Review Article

# Support vector machine applications in the field of hydrology: A review

Sujay Raghavendra. N*, Paresh Chandra Deka

*Department of Applied Mechanics and Hydraulics, National Institute of Technology Karnataka, Surathkal, India*

## ARTICLE INFO

## ABSTRACT

In the recent few decades there has been very significant developments in the theoretical understanding of Support vector machines (SVMs) as well as algorithmic strategies for implementing them, and applications of the approach to practical problems. SVMs introduced by Vapnik and others in the early 1990s are machine learning systems that utilize a hypothesis space of linear functions in a high dimensional feature space, trained with optimization algorithms that implements a learning bias derived from statistical learning theory. This paper reviews the state-of-the-art and focuses over a wide range of applications of SVMs in the field of hydrology. To use SVM aided hydrological models, which have increasingly extended during the last years; comprehensive knowledge about their theory and modelling approaches seems to be necessary. Furthermore, this review provides a brief synopsis of the techniques of SVMs and other emerging ones (hybrid models), which have proven useful in the analysis of the various hydrological parameters. Moreover, various examples of successful applications of SVMs for modelling different hydrological processes are also provided.

© 2014 Elsevier B.V. All rights reserved.

## Contents

* Corresponding author. Tel.: +91 9035219180.
  *E-mail address:* sujayraghavendran@ymail.com (S. Raghavendra. N).

## 1. Introduction

Hydrological models provide us a wide range of significant applications in the multi-disciplinary water resources planning and management activities. Hydrological models can be formulated using deterministic, probabilistic and stochastic approaches for the characterization of surface and ground water systems in conjunction with coupled systems modelling such as hydro-ecology, hydro-geology and climate. However, due to resource constraints and the confined scope of available measurement techniques, there are limitations to the availability of spatial-temporal data [1]; hence a need exists to generalize data procured from the available measurements in space and time using special techniques like support vector machines and its hybrid models.

Hydrological model applications have a wide variety of objectives, depending on the problem that needs to be investigated [1]. To analyze hydrologic data statistically, the user must know basic definitions and understand the purpose and limitations of SVM analysis. The application of SVM methods of hydrological analyses requires measurement of physical phenomena. The modeller has to evaluate the accuracy of the data collected and should have a brief knowledge on how the data are accumulated and processed before they are used in modelling activities. Some of the commonly used data in hydrologic studies include rainfall, snowmelt, stage, streamflow, temperature, evaporation, and watershed characteristics [2].

SVMs have been recently introduced relatively new statistical learning technique. Due to its strong theoretical statistical framework, SVM has proved to be much more robust in several fields, especially for noise mixed data, than the local model which utilizes traditional chaotic techniques [3]. The SVM has brought forth heavy expectations in recent few years as they have been successful when applied in classification problems, regression and forecasting; as they include aspects and techniques from machine learning, statistics, mathematical analysis and convex optimization. Apart from possessing a strong adaptability, global optimization, and a good generalization performance, the SVMs are suitable for classification of small samples of data also. Globally, the application of these techniques in the field of hydrology has come a long way since the first articles began appearing in conferences in early 2000s [4].

This paper aims at discussing the basic theory behind SVMs and existing SVM models, reviewing the recent research developments, and presenting the challenges for the future studies of the hydrological impacts of climate change.

## 2. Theory of support vector machines

The support vector machines (SVMs) are developed based on statistical learning theory and are derived from the structural risk minimization hypothesis to minimize both empirical risk and the confidence interval of the learning machine in order to achieve a good generalization capability. SVMs have been proven to be an extremely robust and efficient algorithm for classification [5] and regression [6,7]. Cortes and Vapnik [8] proposed the current basic SVM algorithm. The beauty of this approach is twofold: it is simple enough that scholars with sufficient knowledge can readily understand, yet it is powerful that the predictive accuracy of this approach overwhelms many other methods, such as nearest neighbours, neural networks and decision tree. The basic idea behind SVMs is to map the original data sets from the input space to a high dimensional, or even infinite-dimensional feature space so that classification problem becomes simpler in the feature space. SVMs have the potential to procreate the unknown relationship present between a set of input variables and the output of the system. The main advantage of SVM is that, it uses kernel trick to build expert knowledge about a problem so that both model complexity and prediction error are simultaneously minimized. SVM algorithms involve the application of the three following mathematical principles:

- Principle of Fermat (1638)
- Principle of Lagrange (1788)
- Principle of Kuhn–Tucker (1951)

### 2.1. Support vector classification

The preliminary objective of SVM classification is to establish decision boundaries in the feature space which separate data points belonging to different classes. SVM differs from the other classification methods significantly. Its intent is to create an optimal separating hyperplane between two classes to minimize the generalization error and thereby maximize the margin. If any two classes are separable from among the infinite number of linear classifiers, SVM determines that hyperplane which minimizes the generalization error (i.e., error for the unseen test patterns) and conversely if the two classes are non-separable, SVM tries to search that hyperplane which maximizes the margin and at the same time, minimizes a quantity proportional to the number of misclassification errors. Thus, the selected hyperplane will have the maximum margin between the two classes, where margin is defined as a summation of the distance between the separating hyperplane and the nearest points on either side of two classes [5].

SVM classification and thereby its predictive capability can be understood by dealing with four basic concepts: (1) the separation hyperplane, (2) the hard-margin SVM (3) the soft-margin SVM and (4) kernel function [9].

SVM models were originally developed for the classification of linearly separable classes of objects. Referring to Fig. 1. Consider
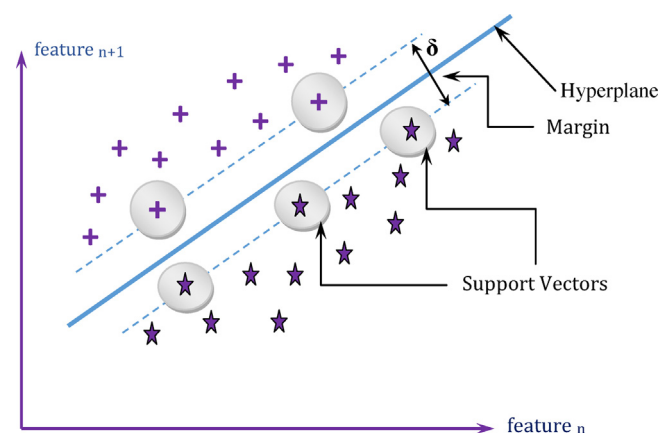


**Fig. 1.** Maximum separation hyperplane.
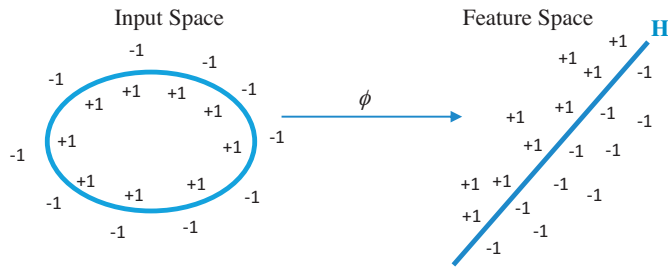Adapted from [61].

**Fig. 2.** Linear separation in feature space.
Adapted from [61].

a two-dimensional plane consisting linearly separable objects of two separate classes {class (+) and class (*)}. The goal is to find a classifier which separates them perfectly. There can be many possible ways to classify/separate those objects but SVM tries to find a unique hyperplane which produces maximum margin (i.e., SVM maximizes the distance between the hyperplane and the nearest data point of each class). The class (+) objects are framed behind hyperplane H1, while the hyperplane H2 frames the class (*) objects. The objects of either class which exactly fall over the hyperplanes H1 and H2 are termed as support vectors. Most "important" training points are support vectors; they define the hyperplane and have direct bearing on the optimum location of the decision surface. The maximum margin obtained is denoted as $\delta$. When only support vectors are used to specify the separating hyperplane, sparseness of solution emerges when dealing with large data sets.

In real time problems it is not possible to determine an exact separating hyperplane dividing the data within the space and also we might get a curved decision boundary in some cases. Hence SVM can also be used as a classifier for non-separable classes (Fig. 2, left). In such cases, the original input space can always be mapped to some higher-dimensional feature space (Hilbert space) using non-linear functions called feature functions $\phi$ (as presented in Figs. 2 and 3). Even though feature space is high dimensional, it could not be practically feasible to use directly the feature functions $\phi$ for classification of the hyperplane. So in such cases, nonlinear mapping induced by the feature functions is used for computation using special nonlinear functions called kernels.

Sound selection of kernels provides us for the excellent generalization performance. The correlation or similarity between two different classes of data points is depicted utilizing kernel functions. Kernels have the advantage of operating in the input space, where the result of the classification problem is a weighted sum of kernel functions evaluated at the support vectors. The Kernel trick allows SVM's to form nonlinear boundaries There are quite a good number of kernels that can be used in SVMs. These include linear, polynomial, radial basis function and sigmoid. The

role of nonlinear kernels provides the SVM with the ability to model complicated separating hyperplanes.

$$K(x, x') = \begin{cases} (x^T \cdot x') & \text{Linear} \\ (x^T \cdot x' + 1)^d & \text{Polynomial} \\ \exp(-\gamma \|x - x'\|^2) & \text{RBF} \\ \tanh(\gamma x \cdot x' + C) & \text{Sigmoidal} \end{cases}$$

Unique characteristics of SVM classification and kernel methods include: (1) their performance is guaranteed as they are purely based on theoretical examples of statistical learning; (2) search space has a unique minimal; (3) training is extremely robust and efficient; (4) the generalization capability enables trade-off between classifier complexity and error.

Detailed theoretical background and basic SVM's formulae can be found in many references especially in [5,9].

Some of the SVM-classifiers include: hard-margin/maximal margin SVM; soft margin SVM; $k$-nearest neighbour SVM, radial basis function-SVM, translational invariant kernels-SVM, tangent distance kernels-SVM. Brief explanations related to these SVM classification techniques can be found in Shigeo [10].

### 2.1.1. Clustering using SVM

Clustering is often formulated as a discrete optimization problem. The main objective of cluster analysis is to segregate objects into groups, such that objects of each group are more "alike" to each other than the objects of other groups (usually represented as a vector of measurements, or a point in a multidimensional space). Clustering is unsupervised learning of a hidden data concept. The applications of clustering often deal with large datasets and data with many attributes. A brief introduction to clustering in pattern recognition is provided in Duda and Hart [11]. Enhancing the SVM training process using clustering or similar techniques has been examined with numerous variations in Yu et al. [12] and Shih et al. [13]. Cluster-SVM accelerates the training process by exploiting the distributional properties of the training data. The cluster-SVM algorithm, first partitions the training data into numerous pairwise separate clusters and then the representatives of these clusters are utilized to train an initial SVM, based on which the support vectors and non-support vectors are identified approximately (refer Fig. 4). The clusters comprising only non-support vectors are replaced with their representatives to reduce the number of training data significantly and thereby the speed of the training process is enhanced.
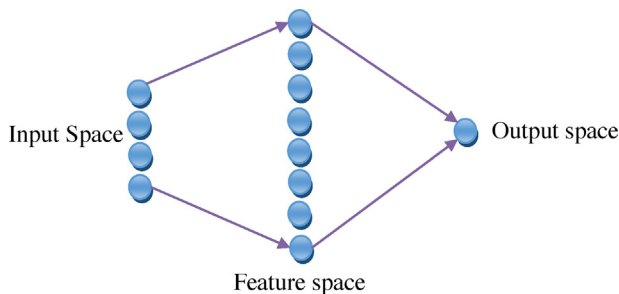


**Fig. 3.** Support vector machines map the input space into a high-dimensional feature space.
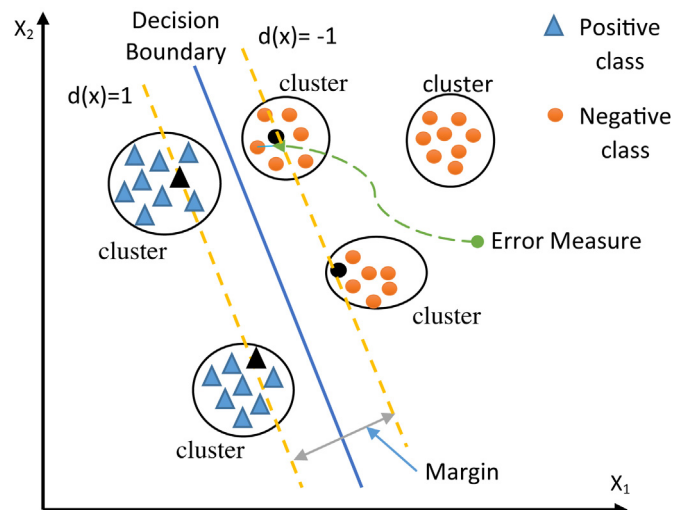Adapted from [61].



**Fig. 4.** SVM-clustering.

Based on a hierarchical micro-clustering algorithm, Yu et al. [12] put forward a scalable algorithm to train SVMs with a linear kernel.

The SVM-Internal approach to clustering was initially proposed by Winters-Hilt et al. [14]. Data points are mapped to a high dimensional feature space by means of a kernel where the minimal enclosing sphere is explored. The width of the kernel function governs the measure at which the data is probed in contrast the soft margin constant assists to supervise outliers and over-lapping clusters. The structure of a data set is surveyed by altering these two parameters, so as to preserve a minimal number of support vectors for substantiate smooth cluster boundaries. Since the SVM-Internal Clustering is faintly biased towards the shape of the clusters in feature space (the bias is for spherical clusters in the kernel space), it lacks robustness. In most of the real-time datasets wherein strongly overlapping clusters are seen, the internal SVM–internal clustering algorithm can only delineate the reasonably small cluster cores. Other drawbacks like initial choice of kernel, excessive geometric constraints, etc. make way for the use of SVM-external clustering algorithm that clusters data vectors without any prior knowledge of data's clustering characteristics, or the number of clusters. $K$-means clustering technique with SVM is an example of SVM-external clustering.

## 2.2. Support vector regression

Let us consider a simple linear regression problem trained on data set $\chi = \{u_i, v_i; \quad i = 1, \ldots, n\}$ with input vectors $u_i$ and linked targets $v_i$. A function $g(u)$ has to be formulated approximately in order to link up inherited relations between the data sets and thereby it can be used in the later part to infer the output $v$ for a new input data $u$.

Standard SVM regression uses a loss function $L_\varepsilon(v, g(u))$ which describes the deviation of the estimated function from the original one. Several types of loss functions can be mined in the literature e.g., linear, quadratic, exponential, Huber's loss function, etc. In the present context the standard Vapnik's – $\varepsilon$ insensitive loss function is used which is defined as

$$L_\varepsilon(v, g(u)) = \begin{cases} 0 & \text{for } |v - g(u)| \le \varepsilon \\ |v - g(u)| - \varepsilon & \text{otherwise} \end{cases} \tag{1}$$

Using $\varepsilon$-insensitive loss function, one can find $g(u)$ that can better approximate the actual output vector $v$ and has the at most error tolerance $\varepsilon$ from the actual incurred targets $v_i$ for all training data, and concurrently as flat as possible. Consider the regression function defined by

$$g(u) = w \cdot u + b \tag{2}$$

where $w \in \chi$, $\chi$ is the input space; $b \in R$ is a bias term and $(w \cdot u)$ is dot product of vectors $w$ and $u$. flatness in Eq. (2) refers to a smaller value of parameter vector $w$. By minimizing the norm $||w||^2$ flatness can be ascertained along with model complexity. Thus regression problem can be stated as the following convex optimization problem.

$$\min_{w, b, \xi, \xi*} \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$
$$v_i - (w \cdot u_i + b) \le \varepsilon + \xi_i \tag{3}$$
$$\text{subject to} \quad (w \cdot u_i + b) - v_i \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \ldots, n$$

where $\xi_i$ and $\xi_i^*$ are slack variables introduced to evaluate the deviation of training samples outside $\varepsilon$-insensitive zone. The trade-off between the flatness of g and the quantity up to which deviations greater than $\varepsilon$ are tolerated is depicted by $C > 0$. $C$ is a positive constant influencing the degree of penalizing loss when a training error

occurs. Underfitting and overfitting of training data are avoided by minimization of the regularization term $w^2/2$ along with the training error term $C\sum_{i=1}^{n}(\xi_i - \xi_i^*)$ in Eq. (3). The minimization problem in Eq. (3) represents the primal objective function.

Now the problem is dealt by constructing a Lagrange function from the primal objective function by introducing a dual set of variables, $\underline{\alpha}_i$ and $\bar{\alpha}_i$ for the corresponding constraints. Optimality conditions are exploited at the saddle points of a Lagrange function leading to the formulation of the dual optimization problem:

$$\max_{\underline{\alpha}_i, \bar{\alpha}_i} \quad -\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i - \bar{\alpha}_i(\alpha_j - \bar{\alpha}_j)\langle u_i \cdot u_j \rangle - \varepsilon\sum_{i=1}^{n}(\alpha_i + \bar{\alpha}_i) + \sum_{i=1}^{n}v_i(\alpha_i - \bar{\alpha}_i)$$
$$\text{subject to} \quad \sum_{i=1}^{n}(\underline{\alpha}_i - \bar{\alpha}_i) = 0$$
$$0 \le \underline{\alpha}_i \le C, \quad i = 1, 2, \ldots, n$$
$$0 \le \bar{\alpha}_i \le C, \quad i = 1, 2, \ldots, n \tag{4}$$

After determining Lagrange multipliers, $\underline{\alpha}_i$ and $\bar{\alpha}_i$; the parameter vectors $w$ and $b$ can be evaluated under Karush–Kuhn–Tucker (KKT) complementarity conditions [15], which are not discussed herein. Therefore, the prediction is a linear regression function that can be expressed as

$$g(u) = \sum_{i=1}^{n}\alpha_i - \bar{\alpha}_i\langle u_i \cdot u \rangle + b \tag{5}$$

Thus SVM regression expansion is derived; where w is depicted as a linear combination of the training patterns $v_i$ and b can be found using primary constraints. For $|g(u)| \ge \varepsilon$ Lagrange multipliers may be non-zero for all the samples inside the $\varepsilon$-tube and these remaining coefficients are termed as support vectors.

### 2.2.1. Nonlinear support vector regression – the kernel trick

Now for making SVM regression to deal with non-linear cases; pre-processing of training patterns $u_i$ has to done by mapping the input space $\chi$ into some feature space $\Im$ using nonlinear function $\varphi = \chi \longrightarrow \Im$ and is then applied to the standard support vector algorithm. Also the dimensionality of $\varphi(x)$ can be very huge, making 'w' hard to represent explicitly in memory, and hard for the quadratic programming optimizer to solve. The representer theorem Kimeldorf and Wabha [74] shows that: $w = \sum_{i=1}^{n}\alpha_i \cdot \phi(x_i)$ for some variables $\alpha$. Instead of optimizing 'w' we can directly optimize $\alpha$. There by the decision function is obtained (kernel trick 1). The representer theorem is exploited to examine the sensitivity properties of $\varepsilon$-insensitive SVR and introduce the concept of approximate degrees of freedom. The degrees of freedom play a vital role in the assessment of the optimism – i.e., the difference between the expected in-sample error and the expected empirical risk.

Let $u_i$ be mapped into the feature space by nonlinear function $\varphi(u)$ and hence the decision function is given by

$$g(w, b) = w \cdot \phi(u) + b \tag{6}$$

This nonlinear regression problem can be expressed as the following optimization problem. Fig. 5 depicts the concept of nonlinear SV regression corresponding to Eq. (7)

$$\min_{w, b, \xi, \xi*} \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$
$$v_i - (w \cdot \phi(u_i) + b) \le \varepsilon + \xi_i \tag{7}$$
$$\text{subject to} \quad (w \cdot \phi(u_i) + b) - v_i \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \ldots, n$$

where $w$ is the vector of coefficients, $\xi_i$ and $\xi_i^*$ are the distances of the training data set points from the region where the errors less
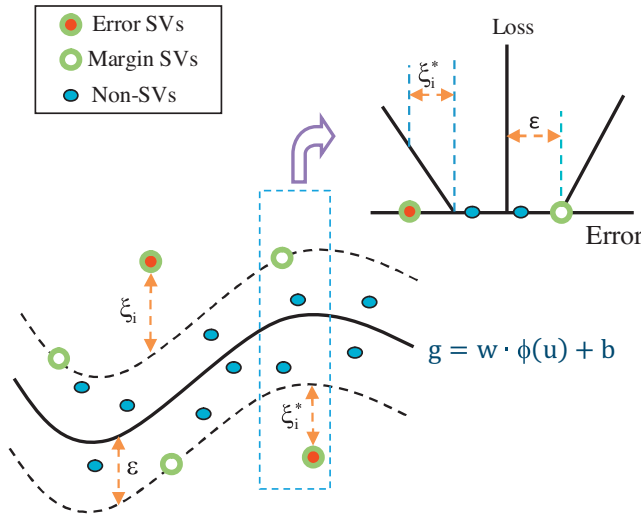
**Fig. 5.** Nonlinear SVR with Vapnik's $\varepsilon$-insensitive loss function. Adapted from [41].



**Fig. 6.** Network architecture of SVM. Adapted from [44,45].

and results in complex learning machine. Greater $\varepsilon$ values result in more flat estimates of the regression function. Determining appropriate values of $C$ and $\varepsilon$ is often a heuristic trial-and-error process. Fig. 6 shows the general network architecture of SVM.

### 2.3. Variants of support vector machines

- Least-squares support vector machines.
- Linear programming support vector machines.
- Nu-support vector machines.

Least-squares support vector machines (LS-SVM) proposed by Suykens et al. [17] is a widely applicable and functional machine learning technique for both classification and regression. The solution of LS-SVM is derived out of linear Karush–Kuhn–Tucker equations instead of a quadratic programming problem of traditional SVM. A major drawback of LS-SVM is that the use of a squared loss function without any regularization leads to less robust estimates. In order to avoid this, the weighted LS-SVM is adopted wherein small weights are assigned to the data and a two-stage training method is proposed.

Originally, Vapnik proposed linear-programming support vector machines (LP-SVM) based on the heuristic that an SVM with less support vectors has the better generalization ability. Support vector machines using linear programming technique have been demonstrated to achieve sparseness and been extensively studied in the framework of feature selection [18–20], scalability of optimization [21], margin maximization, error bound of classification and convergence behaviour of learning, etc. Compared to the conventional QP-SVM, the LP-SVM lacks many features that are unique to the QP-SVM such as a solid theoretical foundation and an elegant dual transformation to introduce kernel functions.

The Nu-Support vector machines ($\nu$-SVM) has been derived so that the soft margin has to lie in the range of zero and one [22], where the SVM employs inhomogeneous separating hyperplanes, that is, it do not essentially includes the origin. The parameter '$\nu$' does not govern the trade-off between the training error and the generalization error but instead it now plays two roles; firstly it is the upper bound on the fraction of margin errors and secondly it is the lower bound on the fraction of support vectors. Parameter '$\nu$' in Nu-SVC/one-class SVM/Nu-SVR approximates the fraction of training errors and support vectors.

### 3. Applications of support vector machines in hydrology

A lot of examples that discuss different applications of SVMs for hydrological prediction and forecasting can be found in the literature. The sections below illustrate a few examples of various

than $\varepsilon$ are ignored and $b$ is a constant. The index $i$ labels the '$n$' training cases. The $y \in \pm 1$ is the class labels and $u_i$ is the independent variable.

Then, the dual form of the nonlinear SVR can be expressed as

$$\max_{\underline{\alpha}_i, \bar{\alpha}_i} \quad -\frac{1}{2}\sum_{i,j=1}^{n}(\underline{\alpha}_i - \bar{\alpha}_i)(\underline{\alpha}_j - \bar{\alpha}_j)\langle\phi(u_i) \cdot \phi(u_j)\rangle$$
$$-\varepsilon\sum_{i=1}^{n}(\underline{\alpha}_i + \bar{\alpha}_i) + \sum_{i=1}^{nl}v_i(\underline{\alpha}_i - \bar{\alpha}_i)$$

$$\text{subject to} \quad \begin{array}{l} \sum_{i=1}^{n}\underline{\alpha}_i - \bar{\alpha}_i = 0 \\ 0 \le \underline{\alpha}_i \le C, i = 1, 2, \ldots, n \\ 0 \le \bar{\alpha}_i \le C, i = 1, 2, \ldots, n \end{array} \tag{8}$$

The "kernel trick" $K(u_i, u_j) = \langle \phi(u_i), \phi(u_j) \rangle$ is used for computations in input space $\chi$ to fetch the inner products into feature space $\Im$. Any function satisfying Mercer's theorem [16] should be used as kernels.

Finally, the decision function of nonlinear SVR with the allowance of the kernel trick is expressed as follows.

$$g(u) = \sum_{i,j=1}^{l}(\underline{\alpha}_i - \bar{\alpha}_i)K\langle u_i \cdot u \rangle + b \tag{9}$$

The parameters that impact over the effectiveness the nonlinear SVR are the cost constant $C$, the radius of the insensitive tube $\varepsilon$, and the kernel parameters. These parameters are mutually dependent over one another and hence altering the value of one parameter affects the other linked parameters also. The parameter $C$ checks for the smoothness/flatness of the approximation function. A smaller value of $C$ yields a learning machine with poor approximation due to underfitting of training data. A greater $C$ value overfits the training data and sets its objective to minimize only the empirical risk making way for more complex learning. The parameter $\varepsilon$ is related with smoothening the complexity of the approximation function and controls the width of the $\varepsilon$-insensitive zone used for fitting the training data. The parameter $\varepsilon$ influences over the number of support vectors, and then both the complexity and the generalization capability of the approximation function is dependent upon its value. It also governs the precision of the approximation function. Smaller values of $\varepsilon$ lead to more number of support vectors
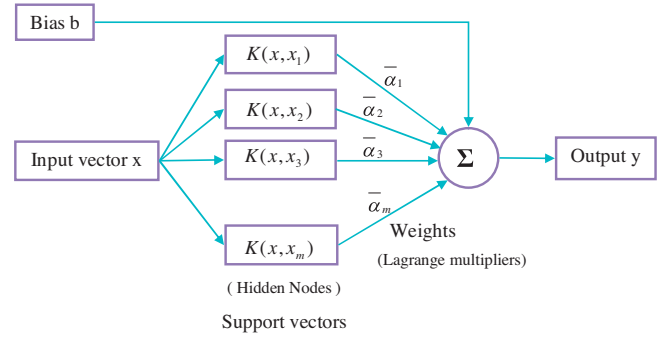
applications of SVMs in the field of surface and ground water hydrology respectively.

### 3.1. Applications in the field of surface water hydrology

#### 3.1.1. Rainfall and runoff forecasting

Several methods/models have been developed by various researchers in order to simulate the rainfall and runoff processes. Hence there is a need for the selection of a suitable robust model for efficient planning and management of watersheds. The intention of the present study is to survey the major proposals in rainfall and runoff forecasting methodology using SVMs.

Many researchers have investigated the potential of SVMs in modelling watershed runoff based on rainfall inputs. Dibike et al. [23] investigated on using SVMs for rainfall-runoff modelling. The daily rainfall, evaporation, and stream-flow data of three different catchments of different rainfall intensities were pre-processed to obtain data format appropriate to SVM and ANN. During SVM training, three types of kernel functions namely – polynomial kernel, RBF kernel and neural network kernel were employed. By trial-and-error process, the parameter $\varepsilon$ corresponding to the error insensitive zone, the capacity factor $C$, and other kernel-specific parameters were set to optimal values. On an average SVM technique achieved a 15% increase of accuracy in the estimation of runoff during the verification period when compared to ANN model. They conclude by remarking the hardship involved during establishing the optimal values for parameters $C$ and $\varepsilon$ and call it as a "heuristic process" and recommends for automation of this process.

Bray and Han [24] emphasize using SVMs for identification of a suitable model structure and its relevant parameters for rainfall runoff modelling. Their testing and training data were composed using rainfall and flow data series of the Bird Creek catchment. Scaling factors were applied to rainfall and flow variables as they had varying units and magnitudes. They have proposed flowcharts for model selection and adaptation of LIBSVM software. An attempt has been made to explore the relationships among various model structures ($\xi$-SV or $\nu$-SV regression), kernel functions (linear, polynomial, radial basis and sigmoid), scaling factor, model parameters (cost $C$, epsilon) and composition of input vectors. Their training results depict that, SVM learnt well with all rain and flow input combinations even though there were slight underestimates in elevated flows and their duration in all cases. They achieved optimized results with three rain observations and four flow observations in each input vector.

Tripathi et al. [25] demonstrated an SVM approach for statistical downscaling of precipitation at monthly time scale. The approach was applied to meteorological sub-divisions (MSDs) in India and tested for its effectiveness. For each MSD, SVM-based downscaling model (DM) was developed for season(s) with significant rainfall using principal components extracted from the predictors as input and the contemporaneous precipitation observed at the MSD as an output. The multi-layer back propagation technique verified that the proposed DM performed well than the conventional downscaling models. Later on to obtain the future projections of precipitation for the MSDs, the SVM-based DM were employed which utilized the second generation coupled global climate model (CGCM2). SVMs proved to be much more superior to conventional artificial neural networks in the field of statistical downscaling for conducting climate impact studies. They concluded that, SVMs was perfectly suited for the downscaling task because of their good generalization performance in capturing non-linear regression relationships between predictors and predictand, in spite of their incapability to understand problem domain knowledge.

Anirudh and Umesh [26] designed an SVM model for the prediction of occurrence of rainy days. The data collected for over a time period of 2061 days encompassing daily values of maximum and minimum temperatures, pan evaporation, relative humidity, wind speed were considered to be model inputs and corresponding days of rainfall were regarded as target variable. They deployed Linear and Gaussian kernel functions for mapping the nonlinear input space to a higher dimensional feature space. Their study stressed the importance of the model parameters namely cost parameter C and gamma parameter $\gamma$ for effective SVM classification. Satisfactory performance was derived out of SVM model and the Gaussian kernel outperformed the linear one.

Chen et al. [27] attempted to evaluate the impact of climate change on precipitation utilizing the potential of smooth support vector machine (SSVM) method in downscaling general circulation model (GCM) simulations in the Hanjiang Basin. The performance of SSVM was compared with the ANN during the downscaling of GCM outputs to daily precipitation. NCEP/NCAR reanalysis data were used to establish the statistical relationship between large-scale circulation and precipitation. The daily precipitation from 1961 to 2000 was used as predictand and a regression function was developed over it. The determining parameters of SSVM were: the width of Radial basis function ($\mu$) and the penalty factor ($C$). The developed SSVM downscaling model showed satisfactory performance in terms of daily and monthly mean precipitation in the testing periods. However, they conclude by stating the drawbacks of the SSVM method in reproducing extreme daily precipitation and standard deviation.

#### 3.1.2. Streamflow and sediment yield forecasting

The simulation of short-term and long-term forecasts of streamflow and sediment yield plays a significant role in watershed management dealing with nature of the storm and streamflow patterns, assessment of onsite rates of erosion and soil loss within a drainage basin, measures for controlling excess runoff, etc. Forecasting streamflow hours, days and months well in advance is very important for the effective operation of a water resources system. The ill effects of natural disaster such as floods can be prevented by prediction of accurate or at least reasonably reliable streamflow patterns.

Jian et al. [28] noticed the importance of accurate time- and site-specific forecasts of streamflow and reservoir inflow for effective hydropower reservoir management and scheduling. They used monthly flow data of the Manwan Reservoir spanning over a time period from January 1974 to December 2003; the data set from January 1974 to December 1998 were used for training and the data sets from January 1999 to December 2003 served for validation. They also had the objective of finding optimal parameters – $C$, $\varepsilon$ and $\sigma$ of RBF kernel. So, a shuffled complex evolution algorithm (SCE-UA algorithm) involving exponential transformation was determined to identify appropriate parameters required for the SVM prediction model. The SVM model gave a good prediction performance when compared with those of ARMA and ANN models. The authors point out, SVMs distinct capability and advantages in identifying hydrological time series comprising nonlinear characteristics and also its potential in the prediction of long term discharges.

Mesut [29] utilized SVMs for predicting suspended sediment concentration/load in rivers. The method was applied to the observed streamflow and suspended sediment data of two rivers in the USA. The predicted suspended sediment values obtained from the SVM model were found to be in good agreement with the original estimated ones. Negative sediment estimates, which were encountered in the soft computing calculations using Fuzzy logic, ANN and Neuro-fuzzy, were not produced by SVM prediction. The best suspended sediment estimates according to the error criteria were obtained with the SVM procedure presented in this study. The suitable selection hyper-parameters and finding out their limiting values is important to model any time series.

Arun and Mahesh [30] researched on predicting the maximum depth of scour on grade-control structures like sluice gates, weirs and check dams, etc. From the available laboratory and field data of earlier published studies, they attempted to explore the potential of SVMs in modelling the scour. Polynomial kernel-based SVMs provided encouraging performance in up scaling results of large scale dataset, several trials were conducted to analyze the effect of various input parameters in predicting the scour depth for Missiga stream data. The results from the SVM based modelling approach indicated an improved performance in contrast to both the empirical relation and back propagation neural network approach with the laboratory data. Its effectiveness was also examined in predicting the scour from a reduced model scale (laboratory data) to the large-scale field data. They concluded that both the RBF and polynomial kernel-based SVMs perform well in comparison to back propagation ANN.

Debasmita et al. [31] emphasized the use of SVMs to simulate runoff and sediment yield from watersheds. They simulated daily, weekly, and monthly runoff and sediment yield from an Indian watershed, with monsoon period data, using SVM, a relatively new pattern-recognition algorithm. In order to evaluate model performance they used correlation coefficient for assessing variability; coefficient of efficiency for assessing efficiency; and the difference of slope of a best-fit line from observed-estimated scatter plots to 1:1 line for assessing predictability. The results of SVM were compared with those of ANN. Runoff estimation was also carried out through the multiple regressive pattern recognition technique (MRPRT). The results of MRPRT did not show any significant improvement when compared to that of SVM; hence, it was ignored for the simulation of sediment yield. They observed significant improvement in performance of the proposed SVM model in training, calibration and validation as compared to ANN. They acknowledge that SVM serves to be an efficient alternative to ANN for runoff and sediment yield predictions.

Hazi et al. [32] proposed an SVM model to predict sediment loads in three Malaysian rivers – Muda, Langat, and Kurau where SVM was employed without any restriction to an extensive database compiled from measurements in these rivers. They employed Neurosolutions 5.0 toolbox, developed by Neurodimension, Inc. for the development of SVM model. They initially fixed model parameters $\alpha_i$ and $\varepsilon$ to be 1 and 0 respectively. The optimal value of $\varepsilon$ was obtained using a genetic algorithm. Their model produced good results compared to other traditional sediment-load methods. The performance of the SVM model demonstrated its predictive capability and the possibility of the generalization of the model to nonlinear problems for river engineering applications. They were successful in predicting total load transport in a great variety of fluvial environments, including both sand and gravel rivers using SVM. The predicted mean total load was almost in perfect agreement with the measured mean total load.

Zahrahtul and Ani [33] investigated the potential of the SVM model for streamflow forecasting at ungauged sites and compared its performance with other statistical method of multiple linear regression (MLR). They used the annual maximum flow series of 88 water level stations in Peninsular Malaysia as data. Radial-basis function (RBF) kernel was used during training and testing as it provided more flexibility with fewer parameters along with mapping nonlinear data into a possibly infinite dimensional space. The performances of both models were assessed by three quantitative standard statistical indices – mean absolute error (MAE), root mean square error (RMSE) and Nash-Sutcliffe coefficient of efficiency (CE). Their study results reveal that SVM model outperformed the prediction ability of the traditional MLR model under all of the designated return periods.

Parag et al. [34] proposed the potential of least square-support vector regression (LS-SVR) approach to model the daily variation of river flow. Their main intention behind using LS-SVM was it offered a higher computational efficiency than that of traditional SVM method and the training of LS-SVM required only the solution of a set of linear equations instead of the long and computationally demanding quadratic programming problem as involved in the traditional SVM. By exploiting the capability of LS-SVR, a couple of issues which hindered the generalization of the relationship between previous and forthcoming river flow magnitudes were dealt. The LS-SVR model development was carried out with pre-construction of river flow regime and its performance was assessed for both pre- and post-construction of the dam for any perceivable difference. A multistep-ahead prediction was carried out in order to investigate the temporal horizon over which the prediction performance was relied upon and the LS-SVR model performance was found to be satisfactory up to 5-day-ahead predictions even though the performance was decreasing with the increase in lead-time. The LS-SVR results were outperforming when compared to that of the ANNs model for all the lead times.

### 3.1.3. Evaporation and evapotranspiration forecasting

Evaporation takes place whenever there is a vapour pressure deficit between a water surface and the overlying atmosphere and sufficient energy is available. The most common and important factors affecting evaporation are solar radiation, temperature, relative humidity, vapour pressure deficit, atmospheric pressure, and the wind speed. Evaporation losses should be considered in the design of various water resources and irrigation systems. In areas with little rainfall, evaporation losses can represent a significant part of the water budget for a lake or reservoir, and may contribute significantly to the lowering of the water surface elevation. Evapotranspiration (ET) is defined as the loss of water to the atmosphere by the combined processes of evaporation from the soil and plant surfaces and transpiration from plants. It is necessary to quantify ET for work dealing with water resource management or environmental studies.

Moghaddamnia et al. [35] presented a preliminary study on evaporation and investigated the abilities of SVMs to enhance the prediction accuracy of daily evaporation in the Chahnimeh reservoirs of Zabol in the southeast of Iran. The SVM technique developed for simulating evaporation from the water surface of a reservoir was evaluated based on criteria such as root mean square error (RMSE); mean absolute error (MAE); mean square error (MSE) and coefficient of determination ($R^2$). A technique called Gamma Test was employed for the selection of relevant variables in the construction of non-linear models for daily (global) evaporation estimations. Linear, polynomial, RBF, and sigmoid kernel functions were used in the study and it was observed that RBF kernel outperformed in all the cases. They compared three types of SVMs namely one-class SVM, epsilon-SVR and nu-SVR; the results reveal that, epsilon-SVR model revealed better results.

Eslamian et al. [36] constituted an SVM and an ANN model for the estimation of monthly pan evaporation using meteorological variables like air temperature, solar radiation, wind speed, relative humidity and precipitation. Here the SVM formulated using RBF kernel preferred a grid search for cross-validation to prevent overfitting problem. Wind speed and air temperature were identified to be the most sensitive variables during the computation. Major observations include: (1) the computation effort was significantly smaller with SVMs than the ANNs algorithm; (2) based on the values of mean square error (MSE), the SVM approach performed better than ANN.

Pijush [37] adopted least square support vector machine (LS-SVM) for prediction of Evaporation Losses (EL) in Anand Sagar reservoir, Shegaon (India). Mean air temperature (°C), average wind speed (m/s), sunshine hours (h/day), and mean relative humidity (%) were used as inputs of LSSVM model. The main objectives

put forth by the author includes – developing an equation for the prediction of EL; determination of the error bar of predicted EL; performing sensitivity analysis for investigating the importance of each of the input parameters. A comparative study has been presented between LSSVM and ANN models. The developed LSSVM proved to be a powerful robust tool in estimation of EL and also depicted some prediction uncertainty. Sensitivity analysis indicated that temperature had the maximum effect on EL.

Hossein et al. [38] studied the potential of SVM, adaptive neuro-fuzzy inference system (ANFIS), multiple linear regression (MLR) and multiple non-linear regression (MNLR) for estimating ETo. In addition, four temperature-based and eight radiation-based ETo equations were tested against the PMF-56 model. The performance of SVM and ANFIS models for ETo estimation were better than those achieved using the regression and climate based models and this confirmed the ability of these techniques in ETo modelling in semi-arid environments. Some prominent observations include – SVM's were capable in selecting the key vectors as its support vectors during the training process and removed the non-support vectors automatically from the model. This made the model cope up well with noisy conditions.

### 3.1.4. Lake and reservoir water level prediction

Khan and Coulibaly [39] examined the potential of the support vector machine (SVM) in long-term prediction of lake water levels. Lake Erie mean monthly water levels from 1918 to 2001 were used to predict future water levels up to 12 months ahead. Here the optimization technique (linearly constrained quadratic programming function) used in the SVM for parameter selection has made it superior to traditional neural networks. They found RBF kernel to be the most appropriate and adopts it with a common width of ($\sigma = 0.3$) for all points in the data set. Further they set the values for regularization parameter $C = 100$ and $\varepsilon$-insensitive loss function with $\varepsilon = 0.005$ chosen by the trial-and-error approach. 80–90% of the input data were identified to be support vectors in the model. The performance was compared with a multilayer perceptron (MLP) and with a conventional multiplicative seasonal autoregressive model (SAR). On the whole, SVM showed good performance and was proved to be competitive with the MLP and SAR models. For a 3- to 12-month-ahead prediction, the SVM model outperformed the two other models based on the root-mean square error and correlation coefficient performance criteria. Furthermore, the SVM demonstrated inherent advantages in formulating cost functions and of quadratic programming during model optimization.

Afiq et al. [40] investigated on using SVM to forecast the daily dam water level of the Klang reservoir, located in peninsular Malaysia. In order to ascertain the best model the authors incorporated four different categories, firstly the best input scenario was applied by considering both rainfall R $(t - i)$ and the dam water level L $(t - i)$; secondly the $v$-SVM regression was selected as the best regression type, thirdly the 5-fold cross-validation was considered to produce the most accurate results and finally, all the results were combined to determine the best time lag. The performance of the $v$-SVM model was assessed and compared with that of ANFIS model using statistical parameters like RMSE, MAE and MAPE. Their study results reveal that $v$-SVM model outperformed the predictions of ANFIS.

### 3.1.5. Flood forecasting

Flood forecasting is a significant non-structural approach for flood mitigation. Floods generally develop over a period of days, when there is too much rainwater to fit in the rivers and water spreads over the land next to it (the 'floodplain'). In practice, the river stage is the parameter that is considered in flood forecasting studies. The current paper presents the examples of real-time flood forecasting models using SVMs.

Yu et al. [41] established a real-time flood stage forecasting model based on SVR employing hydrological concept of the time of response for recognizing lags associated with the input variables. Their input vector accounted for both rainfall and river stage, to forecast the hourly stages of the flash flood. The RBF Kernel was employed during modelling of nonlinear SVR. For finding optimal parameters of SVR, a two-step grid search method was employed. Two model structures, with different relationships between the input and output vectors were formulated to predict one- to six-hour-ahead stage forecasts in real time. Both the model structures yielded almost similar forecasting results based on validation events. Their study depicted the robust performance of SVR in stage forecasting and presented a simple and systematic method for selection of lags of the variables and the model parameters for flood forecasting.

Han et al. [42] applied SVM to forecast flood over Bird Creek catchment and focuses on some significant issues related to developing and applying SVM in flood forecasting. They used LIBSVM for modelling along with 'Gunn' toolbox for data normalization. They demonstrated on how to optimally device a model from among a large number of various input combinations and parameters in real-time modelling. Their results show that the linear and non-linear kernel functions (i.e., RBF) can generate superior performance against each other under different circumstances in the same catchment. Their prominent observation is that SVM also suffers from over-fitting and under-fitting problems and the over-fitting is more damaging than under-fitting.

Wei et al. [43] examined the potential of a Distributed SVR (D-SVR) model for river stage prediction, the commonly used parameter in flood forecasting. The proposed D-SVR implemented a local approximation to training data, since partitioned original training data were independently fitted by each local SVR model. To locate the optimal triplets ($C, \varepsilon, \sigma$) a two-step Genetic Algorithm was employed to model D-SVR. In order to assess the performance of D-SVR, prediction was also arrived at via LR (linear regression), NNM (Nearest-neighbour method), and ANN-GA (genetic algorithm-based ANN). The validation results of the proposed D-SVR model revealed that the river flow prediction were better in comparison with others, and the training time was considerably reduced when compared with the conventional SVR model. The key features influencing over the performance of D-SVR was that it implemented a local approximation method in addition to principle of structural risk minimization.

Chen and Yu [44,45] represented SVM as a network architecture resembling artificial neural networks (multilayer perceptron) and was pruned to acquire model parsimony or improved generalization. Their study established two approaches/methods of pruning the support vector networks, and employing them to a case study of real-time flood stage forecasting. The first approach/method pruned the input variables by the cross-correlation method, while the other approach pruned the support vectors according to the derived relationship among SVM parameters. These pruning methods did not alter the SVM algorithm, and hence the pruned models still had the optimal architecture. The real-time flood stage forecasting performance pertaining to the primary and the pruned SVM models were evaluated, and the comparison results indicated that the pruning reduced the network complexity but did not degrade the forecasting ability.

### 3.1.6. Drought forecasting

Water stress and drought are conditions increasingly faced in many parts of the world. Drought occurrence varies in spatial coverage, frequency, intensity, severity and duration as it is the extended period of dry season or shortfall in the moisture from the surrounding environment. Superficially one can say that drought periods are associated with periods of anomalous atmospheric circulation

pattern, which finally induces risks. The problem of drought forecasting is closely related to the problem of long-term forecasting of temperatures and precipitation. Drought impacts both surface and ground water resources and can lead to reductions in water supply, diminished water quality, crop failure, diminished power generation and as well as host of other economic and social activities.

Shahbazi et al. [46] applied the SVM technique to develop models for forecasting seasonal standardized precipitation index (SPI). Their intent was to evaluate the performance of SVM in recognizing repetitive statistical patterns in variations of meteorological variables in a reasonably large area surrounding Iran, which would also make way for enhancing the prediction capability of seasonal values of SPI-indicator of meteorological drought severity. They used historical time series of the meteorological variables collected from NCEP/NCAR reanalysis dataset as the model predicts. Mutual Information (MI) index was utilized as feature selection filter to decrease the SVM input space dimension. The model was trained to predict seasonal SPIs in autumn, winter, and spring seasons. The performance of the proposed model demonstrated that the seasonal SPI values can be predicted with 2–5 months lead-time with enough accuracy which can be utilized for long-term water resources planning and management. SVM outperformed ANN in almost all scenarios indicating higher accuracy of SVM predictions. Their major observation was that linear kernel function outperformed the other kernel functions in the SPI prediction. They concluded that, the prediction lead-time should be taken into account while evaluating the performance of the proposed models in the prediction of seasonal SPI's.

Belayneh and Adamowski [47] compared the effectiveness of different data-driven models like artificial neural networks (ANN), wavelet neural networks (WNN), and support vector regression (SVR) for forecasting drought conditions in the Awash River Basin of Ethiopia. A 3-month standard precipitation index (SPI) and SPI 12 were forecasted over lead times of 1 and 6 months in each sub-basin. The authors effectively illustrate the changes in the seasonal trends in precipitation using 3-month SPI, a good indicator of agricultural drought. The performance of all the models was assessed and compared using RMSE and coefficient of determination-$R^2$. The WNN forecasts of SPI 3 and 12 in all the sub basins were found to be very effective. WNN models revealed superior correlation between observed and predicted SPI compared to simple ANNs and SVR models.

Ganguli and Reddy [48] studied and analyzed the climate teleconnections with meteorological droughts to develop ensemble drought prediction models utilizing support vector machine (SVM) – copula approach over Western Rajasthan (India). The standardized precipitation index (SPI) was used to identify the meteorological droughts. Regional droughts exhibited interannual as well as interdecadal variability during the analysis of large-scale climate forcing characterized by climate indices such as El Niño Southern Oscillation, Indian Ocean Dipole Mode and Atlantic Multidecadal Oscillation. SPI-based drought forecasting models were developed with up to 3 months lead time taking into consideration that potential teleconnections exist between regional droughts and climate indices. Since the traditional statistical forecast models failed to capture nonlinearity and nonstationarity correlated with drought forecasts, support vector regression (SVR), was adopted to predict the drought index, and the copula method was utilized to simulate the joint distribution of observed and predicted drought index. The ensembles of drought forecast are modelled utilizing the copula-based conditional distribution of an observed drought index trained over the predicted drought index. The authors developed two variants of drought forecast models, namely a single model for all the periods in a year and separate models for each of the four seasons in a year and subsequently mentioned that there was an enhancement in ensemble prediction of drought indices

in combined seasonal model over the single model without seasonal partitions. The performance of developed models is validated for predicting drought time series for 10 years data. The authors conclude that the proposed SVM – copula approach improves the drought prediction capability and caters assessment of uncertainty associated with drought predictions.

### 3.2. Applications in the field of ground water hydrology

#### 3.2.1. Ground water level prediction

The water table is the surface of the saturated zone, below which all soil pores or rock fractures are filled with water. Precise groundwater level modelling and forecasting influence deep foundation designs, land use and ground water resources management. Simulating of groundwater level by means of numerical models requires various hydrological and geological parameters. In these papers below, application of SVM in developing a reliable groundwater level fluctuation forecasting system to generate trend forecasts is being discussed (Table 1).

Jin et al. [49] proposed SVM based dynamic prediction of groundwater level. Taking into account of groundwater level dynamic series length and peak mutation characters, they proposed the least squares support vector machine (LS-SVM) arithmetic based on chaos optimization peak value identification. Considering the effective generalization capability of the sigmoid RBF function, it was employed as a kernel function in their study. Their model performance illustrated that; the fitted values, the tested values and the predicted values had smaller difference from their real values. Their simulation results depicted that, the $C$-$\lambda_i v_i$-SVM model was very effective in reflecting the dynamic evolution process of groundwater level well.

Mohsen et al. [50] made an attempt with SVMs and ANNs for predicting transient groundwater levels in a complex groundwater system under variable pumping and weather conditions. Their objective of modelling was to predict the water level elevations at different time horizons, i.e., daily, weekly, biweekly, monthly, and bimonthly. By minimizing the 10-fold cross validation estimates of the prediction error (arrived by utilizing grid search), the optimal set of hyper-parameters were decided. Radial basis function (RBF) was identified to be a proper kernel function for their study. It was found that SVM outperformed ANN in terms of prediction accuracy and generalization for longer prediction horizons even when fewer data events were available for model development. A high consistency was seen in training and testing phases of SVM modelling when compared to ANN. They conclude that in groundwater problems, particularly when data are relatively sparse, SVM offered a better alternative to ANN with its global optimization algorithm.

Heesung et al. [51] developed two nonlinear time-series models for predicting the groundwater level (GWL) fluctuations using ANNs and SVMs. The objective of their study was to develop and compare data-based time-series forecasting models for short-term GWL fluctuations in a coastal aquifer due to recharge from precipitation and tidal effect. The dilemma of salt water intrusion necessitated for the GWL prediction in a coastal aquifer. The input variables like past GWL, precipitation, and tide levels were utilized for the GWL prediction of two wells of a coastal aquifer. The model performance implicated that SVM was better than ANN during both model training and testing phases. The SVM model generalized well with input structures and lead times than an ANN model. The uncertainty analysis of model parameters detected an equifinality of model parameter sets and higher uncertainty in the ANN model than SVM in this case.

#### 3.2.2. Soil moisture estimation

Gill et al. [52] applied SVM's to study the distribution and variation of soil moisture which is advantageous in predicting and

**Table 1**
Summary of various SVM models in hydrology adopted by different authors.

| | Sl. no. | Authors | Applications | Data used/model developed |
|---|---|---|---|---|
| Surface water hydrology | 1 | Dibike et al. [23] | Rainfall-runoff modelling | The daily rainfall, evaporation, and stream-flow data of three different catchments of different rainfall intensities |
| | 2 | Bray and Han [24] | Rainfall-runoff modelling | Rainfall and flow data series of the Bird Creek catchment. Scaling factors were applied to rainfall and flow variables as they had varying units and magnitudes |
| | 3 | Tripathi et al. [25] | Downscaling of precipitation | Statistical downscaling of precipitation at monthly time scale for each meteorological sub-divisions (MSDs) in India |
| | 4 | Anirudh and Umesh [26] | Predicted occurrence of rainy days | The data collected for over a time period of 2061 days encompassing daily values of maximum and minimum temperatures, pan evaporation, relative humidity, wind speed were considered to be model inputs and corresponding days of rainfall were regarded as target variable |
| | 5 | Chen et al. [27] | Studied the Impact of climate change on precipitation | NCEP/NCAR reanalysis data were used to establish the statistical relationship between large-scale circulation and precipitation. The daily precipitation from 1961 to 2000 was used as predictand |
| | 6 | Jian et al. [28] | Forecasted streamflow and reservoir inflow | Monthly flow data of the Manwan reservoir spanning over a time period from January 1974 to December 2003 |
| | 7 | Mesut [29] | Predicted suspended sediment concentration/load | The method was applied to the observed streamflow and suspended sediment data measured at Quebrada Blanca station during the water years 1990 and 1992–1995 in the USA |
| | 8 | Arun and Mahesh [29] | Predicted the maximum depth of scour | Used laboratory and field data of earlier published studies. Tried to analyze the effect of various input parameters in predicting the scour depth for Missiga stream data |
| | 9 | Debasmita et al. [31] | Simulate runoff and sediment yield | The rainfall, runoff and sediment yield data of monsoon period (June 1–October 31) for 1984–1987 was used for training the SVM model, and the data of 1988–1989 and 1992–1995 for calibration and validation |
| | 10 | Hazi et al. [32] | Predicted sediment loads in three Malaysian rivers | The calibration of the SVM model was performed based on 214 input target pairs of the collected data, composed of sediment data sets, and the parameters that affect the total sediment load |
| | 11 | Zahrahtul and Ani [33] | Forecasted streamflow at ungauged sites | The annual maximum flow series from 88 water level stations in Peninsular Malaysia as data |
| | 12 | Parag et al. [34] | Modelled the daily variation of river flow | LS-SVR model development was carried out with pre-construction of river flow regime and its performance was assessed for both pre- and post-construction of the dam. River flow data of Sandia station, operated by the Water Resources Agency was used |
| | 13 | Moghaddamnia et al. [35] | Enhanced the prediction accuracy of daily evaporation | The daily weather variables of automated weather station namely, Chahnimeh Station of Zabol. The data sample consisted of 11 years (1983–2005) of daily records of air minimum temperature ($T_{min}$), air mean temperature ($T_{mean}$), air maximum temperature ($T_{max}$), wind speed ($W$), saturation vapour pressure deficit (Ed), mean relative humidity (RH mean), 6:30 AM relative humidity (RHAM), 12 noon relative humidity (RH noon), 6:30 PM relative humidity (RHPM), solar radiation (SR) and pan evaporation (E) |
| | 14 | Eslamian et al. [28] | Modelled the estimation of monthly pan evaporation | Meteorological variables like air temperature, solar radiation, wind speed, relative humidity and precipitation data obtained from five meteorological stations located in the east of Esfahan Province |
| | 15 | Pijush [37] | Predicted evaporation losses | Mean air temperature (°C), average wind speed (m/s), sunshine hours (h/day), and mean relative humidity (%) data collected from a reservoir called Anand Sagar, Shegaon (India) were used as inputs of LSSVM model |
| | 16 | Hossein et al. [38] | Estimated reference crop estimation ETo | The data composed of mean, maximum and minimum air temperatures, relative humidity, dew point temperature, water vapour pressure, wind speed, atmospheric pressure, precipitation, solar radiation and sunshine hours for the period 1986–2005 collected from the agency IRIMO, Iran |
| | 17 | Khan and Coulibaly [39] | Long-term prediction of lake water levels | Lake Erie mean monthly water levels from 1918 to 2001 were used to predict future water levels up to 12 months ahead |
| | 18 | Afiq et al. [40] | Forecasted the daily dam water level | The data composed of daily rainfall and water level of the dam collected from the Klang gate dam reservoir |
| | 19 | Yu et al. [41] | Developed real-time flood stage forecasting model | Hourly water stage data (m) from two water level stations, Lan-Yang Bridge and Niu-Tou, and hourly rainfall data (mm) from rainfall stations operated by the Water Resources Agency and the Central Weather Bureau (Taiwan) were collected for this work |
| | 20 | Han et al. [42] | Forecast flood over Bird Creek catchment | The rainfall data used were obtained from 12 rain gauges situated in/near the Bird Creek catchment area for the period October 1955–November 1974. The river flow values were obtained from a continuous stage recorder |
| | 21 | Wu et al. [69] | Modelled river stage prediction | The data of water levels at the Luo-Shan station, a channel reach in the middle stream of the Yangtze River were used to predict the water level of another station at Han-Kou for the next day |
| | 22 | Chen and Yu [44,45] | Pruning of support vector networks on flood forecasting | Hourly water stage data (m) from two water level stations, Lan-Yang Bridge and Niu-Tou, and hourly rainfall data (mm) from rainfall stations operated by the Water Resources Agency and the Central Weather Bureau (Taiwan) were collected for this work |

Table 1 (*Continued*)

| | Sl. no. | Authors | Applications | Data used/model developed |
|---|---|---|---|---|
| | 23 | Shahbazi et al. [46] | Forecasted seasonal standardized precipitation index (SPI) | The historical time series of the meteorological variables collected from NCEP/NCAR reanalysis dataset were used as model predicts |
| | 24 | Belayneh and Adamowski [47] | Forecasted drought conditions in the Awash River Basin | The precipitation data collected from 20 rainfall gauge stations located in the Awash River Basin over the period 1970–2005 were chosen for this study |
| | 25 | Ganguli and Reddy [48] | Developed ensemble drought prediction models | This study used, gridded (0.5° latitude × 0.5° longitude) daily precipitation data for the years 1971–2005 from Western Rajasthan meteorological subdivision of India |
| Groundwater hydrology | 26 | Jin et al. [49] | Predicted dynamic groundwater levels | Historical groundwater level data collected since 1990 from 51 observation wells of the Yichang irrigation district were used in the study |
| | 27 | Mohsen et al. [50] | Predicted transient groundwater levels in a complex groundwater system | In this research, the data used was composed of pumping, weather, and groundwater level data corresponding to the Cooks monitoring well |
| | 28 | Heesung et al. [51] | Predicted groundwater level (GWL) fluctuations | Data was obtained using automated data loggers installed at two observation wells OB2 and BH5 to measure GWL, groundwater temperature, electrical conductivity, and air pressure |
| | 29 | Gill et al. [52] | Analyzed the distribution and variation of soil moisture | Soil moisture and meteorological data from the soil climate analysis network site at Ames, IA |
| | 30 | Wei et al. [43] | Predicted soil water content in the purple hilly area | All the soil samples collected from a depth of 30–40 cm from the study area located in Southwest University in Chongqing were analyzed in the lab and the dataset was prepared. Monitoring was done in every 5 days for a period of December 2001 to April 2004 |
| | 31 | Sajjad et al. [53] | Estimated soil moisture (SM) using remote sensing data | The model inputs were backscatter and incidence angle from tropical rainfall measuring mission (TRMM) and normalized difference vegetation index (NDVI) from advanced very high resolution radiometer (AVHRR) |
| | 33 | Seyyed et al. [75] | Characterized an unknown contamination source in terms of location and amount of leakage | The proposed methodology of characterizing an unknown pollution source was applied to a part of Tehran refinery where the existing groundwater quality data indicated that the Tehran aquifer in the study area was seriously polluted due to the leakage of petrochemical products from storage tanks and pipelines |
| | 34 | Liu et al. [54] | Groundwater quality assessment model | Karst water quality data of two wells and four fountains of the years 1988, 1990, 1994 and 2000 are used to assess and compare with the main karst water source |

| | Sl. no. | Authors | Applications | Data used/model developed | Hybrid technique used |
|---|---|---|---|---|---|
| Hybrid-SVM models | 35 | Sivapragasam et al. [4] | Forecasted two hydrological time series | Daily rainfall data from Station 23 of the total 64 rainfall stations located on the main island of Singapore were used in this research | Singular spectrum analysis |
| | 36 | Remesan et al. [55] | Established a real-time flood forecasting model | The data from a small watershed (the Brue catchment in southwest England UK), for a period of 7 years were available | Wavelet transform |
| | 37 | Ozgur and Mesut [56] | Forecasted monthly streamflow | In this study, the daily precipitation data of two stations located in Aegean Region of Turkiye are used. The observed data was 15 years (5479 day) long with an observation period between January 1987 to December 2001 | Discrete wavelet transform |
| | 38 | Wei and Chih-Chiang [57] | Forecasted hourly precipitation during tropical cyclone | Data from 73 typhoons affecting the Shihmen Reservoir watershed were included in the analysis | Wavelet transform |
| | 39 | Sudheer et al. [58] | Predicted monthly streamflows | In this study, the monthly streamflow data of Polavaram Station on Godavari River and Vijayawada Station on Krishna River of India were used | Quantum behaved particle swarm optimization (QPSO) |

understanding various hydrologic processes like energy and moisture fluxes, weather changes, irrigation scheduling, rainfall/runoff generation and drought. Four and seven days ahead SVM predictions were generated using soil moisture and meteorological data. SVM Predictions were in good agreement with actual soil moisture measurements. The performance of SVM modelling was compared with that obtained from ANN models and the SVM models outperformed the ANN predictions in soil moisture forecasting.

Wei et al. [43] investigated the potential of SVM's in predicting soil water content in the purple hilly area and also evaluated its performance by comparing it with ANN time series prediction model. The feasibility of applying SVM's to soil water content

forecasting is successfully demonstrated. After numerous experiments the authors have proposed a set of SVR parameters for predicting soil water content time series. The model performance also reveals that the SVR predictor significantly outperforms the other baseline predictors like ANN.

Sajjad et al. [53] examined the potential of SVMs regression in soil moisture (SM) estimation using remote sensing data. The model inputs were backscatter and incidence angle from Tropical Rainfall Measuring Mission (TRMM) and Normalized Difference Vegetation Index (NDVI) from Advanced Very High Resolution Radiometer (AVHRR). Simulated SM (%) time series for the study sites were available from the Variable Infiltration Capacity (VIC) Three Layer model for top 10 cm layer of soil. The selected study area had varying vegetation cover comprising of low, medium, and dense vegetation. RBF kernel was utilized in their nonlinear SVM regression. The SVM model was developed and tested for ground soil moisture data, whose results addressed that SVM model was capable of capturing the interrelations among soil moisture, backscatter, and vegetation than ANN and multivariate linear regression (MLR) models. The authors conclude by suggesting that SVM technique proves to be a better alternative to the computationally expensive and data intensive physical models.

### 3.2.3. Ground water quality assessment

Utilizing groundwater quality monitoring data Seyyed et al. [75] investigated a novel methodology for estimating the location and amount of leakage from an unknown pollution source. They developed two probabilistic simulation models, namely probabilistic support vector machines (PSVMs) and probabilistic neural networks (PNNs) in order to identify and evaluate the core characteristics of an unknown groundwater pollution source, in terms of magnitude, location and timing. The groundwater quantity and quality simulation models derived out of MODFLOW and modular three-dimensional transport model (MT3D) were coupled with a multi-objective optimization model, namely Non-dominated Sorting Genetic Algorithm-II (NSGA-II) in their proposed methodology. The validation accuracy of PSVM was found to be superior to that of PNN. Their results reveal that, the probability mass function of an unknown pollution source location and the relative error in estimating the amount of leakage based on the observed concentrations of water quality indicator at the monitoring wells were better assessed by the trained probabilistic simulation models (PSVM and PNN) in real-time groundwater monitoring.

Liu et al. [54] established a multiple-factor water quality assessment model based on SVM. She classified water quality assessment as a multivariate nonlinear system and randomly selected eight different assessment factors (pH, total hardness, TDS, $SO_4^{2-}$, $Cl^-$, $NO_3$-N, $NO_2$-N and $F^-$) to generate sample set. All the test samples of the model were classified correctly after training using SVM and were applied for the assessment of the karst groundwater sample. The proposed SVM exhibited constructive features in pattern recognition and suggested that function fitting problem and multi-category classification problems can be solved together by using an SVM regression algorithm. The performance of SVM prediction explained the complex nonlinear relationship between assessment factor and water quality grade.

### 3.3. Applications of hybrid SVM models

A hybrid system has the benefit of encompassing a larger class of systems within its structure, allowing for more flexibility in modelling dynamic phenomena. Hybrid models take advantage of: (1) the unique parameterization of metric models and their ability to efficiently characterize the observational data in statistical terms, and (2) other prior knowledge to test hypotheses about the structure of component hydrological stores.

For the purpose of parameter tuning of traditional SVM, there are many approaches that have been presented regardless of the application domain. As compared to the man-made selection approach in determining the value of parameters of interest, the hybridization of traditional SVM with existing optimization techniques seems much reliable and systematic. Also the other drawbacks of the traditional SVM technique, that is training time consumption and unsystematic, which consequently affect the generalization performance can be efficiently tackled by hybridizing.

The study by Sivapragasam et al. [4] established the application of support vector machine (SVM), for forecasting two hydrological time series: (1) the daily rainfall data; and (2) the daily runoff data. They proposed an efficient prediction technique based on singular spectrum analysis (SSA) (i.e. for pre-processing the data) coupled with SVM method. The information hidden behind short and noisy time series were extracted using SSA without using prior knowledge about the underlying physics of the system. The forecasts are made for both the raw data and the pre-processed data (using SSA). These forecast results were then compared with those achieved (for the raw data) using the nonlinear prediction (NLP) method, which used the concept of phase-space reconstruction. Based on such a comparison, the authors reported that the SSA–SVM method performed significantly better than the NLP method in forecasting and it was seen that the proposed technique yielded a significantly higher accuracy in prediction in the case study on the Singapore rainfall prediction with a correlation coefficient of 0.70 as opposed to 0.51 obtained by NLP.

Remesan et al. [55] proposed the SVM in conjunction with wavelets to establish a real-time flood forecasting model. They compared this model with another hybrid model called the Neuro-wavelet model (NW). The methods were tested using the data from a small watershed (the Brue catchment in south-west England UK), for which 7 years of data records were available. The model results showed that the wavelet-based hybrid models can offer more accurate runoff estimates for flood forecasting in the Brue catchment. The gamma test, another innovative technique was also utilized for training the data length and determining input data structure.

Ozgur and Mesut [56] investigated the accuracy of wavelet and support vector regression (SVR) conjunction model in monthly streamflow forecasting. The conjunction model was formulated by combining two methods, discrete wavelet transforms (DWT) and support vector machine, and was compared with the single SVR. The WSVR model was an SVR model which used sub-time series components obtained using DWT on original data. The Gaussian radial basis function (GRBF) was taken into consideration in this study. The accuracies of WSVR and SVR model were assessed using the root mean square error (RMSE), mean absolute error (MAE), Nash–Sutcliffe coefficient (NS) and correlation coefficient (R) statistics. The test performance indicated that the WSVR approach provided a superior alternative to the SVR model for developing input–output simulations and forecasting monthly streamflows in situations that do not require modelling of the internal structure of the watershed.

Wei and Chih-Chiang [57] established two support vector machine (SVM) based models: (1) the traditional Gaussian kernel SVMs (GSVMs) and (2) the advanced wavelet kernel SVMs (WSVMs) for forecasting hourly precipitation during tropical cyclone (typhoon) events. Historical data of 73 typhoons which affected the Shihmen reservoir watershed were included in the analysis. Their study considered six attribute combinations with different lag times for the forecast target. The prediction outcomes were assessed by employing modified RMSE, bias, and estimated threat score (ETS) results. The model results depicted, improved attribute combinations for typhoon climatologic characteristics and typhoon precipitation predictions occurred at 0-h lag time with modified RMSE values of 0.288, 0.257, and 0.296 in GSVM, WSVM,

and the regression, respectively. Furthermore, WSVM bearing average bias and ETS values close to 1.0 gave better predictions than that of GSVM and regression models.

Sudheer et al. [58] investigated the accuracy of the hybrid SVM-QPSO model (support vector machine-quantum behaved particle swarm optimization) in predicting monthly streamflows. The SVM model with various input structures was designed and the efficient model structure was determined using normalized mean square error (NMSE) and correlation coefficient ($R$). In order to determine the optimal values of SVM parameters and thereby minimize NMSE, another state-of-the-art technique called quantum behaved particle swarm optimization was adapted in this study. Moreover, the institution of mean best position into QPSO was an innovative method incorporated. Subsequently, the performance of the SVM-QPSO model was evaluated and compared thoroughly with the popular forecasting models. The results revealed that SVM-QPSO was considerably a better technique for predicting monthly streamflows as it offered a high degree of accuracy and reliability.

## 4. Selection of optimization algorithm, kernels and related hyper-parameters

We can summarize the SVM algorithm as follows: (1) selection of a suitable kernel function; (2) assigning the value for regularization parameter-$C$; (3) solving the quadratic programming (QP) problem; (4) constructing the discriminant function from the support vectors. In the context of designing a robust SVM model one has to have a clear understanding regarding the terminologies: "training" and "tuning". The term training refers to the process of optimizing the coefficients in the decision function by solving the quadratic programming problem. The term tuning entails the process of selecting the appropriate values for kernel parameters, trade-off/regularization parameter-$C$ and width of $\varepsilon$-insensitive tube. The most efficient SVR model is derived out of tuning these three interdependent parameters-$C$, $\varepsilon$ and kernel parameter for which their (near) optimal values are often obtained by a trial and error method [59]. Grid search and cross-validation error estimation are the standard ways of performing hyper-parameter optimization, but they often suffer from the curse of dimensionality during data analysis. Hence using naive heuristics like nested grid search, where the search resolution increases iteratively, a four- or perhaps even five-dimensional SVM hyper-parameter space can be still be tested sufficiently.

The value $C$ can be regarded as the error penalty which prevents classification errors during SVM training. A larger $C$ will provide us with a larger search space for the QP optimizer which generally increases the duration of the QP search.

The solution of an SVM is the optimum of a well-defined convex optimization problem. Quadratic optimization packages are designed to take advantage of sparsity in the quadratic part of the objective function. The sequential minimal optimization (SMO) introduced by Platt [60] divides large quadratic programming problems into smaller ones thereby solving them separately and analytically. Generic optimization packages sometimes provide optimum results with greater accuracy. Nowadays particle swarm optimization (PSO) and its advanced part quantum-PSO is being incorporated in some case studies for SVM optimization.

Another significant problem in the framework of designing an SVM model is the choice of kernel function. Kernel functions explicitly define the feature space and henceforth the performance of the algorithm is dependent over them. There are several result oriented kernels in SV learning. The list of basic kernels used is presented in the above SVM theory part. In the majority of case studies dealt herein polynomial and Gaussian RBF kernels were adopted for nonlinear SVM modelling. Polynomial kernels are very much

advantageous for increasing the dimensionality as the order of the polynomial defines the smoothness of the function. The polynomial kernel has more hyperparameters than the RBF kernel. For a standard Gaussian RBF kernel the bandwidth parameter $\gamma$ (or $\sigma$) is the only kernel parameter to be defined. When the RBF kernel is used, the SVM algorithm spontaneously determines threshold, centres, and weights that reduces an upper bound on the expected test error [61,62].

## 5. Advantages of support vector machines

- SVMs are capable of producing accurate and robust classification results, even when input data are non-monotone and non-linearly separable. So they can help to evaluate more relevant information in a convenient way [9].
- The structural risk minimization principle provides SVM, the desirable property to maximize the margin and thereby the generalization ability does not deteriorate and is able to predict the unseen data instances [63,64].
- By properly setting the value of $C$ – regularization parameter one can easily suppress the outliers and hence SVM's are robust to noise [5].
- A key feature of SVM is that it automatically identifies and incorporates support vectors during the training process and prevents the influence of the non-support vectors over the model. This causes the model to cope well with noisy conditions [42].
- With some key actual training vectors embedded in the models as support vectors, the SVM has the potential to trace back historical events so that future predictions can be improved with the lessons learnt from the past [42,65].
- Input vectors of SVM are quite flexible; hence various other influential factors (such as temperature, relative humidity, and wind speed) can be easily incorporated into the model [35,37].

## 6. Disadvantages of support vector machines

1. Since SVMs linearize data on an implicit basis by means of kernel transformation, the accuracy of results does not rely on the quality of human expertise judgment for the optimal choice of the linearization function of non-linear input data [66].
2. The main drawback of SVM is, for instance, selection of the suitable kernel function and hyper parameters are heuristic and depend on trial and error process which is a time-consuming approach [41,67].
3. The behaviour of the nonlinear SVR model cannot be easily understood and interpreted due to the inherent complexity involved in mapping nonlinear input space into a high dimensional feature space. Hence the training process becomes much slower when compared to that of linear models [25,68].
4. Poor model extrapolation prevails in case of past data inconsistency as the model completely depends on the past records as support vectors.
5. SVM produces only point predictions and is not designed for probabilistic forecasts.

## 7. Conclusions

Support vector machines represent the most important development in the field of tackling hydrological parameters after (chronologically) fuzzy and artificial neural networks. This paper explores extensively the literature of SVM applications related to hydrology. Traditional predictive regression models were found to be ineffective when compared to SVM models in most of the cases as reviewed in literature [28,69]. While the ANN is sensitive to the number of hidden nodes, the SVM is sensitive to the choice of the

mapping kernel. By optimizing these factors, it is possible to obtain comparable results with the two models. [50,51]. Several variants of SVM like epsilon-SVR, nu-SVR and LS-SVM offers more flexibility in modelling specific data oriented problems. More studies have to be carried out in order to develop strategies that can accommodate many of the problems that are still challenging for traditional predictive regression models. An SVM model is largely characterized by the choice of its kernel; as a result it is required to select the appropriate kernel for each application case in order to get promising results [42].

Quite a few examples in this paper demonstrate the SVM capabilities for both classification and regression related to hydrological problems [26,31]. These examples also show that the nonlinear features of SVM should be used with caution, because there can be possibilities of overfitting [41,42]. The literature results reviewed here show that SVMs have a numerous applications in computational hydrology. Future researchers can base their studies on the framework to develop more rigorous hybrid mechanisms and extend the support vector machine techniques in more complex hydrological forecasting.

## 8. Future scope of work

Understanding hydrologic systems at the scale of large watersheds and river basins is critically important to society when faced with extreme events, such as floods and droughts, or with concerns about water quality. Next-generation hydrology modelling will be increasingly sophisticated, encompassing a wide range of natural phenomena.

In future, one can extend SVMs to address hydrologic inverse problems that incorporates a physical understanding of the geological processes that form a hydrologic system. For example, density estimation [65,70] is a classical inverse problem that has been tackled successfully using SVMs.

Several evolutionary algorithms, such as genetic algorithms and simulated annealing algorithms have been used in parameter selection, but these algorithms often suffer from the possibility of being trapped in local optima. Hence there has been an avalanche of the above studies with the emerging of Swarm Intelligence (SI) technique and the hybridization between SI with SVM has become increasingly popular during the last decade. The SI technique mimics the intelligent behaviour of swarm of social insect, flocks of birds, or school of fish. Recent research interests are focussed towards hybridizing these relatively new optimization techniques, namely artificial bee colony (ABC) [71], artificial fish swarm algorithm (AFSA), ant colony optimization (ACO) [72] with SVM to obtain satisfactory global optimal results, which facilitates the selection of optimal parameters for SVM.

In the conventional method, support vector machine (SVM) algorithms are trained over pre-configured intranet/internet environments to find out an optimal classifier. If the datasets are large, SVM models become complicated and costly. Hence research discussions are going on regarding the Cloud SVM training mechanism (Cloud SVM) in a cloud computing environment with MapReduce technique for distributed machine learning applications where in we can achieve speed-ups in cloud-based watershed model calibration system which would pave way for real-time interactive model creation with continuous calibration, a new paradigm for watershed modelling [73].

## References

[1] I.G. Pechlivanidis, B.M. Jackson, N.R. Mcintyre, H.S. Wheater, Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods, Global NEST Journal 13 (3) (2011) 193–214.

[2] National Engineering Handbook, in: United States Department of Agriculture, Selected Statistical Methods, Part 630-Hydrology, NCRS, Washington, D.C, 2012 (chapter 18).

[3] Y.S. Xinying, V. Liong, Babovic, EC-SVM approach for real-time hydrologic forecasting, Journal of Hydroinformatics 6 (3) (2004) 209–223.

[4] C. Sivapragasam, S.Y. Liong, M.F.K. Pasha, Rainfall and runoff forecasting with SSA–SVM approach, Journal of Hydroinformatics 3 (3) (2001) 141–152.

[5] V. Vapnik, in: The Nature of Statistical Learning Theory, Springer, New York, 1995.

[6] V. Vapnik, S. Golwich, A.J. Smola, Support vector method for function approximation, regression estimation and signal processing, in: M. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, vol. 9, MIT Press, Cambridge, MA, USA, 1997, pp. 281–287.

[7] R. Collobert, S. Bengio, SVMTorch: support vector machines for large-scale regression problems, Journal of Machine Learning Research 1 (2001) 143–160.

[8] C. Cortes, V. Vapnik, Support vector networks, Machine Learning 20 (1995) 273–297.

[9] N. Cristianini, J. Shawe-Tylor, in: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, UK, 2000.

[10] A. Shigeo, in: Support Vector Machines for Pattern Classification, second ed., Springer-Verlag, London, 2010, http://dx.doi.org/10.1007/978-1-84996-098-4.

[11] R.O. Duda, P.E. Hart, in: Pattern Classification and Scene Analysis, John Wiley and Sons, Newyork, NY, 1973 (Nineteen printings)(chapter 4).

[12] H. Yu, J. Yang, J. Han, Classifying large data sets using SVM with hierarchical clusters, Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2003).

[13] L.J. Shih, D.M. Rennie, Y.-H. Chang, D.R. Karger, Text bundling: statistics-based data reduction, Twentieth International Conference on Machine Learning (2003).

[14] S. Winters-Hilt, A. Yelundur, C. McChesney, M. Landry, Support Vector Machine Implementations for Classification and Clustering, BMC Bioinformatics 7 (Suppl. 2) (2006) S4.

[15] R. Fletcher, in: Practical Methods of Optimization, second ed., Wiley, New York, 1987.

[16] V. Vapnik, An overview of statistical learning theory, IEEE Transactions on Neural Networks 10 (5) (1999) 988–999.

[17] J.A. Suykens, K. Van Gestel, T. De Brabanter, J. De Moor, B.J. Vandewalle, in: Least Squares Support Vector Machines, World Scientific Publishing, Singapore, 2002.

[18] P.S. Bradley, O.L. Mangasarian, Feature selection via concave minimization and support vector machines, in: J. Shavlik (Ed.), The Fifteenth International Conference of Machine Learning, Morgan Kaufmann, San Francisco, CA, 1998, pp. 82–90.

[19] J. Bi, K. Bennett, M. Embrechts, C. Breneman, M. Song, Dimensionality reduction via sparse support vector machines, Journal of Machine Learning Research 3 (2003) 1229–1243.

[20] J. Zhu, S. Rosset, T. Hastie, R. Tibshirani, 1-Norm Support Vector Machines, in: Neural Information Processing Systems, MIT Press, 16, Massachusettes, 2003.

[21] P.S. Bradley, O.L. Mangasarian, Massive data discrimination via linear support vector machines, Optimization Methods and Software 13 (2000) 1–10.

[22] B. Schölkopf, A. Smola, R. Williamson, P.L. Bartlett, New support vector algorithms, Neural Computation 12 (5) (2000) 1207–1245.

[23] Y.B. Dibike, S. Velickov, D. Solomatine, M.B. Abbott, Model induction with support vector machines: introduction and application, Journal of Computing in Civil Engineering 15 (3) (2001) 208–216.

[24] M. Bray, D. Han, Identification of support vector machines for runoff modelling, Journal of Hydroinformatics 6 (4) (2004) 265–280.

[25] S.H. Tripathi, V.V. Srinivas, R.S. Nanjundiah, Downscaling of precipitation for climate change scenarios: a support vector machine approach, Journal of Hydrology 330 (2006) 621–640.

[26] V. Anirudh, C.K. Umesh, Classification of rainy days using SVM, in: Proceedings of Symposium at HYDRO, Norfolk, Virginia, USA, 2007.

[27] H. Chen, J. Guo, X. Wei, Downscaling GCMs using the smooth support vector machine method to predict daily precipitation in the Hanjiang basin, Advances in Atmospheric Sciences 27 (2) (2010) 274–284, http://dx.doi.org/10.1007/s00376-009-8071-1.

[28] Y.L. Jian, T.C. Chun, W.Ch. Kwok, Using support vector machines for long term discharge prediction, Hydrological Sciences Journal 51 (4) (2006) 599–612, http://dx.doi.org/10.1623/hysj.51.4.599.

[29] C. Mesut, Estimation of daily suspended sediments using support vector machines, Hydrological Sciences Journal 53 (3) (2008) 656–666, http://dx.doi.org/10.1623/hysj.53.3.656A.

[30] G. Arun, P. Mahesh, Application of support vector machines in scour prediction on grade-control structures, Engineering Applications of Artificial Intelligence 22 (2009) 216–223, http://dx.doi.org/10.1016/j.engappai.2008.05.008.

[31] M. Debasmita, O. Thomas, A. Avinash, K.M. Surendra, M.T. Anita, Application and analysis of support vector machine based simulation for runoff and sediment yield, Bio Systems Engineering 103 (2009) 527–535, http://dx.doi.org/10.1016/j.biosystemseng.2009.04.017.

[32] Md.A. Hazi, A.G. Aminuddin, K. Chun, Chang, A.H. Zorkeflee, N.A. Zakaria, Machine learning approach to predict sediment load – a case study, Journal of Clean – Soil, Air, Water 38 (10) (2010) 969–976.

[33] A.Z. Zahrahtul, S. Ani, Streamflow forecasting at ungaged sites using support vector machines, Applied Mathematical Sciences 6 (60) (2012) 3003–3014.

[34] P. Parag, Bhagwat, M. Rajib, Multistep-ahead river flow prediction using LS-SVR at daily scale, Journal of Water Resource and Protection 4 (2012) 528–539, http://dx.doi.org/10.4236/jwarp.2012.47062.

[35] Moghaddamnia, A. Ghafari, M. Piri, J.D. Han, Evaporation estimation using support vector machines technique, World Academy of Science, Engineering and Technology 43 (2008) 14–22.

[36] S.S. Eslamian, S.A. Gohari, M. Biabanaki, R. Malekin, Estimation of pan evaporation using ANNs and SVMS, Journal of Applied Sciences 8 (19) (2008) 3497–3502.

[37] S. Pijush, Application of least square support vector machine (LSSVM) for determination of evaporation losses in reservoirs, Engineering 3 (4) (2011) 431–434, http://dx.doi.org/10.4236/eng.2011.34049.

[38] T. Hossein, K. Ozgur, E. Azadeh, P.T. Hosseinzadeh, SVM, ANFIS, regression and climate based models for reference evapotranspiration modelling using limited climatic data in a semi-arid highland environment, Journal of Hydrology 444–445 (2012) 78–89.

[39] S.M. Khan, P. Coulibaly, Application of support vector machine in lake water level prediction, Journal of Hydrologic Engineering ASCE 11 (2006) 199–205, http://dx.doi.org/10.1061/(ASCE)1084-0699(2006)11:3(199).

[40] H. Afiq, E. Ahmed, N. Ali, A. Othman, K. Aini, H.M. Mukhlisi, Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS), Water Resources Management 27 (2013) 3803–3823, http://dx.doi.org/10.1007/s11269-013-0382-4.

[41] P.S. Yu, S.T. Chen, I.F. Chang, Support vector regression for real-time flood stage forecasting, Journal of Hydrology 328 (2006) 704–716.

[42] D. Han, L. Chan, N. Zhu, Flood forecasting using support vector machines, Journal of Hydroinformatics 9 (4) (2007) 267–276.

[43] W. Wei, X. Wang, D. Xie, H. Liu, Soil water content forecasting by support vector machine in purple hilly region. Computer and computing technologies in agriculture, International Federation for Information Processing 258 (2007) 223–230, http://dx.doi.org/10.1007/978-0-387-77251-6_25.

[44] S.T. Chen, P.S. Yu, Pruning of support vector networks on flood forecasting, Journal of Hydrology 347 (2007) 67–78.

[45] S.T. Chen, P.S. Yu, Real-time probabilistic forecasting of flood stages, Journal of Hydrology 340 (1–2) (2007) 63–77.

[46] A.R. Shahbazi, N. Banafsheh, Z. Hossein, S. Manshouri, M.N. Mohsen, Seasonal meteorological drought prediction using support vector machine, World Applied Sciences Journal 13 (6) (2011) 1387–1397.

[47] A. Belayneh, J. Adamowski, Drought forecasting using new machine learning methods, Journal of Water and Land Development 18 (2013) 3–12.

[48] P. Ganguli, M.J. Reddy, Ensemble prediction of regional droughts using climate inputs and the SVM – copula approach, Hydrological Processes (2013), http://dx.doi.org/10.1002/hyp.9966.

[49] L. Jin, J.X. Chang, W.G. Zhang, Groundwater level dynamic prediction based on chaos optimization and support vector machine, in: Proceedings of the 3rd International Conference on Genetic and Evolutionary Computing, IEEE Transaction, Zhengzhou, China, 2009, http://dx.doi.org/10.1109/WGEC.2009.25.

[50] B. Mohsen, A. Keyvan, A. Emery, Coppola Jr., Comparative study of SVMs and ANNs in aquifer water level prediction, Journal of Computing in Civil Engineering 24 (5) (2010) http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000043.h.

[51] Y. Heesung, S.C. Jun, H. Yunjung, G.O. Bae, K.L. Kang, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, Journal of Hydrology 396 (2011) 128–138, http://dx.doi.org/10.1016/j.jhydrol.2010.11.002.

[52] M.K. Gill, T. Asefa, M.W. Kemblowski, M. McKee, Soil moisture prediction using support vector machines, JAWRA Journal of the American Water Resources Association 42 (2006) 1033–1046, http://dx.doi.org/10.1111/j.1752-1688.2006.tb04512.x.

[53] A. Sajjad, K. Ajay, S. Haroon, Estimating soil moisture using remote sensing data: a machine learning approach, Advances in Water Resources 33 (2010) 69–80, http://dx.doi.org/10.1016/j.advwatres.2009.10.008.

[54] J. Liu, M. Chang, M.A. Xiaoyan, Groundwater quality assessment based on support vector machine, in: Report of "Introducing Intelligence Project" (B08039),

[55] funded by Global Environment Fund (GEF) Integral Water Resource and Environment Management of Haihe River basin (MWR-9-2-1), 2011, p. 1.

[55] R. Remesan, Bray. Michaela, M.A. Shamim, D. Han, Rainfall-runoff modelling using a wavelet-based hybrid SVM scheme, Proceedings of Symposium JS. 4 at the Joint Convention of the International Association of Hydrological Sciences (IAHS) and the International Association of Hydrogeologists (IAH) (2009).

[56] K. Ozgur, C. Mesut, Precipitation forecasting by using wavelet-support vector machine conjunction model, Engineering Applications of Artificial Intelligence 25 (2012) 783–792.

[57] C.-C. Wei, Wavelet support vector machines for forecasting precipitation in tropical cyclones: comparisons with GSVM, regression, and MM5, Weather and Forecasting 27 (2012) 438–450.

[58] C.H. Sudheer, A.B.K. Nitin, Panigrahi, M. Shashi, Streamflow forecasting by SVM with quantum behaved particle swarm optimization, Neurocomputing 101 (2013) 18–23, http://dx.doi.org/10.1016/j.neucom.2012.07.017.

[59] H. Drucker, C.J. Burges, C. Kaufman, L. Smola, A.V. Vapnik, Support vector regression machines, in: M. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, 9, MIT Press, Cambridge, MA, 1997, pp. 155–161.

[60] J.C. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector. (1998).

[61] I. Ovidiu, Applications of support vector machines in chemistry, in: K.B. Lipkowitz, T.R. Cundari (Eds.), Reviews in Computational Chemistry, vol. 23, Wiley-VCH, Weinheim, 2007, pp. 291–400.

[62] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (2004) 113–126.

[63] A.J. Smola, in: Regression Estimation with Support Vector Learning Machines, Diplomarbeit, Technische Universitat Munchen, 1990.

[64] C. Saunders, M.O. Stitson, J. Weston, L. Bottou, B. Schölkopf, A. Smola, Support vector machine – Reference manual Royal Holloway Technical Report CSD-TR-98-03. Royal Holloway (1998).

[65] S. Mukherjee, V. Vapnik, Support vector method for multivariate density estimation, in: Center for Biological and Computational Learning. Department of Brain and Cognitive Sciences, MIT. C. B. C. L. No. 170, 1999.

[66] S. Gunn, Support vector machines for classification and regression, ISIS Technical Report. (1998).

[67] C. Yuri, Applications of SVMs in the exploratory phase of petroleum and natural gas: a survey, International Journal of Engineering and Technology 2 (2) (2013) 113–125.

[68] B.D. Yonas, V. Slavco, S. Dimitri, Support vector machines: Review and applications in civil engineering, in: Proceedings of the 2nd Joint Workshop on Application of AI in Civil Engineering March 2000, Cottbus, Germany, 2000.

[69] C.L. Wu, K.W. Chau, Y.S. Li, River stage prediction based on a distributed support vector regression, Journal of Hydrology 358 (2008) 96–111.

[70] J. Weston, A. Gammerman, M.O. Stitson, V. Vapnik, V. Vovk, C. Watkins, Support vector density estimation, in: B. Scholkopf, J.C. Burges, A. Smola (Eds.), Advances in Kernel Methods: Support Vector Learning, MIT Press, Cambridge, MA, 1999, pp. 293–305.

[71] D. Karaboga, B. Akay, Artificial bee colony (ABC), Harmony Search and Bees Algorithm on Numerical Optimization, in: Proceedings of Innovative Production Machines and Systems virtual Conference, IPROMS, 2009.

[72] M. Dorigo, C. Blum, Ant colony optimization theory: a survey, Theoretical Computer Science 344 (2005) 243–278.

[73] H. Marty, N. Beekwilder, J.L. Goodall, M.B. Ercan, Calibration of watershed models using cloud computing, in: Proceedings of the 8th IEEE International Conference on eScience, Chicago, IL, 8–12 October, 2012.

[74] Kimeldorf, G. Wabha, Some results on Tchebycheffian spline functions, Journal of Mathematical Analysis and Applications 33 (1) (1971) 82–95.

[75] N.B.A. Seyyed, K. Reza, R.B.L. Mohammad, S. Kazem, Characterizing an unknown pollution source in groundwater resources systems using PSVM and PNN, Journal Expert Systems with Applications, An International Journal archive 37 (10) (2010) 7154–7161.