood events except 1, 3 and 9. Though, it is evident from the statistical analysis that the VPMM method shows superiority over SVM and WA-SVM, the reproduction of the downstream hydrographs for all the flood events by the data based models are very close to the observed hydrographs. This argument is well supported by the graphical representation of the observed and simulated hydrographs by VPMM, SVM and WA-SVM (Fig. 5-13). It is also evident from these figures that the absolute average relative error (AARE) of VPMM is very low. The AARE of SVM and WA-SVM is significantly higher than the VPMM, however WA-SVM shows relatively less error than the SVM. These figures reveal that, under significant lateral flow conditions, the rising limb, recession limb, and the peaks of the event-based flood hydrographs are all most well-reproduced by the VPMM, SVM and WA-SVM model.

Table. 3 Performance of VPMM, SVM and WA-SVM during the discharge prediction at the Oberndorf gauging station

Flood event	Method	NMSE	R^2	RMSE (m^3/s)	NSE
1	VPMM	**0.028**	**0.984**	3.421	0.948
	SVM	0.049	0.966	**3.316**	**0.951**
	WASVM	0.052	0.962	3.434	0.948
2	VPMM	**0.083**	**0.980**	**4.197**	**0.916**
	SVM	0.130	0.922	5.244	0.869
	WASVM	0.118	0.928	5.009	0.881
3	VPMM	**0.018**	**0.987**	**2.195**	**0.982**
	SVM	0.129	0.948	5.918	0.870

	WASVM	0.145	0.943	6.261	0.855
4	VPMM	**0.020**	**0.981**	**1.471**	**0.979**
	SVM	0.176	0.831	4.356	0.822
	WASVM	0.175	0.835	4.345	0.823
5	VPMM	**0.071**	**0.968**	**4.301**	**0.928**
	SVM	0.106	0.919	5.254	0.893
	WASVM	0.099	0.926	5.061	0.901
6	VPMM	**0.030**	**0.987**	**2.415**	**0.970**
	SVM	0.081	0.951	4.005	0.919
	WASVM	0.069	0.958	3.694	0.931
7	VPMM	**0.035**	**0.970**	**3.579**	**0.965**
	SVM	0.046	0.954	4.106	0.954
	WASVM	0.046	0.954	4.093	0.954
8	VPMM	0.128	0.976	4.545	0.872
	SVM	**0.053**	0.953	2.932	**0.947**
	WASVM	**0.053**	**0.955**	**2.920**	**0.947**
9	VPMM	**0.015**	**0.997**	**1.469**	**0.985**
	SVM	0.062	0.964	3.011	0.938
	WASVM	0.070	0.955	3.191	0.929

Further analysis of the results indicates that the VPMM model works well in both the cases of single or multi-peak peak flood events, however, data based models simulates the multi-peak flood events (Events 1 and 8) better than VPMM. The reason for such outcome can be attributed to the fact that the data based model performance primarily depends on the data length. In case of flood event 8, the discharge time series length is around 800 hrs with multiple peaks, which allowed the model to learn such occurrence properly. The study suggest that, if the data based models are fed with sufficient length of discharge time series data which encompass the variability in nature, they can simulate the discharge process with reasonable accuracy. On the other hand, the reduction in accuracy of VPMM for these flood events derives from the uncertainty in estimating the lateral flow which, mainly depends on the initial soil moisture conditions. The spatial and temporal variability of soil moisture content can have significant impact on the lateral flow estimation which in turn will reflect in the simulation accuracy of the VPMM.

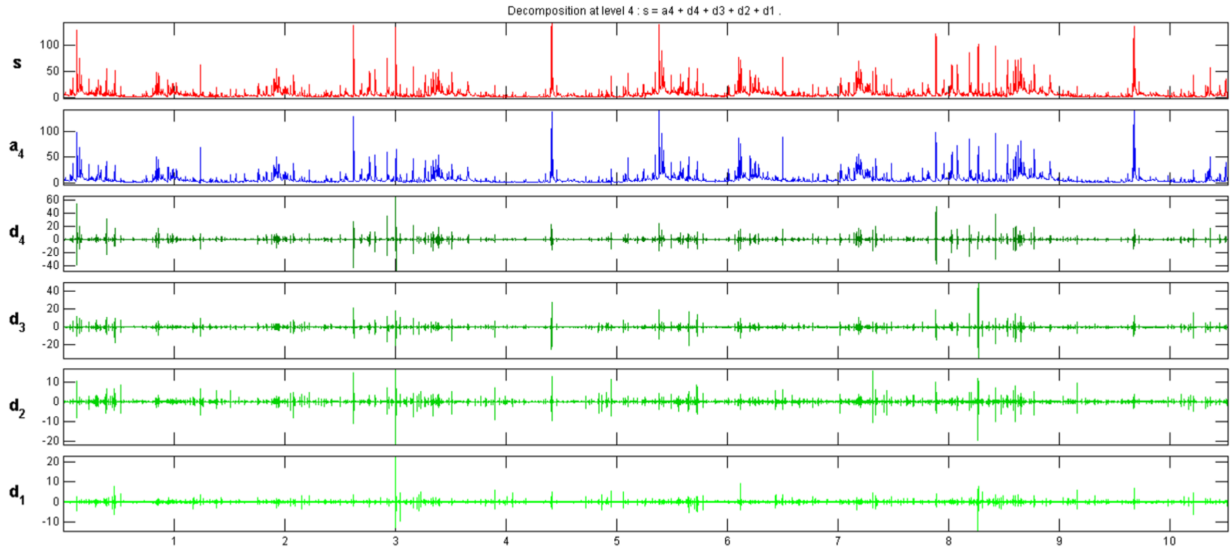

Fig. 4 Decomposed time series of the for the period of 1999 to 2000

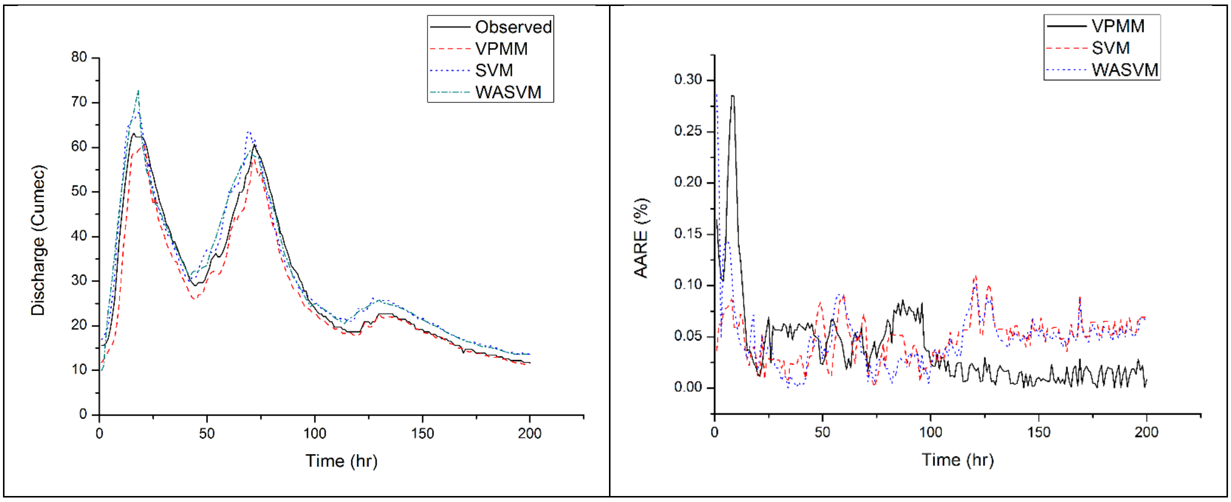

Fig. 5 Routed hydrograph and AARE for flood event 1 using VPMM, SVM and WASVM.

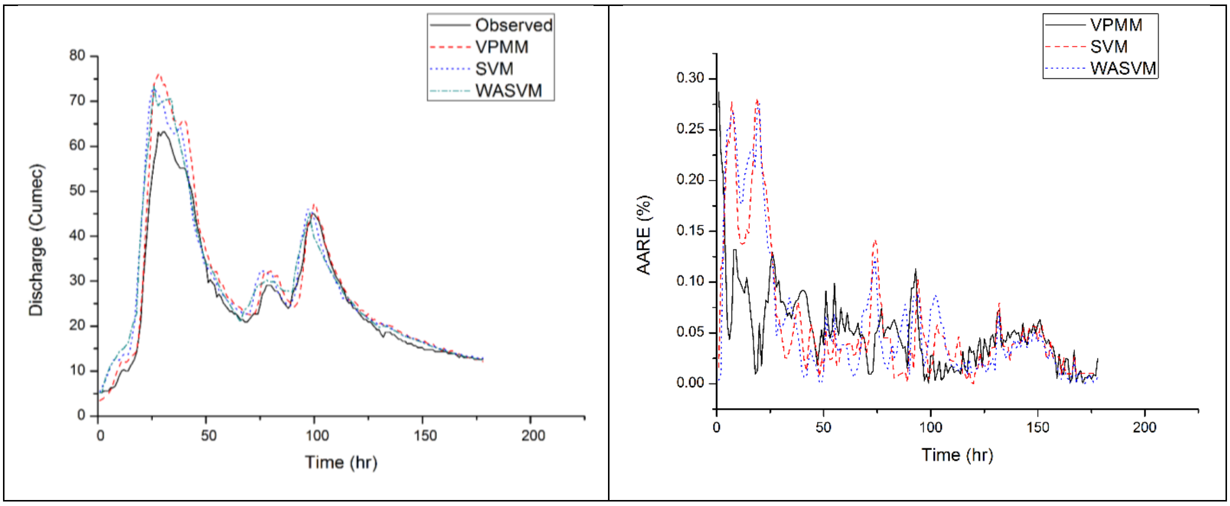

Fig. 6 Routed hydrograph and AARE for flood event 2 using VPMM, SVM and WASVM.

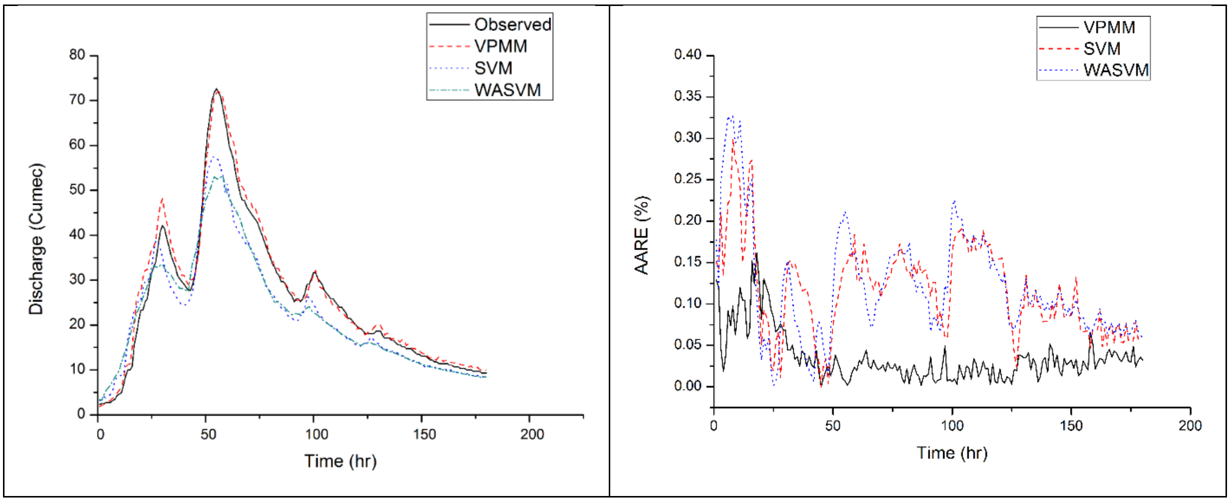

Fig. 7 Routed hydrograph and AARE for flood event 3 using VPMM, SVM and WASVM.

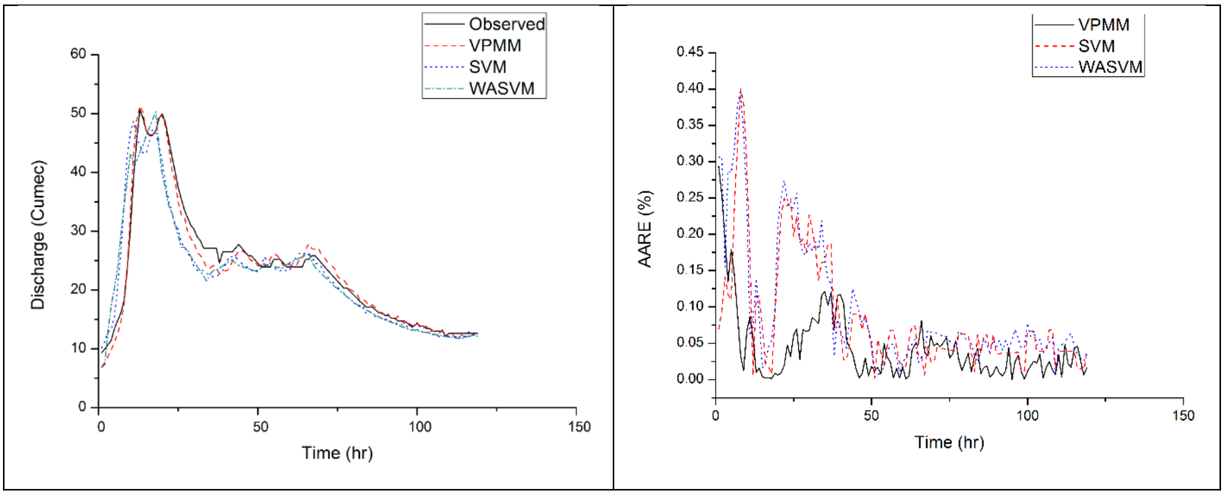

Fig. 8 Routed hydrograph and AARE for flood event 4 using VPMM, SVM and WASVM.

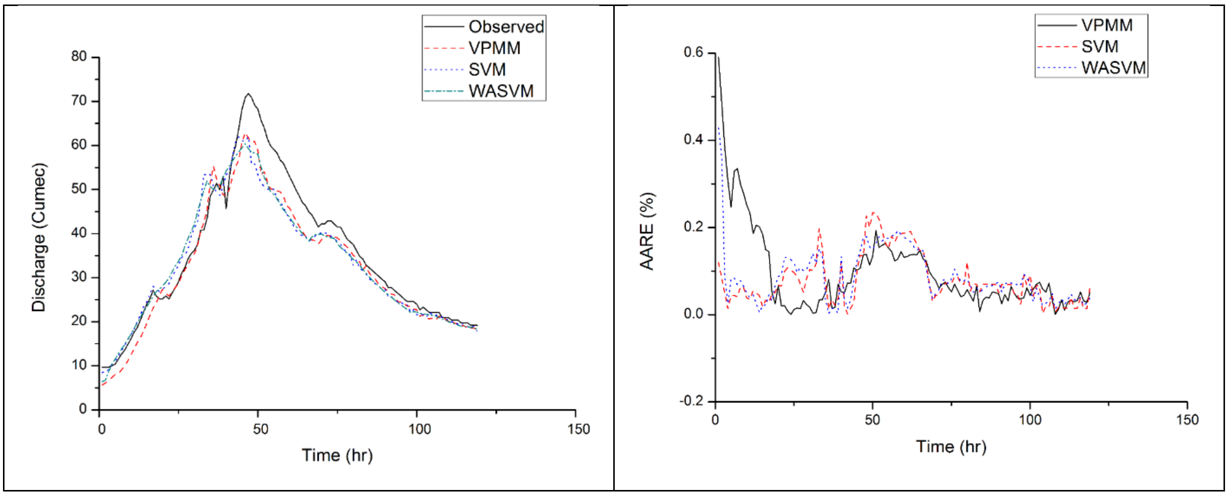

Fig. 9 Routed hydrograph and AARE for flood event 5 using VPMM, SVM and WASVM.

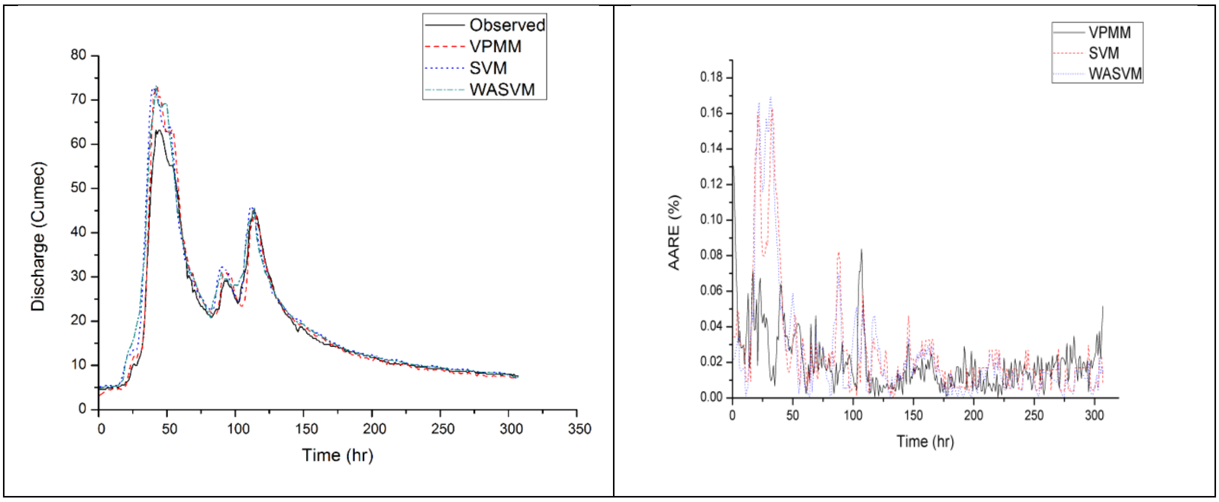

Fig. 10 Routed hydrograph and AARE for flood event 6 using VPMM, SVM and WASVM.

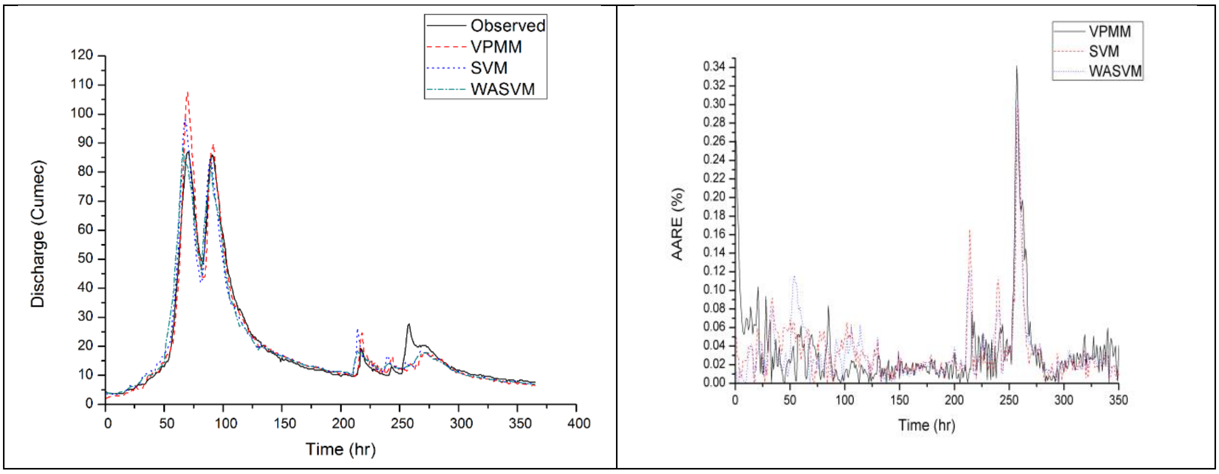

Fig. 11 Routed hydrograph and AARE for flood event 7 using VPMM, SVM and WASVM.

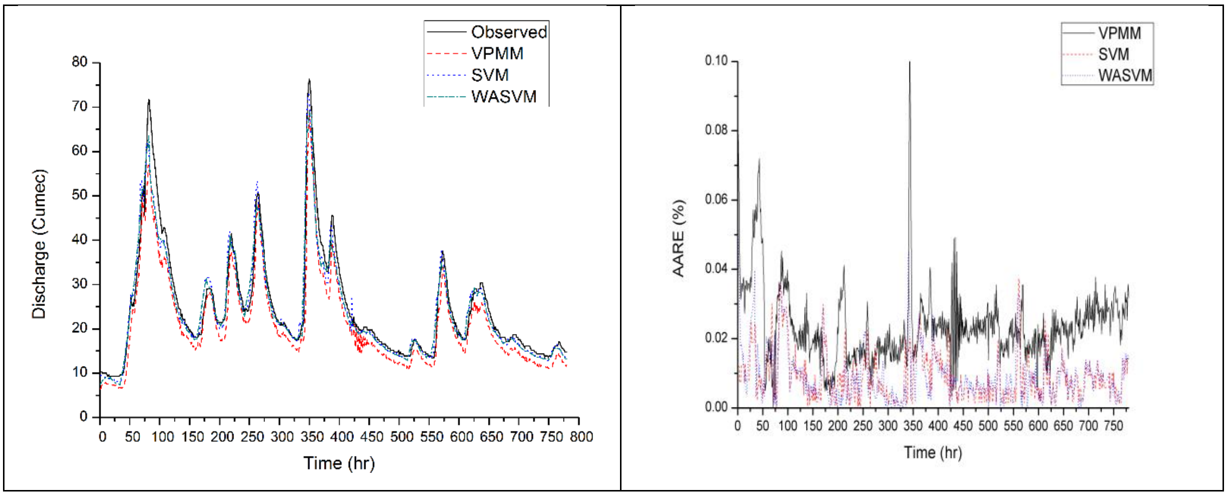

Fig. 12 Routed hydrograph and AARE for flood event 8 using VPMM, SVM and WASVM.

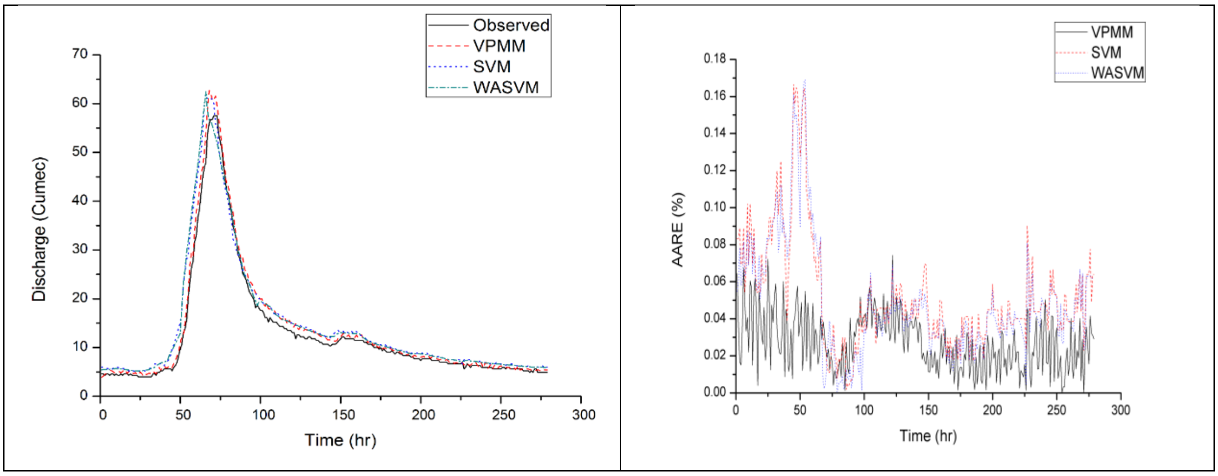

Fig. 13 Routed hydrograph and AARE for flood event 9 using VPMM, SVM and WASVM.

Further, considered methods in this study were also evaluated using the original criteria used for the development of VPMM. The percentage error in the peak discharge (Q_{er} in %), the error in the time-to-peak discharge (t_{Qe} in hr), and the percentage error in the volume ($EVOL$ in %) for all the 9 flood events has been depicted in Figs. 14, 15 and 16. It is evident from the Fig. 14 that the VPMM method predicts most of the peak values (5 out of 9) within $\pm 10\%$ error and just 2 above the 20% error. However, in case of SVM and WA-SVM Q_{er} is well above the $\pm 10\%$ range for most of the flood events. Which suggest that the data based models may requires more training to predict such high discharge values which comes rarely in a discharge time series but extremely important in case of flood forecasting. Similarly, Fig. 15 presents the error in time to peak discharge and VPMM predicts the peak value very close to its time of arrival in observed flood event. In fact, 7 out of 9 peaks has error 1 hour or less, and just two with the error range of $\pm 4\,(\text{hr})$. SVM also produced most of the peak discharge values with $\pm 2\,(\text{hr})$ error, however WA-SVM shows the higher error variation ranging from -2 to $+2\,(\text{hr})$. Further, the percentage error in the volume ($EVOL$ in %) is depicted in Fig. 16, which shows that the VPMM method despite receiving significant amount of lateral flow from the intervening catchment could reproduce the downstream hydrograph with just $\pm 10\%$ error in volume for 8 out of 9 flood events. Though, the error in the volume for SVM and WA-SVM is also within the same range as it was for the VPMM, however some flood events showed higher error.

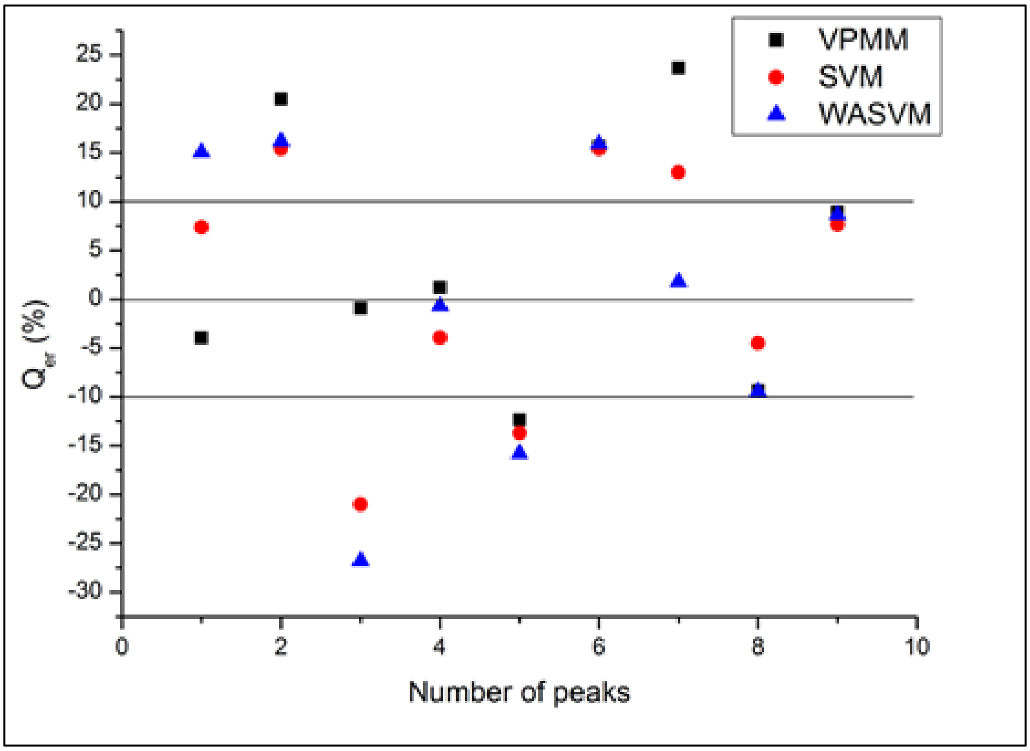

Fig. 14 Error in peak discharge prediction while using VPMM, SVM and WASVM

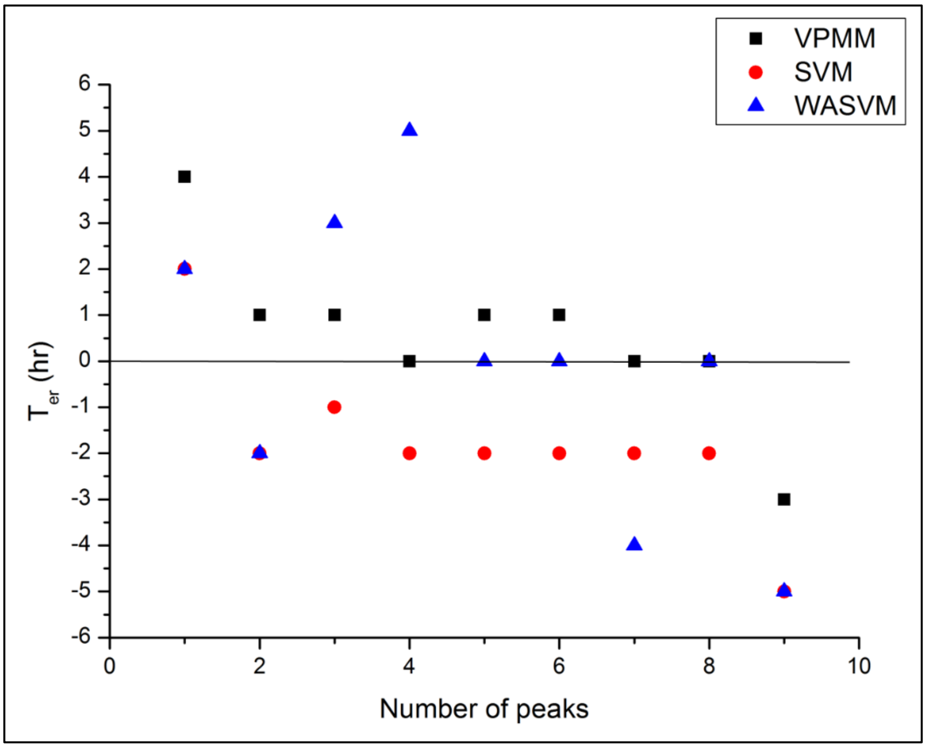

Fig. 15 Error in time to peak discharge prediction while using VPMM, SVM and WASVM

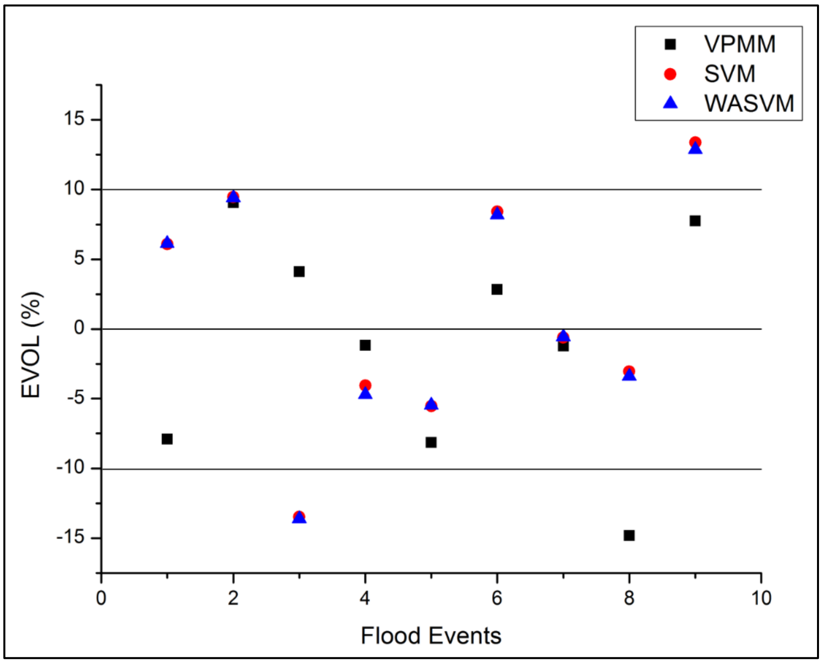

Fig. 16 variation of error in volume while using VPMM, SVM and WASVM

4.2 Level of Complexity in VPMM, SVM amd WA-SVM

The method under consideration were also evaluated to assess the level of complexity while desgining the model for discharge prediction. Table 4 presents the model complexity analysis of VPMM, SVM and WA-SVM based on the number of parameter each model requires to be tuned while designing the model for a specific application. The VPMM method has only two parameters that is K and θ while the SVM has three parameters namely regularization constant (C), insensitive loss function (ε), and parameter of radial basis function (γ). It is evident from the table that the Akaike information criterion (AIC) is lowest while using VPMM, in comparison to SVM and WASVM for most of the flood events. Similarly the model selection criteria (MSC) value is highest for 8 out of 9 flood events when VPMM is used, however it decreased significantly for SVM and WA-SVM.

Table 4. Akaike information criterion (AIC) and model selection criteria (MSC) for VPMM, SVM and WA-SVM

Flood Event	AIC			MSC		
	VPMM	SVM	WA-SVM	VPMM	SVM	WA-SVM
1	1555.64	**1545.19**	1559.21	**-0.02**	-0.030	-0.030
2	**1437.02**	1518.28	1501.94	**-0.022**	-3.203	-5.220
3	**1221.69**	1580.78	1601.07	**-0.022**	0.233	-4.822
4	**664.62**	924.92	924.34	**-0.033**	-0.391	-2.030
5	**919.94**	969.57	960.66	**-0.033**	-0.805	-4.307
6	**2303.60**	2616.07	2566.52	**-0.013**	-0.610	-3.529
7	**3088.30**	3190.54	3188.27	**-0.011**	-0.040	-2.250
8	7560.23	6878.02	**6871.93**	-0.005	**1.348**	-2.286
9	**1789.87**	2192.25	2224.57	**-0.014**	-0.889	-3.288

5. Conclusion

In this study two approaches were used to predict the downstream discharge of Neckar River in which VPMM is a physically based method and the SVM is data based method. Further, wavelet analysis was also used to develop a hybrid WA-SVM model. The study was conducted using 9 flood events from the year 2002 which is characterised of having significant lateral flow joining from the intervening catchment, which in general is difficult to model due to its spatial and temporal variability. Based on the analysis of statistical and graphical results, it is inferred that the extended physically based variable parameter Muskingum routing method (VPMM) is more robust and reliable than the data based models like SVM and WA-SVM, when used to predict the discharge in a river reach with significant lateral flow joining between the upstream and downstream gauging stations. However, it is also evident from the analysis that the data based models successfully captured the flood wave moment phenomenon and were able to map the process even with lateral flow, hence reproduced the discharge hydrograph close to the observed hydrograph at the downstream location. Further, based on the Akaike information criterion (AIC) and model selection criteria (MSC), it can be concluded that the VPMM model is relatively less complex than the SVM and WA-SVM. Lastly, it can be summarised that the physically based extended VPMM method can predict the discharge hydrograph better than the data based mode, however, in case of multi-peak flood events with sufficient discharge data, the later performed better than VPMM method.

Acknowledgement

The author thankfully acknowledge the support and motivation provided by Prof. M. Perumal. The necessary data to conduct this study was provided by TU Stuttgart, Germany. The financial support to conduct this study was provided by IIT Delhi, India.

Conflict of Interest Statement

We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal. All the authors have approved the manuscript and agree with the submission to Neural Computing and Applications Journal. The financial support was provided by IIT Delhi. The authors have no conflict of interest to declare.

References

Adamowski, J., & Sun, K., 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. Journal of Hydrology, 390(1), 85-91.

Adamowski, J., & Chan, H. F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology, 407(1), 28-40.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000.

Agarwal, A., Maheswaran, R., Kurths, J., & Khosa, R. 2016. Wavelet Spectrum and self-organizing maps-based approach for hydrologic regionalization-a case study in the western United States. Water Resources Management, 30(12), 4399-4413.

Badrzadeh, H., Sarukkalige, R., & Jayawardena, A. W., 2013. Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. Journal of Hydrology, 507, 75-85.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of hydrology, 320(1), 18-36.

CC-HYDRO, 1999. Impact of Climate Change on River Basin Hydrology under Different Climatic Conditions, March 1999. Universität Stuttgart, Germany.

Borges, R. V., Garcez, A. D. A., & Lamb, L. C., 2011. Learning and representing temporal knowledge in recurrent networks. IEEE Transactions on Neural Networks, 22(12), 2409-2421.

Cannas, B., Fanni, A., See, L., & Sias, G., 2006. Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Physis and Chemistry of the Earth 31 (18), 1164–1171.

Chang C. C., & Lin C. J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (3) 27.

Chow, V.T., Maidment, D.R., & Mays, L.W., 1988. Applied Hydrology. McGraw-Hill, New York.

Choy K., & Chan C., 2003. Modelling of river discharges and rainfall using radial basis function networks based on support vector regression, Int. Journal of System Science, 34 (1) 763–773

Das, T., 2006. The Impact of Spatial Variability of Precipitation on the Predictive Uncertainty of Hydrological Models, Ph.D. Dissertation No. 154, University of Stuttgart.

Dawson C.W., & Wilby R., 1998. An artificial neural network approach to rainfall–runoff modeling. Hydrological Sciences Journal, 43 (1), pp. 47–66.

Ghalkhani, H., Golian, S., Saghafian, B., Farokhnia, A., & Shamseldin, A., 2013. Application of surrogate artificial intelligent models for real-time flood routing. Water and Environment Journal, 27(4), 535-548.

Harpham, C., & Dawson, C. W., 2006. The effect of different basis functions on a radial basis function network for time series prediction: a comparative study. Neurocomputing, 69(16), 2161-2170.

Kalteh, A. M., 2013. Monthly River flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Computers & Geosciences, 54, 1-8.

Karahan, H., Gurarslan, G., & Geem, Z. W., 2015. A new nonlinear Muskingum flood routing model incorporating lateral flow. Engineering Optimization, 47(6), 737-749.

Kasiviswanathan, K. S., He, J., Sudheer, K. P., & Tay, J. H., 2016. Potential application of wavelet neural network ensemble to forecast streamflow for flood management. Journal of Hydrology, 536, 161-173

Kisi, O., 2008. River flow forecasting and estimation using different artificial neural network techniques. Hydrological Research, 39 (1), 27–40.

Koza, J. R., 1992. Genetic programming: on the programming of computers by means of natural selection (Vol. 1). MIT press.

Lin, T., Guo, T., & Aberer, K., 2017. Hybrid neural networks for learning the trend in time series. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17 (pp. 2273-2279).

Loague, K., & VanderKwaak, J. E., 2004. Physics-based hydrologic response simulation: Platinum bridge, 1958 Edsel, or useful tool. Hydrological Processes, 18(15), 2949-2956.

Maier, H. R., & Dandy, G. C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental modelling & software, 15(1), 101-124.

McCarthy, G.T., 1938. The unit hydrograph and flood routing. In: Conference of North Atlantic Div., U.S. Army Corps of Engineers.

Nourani, V., Baghanam, A. H., Adamowski, J., & Kisi, O., 2014. Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. Journal of Hydrology, 514, 358-377.

O'Donnell, T., 1985. A direct three-parameter Muskingum procedure incorporating lateral inflow. Hydrological Science Journal. 30 (4), 479–496.

Nayak, P. C., Sudheer, K. P., & Jain, S. K., 2007. Rainfall-runoff modeling through hybrid intelligent system. Water Resources Research, 43(7).

Perumal, M., 1994a. Hydrodynamic derivation of a variable parameter Muskingum method: 1. Theory and solution procedure. Hydrological sciences journal, 39(5), 431-442.

Perumal, M., 1994b. Hydrodynamic derivation of a variable parameter Muskingum method: 2. Verification. Hydrological sciences journal, 39(5), 443-458.

Perumal, M., O'Connell, P.E., & Ranga Raju, K.G., 2001. Field applications of a variable parameter Muskingum method. Journal of Hydraulic Engineering. 6 (3), 1084-0699/01/0003-0196-0207.

Perumal, M., & Sahoo, B., 2008. Volume conservation controversy of the variable parameter Muskingum–Cunge method. Journal of Hydraulic Engineering, 134(4), 475-485.

Perumal, M., & Price, R.K., 2013. A fully volume conservative variable parameter McCarthy-Muskingum method: theory and Verification. Journal of Hydrology. 502, 89–102.

Perumal, M., Tayfur, G., Rao, C. M., & Gurarslan, G., 2017. Evaluation of a physically based quasi-linear and a conceptually based nonlinear Muskingum methods. Journal of Hydrology, 546, 437-449.

Ponce, V. M., & Yevjevich, V., 1978. Muskingum-Cunge method with variable parameters. Journal of the Hydraulics Division, 104(12), 1663-1667.

Price, R. K., 2009. Volume-conservative nonlinear flood routing. Journal of Hydraulic Engineering, 135(10), 838-845.

Rezaeianzadeh, M., Tabari, H., Yazdi, A. A., Isik, S., & Kalin, L., 2014. Flood flow forecasting using ANN, ANFIS and regression models. Neural Computing and Applications, 25(1), 25-37.

Sang, Y. F., 2013. A review on the applications of wavelet transform in hydrology time series analysis. Atmospheric research, 122, 8-15.

Shiri, J., & Kisi, O., 2010. Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. Journal of hydrology, 394(3), 486-493.

Shiri, J., Kişi, Ö., Makarynskyy, O., Shiri, A. A., & Nikoofar, B., 2012. Forecasting daily stream flows using artificial intelligence approaches. ISH Journal of Hydraulic Engineering, 18(3), 204-214.

Singh, S.K., 2008. Robust Parameter Estimation in Gauged and Ungauged Basins, Ph.D. dissertation No. 198, University of Stuttgart.

Sudheer, K. P., Gosain, A. K., & Ramasastri, K. S., 2002. A data-driven algorithm for constructing artificial neural network rainfall–runoff models. Hydrological Processes, 16, 1325–1330.

Suryanarayana, C., Sudheer, C., Mahammood, V., & Panigrahi, B. K., 2014. An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. Neurocomputing, 145, 324-335.

Swain, R., & Sahoo, B., 2015. Variable parameter McCarthy–Muskingum flow transport model for compound channels accounting for distributed non-uniform lateral flow. Journal of Hydrology, 530, 698-715.

Tehrany, M. S., Pradhan, B., & Jebur, M. N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of Hydrology, 512, 332-343.

Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N., 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena, 125, 91-101.

Tiwari, M.K., & Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. Journal of Hydrology 1 (394), 458–470.

Todini, E., 2007. A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach. Hydrology and Earth System Sciences Discussions, 4(3), 1549-1592.

Uhlenbrook, S., Seibert, J., Leibundgut, C., & Rodhe, A., 1999. Prediction uncertainty of conceptual rainfall–runoff models caused by problems to identify model parameters and structure. Hydrological Sciences Journal 44, 779–798.

Vapnik, V. N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York, USA. 314 p.

Yadav, B., Mathur, S., & Yadav, B. K. 2018. Data-based modelling approach for variable density flow and solute transport simulation in a coastal aquifer. Hydrological Sciences Journal, 1-17.

Yadav, B., & Eliza, K., 2017. A hybrid wavelet-support vector machine model for prediction of Lake water level fluctuations using hydro-meteorological data. Measurement, 103, 294-301.

Yadav, B., Ch, S., Mathur, S., & Adamowski, J., 2016a. Estimation of in-situ bioremediation system cost using a hybrid Extreme Learning Machine (ELM)-particle swarm optimization approach. Journal of Hydrology, 543, 373-385.

Yadav, B., Ch, S., Mathur, S., & Adamowski, J. 2016b. Discharge forecasting using an online sequential extreme learning machine (OS-ELM) model: a case study in Neckar River, Germany. Measurement, 92, 433-445.

Yadav, B., Perumal, M., & Bardossy, A., 2015. Variable parameter McCarthy–Muskingum routing method considering lateral flow. Journal of Hydrology, 523, 489-499.

Yao, X., Tham, L., & Dai, F., 2008. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. Geomorphology 101, 572–582.

Yang, J., L., Y., Tian, Y., Duan, L., & Gao, W. Group-sensitive multiple kernel learning for object categorization. In ICCV, 2009.

Yu X., Liong S., & Babovic V., 2004. EC-SVM approach for real-time hydrologic forecasting. Journal of Hydroinformatics, 6 (3) 209–223.