RESEARCH ARTICLE

Coupling machine learning with signal process techniques and particle swarm optimization for forecasting flood routing calculations in the Eastern Black Sea Basin, Türkiye

Okan Mert Katipoğlu[1] [ORCID] · **Metin Sarıgöl[2]** [ORCID]

Abstract

With the effect of global warming, the frequency of floods, one of the most important natural disasters, increases, and this increases the damage it causes to people and the environment. Flood routing models play an important role in predicting floods so that all necessary precautions are taken before floods reach the region, loss of life and property in the region is prevented, and agricultural lands are protected. This research aims to compare the performance of hybrid machine learning models such as least-squares support vector machine technique hybridized with particle swarm optimization, empirical mode decomposition, variational mode decomposition, and discrete wavelet transform processes for flood routing estimation models in Ordu, Eastern Black Sea Basin, Türkiye. In addition, it is aimed to examine the effect of data division in flood forecasting. Accordingly, 70%, 80%, and 90% of the data were used for training, respectively. For this purpose, the flood data of 2009 and 2013 in Ordu were used. The performance of the established models was evaluated with the help of statistical indicators such as mean bias error, mean absolute percentage error, determination coefficient, Nash–Sutcliffe efficiency, Taylor Diagrams, and boxplot. As a result of the study, the particle swarm optimization least-squares support vector machine technique was chosen as the most successful model in predicting flood routing results. In addition, the optimum data partition ratio was found to be Train:70:Test:30 in the flood routing calculation. The findings are essential regarding flood management and taking necessary precautions before the flood occurs.

Keywords Flood routing · Flood management · Particle swarm optimization · Machine learning · Least-squares support vector machine

Introduction

Floods cause significant damage to people, and major economic and social effects on the environment, including flooding homes, disrupting transportation systems, eroding

Responsible Editor: Philippe Garrigues

Okan Mert Katipoğlu and Metin Sarıgöl contributed equally to the publication.

✉ Okan Mert Katipoğlu
okatipoglu@erzincan.edu.tr

[1] Department of Civil Engineering, Faculty of Engineering and Architecture, Erzincan Binali Yıldırım University, Erzincan, Turkey

[2] Design Department, Erzincan Uzumlu Vocational School, Erzincan Binali Yildirim University, Erzincan, Turkey

fertile land, and damaging crops (Bagatur and Onen 2018; Zhou et al. 2018; Mazzoleni et al. 2018; Hadidi et al. 2020; Ball 2022). Therefore, flood forecasting, flood control, and rapid response management are crucial issues for minimizing flood damage and taking preventive measures. Also, accurately predicting future floods reduces flood losses (Yaseen et al. 2015; Yuan et al. 2020; Hassanvand et al. 2018; Pant et al. 2018).

Flood routing is defined as the calculation of changes in the velocity and size of a flood wave over time in any section of a stream and is necessary for the design and implementation of structural and nonstructural flood control measures (Hamedi et al. 2016; Barati et al. 2018).

The first one of the flood routing methods used in the literature is the hydraulic method. In hydraulic methods, the flow is calculated as a function of time and space, and the continuity and momentum equations are used for this calculation. This method is based on the solution of Saint

Venant differential equations, which do not have analytical solutions representing the motion of the flood wave. Since these equations are difficult to solve, they are solved using numerical methods. The main hydraulic routing methods are kinematic wave, diffusion wave, Muskingum-Cunge, and dynamic wave. The second is hydrological methods, such as the Muskingum method and SCS method, which rely solely on the principles of conservation of mass. Although hydrological processes are simpler than hydraulic methods, they generally do not give good results due to the effect of counter currents and swells in flooding along a river. For the solution of both basic techniques, many parameters such as roughness, velocity, cross-sectional area, and flow depth are needed. Obtaining and applying these data is costly and leads to lengthy calculations (Chow et al. 1988). In addition, when the lateral flow is added to the main tributaries of the rivers, the flood routing analysis results give high errors (Barbetta et al. 2017).

Artificial neural networks (ANNs) are one of the most important sub-disciplines of machine learning. The ANN technique started with McCulloch and Pitts (1943) modeling of neuronal activity in the brain, and Hebb (1949) proposed a reinforcement-based learning mechanism to express the brain's learning ability. In the following, in the 1990s and 2000s, new ML algorithms, such as gene expression programming and genetic programming algorithms, were developed (ASCE 2000). Due to the disadvantages of physically based and statistical methods, data-based models in ML algorithms have become widespread. Data-based forecasting models established with ML algorithms give faster results with minimum inputs and provide more straightforward implementation with low computational costs (Mosavi et al. 2018).

Differential equations are used to solve conceptual and physically based models, and many parameters are needed. However, ML algorithms are data-driven and do not need parameters as physically based hydrological models (Tayfur et al. 2007; Tayfur and Moramarco 2008; Tayfur 2017). Therefore, in ML algorithms, input and output data are sufficient, and high computational efficiency is one of its most essential advantages (Zounemat-Kermani et al. 2020; Dazzi et al. 2021).

ML algorithms have been used frequently in the field of hydrology in recent years and especially least squares support vector machine (LSSVM), wavelet transform (WT), variational mode decomposition (VMD), empirical mode decomposition (EMD), and particle swarm optimization (PSO) have wide usage areas. Okkan and Serbes (2012) determined the applicability of LSSVM to predict runoff values and compared with ANN and other algorithms. It was found that the LSSVM algorithm is the most successful model. Shabri and Suhartono (2012) investigated the ability of the LSSVM algorithm to improve the accuracy of flow

prediction and determined that the LSSVM algorithm is better than autoregressive integrated moving average, ANN, and SVM models for monthly flow prediction. Sudheer et al. (2014) used the SVM-PSO model to predict streamflow values and found that the SVM-PSO model gave the most accuracy predicts peaks with 0.24 error, whereas ARMA predicts peaks with 0.355 error and ANN has 0.266 error. Noury et al. (2014) used wavelet function SVM and neural wavelet network (NWN) models to simulate lake water level fluctuation and determined that the SVM model gave SSE 0.43, RMSE 0.23, and R^2 0.97 better statistics results than SSE 1.33, RMSE 0.41, and R^2 0.95 statistics of NWN model. Huang et al. (2014) used the EMD-SVM, ANN, and SVM algorithms for flow estimation and determined that among all the models, the EMD-SVM algorithm has the best MAE, RMSE, and MAPE statistics of 0.51%, 1.07%, and 17.4%, respectively. Granata et al. (2016) researched a comparative study of the storm water management model (SWMM) and the SVM-based precipitation flow model. It was found that the SVM model results had higher coefficient of determination and lower RMSE error values than the SWMM model. Also, the SVM underestimates peak flow by up to 10%, while SWMM overestimates peak flow by up to 20%. After all, the SVM algorithm gave quite good results. Seo et al. (2016) used SVM and wavelet packet, adaptive neuro-fuzzy inference system and wavelet packet, and ANN and wavelet packet to forecast river stage. It was revealed that WPANFIS has the best MAE, RMSE, and MSRE statistics of 0.062, 0.0124, and 0.139, respectively, giving the best results among all the models. Zhao et al. (2017) used a new hybrid model EMD-based chaotic LSSVM (EMD-CLSSVM) and CLSSVM for annual flow prediction and found that this algorithm reduced the RMSE, MARE, and MAE error results by 39%, 28.6%, and 25.6%, respectively. Eventually, the proposed hybrid model is superior to the CLSSVM hybrid model. Ehteram et al. (2018) developed a hybrid bat swarm algorithm (HBSA) model, which is a hybrid of PSO algorithm and bat algorithm (BA), for the optimal detection of the four parameters of the Muskingum method. It has been determined that the SSQ error values of the proposed HBSA algorithm are 65% and 72% lower than those of the BA and PSO algorithms and give perfect flood routing results compared to the other methods. Seo et al. (2018) applied two hybrid ML models consisting of VMD-based least squares support vector regression and VMD-based extreme learning machine (VMD-ELM) to improve the accuracy of the daily precipitation-flow model. It was determined that VMD-LSSVR has MAE, RMSE, and MSRE statistics of 1.418, 2.887, and 0.042, and VMD-ELM has MAE, RMSE, and MSRE statistics of 1.376, 3.193, and 0.036, respectively. Both hybrid algorithms gave the best results in daily precipitation-flow modeling among ANN, ELM, LSSVR, VMD-ANN, DWT-ANN, DWT-ELM, and

DWT-LSSVR models. Ma et al. (2019) used LSSVM and the classical canonical method of logistic regression (LR) on historical flash flood records to evaluate the flash flood risk. It has been seen that the LSSVM algorithm works with 0.79 accuracy and the LR algorithm works with 0.75 accuracy. As a result, it was determined that the LSSVM algorithm gave better results. Okkan and Kirdemir (2020) used the PSO algorithm for each random experiment for four different flood data. The hybrid use of PSO with the Levenberg–Marquardt (LM) model was discussed they found that the hybrid PSO-LM algorithm gave stable global solutions. Sahana et al. (2020) examined the effectiveness of the modified frequency ratio, conventional frequency ratio, and SVM model in storm flooding susceptibility assessment. The success rate was calculated as the SVM algorithm (0.8221) according to the modified frequency ratio model (0.797) and the traditional frequency ratio model (0.753). As a result, it was found that the SVM algorithm best predicted storm flooding in the Sundarban Biosphere. Norouzi and Bazargan (2020) calculated the parameters of the linear Muskingum model for calculating the downstream hydrograph using the PSO model. It has been determined that the calculation error is reduced by 66% and 43%, respectively, when the proposed model is compared with the case where the first and second floods are used. Akbari et al. (2020) used a new nonlinear Muskingum method with four variable parameters to improve the accuracy of the outflow estimation that calculates the values of hydrological parameters by combining both PSO and GA. Compared with the three-parameter and best-variable parameter nonlinear Muskingum model, the SSQ decreased by 52% and 6.9% for the first and second case studies, respectively, and by 76% and 62% for the third and fourth case studies, respectively. It has been determined that the proposed model gives better results than other nonlinear Muskingum models. Şenel et al. (2020) estimated the flow data for the next period in the Yeşilırmak River using the ANN model. In this modeling, the time delay size was optimized using the Ant Lion Algorithm, and the number of hidden layers in the artificial neural networks was added to the model and the optimization process was completed. Thus, they determined that using heuristic optimization techniques together with ANN gives better results in the established model. Wang et al. (2021) proposed a new hybrid algorithm (VMD-LSTM-PSO) by combining VMD with long short-term memory (LSTM) and PSO for daily runoff prediction. They determined that the new model could be used in practice for hydrological forecasting based on higher forecast accuracy than LSTM-PSO, complementary ensemble empirical mode decomposition-LSTM-PSO, and EEMD-LSTM-PSO model results. Zhang et al. (2022) applied the multivariate empirical mode decomposition (MEMD) algorithm to analyze the factors affecting monthly flow. They concluded that the MEMD model outcome R^2 value

increased by 24.2% compared to the stepwise multiple linear regression models using the original time series and was effective. Cai et al. (2022) used the PSO back-propagation algorithm on Sentinel-1A satellite data to predict surface subsidence. As a result of the study, they determined that the MAE value of the results obtained by this algorithm is 0.17, which is better than six algorithms, such as SVM and is an effective tool for providing early warning of surface subsidence. Xu et al. (2022) used deep learning neural network model based on LSTM networks and PSO algorithm to forecast flood accuracy. It has been determined that the PSO-LSTM algorithm has higher forecasting accuracy than M-EIES (physical model), ANN, PSO-ANN, and LSTM models at all stations in the watersheds and can be used to improve accuracy in the short-term flood forecasting model.

As a result of the literature review, flood routing studies were carried out with the SVM and ANN models, which are the most widely used in research (Okkan and Serbes 2012; Shabri and Suhartono 2012; Sudheer et al. 2014; Huang et al. 2014). On the other hand, studies such as flood level estimation, flood risk, and flood mapping were carried out with ML algorithms. Flood routing research topics made with hybrid algorithms in the literature have significantly increased recently but are limited to specific algorithms (Tayfur et al. 2018; Norouzi et al. 2021). For this reason, in the literature, it was first investigated how flood routing estimations give results for a region with the LSSVM algorithm which combines various optimization and signal decomposition techniques, such as PSO, Wavelet, EMD, and VMD, respectively, to improve streamflow estimation accuracy. In addition, this research adds innovation to the literature by comparing the performance of hybrid algorithms for flood routing against various statistical performances and investigating how the estimating performance changes when the parameters of the models are optimized. The contribution of this article to the literature is to show the latest state of artificial intelligence models in flood estimating and to evaluate the flood estimating performances of some hybrid algorithms by parameter optimization.

This study aims to estimate the hourly flood hydrograph of the downstream region by ML models, which are LSSVM, PSO-LSSVM, EMD-LSSVM, Wavelet-LSSVM, and VMD-LSSVM hybrid techniques. It is hoped that the models obtained from the study can be applied to different regions to predict flood routing results. In the study's first phase, a training and test model was created with the flood hydrograph ML methods measured in the upstream and downstream regions. Then, by applying the upstream flood data to the established model, the flood values in the downstream region were tried to be estimated.

Thus, it is hoped that a future flood will be easily predicted by applying only input data to the established model, without requiring much data such as friction coefficient,

slope, and cross-sectional area. For this reason, performing analysis processes quickly and in a short time will be vital in terms of saving time.

Material and method

Study data and area

Turnasuyu Stream is an Ordu stream originating from Giresun Mountains and pouring into the Black Sea. Its length is 56 km, its catchment area is 278 km^2, and the average flow is 7.2 m^3/s (Bostanci et al. 2016). Observation Stations on River are located in Eastern Black Sea Basin, and the locations of dïscharge observatïon statïon (DOS) are shown in Fig. 1.

The Eastern Black Sea Region is located on the northeastern coast of Turkey and is bordered by the Eastern Black Sea Mountains to the south and the Black Sea to the North. The most important characteristic of this region is that it receives abundant precipitation in all seasons, and has very sharp valleys and many steep streams with high flow rates (Capik et al. 2012).

Kapçullu (E22A063) DOS is located at an altitude of 18 m at the coordinates 40:57:19 North, 38:00:07 East, and Cumhuriyet Köyü (D22A093) DOS at 40:49:54 North, 37:57:42 East, at the height of 375 m stations, were used in Ordu. The flood data of Ordu for the years 2009 and 2013 were used.

In this study, hourly flood hydrograph was used. The input values of the model were selected as D22A093 DOS upstream data, and the output values were chosen as E22A063 DOS downstream data. Using the upstream and downstream flood hydrograph data was modeled with LSSVM, PSO-LSSVM, EMD-LSSVM, Wavelet-LSSVM, and VMD-LSSVM algorithms and passed through training and testing stages, then evaluated with mean bias error (MBE), MAPE, determination coefficient (R^2), Nash–Sutcliffe efficiency (NSE), Taylor diagrams, and boxplot analysis. In addition, it is examined the effect of data division in flood estimating.

In this study, 286 hourly discharge data were used for Ordu. These data were trained and tested on various ML models. The parameters of DOS are shown in Table 1, and DOS hydrographs are plotted in Fig. 2.

Machine learning models

The comparison of the performance of flood routing methods is established by using ML algorithms such as LSSVM, PSO-LSSVM, EMD-LSSVM, Wavelet-LSSVM, and VMD-LSSVM algorithms which are widely used in the literature investigated. While installing the ML models, 70%, 80%, and 90% of the data were used for training and testing, respectively, and the flood data of the upstream DOS was


Fig. 1 Location of E22A063-D22A093 DOS

Table 1 Discharge observation stations

Province/district name	Station number	Station code	Station name	Basin name	River name	Station elevation	Area (km²)
Ordu	1	E22A063	Kapçullu	Doğu Karadeniz	Turnasuyu	18	274,50
	2	D22A093	Cumhuriyet Köyü	Doğu Karadeniz	Turnasuyu	375	210

Fig. 2 Hourly hydrograph of E22A063-D22A093 for 2009 and 2013 flood
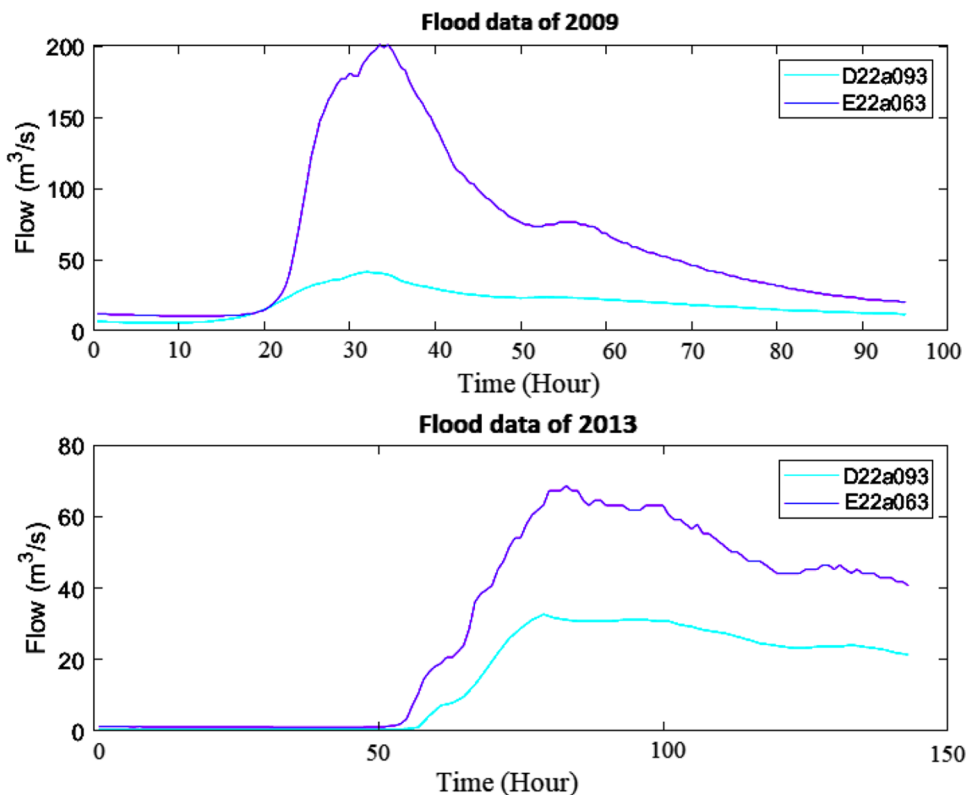
used as input, and the flood data of the downstream DOS was used as output. In the setup of hybrid models, the input series decomposed into various sub-signals and the target series is predicted. Decomposition allows you to understand the underlying structure of time series and reveal patterns or trends that may be present. It is important to make predictions about future values and detect unusual data anomalies. Also, when a time series does not show significant fluctuations or white noise, underlying trends or seasonality may occur in the data. Decomposition can help express these patterns and reveal their effects (Zhang and Qi 2005; Wen et al. 2019).

Least-squares support vector machine

LSSVM is a type of SVM used for classification problems and regression analysis. The main advantages of LSSVM are mathematical traceability, high precision, and direct geometric interpretation, which converts the nonlinear relationship between outputs and inputs into a linear relationship and uses the following equation.

$$M = \sum_{i=1}^{n} k(x, x_i)\alpha_i + b \tag{1}$$

In Eq. 1, α_i is the weighting coefficient of input data, M is the output value, b is the bias, and $k(x)$ is the nonlinear mapping function. The LSSVM algorithm minimizes the difference between predicted and measured data. Calculation of α_i and b parameters is shown in the equations as follows.

$$\begin{bmatrix} kernel & \overline{1}^T \\ \overline{1} & kernel + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ M \end{bmatrix} \tag{2}$$

Here, C is the regulation parameter. α, I, M, and $\overline{1}$ parameters are calculated as follows:

$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \overline{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, M = \begin{bmatrix} M_1 \\ \vdots \\ M_n \end{bmatrix}, I = (1, 1 \cdots 1) \qquad (3)$$

The following equation uses the radial basis function as a kernel function.

$$k(x, x_1) = exp\left(\frac{-\|x - x_i\|}{2\sigma^2}\right) \qquad (4)$$

The LSSVM algorithm is shown schematically in Fig. 3 (Kadkhodazadeh and Farzin 2021).

Particle swarm optimization

This model is a stochastic population-based technique, and each i is a candidate solution whose motion in the search space is governed by four vectors. The corresponding Eqs. 5 and 6 are expressed below.

$$v_i = \omega v_i + c_1 \eta_1 \circ (p_i - x_i) + c_2 \eta_2 \circ (g_{best} - x_i) \qquad (5)$$

$$x_i = x_i + v_i \qquad (6)$$

x_i is the position is defined as v_i is the velocity, p_i is the best individual solution position, g_{best} is the best neighbor solution position, c_1 and c_2 are the acceleration constants, ω is the inertia weight, η_1 and η_2 are the random vectors, and $\circ$ is the entry-wise products (Novoa-Hernández et al. 2011).

In the PSO algorithm, while the parameters are determined according to the type of the problem, the initial positions and velocities of each parameter are calculated respectively. Then the fit values between the limit values of all particles are calculated. While the local best (pbest) values are estimated at the end of each iteration, the global best (gbest) is determined from the current values and updated within the velocities of all particles. Accordingly, the resulting flow chart is shown in Fig. 4 (Saplioglu et al. 2020).

Empirical mode decomposition

This algorithm offers widely used and adaptable instantaneous frequency-based intrinsic mode functions (IMFs) to flexibly analyze multi-channel data with linear and non-stationary time series. Their results are processed into energy-frequency-time distributions (Huang et al. 1998). Each IMF represents a simple oscillation in the signal. The amplitudes and frequencies of the IMFs are not fixed but may change over time. If $x(n)$ is expressed as a time series:

$$x_n = \sum_{i=1}^{m} IMF(n) + r(n) \qquad (7)$$

In Eq. 7, IMF(n) is the components of decomposed sub-series, and $r(n)$ is the components of decomposed residual (Başakın et al. 2021).

Wavelet transform

This algorithm uses a mathematical structure and analyzes local changes in time series and information from various data sources. This transformation improves data quality by providing reliable decomposition of an original time series, and prediction accuracy is increased by discrete WT banding of the data. In addition, discrete wavelet transform (DWT) splits the initial dataset into different resolution levels to extract higher-quality data when building the model. It is widely used in flood time series predicting due to its valuable properties (Mosavi et al. 2018).

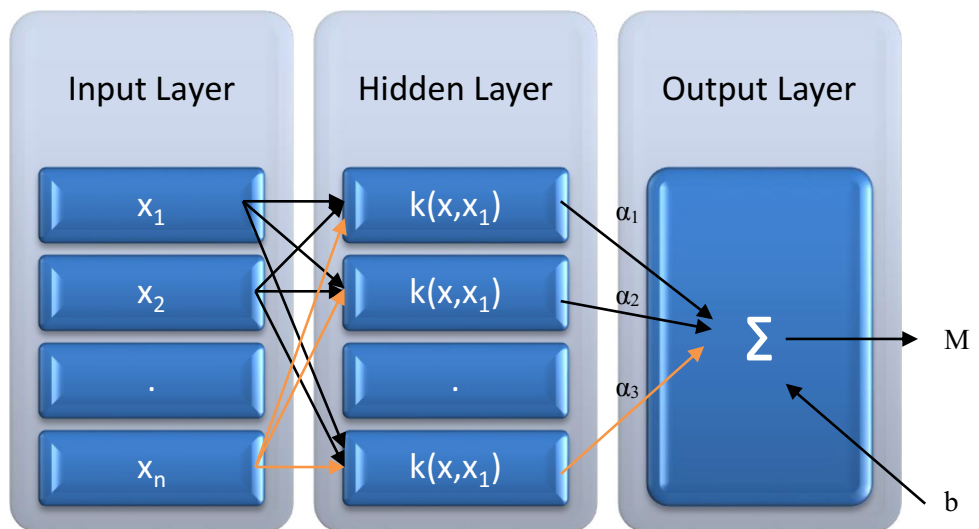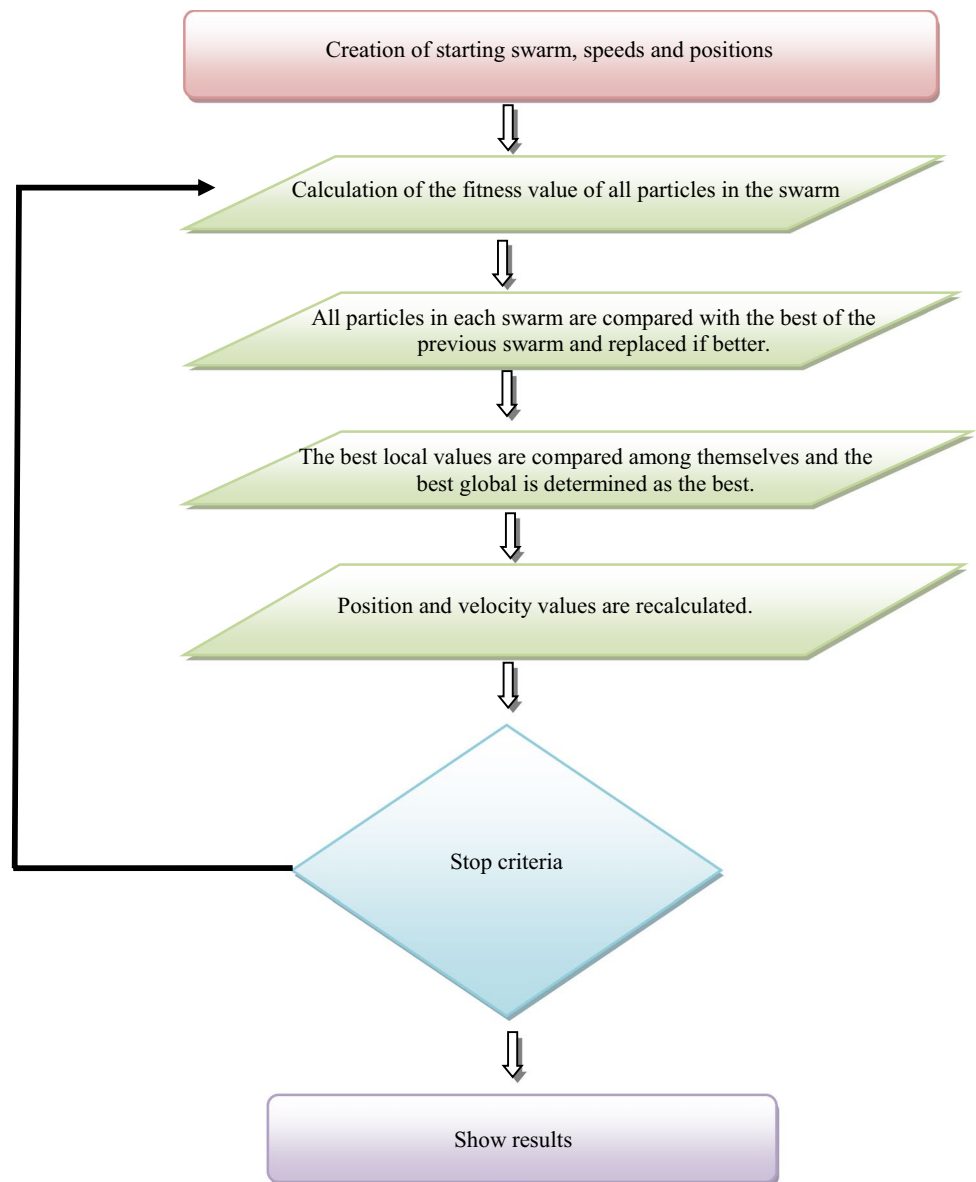**Fig. 3** The schematic structure of the LSSVM algorithm

Fig. 4 PSO algorithm flowchart

To specify the optimal number of decomposition levels in the wavelet transform, the formula based on the signal length in Eq. 8 is used.

$$L = \text{int}\left[\log(N)\right] \tag{8}$$

N shows data length, and L is the decomposition level in Eq. 8. DWT is obtained by applying two filters to the original signal, where the first filter captures the trend of the signal and the second filter captures the trend deviations. If the equation related to this is expressed as follows:

$$w = A_1 + D_1 \tag{9}$$

In Eq. 9, A is the scale, B is the translation, A_1 is the low frequency, and B_1 is the high frequency. This

decomposition process is done for the next levels and continues until the desired resolution is obtained, the positions and scales are based on the powers of the two and discrete time series at any time t, and the related equations are expressed as follows:

$$W = A_n + \sum_{i=1}^{n} D_i \tag{10}$$

$$DWT_{\varphi}(j,k) = a_0^{-\frac{i}{2}} \int_R \varphi\left(a_0^{-j}t - kb_0\right)f(t)dt \tag{11}$$

In Eqs. 10 and 11, a is the dilation factor, b is the time factor, R is the reel numbers domain, j is the integer, and k is the integer (Başakın et al. 2022).

Variational mode decomposition

This algorithm is a signal decomposition algorithm, and each IMF transforms the signal decomposition into an iterative solution to the variation problem. It is also a variation problem with a finite bandwidth with a different center frequency, and the original signal is decomposed into K mode functions $u_k(t)$ ($k = 1, 2, \cdot\cdot, K$). The sum of the estimated bandwidths for each mode function here is minimized. Thus, the variation model equation can be written as:

$$\begin{cases} min \left\{ \sum_1^K \| \partial t \left[\left(\delta(t) + \frac{i}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \|_2^2 \right\} \\ \{u_k\}, \{w_k\} \\ s.t \sum_{k=1}^K u_k(t) = f(t) \end{cases} \qquad (12)$$

w_k is the frequency center of each IMF, $\{w_k\} = \{w_1, w_2, \ldots, w_k\}; \{u_k\} = \{u_1, u_2, \ldots, u_k\}$ in Eq. 12 (Li et al. 2018).

Testing routing success

The first method used to measure the model's success is MAPE, calculated by dividing the absolute error in each period by the values observed for that period.

$$MAPE = \frac{\sum_{t=1}^n \frac{|Q_t - P_t|}{Q_t}}{n} \qquad (13)$$

where Q_t is the observed discharge, P_t is the predicted discharge, and n is the number of data (Widiasari et al. 2017).

The second criterion used to measure the model's success, NSE, is calculated by subtracting the ratio of the mean squared errors and the variance of the observed values from 1. The NSE equation is written below:

$$NSE = 1 - \left[\frac{\sum_{i=1}^n \left(Q_{pi} - Q_{oi} \right)^2}{\sum_{i=1}^n \left(Q_{oi} - \overline{Q_o} \right)^2} \right] \qquad (14)$$

where n is the the number of predictions and observations, Q_{oi} is the ith observed discharge, Q_{pi} is the ith predicted discharge, and Q_o is the average of observed discharge (Başakın et al. 2021).

The third criterion used to measure the model's success, R^2, represents the linear regression between the predicted and actual values and is expressed by the formula below. In trend analysis, it gains weight and is defined as a value between 0 and 1. If the result value approaches 1, the better the harmony or relationship between the two factors (Zare and Koch 2014).

$$R^2 = \left(\frac{\sum_{i=1}^n \left(Qi_{observed} - Q^{mean}_{observed} \right) \left(Qi_{calculated} - Q^{mean}_{calculated} \right)}{\sqrt{\sum_{i=1}^n \left(Qi_{observed} - Q^{mean}_{observed} \right)^2 \sum_{i=1}^n \left(Qi_{predicted} - Q^{mean}_{predicted} \right)^2}} \right)^2 \qquad (15)$$

In Eq. 15, $Qi_{Predicted}$ is the predicted discharge, $Q^{mean}_{Predicted}$ is the average of predicted discharge, $Qi_{Observed}$ is the observed discharge, $Q^{mean}_{Observed}$ is the average of observed discharge, and n is the number of data.

The fourth criterion used to measure the model's success, MBE, is shown in Eq. 16. This formula is expressed as the mean deviation of the predicted values from the observed data and provides information about the long-term performance of the models (El Boujdaini et al. 2021).

$$MBE = \frac{1}{n} \sum_{i=1}^n \left| Q_{ipredicted} - Q_{iobserved} \right| \qquad (16)$$

In Eq. 16, $Q_{ipredicted}$ is the predicted discharge, $Q_{imeasaured}$ is the observed discharge, and n is the number of data.

The fifth criterion used to measure the model's success is boxplots which gives detailed information about the data distribution and maximum and minimum data for the established models. The advantage of this graph is that it can show how a model predicts the maximum, minimum, median, and quintile values (Nhu et al. 2020; Dehghani et al. 2022).

The sixth criterion used to measure the model's success is the Taylor diagram which provides a graphical summary of the results. This diagram includes the correlations, RMSE, and standard deviation values of the data obtained from the models. The x and y axes show the standard deviation values, and quarter circle arcs show the diagram's correlation coefficient and RMSE values (Taylor 2001).

Results and Discussion

This study compares LSSVM, PSO-LSSVM, EMD-LSSVM, Wavelet-LSSVM, and VMD-LSSVM techniques for modeling floods in Ordu in 2009 and 2013. In addition, it is aimed to examine the effect of data division in flood estimating. Accordingly, 70%, 80%, and 90% of the data were used for training, respectively. Model results were evaluated visually using various statistical indicators, Taylor diagrams, and a boxplot. The optimal decomposition level (L) must first be determined to decompose input variables with DWT. Therefore, Eq. (5) was used to determine the L value in this study. Then, using Eq. (5), $L = 3$ was determined. Although the level of decomposition can be determined by trial and error, this process is grueling and time-consuming. For this reason, three commonly used

Fig. 5 Subcomponents of the flood data of 2009 separated by various signal decomposition techniques: **a** dmey wavelet, **b** db 10 wavelet, **c** EMD, **d** VMD
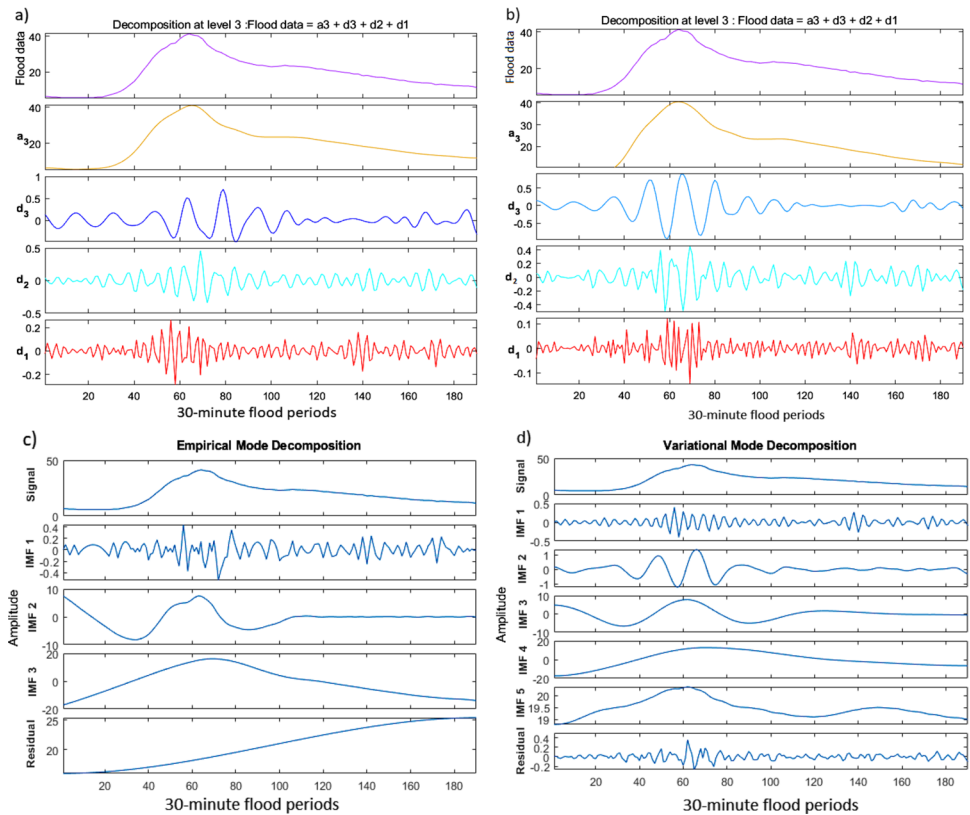
decomposition levels were chosen in this study (Nourani et al. 2009; Seo et al. 2018; Shafaei and Kisi 2016). Figure 5 shows flood data divided into subcomponents by WT, EMD, and VMD techniques.

PSO uses global best result (gbest) values and pbest when updating particle velocities and positions. This search process is terminated when the maximum number of iterations is reached. In this study, PSO was modeled

with 100 iterations. Since the number of particles (N) and c constants changes according to the problem to be solved, different experiments have been made. Accordingly, $N = 20$ and $c1 = c2 = 2$ values were chosen because they showed very successful results. The radial basis kernel function was used when setting up the SVM model.

Table 2 shows the estimation results of the 2009 floods of the test set obtained by the LSSVM, W-LSSVM,

Table 2 Flood routing results for 2009 in test data

	Train-test ratio (70:30%)**				Train-test ratio (80:20%)				Train-test ratio (90:10%)			
	NSE	R^2	MAPE	MBE	NSE	R^2	MAPE	MBE	NSE	R^2	MAPE	MBE
LSSVM					LSSVM				LSSVM			
	0.99	0.99	0.03	0.001	0.98	0.98	0.02	−0.03	0.85	0.87	0.02	0.01
Dmey wavelet LSSVM					Dmey wavelet LSSVM				Dmey wavelet LSSVM			
	0.85	0.86	0.09	0.14	0.28	0.43	0.11	0.54	−4.05	0	0.15	−1.59
DB 10 wavelet LSSVM					DB 10 wavelet LSSVM				DB 10 wavelet LSSVM			
	0.86	0.90	0.09	0.20	0.72	0.77	0.08	0.19	−0.23	0.42	0.07	−0.03
EMD-LSSVM					EMD-LSSVM				EMD-LSSVM			
	0.74	0.87	0.08	2.88	0.81	0.82	0.04	0.49	−0.99	0.75	0.05	1.37
VMD-LSSVM					VMD-LSSVM				VMD-LSSVM			
	0.48	0.57	0.21	−1.68	0.13	0.35	0.14	−2.27	−0.56	0.61	0.06	0.12
PSO-LSSVM					PSO-LSSVM				PSO-LSSVM			
	0.99*	**0.99***	**0.02***	**0.07***	**0.98**	**0.98**	**0.02**	**0.001**	**0.86**	**0.87**	**0.07**	**-0.01**

* mark indicates the best model among all models, the ** mark indicates the most appropriate data division value, and bold characters present the best algorithm of the selected data division ratio

EMD-LSSVM, VMD-LSSVM, and PSO-LSSVM approaches. The results were evaluated according to the NSE and R^2 values closest to 1 and the lowest MAPE and MBE values. Accordingly, the most successful model for all data partition ratios was PSO-LSSVM. In addition, the weakest flood estimation was made with the VMD-LSSVM hybrid model. Finally, when the effect of data division ratios on model performance was evaluated, the most successful estimation performance was obtained with the train-test ratio (70:30%).

In Fig. 6, test results were evaluated with Taylor diagrams to visually assess the success of stand-alone and hybrid techniques in estimating floods in 2009. In these diagrams, model successes were assessed according to the closeness of the estimation models to the reference point, which is the actual value, and RMSE, correlation, and standard deviation

values. These diagrams determined the best model for all data division ratios as the PSO-LSSVM hybrid model. In addition, it is noteworthy that the single LSSVM model shows close estimation accuracy to the PSO-LSSVM model.

In Fig. 7, boxplots are shown to evaluate the success of the models used in estimating the floods of 2009. It is aimed to determine the best model by comparing the distribution, median, mean, and percentile slices of the predicted values with the actual values with boxplots. When Fig. 7 is examined, the most overlapping model is the PSO-LSSVM hybrid in terms of overlapping of data parameters such as median, quartiles, and maximum and minimum values of real data and predicted data. These graphs decided the best model for all data division ratios as the PSO-LSSVM hybrid model. In addition, it has been seen that the single LSSVM model produces realistic estimations.
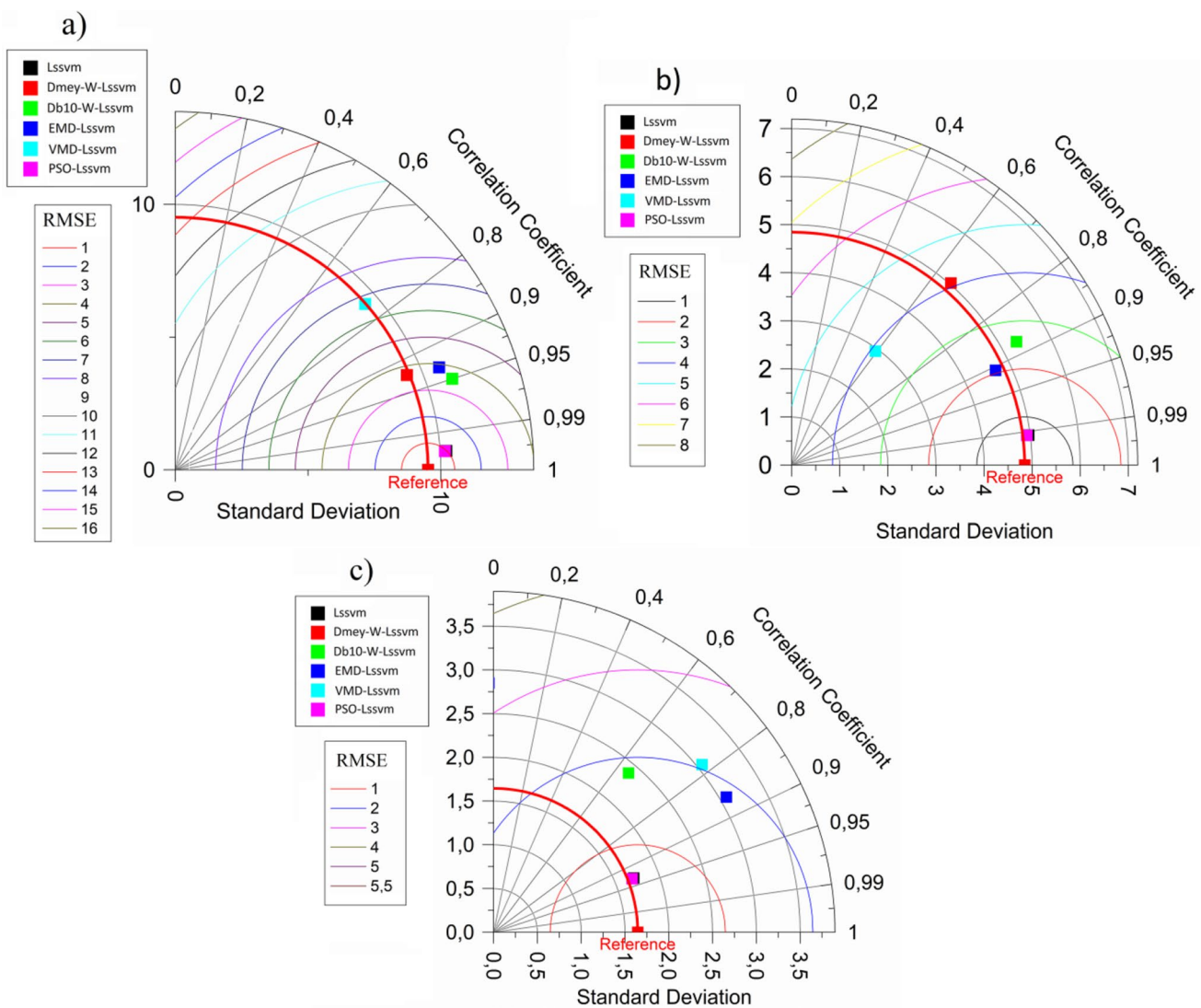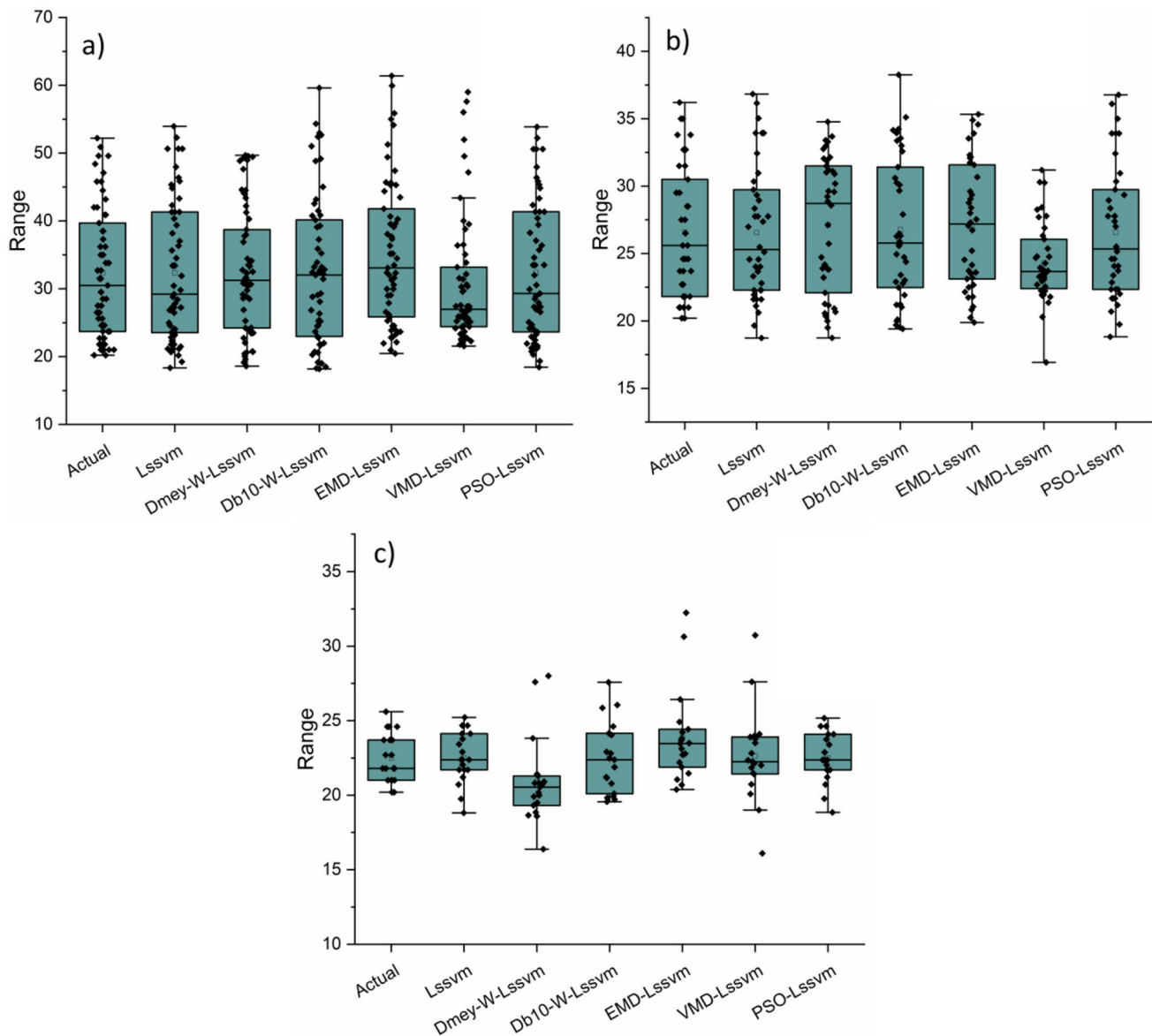

Fig. 6 Taylor diagrams of test results of 2009 flood. **a** Train-test ratio 70–30%, **b** train-test ratio 80–20%, **c** train-test ratio 90–10%

Fig. 7 Boxplot diagrams of the test set of 2009 flood. **a** Train-test ratio 70–30%, **b** train-test ratio 80–20%, **c** train-test ratio 90–10%

Figure 8 shows the scatter plots for the best flood routing models in 2009. Again, graphs are used for comparison of model accuracy. Scatter plots show the degree of correlation and distribution between predicted and observed values. When the scatter plots were analyzed, it was deduced that the PSO-LSSVM models produced close to the truth estimates because the predicted values were collected on the 45-degree regression line of the actual values.

Table 3 presents the estimation results of the 2013 floods of the test set obtained by ML, bio-inspired algorithm, and signal process approaches. When the results were interpreted according to the NSE and R^2 values closest to 1 and the lowest MAPE and MBE values, it was seen that the

PSO-LSSVM hybrid model was superior to the other models. In addition, the weakest flood estimation was produced with the VMD-LSSVM hybrid model. When the effect of data division ratios on model performance is examined, the most successful estimation performance was obtained with the train-test ratio (70:30%).

In Fig. 9, the test results are compared with Taylor diagrams for the visual analysis of the performance of the approaches used in predicting the 2013 floods. In these diagrams, the model performances are compared according to the closeness of the estimation models to the reference point, which is the true value, and the correlation, RMSE, and standard deviation values. According to these diagrams,
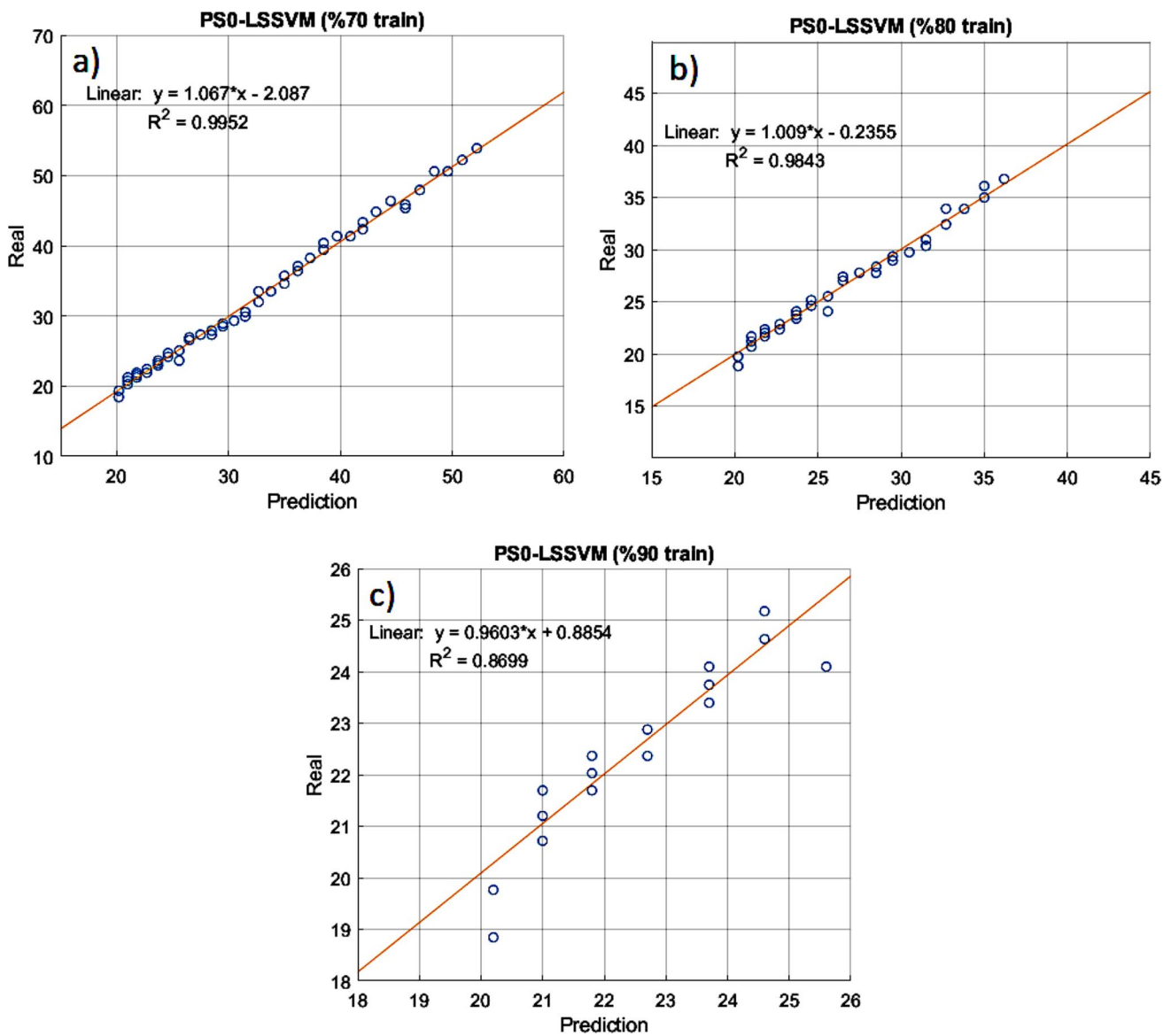
Fig. 8 Scatter plots of flood routing of the 2009 year. **a** Train-test ratio 70–30%, **b** train-test ratio 80–20%, **c** train-test ratio 90–10%

LSSVM and PSO-LSSVM models made the closest estimations for the train-test ratio (70:30%) and (80:20%). However, in the train-test ratio (90:10%), the EMD-LSSVM hybrid model showed the best results.

In Fig. 10, boxplots used to rank the estimation success of the models used in the estimation of the 2013 floods are shown. Boxplots effectively reveal the best model by comparing the distribution, median, mean, and percentiles of the predicted and actual values. According to these diagrams, the PSO-LSSVM model made the closest estimations for all train-test ratios.

Figure 11 shows the scatter diagrams and the spread of errors for the best model for modeling the floods of 2013. Scatter plots show the correlation between predicted and observed values and the degree of scatter. It is used to express the intervals at which the errors increase. When the scatter plots are analyzed, it is seen that the estimated values are gathered around the 45-degree line of the actual values. This plot indicates that the PSO-LSSVM models make estimations of actual values.

This study aims to estimate the hourly flood hydrograph of the downstream region by ML models, which are LSSVM, PSO-LSSVM, EMD-LSSVM, Wavelet-LSSVM, and VMD-LSSVM hybrid techniques. For this purpose, upstream data as input and downstream data as output are presented to the model in Turnasuyu River. At the end of the study, ML algorithms allow effective and reliable use of flood estimation. Furthermore, the results of the Zhao

**Table 3** Flood routing results for 2013 in test data

	Train-test ratio (70:30%)**				Train-test ratio (80:20%)				Train-test ratio (90:10%)			
	NSE	R^2	MAPE	MBE	NSE	R^2	MAPE	MBE	NSE	R^2	MAPE	MBE
LSSVM					LSSVM				LSSVM			
	0.771	0.98	0.05	2.46	0.32	0.76	0.03	1.11	0.421	0.78	0.02	0.89
Dmey wavelet LSSVM					Dmey wavelet LSSVM				Dmey wavelet LSSVM			
	0.32	0.96	0.08	4.31	−1.04	0.31	0.04	0.58	−1.12	0.54	0.04	1.69
DB 10 wavelet LSSVM					DB 10 wavelet LSSVM				DB 10 wavelet LSSVM			
	0.46	0.93	0.07	3.69	−0.28	0.57	0.03	1.01	−0.28	0.65	0.03	0.84
EMD-LSSVM					EMD-LSSVM				EMD-LSSVM			
	0.42	0.80	0.08	0.93	−3.00	0.74	0.07	−2.12	0.79	0.90	0.01	0.09
VMD-LSSVM					VMD-LSSVM				VMD-LSSVM			
	0.001	0.87	0.11	−4.39	−19.4	0.28	0.20	−7.20	0.71	0.80	0.02	0.25
PSO-LSSVM					PSO-LSSVM				PSO-LSSVM			
	0.94*	**0.97***	**0.02***	**0.15***	**0.73**	**0.73**	**0.01**	**0.005**	**0.80**	**0.80**	**0.01**	**−0.01**

* indicates the best model among all models, ** indicates the most appropriate data division value, and bold characters present the best algorithm for the selected data division ratio
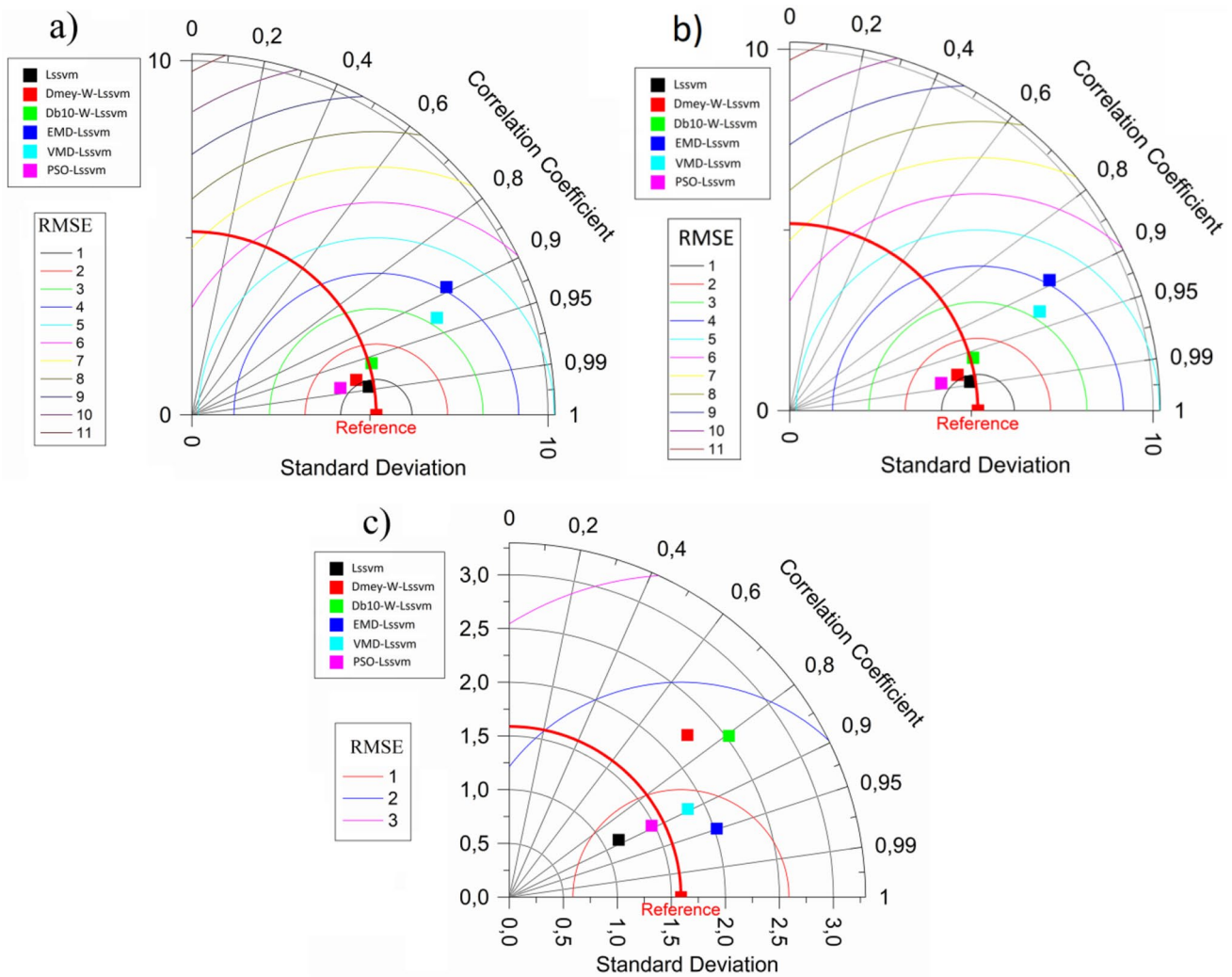
Fig. 9 Taylor diagrams of test results of 2013 flood. **a** Train-test ratio 70–30%, **b** train-test ratio 80–20%, **c** train-test ratio 90–10%
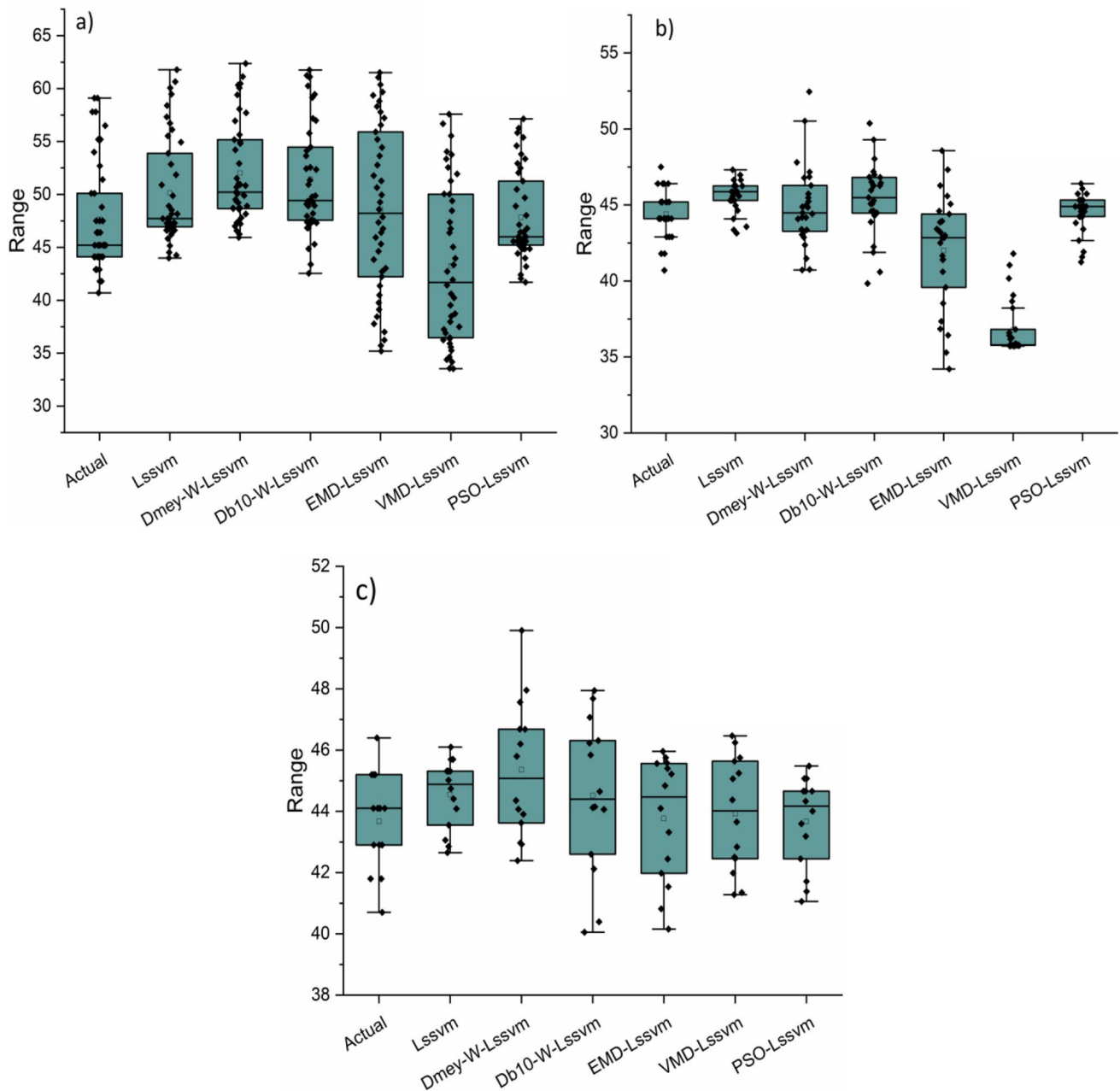
Fig. 10 Boxplots of the 2013 floods test set. **a** Train-test ratio 70–30%, **b** train-test ratio 80–20%, **c** train-test ratio 90–10%

et al. (2017), Alizadeh et al. (2021), Tayfur et al. (2018), and Wang et al. (2021) studies in the literature are compatible with the presented research.

Tayfur et al. (2018) used ant colony optimization, artificial neural network (ANN), genetic algorithm (GA), and PSO algorithms for flood hydrograph estimating. It has been determined that the models make flood hydrograph estimating effective and can be used profitably for flood hydrograph estimating. Furthermore, this study is compatible with Tayfur et al. (2018) flood routing study in establishing the PSO model and obtaining generally successful results in estimating flood routing results.

As a result of the models made by combining the LSSVM algorithm with PSO and signal processing techniques, it has been concluded that it is effective and reliable in streamflow estimations. Zhao et al. (2017) applied a new hybrid algorithm based on the EMD model for annual streamflow estimation. It was identified that the EMD-based chaotic LSSVM (EMD-CLSSVM) hybrid algorithm gives better results than CLSSVM hybrid model for estimating annual streamflow. The study's results significantly overlap with Zhao et al. (2017) hybrid models in producing effective streamflow estimation results. Alizadeh et al. (2021)
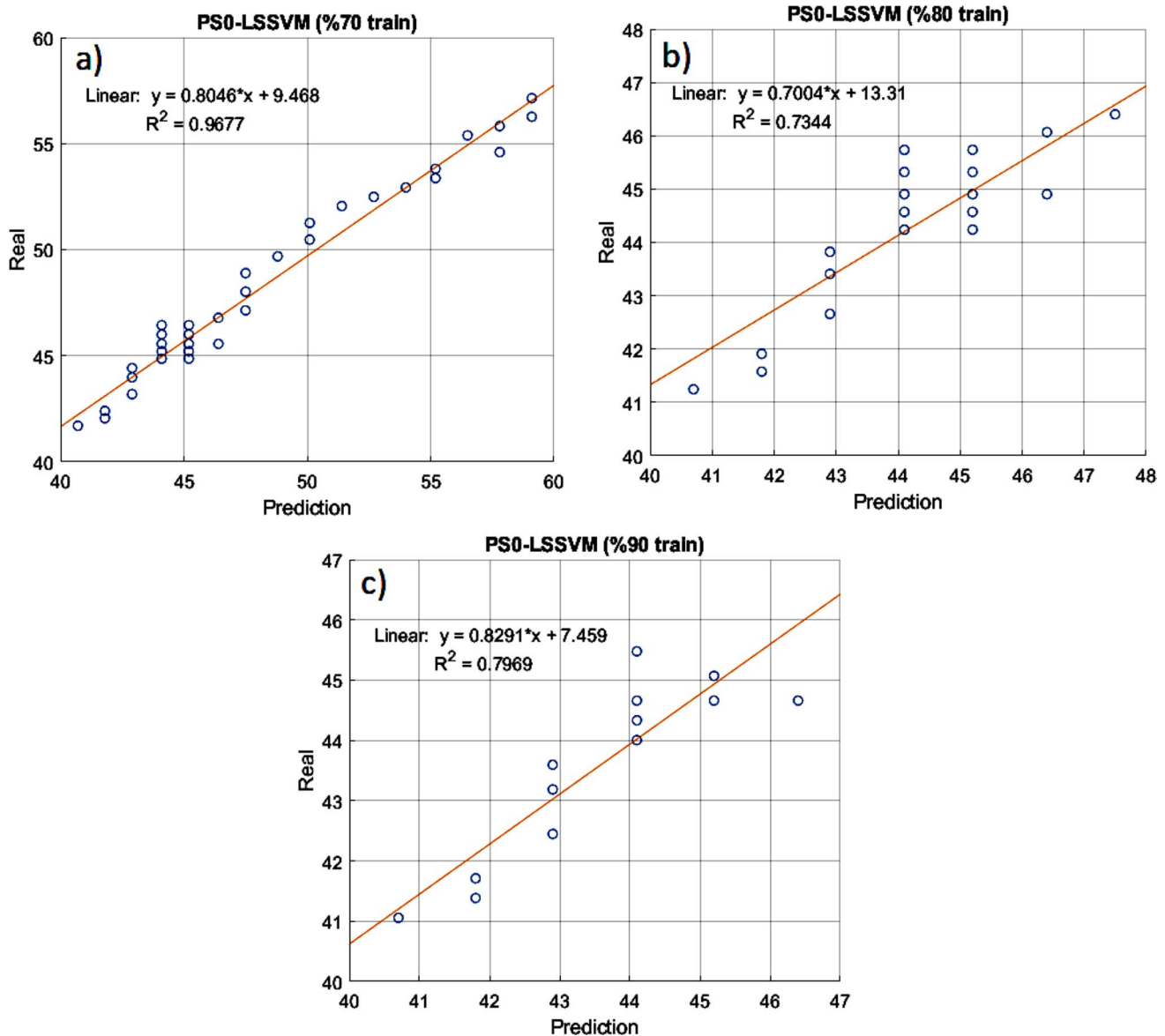
Fig. 11 Scatter plots of flood routing of the 2013 year. **a** Train-test ratio 70–30%, **b** train-test ratio 80–20%, **c** train-test ratio 90–10%

used wavelet transform (WT), ensemble EMD (EEMD), and mutual information (MI) to estimate river-level streamflow as a hybrid preprocessing approach and determined that the proposed WTEEMD-MI hybrid algorithm improves the accuracy of different modeling strategies. Wang et al. (2021) used a new hybrid algorithm (VMD-LSTM-PSO) by combining VMD with LSTM and PSO for daily runoff prediction. It was identified that the new model could be used in practice for hydrological estimating based on high estimating accuracy. Although decomposition techniques are successful in estimating river-level streamflow and daily runoff prediction, it contradicts Alizadeh et al. (2021) and Wang et al. (2021) studies regarding low performance in

flood routing calculations. However, it overlaps with the literature in optimizing the machine learning model with PSO. It is thought that the occurrence of this contradiction depends on the data length and data type used.

Conclusion

This study combined the LSSVM technique with PSO and various signal decomposition techniques to model two flood events in Ordu. In addition, it is aimed to evaluate the effect of the training-test ratio on flood routing. The main results obtained in the study are listed as follows:

- The most successful technique in flood routing was determined as PSO-LSSVM.
- ML parameters optimized with the PSO algorithm showed higher estimation accuracy than hybrid models established with inputs separated by signal processing techniques.
- The highest success in flood routing models was obtained using train-test ratio (70:30%).
- The study outputs are essential for various government agencies, hydrologists, agronomists, civil engineers, and urban and regional planners.

It has been shown in the literature that AI techniques have successful results in flood estimation and the estimation accuracy of optimized algorithms has been increased. These outputs are largely in line with the work done, and it is concluded that AI techniques produce highly accurate and cost-effective solutions for modeling complex mathematical expressions of floods. As a result, it has been proven that the loss of life and property against floods can be reduced by disseminating artificial intelligence techniques for developing flood estimating systems. In addition, thanks to the flood routing estimation model obtained in the study, it is possible to carry out early warning, awareness, and preparation stages against floods effectively and to implement flood risk management successfully.

The study's main limitation is that only two DOS were used on the same stream. Usually, there is one DOS on the same stream. It is thought that ML models established for different regions by increasing the number of DOS will produce effective results. It is hoped that the effects of a flood on the environment can be predicted thanks to a network to be established that can be controlled from a central location and that the measures to be taken by informing the relevant units will save many lives.

In future studies, flood estimation accuracy can be investigated by combining bio-inspired algorithms such as whale optimization algorithm, gray wolf optimizer and artificial bee colony, ant colony optimization, chaos game optimization, robust empirical mode decomposition, discrete multiresolution analysis, and various signal decomposition techniques such as empirical wavelet transform and variational mode decomposition with tree-based and neural network–based ML techniques.

Data availability Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval The manuscript complies with all the ethical requirements. The paper was not published in any journal.

Consent for publication Not applicable.

Consent to participate Not applicable.

Conflict of interest The authors declare no competing interests.

References

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000) Artificial neural networks in hydrology I: Preliminary concepts. J Hydrol Eng 5(2):115–123. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115)

Akbari R, Hessami-Kermani MR, Shojaee S (2020) Flood routing: improving outflow using a new nonlinear muskingum model with four variable parameters coupled with PSO-GA algorithm. Water Resour Manage 34(10):3291–3316. https://doi.org/10.1007/s11269-020-02613-5

Alizadeh F, Faregh Gharamaleki A, Jalilzadeh R (2021) A two-stage multiple-point conceptual model to predict river stage-discharge process using machine learning approaches. J Water Clim Change 12(1):278–295. https://doi.org/10.2166/wcc.2020.006

Bagatur T, Onen F (2018) Development of predictive model for flood routing using genetic expression programming. J Flood Risk Manag 11:S444–S454. https://doi.org/10.1111/jfr3.12232

Ball JE (2022) Modelling accuracy for urban design flood estimation. Urban Water J 19(1):87–96. https://doi.org/10.1080/1573062X.2021.1955283

Barati R, Badfar M, Azizyan G, Akbari GH (2018) Discussion of "Parameter Estimation of Extended Nonlinear Muskingum Models with the Weed Optimization Algorithm" by Farzan Hamedi, Omid Bozorg-Haddad, Maryam Pazoki, Hamid-Reza Asgari, Mehran Parsa, and hugo a. Loáiciga J Irrig Drain Eng 144:7017021. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001095

Barbetta S, Moramarco T, Perumal M (2017) A Muskingum-based methodology for river discharge estimation and rating curve development under significant lateral inflow conditions. J Hydrol 554:216–232. https://doi.org/10.1016/j.jhydrol.2017.09.022

Başakın EE, Ekmekcioğlu Ö, Özger M (2021) Drought prediction using hybrid soft-computing methods for semi-arid region. Model Earth Syst Environ 7(4):2363–2371. https://doi.org/10.1007/s40808-020-01010-6

Başakın EE, Ekmekcioğlu Ö, Çıtakoğlu H, Özger M (2022) A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on preprocessing uncertainty assessment. Neural Comput Appl 34(1):783–812. https://doi.org/10.1007/s00521-021-06424-6

Bostanci D, İskender R, Helli S, Polat N (2016) The determination of fish fauna of Turnasuyu stream (Ordu). Ordu Univ J Sci Technol 5(2):1–9

Cai H, Wang Y, Song C, Wang T, Shen Y (2022) Prediction of surface subsidence based on PSO-BP neural network. J Phys Confer

Ser 2400(1):012046. https://doi.org/10.1088/1742-6596/2400/1/012046

Capik M, Yılmaz AO, Cavusoglu İ (2012) Hydropower for sustainable energy development in Turkey: the small hydropower case of the Eastern Black Sea Region. Renew Sustain Energy Rev 16(8):6160–6172. https://doi.org/10.1016/j.rser.2012.06.005

Chow VT, Maidment DR, Mays LW (1988) Applied hydrology. McGraw-Hill, New York

Dazzi S, Vacondio R, Mignosa P (2021) Flood stage forecasting using machine-learning methods: a case study on the Parma River (Italy). Water 13(12):1612. https://doi.org/10.3390/w13121612

Dehghani R, TorabiPoudeh H, Izadi Z (2022) Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. Model Earth Syst Environ 8(2):2599–2613. https://doi.org/10.1007/s40808-021-01253-x

Ehteram M, Binti Othman F, Mundher Yaseen Z, Abdulmohsin Afan H, Falah Allawi M, Bt. Abdul Malek M, ..., El-Shafie A (2018) Improving the Muskingum flood routing method using a hybrid of particle swarm optimization and bat algorithm. Water 10(6):807 https://doi.org/10.3390/w10060807

El Boujdaini L, Mezrhab A, Moussaoui MA (2021) Artificial neural networks for global and direct solar irradiance forecasting: a case study. Energy Sour A: Recover Util Environ Effects 1–21. https://doi.org/10.1080/15567036.2021.1940386

Granata F, Gargano R, De Marinis G (2016) Support vector regression for rainfall-runoff modeling in urban drainage: a comparison with the EPA's storm water management model. Water 8(3):69. https://doi.org/10.3390/w8030069

Hadidi A, Holzbecher E, Molenaar RE (2020) Flood mapping in face of rapid urbanization: a case study of Wadi Majraf-Manumah, Muscat Sultanate of Oman. Urban Water J 17(5):407–415. https://doi.org/10.1080/1573062X.2020.1713172

Hamedi F, Bozorg-Haddad O, Pazoki M, Asgari HR, Parsa M, Loáiciga HA (2016) Parameter estimation of extended nonlinear Muskingum models with the weed optimization algorithm. J Irrig Drain Eng 142(12):04016059. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001095

Hassanvand MR, Karami H, Mousavi SF (2018) Investigation of neural network and fuzzy inference neural network and their optimization using meta-algorithms in river flood routing. Nat Hazards 94:1057–1080. https://doi.org/10.1007/s11069-018-3456-z

Hebb DO (1949) The first stage of perception: growth of the assembly. Organ Behav 4:60–78

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Royal Soc London Ser A: Math Phys Eng Sci 454(1971):903–995. https://doi.org/10.1098/rspa.1998.0193

Huang S, Chang J, Huang Q, Chen Y (2014) Monthly streamflow prediction using modified EMD-based support vector machine. J Hydrol 511:764–775. https://doi.org/10.1016/j.jhydrol.2014.01.062

Kadkhodazadeh M, Farzin S (2021) A novel LSSVM model integrated with GBO algorithm to assessment of water quality parameters. Water Resour Manage 35(12):3939–3968. https://doi.org/10.1007/s11269-021-02913-4

Li G, Ma X, Yang H (2018) A hybrid model for monthly precipitation time series forecasting based on variational mode decomposition with extreme learning machine. Information 9(7):177. https://doi.org/10.3390/info9070177

Ma M, Liu C, Zhao G, Xie H, Jia P, Wang D, ..., Hong Y (2019) Flash flood risk analysis based on machine learning techniques in the Yunnan Province, China. Remote Sens 11(2):170. https://doi.org/10.3390/rs11020170

McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133. https://doi.org/10.1007/BF02478259

Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. Water 10(11):1536. https://doi.org/10.3390/w10111536

Mazzoleni M, Noh SJ, Lee H, Liu Y, Seo DJ, Amaranto A, ..., Solomatine DP (2018) Real-time assimilation of streamflow observations into a hydrological routing model: effects of model structures and updating methods. Hydrol Sci J 63(3):386–407. https://doi.org/10.1080/02626667.2018.1430898

Nhu VH, Shahabi H, Nohani E, Shirzadi A, Al-Ansari N, Bahrami S, ..., Nguyen H (2020) Daily water level prediction of Zrebar Lake (Iran): a comparison between M5P, random forest, random tree and reduced error pruning trees algorithms. ISPRS Int J Geo-Inf 9(8):479. https://doi.org/10.3390/ijgi9080479

Norouzi H, Bazargan J (2020) Flood routing by linear Muskingum method using two basic floods data using particle swarm optimization (PSO) algorithm. Water Supply 20(5):1897–1908. https://doi.org/10.2166/ws.2020.099

Norouzi H, Karimi V, Bazargan J, Hemmati H (2021) Different types of optimizing the parameters of hydrological routing methods using particle swarm optimization (PSO) algorithm for flood routing in the Karun River. J Watershed Manag Res 12(23):285–295

Nourani V, Alami MT, Aminfar MH (2009) A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng Appl Artif Intell 22:466–472. https://doi.org/10.1016/j.engappai.2008.09.003

Noury M, Sedghi H, Babazedeh H, Fahmi H (2014) Urmia lake water level fluctuation hydro informatics modeling using support vector machine and conjunction of wavelet and neural network. Water Resour 41(3):261–269. https://doi.org/10.1134/S0097807814030129

Novoa-Hernández P, Corona CC, Pelta DA (2011) Efficient multiswarm PSO algorithms for dynamic environments. Memetic Computing 3(3):163–174. https://doi.org/10.1007/s12293-011-0066-7

Okkan U, Serbes ZA (2012) Rainfall–runoff modeling using least squares support vector machines. Environmetrics 23(6):549–564. https://doi.org/10.1002/env.2154

Okkan U, Kirdemir U (2020) Locally tuned hybridized particle swarm optimization for the calibration of the nonlinear Muskingum flood routing model. J Water Clim Change 11(S1):343–358. https://doi.org/10.2166/wcc.2020.015

Pant R, Thacker S, Hall JW, Alderson D, Barr S (2018) Critical infrastructure impact assessment due to flood exposure. J Flood Risk Manag 11(1):22–33. https://doi.org/10.1111/jfr3.12288

Shafaei M, Kisi O (2016) Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. Water Resour Manag 30:79–97. https://doi.org/10.1007/s11269-015-1147-z

Sahana M, Rehman S, Sajjad H, Hong H (2020) Exploring effectiveness of frequency ratio and support vector machine models in storm surge flood susceptibility assessment: a study of Sundarban Biosphere Reserve. India. Catena 189:104450. https://doi.org/10.1016/j.catena.2019.104450

Saplioglu K, Ozturk TSK, Acar R (2020) Optimization of open channels using particle swarm optimization algorithm. J Intell Fuzzy Syst 39(1):399–405. https://doi.org/10.3233/JIFS-191355

Seo Y, Kim S, Kisi O, Singh VP, Parasuraman K (2016) River stage forecasting using wavelet packet decomposition and machine learning models. Water Resour Manage 30(11):4011–4035. https://doi.org/10.1007/s11269-016-1409-4

Seo Y, Kim S, Singh VP (2018) Machine learning models coupled with variational mode decomposition: a new approach for modeling

daily rainfall-runoff. Atmosphere 9(7):251. https://doi.org/10.3390/atmos9070251

Shabri A, Suhartono (2012) Streamflow forecasting using least-squares support vector machines. Hydrol Sci J 57(7):1275–1293. https://doi.org/10.1080/02626667.2012.714468

Sudheer C, Maheswaran R, Panigrahi BK, Mathur S (2014) A hybrid SVM-PSO model for forecasting monthly streamflow. Neural Comput Appl 24(6):1381–1389. https://doi.org/10.1007/s00521-013-1341-y

Şenel FA, Öztürk TSK, Saplioğlu K (2020) Optimization of time delay dimension by ant lion algorithm using artificial neural networks for estimation of Yeşilırmak river flow data. Afyon Kocatepe Univ J Sci Eng 20:310–318. https://doi.org/10.35414/akufemubid.669602

Tayfur G (2017) Modern optimization methods in water resources planning, engineering and management. Water Resour Manage 31(10):3205–3233. https://doi.org/10.1007/s11269-017-1694-6

Tayfur G, Moramarco T (2008) Predicting hourly-based flow discharge hydrographs from level data using genetic algorithms. J Hydrol 352(1–2):77–93. https://doi.org/10.1016/j.jhydrol.2007.12.029

Tayfur G, Moramarco T, Singh VP (2007) Predicting and forecasting flow discharge at sites receiving significant lateral inflow. Hydrol Process: Int J 21(14):1848–1859. https://doi.org/10.1002/hyp.6320

Tayfur G, Singh VP, Moramarco T, Barbetta S (2018) Flood hydrograph prediction using machine learning methods. Water 10(8):968. https://doi.org/10.3390/w10080968

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res: Atmos 106(D7):7183–7192. https://doi.org/10.1029/2000JD900719

Wang X, Wang Y, Yuan P, Wang L, Cheng D (2021) An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach. Hydrol Sci J 66(9):1488–1502. https://doi.org/10.1080/02626667.2021.1937631

Wen Q, Gao J, Song X, Sun L, Xu H, Zhu S (2019) RobustSTL: a robust seasonal-trend decomposition algorithm for long time series. Proc AAAI Conf Artif Intell 33(01):5409–5416. https://doi.org/10.1609/aaai.v33i01.33015409

Widiasari IR, Nugroho LE, Widyawan (2017) Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. In: 2017 International Conference on Innovative and Creative Information Technology (ICITech). IEEE, pp 1–5. https://doi.org/10.1109/INNOCIT.2017.8319150

Xu Y, Hu C, Wu Q, Jian S, Li Z, Chen Y, ..., Wang S (2022) Research on particle swarm optimization in LSTM neural networks for

rainfall-runoff simulation. J Hydrol 608:127553. https://doi.org/10.1016/j.jhydrol.2022.127553

Yaseen ZM, El-Shafie A, Jaafar O, Sayl AHA, KN, (2015) Artificial intelligence-based models for streamflow forecasting: 2000–2015. J Hydrol 530:829–844. https://doi.org/10.1016/j.jhydrol.2015.10.0384

Yuan X, Zhang X, Tina F (2020) Research and application of an intelligent networking model for flood forecasting in the arid mountainous basins. J Flood Risk Manag 13:e12638. https://doi.org/10.1111/jfr3.12638

Zare M, Koch M (2014) An analysis of MLR and NLP for use in river flood routing and comparison with the Muskingum method. In: ICHE 2014. Proceedings of the 11th International Conference on Hydroscience & Engineering, September 28–October 2, 2014, Hamburg, Germany, pp 505–514

Zhang GP, Qi M (2005) Neural network forecasting for seasonal and trend time series. Eur J Oper Res 160:501–514. https://doi.org/10.1016/j.ejor.2003.08.037

Zhang H, Liu L, Jiao W, Li K, Wang L, Liu Q (2022) Watershed runoff modeling through a multi-time scale approach by multivariate empirical mode decomposition (MEMD). Environ Sci Pollut Res 29(2):2819–2829. https://doi.org/10.1007/s11356-021-13676-1

Zhao X, Chen X, Xu Y, Xi D, Zhang Y, Zheng X (2017) An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. Water 9(3):153. https://doi.org/10.3390/w9030153

Zhou Y, Guo S, Chang FJ, Liu P, Chen AB (2018) Methodology that improves water utilization and hydropower generation without increasing flood risk in mega cascade reservoirs. Energy 143:785–796. https://doi.org/10.1016/j.energy.2017.11.035

Zounemat-Kermani M, Matta E, Cominola A, Xia X, Zhang Q, Liang Q, Hinkelmann R (2020) Neurocomputing in surface water hydrology and hydraulics: a review of two decades retrospective, current status and future prospects. J Hydrol 588:125085. https://doi.org/10.1016/j.jhydrol.2020.125085