Results in Engineering 22 (2024) 102104

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

A novel additive regression model for streamflow forecasting in German rivers

Francesco Granata [a], Fabio Di Nunno [a,*], Quoc Bao Pham [b]

[a] University of Cassino and Southern Lazio, Department of Civil and Mechanical Engineering (DICEM), Via Di Biasio, 43, 03043, Cassino, Frosinone, Italy
[b] University of Silesia in Katowice, Institute of Earth Sciences, Faculty of Natural Sciences, Będzińska Street 60, 41-200, Sosnowiec, Poland

ARTICLE INFO

Keywords:
Streamflow forecasting
Machine learning
Hybrid models
Medium-term forecast
Germany

ABSTRACT

Forecasting streamflows, essential for flood mitigation and the efficient management of water resources for drinking, agriculture and hydroelectric power generation, presents a formidable challenge in most real-world scenarios. In this study, two models, the first based on the Additive Regression of Radial Basis Function Neural Networks (AR-RBF) and the second based on the stacking with the Pace Regression of the Multilayer Perceptron and Random Forest (MLP-RF-PR), were compared for the prediction of short-term (1–3 days ahead) and medium-term (7 days ahead) daily streamflow rates of three different rivers in Germany: the Elbe River at Wittenberge, the Leine River at Herrenhausen, and the Saale River at Hof. The lagged values of streamflow rate, precipitation and temperature were considered for the modeling. Moreover, the Bayesian Optimization (BO) algorithm was used to assess the optimal number of lagged values and hyperparameters. Both models showed accurate predictions for short-term forecasting, with R^2 for 1-day ahead predictions ranging from 0.939 to 0.998 for AR-RBF and from 0.930 to 0.996 for MLP-RF-PR, while MAPE ranged from 2.02 % to 8.99 % for AR-RBF and from 2.14 % to 9.68 % for MLP-RF-PR, when exogeneous variables were included. As the forecast horizon increased, a reduction in forecasting accuracy was observed. However, both models could still predict the overall flow pattern, even for 7-day-ahead predictions, with R^2 ranging from 0.772 to 0.871 for AR-RBF and from 0.703 to 0.840 for MLP-RF-PR, while MAPE ranged from 10.60 % to 20.45 % for AR-RBF and from 10.44 % to 19.65 % for MLP-RF-PR. Overall, the outcomes of this study suggest that both AR-RBF and MLP-RF-PR can be reliable tools for the short- and medium-term streamflow rate prediction, requiring a short number of parameters to be optimized, making them easy to implement while reducing the calculation time required.

1. Introduction

Streamflow forecasting has a pivotal role in multifaceted domains such as water resource management, flood alertness, and environmental conservation [1]. The accuracy of predictions facilitates optimized water allocation, hydropower planning, and agricultural scheduling, thereby accentuating the efficacy of water resource utilization [2]. Furthermore, its ability to provide timely alerts regarding potential flood events ameliorates disaster impacts and serves as a custodian of lives and property. The ecological domain also stands to benefit, as streamflow forecasts aid in preserving aquatic ecosystems by maintaining critical flow conditions and supporting biodiversity. Underscoring the profound importance of this practice is fundamental in nurturing sustainable water management and fostering resolute decision-making amidst the dynamic backdrop of evolving hydrological patterns.

Several types of models can be employed in streamflow prediction: empirical models, physical-based models, conceptual models, data-driven models, and ensemble forecasting models. Empirical models rely on historical relationships between meteorological and hydrological variables. Their simplicity and computational efficiency make them ideal for quick forecasts, especially in data-limited regions. However, their disadvantage lies in potential limitations in capturing complex hydrological processes accurately, as they lack a complete physical understanding. Consequently, their performance may be compromised during extreme events or in catchments with significant nonlinearity.

Physical-based models employ mathematical representations of the hydrological cycle to simulate streamflow [3,4]. They offer a comprehensive understanding of catchment behavior and prove valuable in ungauged basins. However, their complexity demands extensive data

* Corresponding author.
E-mail addresses: f.granata@unicas.it (F. Granata), fabio.dinunno@unicas.it (F. Di Nunno), quoc_bao.pham@us.edu.pl (Q.B. Pham).

https://doi.org/10.1016/j.rineng.2024.102104
Received 7 February 2024; Received in revised form 16 March 2024; Accepted 3 April 2024
Available online 10 April 2024

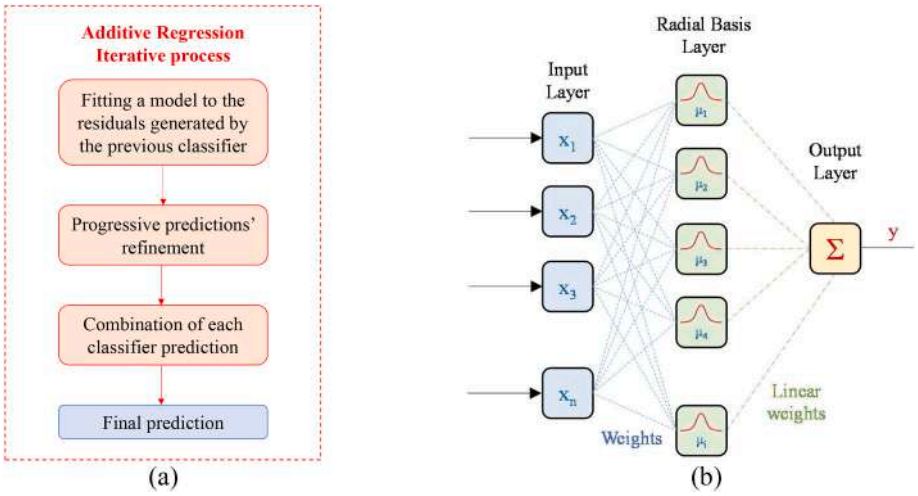

Fig. 1. Schemes related to the Additive Regression (a) – Radial Basis Function (b) model.

inputs and computational resources. Calibrating physical models can be challenging, and uncertainties in parameter estimation can impact forecast accuracy.

Conceptual models provide a compromise between simplicity and process representation in streamflow generation [5]. They suit a wide range of watersheds and can be calibrated to site-specific conditions, enhancing accuracy compared to empirical models. Nevertheless, conceptual models require substantial effort in parameterization and calibration, and their performance may degrade under significant changes in catchment characteristics.

Data-driven models, which include models based on machine learning algorithms and statistical models, aim to discover patterns and relationships within historical data for streamflow prediction [6]. The increasing availability of data, the ability to reproduce complex nonlinear relationships between inputs and outputs, the capacity for progressive adaptation to newly available information, and high predictive accuracy increasingly drive scientists and engineers to use models based on Machine Learning (ML) and Deep Learning (DL) algorithms for streamflow forecasting [7–13].

The choice of an appropriate model depends on the specific needs, the level of knowledge of the watercourse and its basin, the availability of data, and the desired level of prediction accuracy.

In recent years, researchers have frequently focused on creating hybrid or ensemble models [14–18]. These models offer several benefits compared to individual ML models, including improved accuracy, enhanced stability, reduced overfitting, the ability to handle diverse data characteristics, and heightened model robustness. However, it is essential to acknowledge that ensemble models also present challenges, such as increased computational complexity, potential difficulties in interpreting the model, and the necessity for careful selection and calibration of the individual models. Lin et al. [19] proposed a hybrid DIFF–FFNN–LSTM model, based on the first-order difference (DIFF), feedforward neural network (FFNN) and long short-term memory network (LSTM), to predict hourly streamflow in the Andun basin, China. The authors compared the prediction performed with the hybrid DIFF–FFNN–LSTM model with those achieved with individual ML- or DL-based models, demonstrating the superiority of the hybrid model. Tyralis et al. [14] employed super ensemble learning for a one-step-ahead daily streamflow forecasting, considering 511 basins in the USA. The authors combined 10 different machine learning algorithms, showing how the super ensemble learning model outperformed all individual algorithms, the best of which was the ANN model. Granata et al. [15] compared the performance of a stacked model based on MLP and RF algorithms with a more complex model based on bi-directional LSTM. The authors demonstrated how the performance of the two models was comparable, with the stacked model that outperformed the bi-directional LSTM network model in the prediction of the flow rate peaks while maintaining the advantage of shorter computation times.

This study introduces a novel forecasting model for streamflow based on Additive Regression of Radial Basis Function Neural Networks (AR-RBF), and assesses its capabilities in providing short-term (1–3 days) and medium-term (7 days) predictions. To the best of the authors' knowledge, the AR-RBF algorithm is herein employed for the first time to address a forecasting problem within the domain of hydrology. In addition, the Bayesian Optimization (BO) method was employed to optimize the optimal number of lagged values for each variable and the hyperparameters of the algorithms. The performance of this new model is compared with that of another predictive model of recognized high accuracy: the MLP-RF Stacked Model, in a new variant that uses Pace Regression as a meta-learner. This variant based on the Pace Regression of the MLP-RF Stacked Model is also employed here for the first time. Special attention is paid to the ability to predict flood events accurately. Special attention is paid to the ability to predict flood events accurately. The impact of the absence of exogenous predictors, such as rainfall and temperature, is also investigated. Three rivers in Germany, characterized by different hydrological characteristics and extent of catchment areas, were chosen as case studies: Elbe, Leine and Saale. To date, German rivers have been poorly considered by forecasting models based on Artificial Intelligence algorithms. Furthermore, this study underscores how daily monitoring data enables the development of accurate predictive models that consider meteorological factors' influence on hydrological processes. Comparative analysis illuminates the effectiveness of the utilized methodologies in predicting streamflow. The findings contribute to propel statistical modeling [20] objectives within hydrology research, providing valuable insights to inform efficient water resource management strategies.

2. Materials and methods

2.1. Additive Regression of radial basis function networks

Additive Regression (AR) stands as a prominent ensemble learning technique, exerting considerable influence as a metaclassifier, particularly in enhancing the predictive capabilities of foundational regression models. The fundamental premise of AR lies in its iterative methodology, whereby successive models are fit to the residuals left unexplained by preceding classifiers, thereby facilitating a progressive refinement of predictive outcomes. This iterative refinement culminates in aggregating predictions rendered by each constituent classifier (Fig. 1a). Central to the efficacy of AR is the integration of a shrinkage parameter,

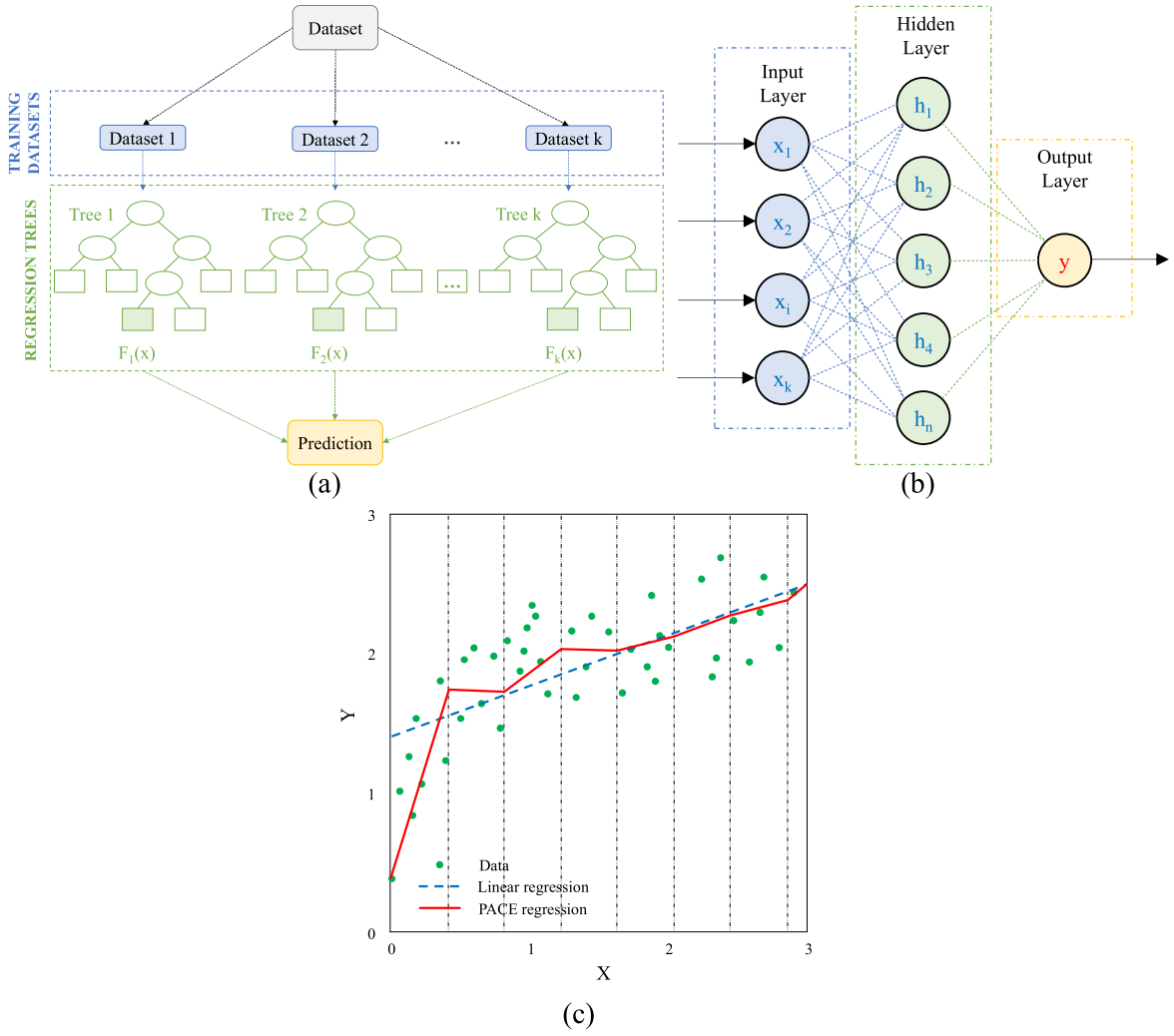

Fig. 2. Schemes related to the Multilayer Perceptron (a) – Random Forest (b) – Pace Regression (c) model.

which assumes a pivotal role in constraining overfitting tendencies while concurrently imparting a smoothing effect on the resultant predictions.

The AR algorithm unfolds in a sequence of steps, commencing with the initialization of the prediction vector as the mean response across all observations. Subsequently, individual base learners are fitted sequentially to the residuals, with each iteration contributing incrementally to the refinement of predictive accuracy. At the core of AR's functionality lies the concept of model stacking, where the predictions generated by each base learner are amalgamated to formulate a composite prediction. This amalgamation is achieved through weighted summation, where the weighting coefficients are governed by the shrinkage parameter, thus facilitating optimal balance between model complexity and predictive performance.

Noteworthy is the pivotal role played by the shrinkage parameter in modulating the trade-off between bias and variance. By constraining the magnitude of individual classifier contributions, the shrinkage parameter serves as a potent mechanism for mitigating overfitting, thereby enhancing the generalizability of the resultant model. However, it is imperative to acknowledge the inherent computational overhead associated with fine-tuning the shrinkage parameter, as reductions in its value necessitate a proportional increase in learning time.Despite its demonstrable efficacy in diverse domains, AR has been relatively underutilized in the domain of hydrological modeling, and in the available literature studies, tree models have been most frequently used as basic regressors (e. g. Ref. [21]).

In this study, the RBF-NN [22] algorithm was adopted as the base regressor, which represents a novel solution for studies concerning streamflow prediction. The RBF-NN is a feedforward network consisting of three layers: an input layer, a hidden layer, and an output layer. The hidden layer employs radial basis functions to transform input data into a higher-dimensional space, simplifying the regression task. Radial basis functions are mathematical functions centered at specific points, exhibiting exponential decay as the distance from the center increases. The initial centers for the Gaussian radial basis functions are determined using the K-Means algorithm. Among various choices, the Gaussian function is commonly adopted as the activation function for the hidden layer. The output layer of the RBF neural network typically consists of a linear combination of the hidden layer activations (Fig. 1b). The weights for this combination are determined through either a least-squares approach or gradient descent. An advantage of the RBF-NN lies in its reduced dependency on extensive training data when compared to other neural networks. This is due to the feature extraction performed by the hidden layer, which effectively reduces the dimensionality of the input data. Additionally, the RBF-NN demonstrates enhanced resilience to overfitting. This property further solidifies the appeal of the RBF-NN in various modelling tasks [23].

2.2. *Stacked model of Multilayer Perceptron and Random Forest*

Another powerful type of ensemble model is the stacked model. The stacking process typically consists of two or more layers. In the first

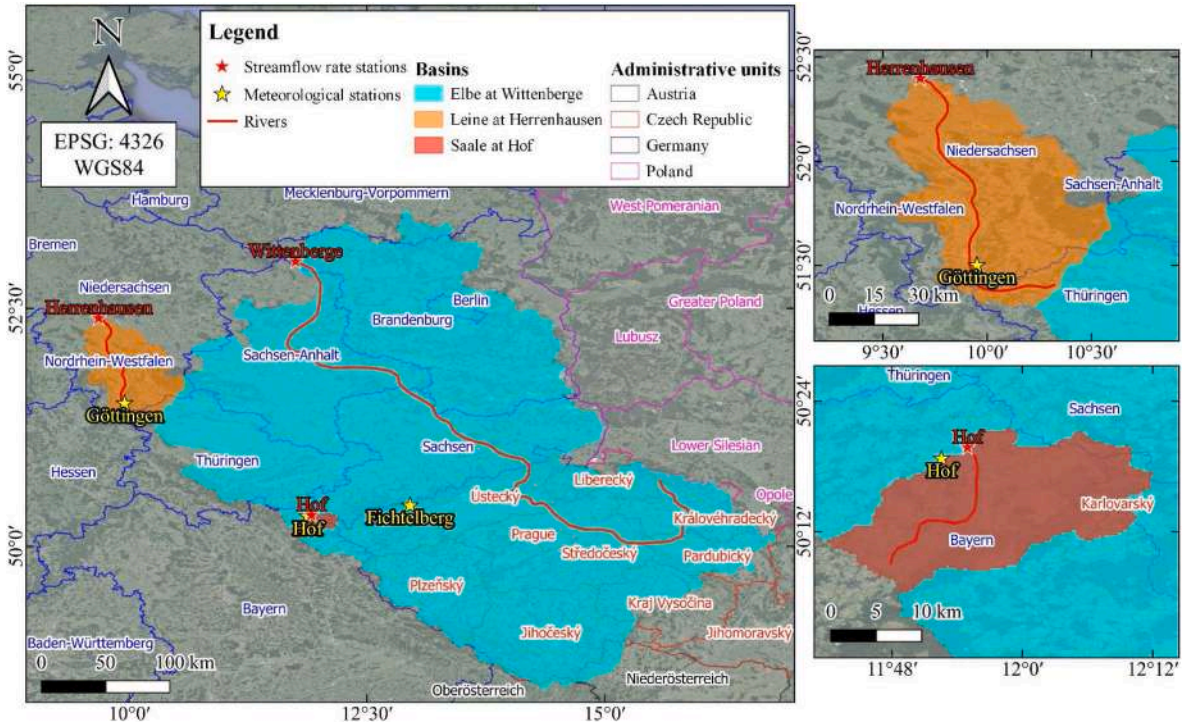

Fig. 3. Location of the investigated rivers and catchment areas.

layer, base models with different algorithms are trained on the training data. The predictions from these base models serve as input features for the second layer, which is the meta-learner and combines the outputs from the base models to generate the final prediction. The stacked model employed in this study includes the MLP and RF algorithms as base models, while the PR algorithm was chosen as a meta-learner.

MLP is a feedforward neural network characterized by its layered architecture [24]. It comprises an input layer, one or more hidden layers, and an output layer [25,26]. Each layer consists of interconnected neurons, and data flows only in a forward direction, from input to output (Fig. 2a). The first layer of the MLP receives input data, typically represented as feature vectors. Each neuron in this layer corresponds to a specific input data feature. The MLP may contain one or more hidden layers. These layers are responsible for hierarchical feature extraction and abstract representation learning. Neurons within the hidden and output layers perform a weighted summation of inputs, followed by the application of an activation function. The activation function introduces non-linearity, enabling the network to learn complex patterns. The connections between neurons in adjacent layers are associated with learnable parameters known as weights. Additionally, each neuron typically has a bias term, contributing to the overall transformation. The weights and biases are usually initialized randomly at the beginning of training and adjusted during the optimization process. Common activation functions include sigmoid, ReLU (Rectified Linear Unit), and tanh (hyperbolic tangent), each offering different properties affecting the network's performance and training speed. The last layer of the MLP produces the final predictions or representations of the input data, depending on the task at hand. MLPs learn by minimizing a loss function quantifying the difference between predicted and actual outputs. The backpropagation algorithm updates the weights based on the gradients of the loss function, iteratively improving the model's performance. Standard optimization methods like stochastic gradient descent (SGD) or its variants are employed to optimize the network's weights and biases. Training data is often divided into batches to enhance computational efficiency during optimization.

An RF [27] is an ensemble learning method composed of multiple regression trees, where each tree is constructed independently and contributes to the final prediction through averaging. The basic building blocks of an RF are regression trees, which partition the input data space into regions based on feature values to make predictions (Fig. 2b). Constructing a regression tree entails iteratively partitioning the input dataset into subsets and employing a multivariable linear regression model to make predictions within each subset. As the tree advances, each branch is subsequently divided into smaller partitions by assessing all possible subdivisions in each field. At each stage, the division minimizing the Least-Squared Deviation is chosen. This deviation is computed using the formula:

$$R(t) = \frac{1}{N(t)} \sum_{i \in t} (y_i - y_m(t))^2 \quad (1)$$

where N(t) represents the number of units in the node, y_i denotes the value of the target variable in the i-th unit, and y_m signifies the average value of the target variable in node t. R(t) estimates the impurity in each node, and the tree-building process persists until the minimum impurity is reached or another stopping criterion is satisfied. To counteract overfitting, a pruning procedure is employed.

In addition, combining multiple regression trees further reduces the risk of overfitting and makes RFs more resilient to noisy data. RFs perform well on high-dimensional and large-scale datasets and exhibit good generalization capabilities.

PR [28] is a non-parametric regression technique that can be used to approximate a non-linear relationship between two variables when the errors are non-normal. In PR, the essential operation is to partition the range of the independent variable into equally sized segments and analyze the relationship between the dependent and independent variables within each segment. These relationships are known as "paces". The number of intervals is generally determined based on the sample size and the complexity of the variable relationship. To obtain the overall estimated relationship between the variables in an ensemble model, the paces are combined using a weighted average, where the weights are influenced by the number of observations in each interval. As a result, the estimated relationship takes the form of a piecewise function, effectively approximating the non-linear relationship between

Table 1
Daily streamflow rate, precipitation and mean air temperature statistics for the investigated rivers.

	Elbe at Wittenberge	Leine at Herrenhausen	Saale at Hof
Streamflow rate (m^3/s)			
Mean	679.87	49.85	5.11
Median	547.00	35.80	3.27
Max	4200.00	924.00	104.00
Min	131.00	8.90	0.16
Std Deviation	444.82	43.02	5.73
CV	0.65	0.86	1.12
1st Quartile	376.00	23.00	1.90
3rd Quartile	846.00	59.80	6.19
Skew	0.90	0.98	0.96
Precipitation (mm)			
Mean	3.14	1.74	1.96
Median	0.50	0.10	0.10
Max	137.80	72.30	67.70
Std Deviation	6.00	4.01	4.24
CV	1.91	2.30	2.17
3rd Quartile	3.80	1.70	2.00
Skew	1.32	1.23	1.31
Mean air temperature (°C)			
Mean	3.46	9.03	7.02
Median	3.55	9.35	7.30
Max	25.25	28.25	27.50
Min	−24.85	−20.00	−23.45
Std Deviation	7.63	7.21	7.79
CV	2.20	0.80	1.11
1st Quartile	−2.45	3.80	1.20
3rd Quartile	9.40	14.70	13.20
Skew	−0.03	−0.13	−0.11
Basin			
Catchment area (km^2)	123532	5304	521
Mean altitude (m.a.s.l.)	17	43.81	470.82

the variables (Fig. 2c).

2.3. Case studies and dataset

This study investigated three different rivers that flow through various areas of Germany (Fig. 3). The first is the Elbe River, which is one of Central Europe's longest rivers, with a significant drainage basin covering parts of the Czech Republic and Germany. The Elbe River has a total length of 1094 km, rising in the Giant Mountains of the northern Czech Republic before flowing through much of Bohemia (The western half of the Czech Republic) and then Germany, reaching the North Sea at Cuxhaven. The streamflow rate data were collected at the Wittenberge station, in northern Germany, approximately 250 km from the mouth of the river, draining a basin of 123532 km^2. Moreover, the data on meteorological variables were collected from the Fichtelberg station, located within the basins of the Elbe River. The streamflow of the Elbe River exhibits higher discharges in the late winter and the early spring. Conversely, during the late summer and early autumn, the river experiences lower flow rates. The Elbe River plays a crucial role in the region's hydrology, serving as a vital waterway for transportation and supporting various economic activities. However, the Elbe faces several challenges, including pollution from industrial and agricultural sources, which impacts water quality and aquatic ecosystems [29]. The second river investigated is the Leine, which flows through Lower Saxony and Thuringia. The streamflow rate data were collected at the Herrenhausen station, in Lower Saxony, approximately 75 km from its mouth into the Aller River to the north, draining a basin of 5304 km^2. Moreover, the data on meteorological variables were collected from the Göttingen station, located within the basins of the Leine River. The Leine exhibits a relatively stable streamflow throughout the year. However, it can experience variations during periods of heavy rainfall or drought. Despite its relatively stable flow regime, the Leine is not immune to water quality problems. Urbanization and agricultural activities affect water clarity and aquatic habitat [30]. The third case study is the Saale River, which originates in the Fichtel Mountains and flows through Thuringia, Saxony-Anhalt, and Saxony. Both streamflow rate and meteorological data were collected at the Hof station, in Upper Franconia, approximately 200 km south of its mouth into the Elbe River in the municipality of Barby, Saxony-Anhalt, draining a basin of 521 km^2. Similar to the Elbe and Leine rivers, the streamflow rate of the Saale River exhibits a comparable pattern with higher discharges during late winter and early spring and lower ones in late summer and early autumn. Additionally, the Saale faces water quality issues, including the presence of microplastic pollution in its waters [31].

It should be noted that the selection of the three watercourses, in addition to being connected to the issues described for each river, is also driven by the desire to test predictive models on watercourses with basins of different sizes, resulting in different hydrological regimes. Indeed, the Elbe at Wittenberge is characterized by a basin approximately 23 times larger than the Leine at Herrenhausen and 240 times larger than the Saale at Hof. Consequently, there are significantly different average flow rates among the rivers, as described in more detail below, along with much more pronounced flow fluctuations for Saale at Hof, with a coefficient of variation (CV) equal to 1.12, compared to Leine at Herrenhausen (CV = 0.86), and even more so for Elbe at Wittenberge (CV = 0.65).

For the modeling, daily data pertaining to streamflow rate, precipitation, and air temperature were considered. Table 1 provides

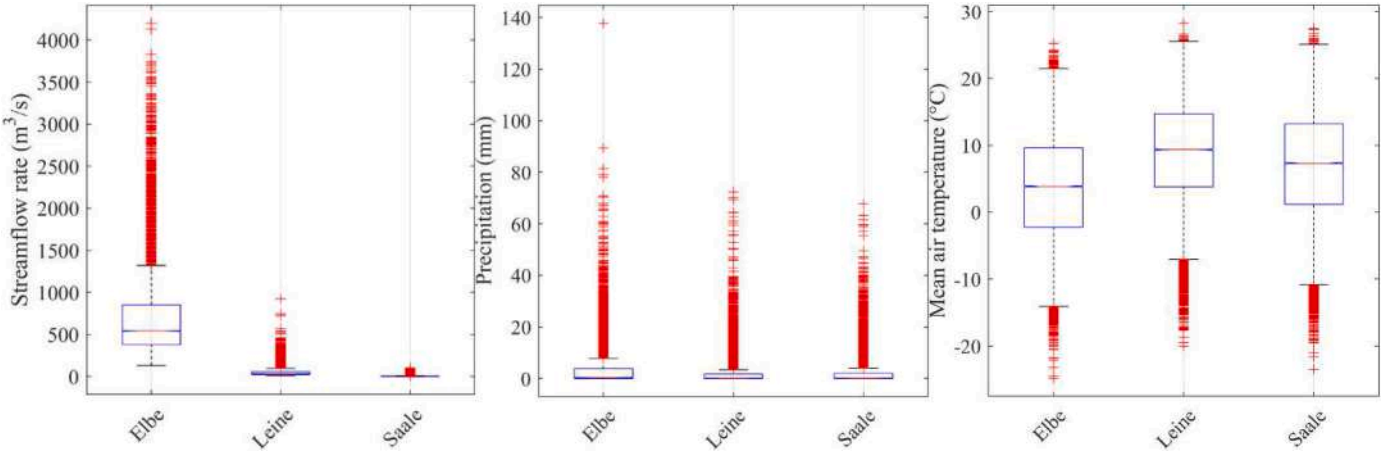

Fig. 4. Box plots of the daily streamflow rate, precipitation and mean air temperature.

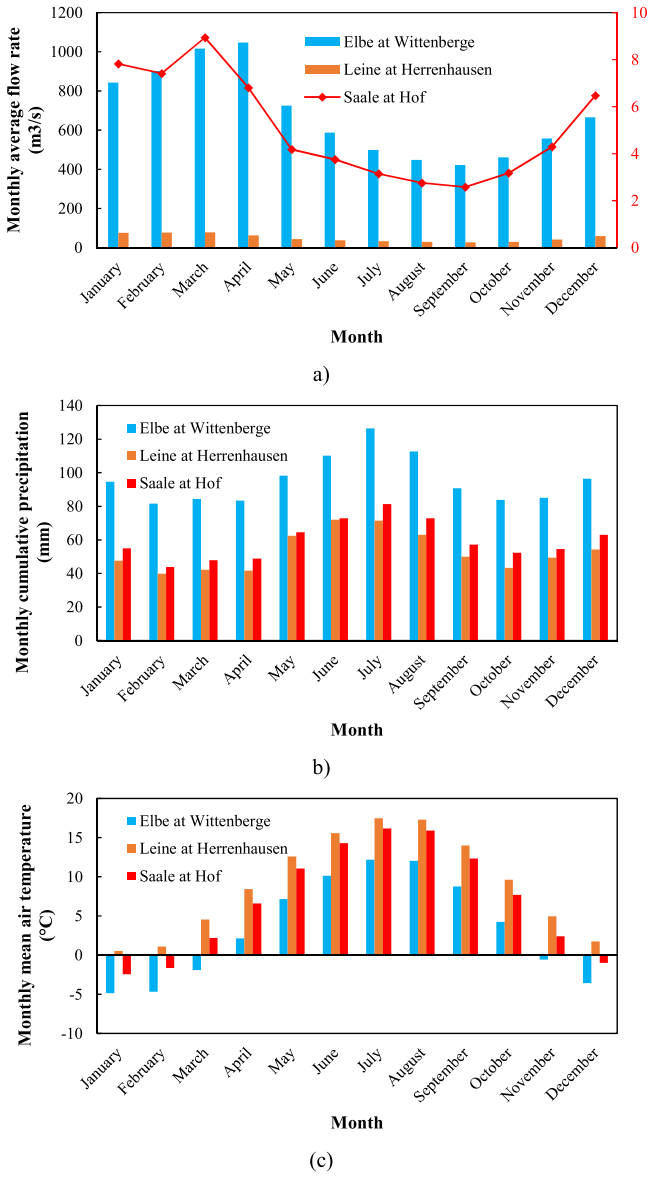

Fig. 5. Monthly average values of: streamflow rate (a), cumulative precipitation (b) and mean air temperature (c).

information on each river's catchment area, recording period, and statistics concerning streamflow rate, precipitation, and air temperature. Moreover, Fig. 4 provides the box plots of the daily streamflow rate, precipitation, and air temperature, while Fig. 5 illustrates the monthly mean values of streamflow rate, cumulative precipitation and air temperature.

As mentioned earlier, the highest monthly average flow rates were recorded in the late winter and early spring and are equal to 1046.97 m^3/s (April), 78.36 m^3/s (March) and 8.95 m^3/s (March) for Elbe at Wittenberge, Leine at Herrenhausen and Saale at Hof, respectively. Conversely, the lowest flow rates were observed in September, with values equal to 422.90 m^3/s, 26.69 m^3/s and 2.58 m^3/s, respectively. Furthermore, the box plots of the daily flow rate showed the highest and lowest interquartile range for Elbe at Wittenberge (470 m^3/s) and Saale at Hof (4.29 m^3/s), respectively.

The lowest monthly cumulative precipitations were observed in February, with values equal to 81.63 mm, 39.81 mm and 43.84 mm for Elbe at Wittenberge, Leine at Herrenhausen and Saale at Hof, respectively. The highest monthly cumulative precipitations were instead observed in summer, with values equal to 126.35 mm (July), 71.99 mm

Table 2
Evaluation metrics.

Coefficient of determination Provides a statistical measure that quantifies the goodness of fit of a regression model to experimental data. It ranges between 0 (the model does not predict the outcome) to 1 (the model perfectly predicts the outcome).	$R^2 = 1 - \frac{\sum_{i=1}^{n}(Q_P^i - Q_M^i)^2}{\sum_{i=1}^{n}(\overline{Q_M} - Q_M^i)^2}$	(2)
Kling Gupta Efficiency Dimensionless metric depending on correlation coefficient and on the ratio of standard deviation and mean of the observed and predicted streamflow rate. In contrast to R^2, which solely considers the relative magnitude of errors, KGE penalizes models that either overestimate or underestimate the mean or variability of observed values. It ranges between -∞ (the model does not predict the outcome) to 1 (the model perfectly predicts the outcome).	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_P}{\sigma_M} - 1\right)^2 + \left(\frac{\overline{Q_P}}{\overline{Q_M}} - 1\right)^2}$	(3)
Root Mean Square Error The root of the total squared error between the predicted and measured streamflow rate, normalized by the number of samples. It ranges between 0 and +∞, with lower values indicating more accurate models.	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_P^i - Q_M^i)^2}{s}}$	(4)
Relative error The ratio between the difference of predicted and measured GWL and the measured GWL. Range: 0 – ∞, with lower values indicating more accurate models.	$Relative\ error = \frac{Q_P^i - Q_M^i}{Q_M^i}$	(5)
Mean Absolute Error The absolute error between the predicted and measured streamflow rate, normalized by the number of samples. It ranges between 0 and +∞, with lower values indicating more accurate models.	$MAE = \frac{\sum_{i=1}^{n}\lvert Q_P^i - Q_M^i\rvert}{s}$	(6)
Mean Absolute Percentage Error The relative error between the predicted and measured streamflow rate, normalized by the number of samples. It ranges between 0 and +∞, with lower values indicating more accurate models.	$MAPE = \frac{\sum_{i=1}^{n}\left\lvert\frac{Q_P^i - Q_M^i}{Q_M^i}\right\rvert}{s}$	(7)

(June) and 81.44 mm (July), respectively. The box plots of the daily precipitation showed the highest and lowest interquartile range for Elbe at Wittenberge (3.8 m^3/s) and Leine at Herrenhausen (1.7 m^3/s), respectively.

Moreover, the lowest and highest monthly mean air temperatures were recorded in January and July, respectively. In particular, temperatures below 0 °C were observed in January for Elbe at Wittenberge and Saale at Hof, and slightly above 0 °C for Leine at Herrenhausen. In July, the average temperatures were around 12 °C for Elbe at Wittenberge and 16–17 °C for Leine at Herrenhausen and Saale at Hof. The box plots of the daily mean air temperature showed the highest and lowest interquartile range for Saale at Hof (12 °C) and Leine at Herrenhausen

Table 3
Optimal input variables for the different case studies.

Input variables	Elbe at Wittenberge	Leine at Herrenhausen	Saale at Hof
Number of lagged values of daily flow rates	20	30	12
Number of lagged values of daily cumulative rainfall	7	7	5
Number of lagged values of daily average temperature	7	7	5

(10.9 °C), respectively.

2.4. Evaluation metrics

The accuracy of the forecast models was assessed using six different evaluation metrics: the coefficient of determination (R^2), the Kling Gupta Efficiency (KGE), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Percentage Absolute Error (MAPE). A description of the evaluation metrics is provided in Table 2.

where Q_M^i = measured streamflow rate for the ith data and Q_P^i = predicted streamflow rate for the ith data, $\overline{Q_M}$ = mean of the measured streamflow rate, $\overline{Q_P}$ = mean of the predicted streamflow rate, r = correlation coefficient between measured and predicted streamflow rate, σ_P = standard deviation of the predicted streamflow rate, σ_M = standard deviation of the measured streamflow rate, s = number of samples.

2.5. Input selection and model optimization

The study's predictors were lagged flow rate values, precipitation, and temperature. The optimal number of them was chosen through a BO method, simultaneously with the hyperparameters of the algorithms. BO is a procedure for optimizing functions with a complex structure. It iteratively updates a probabilistic model based on the acquired data to understand the characteristics of the function, pursuing a balance between exploring new search areas and exploiting regions that can provide high performance, allowing the global optimum to be found with minimal evaluations of the objective function. Please refer to the relevant literature for a more detailed explanation of this method [32]. It is essential to highlight that the optimization of the number of lagged values and hyperparameters of the model was explicitly done for 1-day-ahead forecasts, while for extended forecast horizons, previously determined values were used (Table 3). In order to evaluate the impact of the exogenous variables precipitation and temperature, alternative models based only on the lagged values of flow rates have also been developed.

The models were trained using 80 % of the time series data, and the remaining 20 % was reserved for testing. Since this is a time series prediction problem, keeping temporal continuity in the training intervals was crucial. Input values were normalized according to the equation:

$$x_{t\_norm} = \frac{x_t - x_{min}}{x_{max} - x_{min}} \tag{8}$$

For multi-step ahead prediction, a recursive methodology was employed. This means that the predictions from previous steps were used to generate new lagged flow rates for forecasting future flow rate values. Since the future values are unknown during the prediction process, the earlier predicted values serve as reasonable estimates of the actual values.

In addition, in order to assess the impact of exogenous variables on the accuracy of the forecast models, a comparison was made between

Table 4
Evaluation metrics – Training stage.

		AR – RBF						MLP – RF – PR					
		Model A			Model B			Model A			Model B		
		1 day	3 days	7 days	1 day	3 days	7 days	1 day	3 days	7 days	1 day	3 days	7 days
Elbe at Wittenberge	R^2	0.998	0.974	0.883	0.997	0.963	0.788	0.998	0.975	0.888	0.998	0.967	0.811
	KGE	0.998	0.975	0.887	0.997	0.973	0.868	0.997	0.978	0.915	0.994	0.964	0.870
	RMSE (m^3/s)	21.07	69.40	137.01	23.12	83.62	189.48	21.09	68.63	138.69	21.67	77.95	176.18
	MAE (m^3/s)	11.82	39.04	83.55	12.90	46.90	116.04	11.91	38.45	82.06	12.34	44.37	110.56
	MAPE (%)	1.80	5.32	10.60	1.94	6.35	14.82	1.79	5.20	10.44	1.85	5.99	14.15
Leine at Herrenhausen	R^2	0.931	0.864	0.798	0.927	0.741	0.482	0.974	0.919	0.859	0.960	0.790	0.532
	KGE	0.952	0.880	0.816	0.935	0.786	0.555	0.968	0.922	0.871	0.952	0.819	0.619
	RMSE (m^3/s)	11.34	14.86	17.14	11.43	19.11	22.11	6.92	11.86	14.93	8.52	17.59	22.34
	MAE (m^3/s)	3.54	6.74	9.13	4.34	10.12	15.56	2.77	5.25	7.41	3.55	8.87	15.01
	MAPE (%)	6.14	10.81	14.62	7.41	15.85	24.08	4.84	8.40	11.89	6.15	13.93	24.34
Saale at Hof	R^2	0.927	0.846	0.783	0.810	0.515	0.324	0.941	0.860	0.779	0.848	0.562	0.353
	KGE	0.936	0.846	0.792	0.825	0.582	0.406	0.946	0.872	0.809	0.838	0.616	0.471
	RMSE (m^3/s)	1.50	2.01	2.27	2.19	2.87	2.90	1.35	1.97	2.35	1.97	2.78	3.12
	MAE (m^3/s)	0.57	0.91	1.15	0.89	1.65	2.22	0.50	0.84	1.12	0.77	1.53	2.24
	MAPE (%)	10.36	15.96	20.45	14.28	26.80	39.15	9.09	14.59	19.65	12.36	24.64	41.15

Table 5
Evaluation metrics – Testing stage.

		AR – RBF						MLP – RF – PR					
		Model A			Model B			Model A			Model B		
		1 day	3 days	7 days	1 day	3 days	7 days	1 day	3 days	7 days	1 day	3 days	7 days
Elbe at Wittenberge	R^2	0.998	0.971	0.871	0.995	0.943	0.708	0.996	0.958	0.840	0.996	0.950	0.776
	KGE	0.989	0.939	0.806	0.994	0.969	0.793	0.983	0.931	0.840	0.979	0.916	0.791
	RMSE (m^3/s)	21.98	73.18	134.91	30.59	109.68	278.47	26.53	87.67	157.22	27.59	93.81	179.15
	MAE (m^3/s)	12.02	39.59	82.39	15.21	56.94	146.52	13.23	43.75	89.90	14.37	51.42	116.50
	MAPE (%)	2.02	5.77	10.64	2.39	7.95	18.71	2.14	6.35	12.51	2.31	7.46	16.18
Leine at Herrenhausen	R^2	0.970	0.883	0.786	0.946	0.731	0.457	0.967	0.893	0.811	0.954	0.747	0.485
	KGE	0.940	0.813	0.652	0.967	0.843	0.671	0.963	0.895	0.842	0.948	0.784	0.528
	RMSE (m^3/s)	5.98	10.46	11.84	8.28	17.87	25.49	6.36	10.87	13.89	7.45	15.35	18.43
	MAE (m^3/s)	2.93	6.05	9.05	3.88	9.26	14.61	2.87	5.58	7.91	3.83	10.53	20.35
	MAPE (%)	5.49	10.18	14.44	7.36	16.13	24.52	5.46	9.98	14.51	8.35	24.30	56.29
Saale at Hof	R^2	0.939	0.836	0.772	0.844	0.538	0.327	0.930	0.814	0.703	0.838	0.532	0.350
	KGE	0.913	0.796	0.714	0.864	0.638	0.433	0.859	0.719	0.621	0.864	0.679	0.545
	RMSE (m^3/s)	1.19	1.74	1.88	1.85	2.76	2.87	1.19	1.67	1.88	1.88	2.96	3.73
	MAE (m^3/s)	0.49	0.84	1.11	0.74	1.44	1.98	0.51	0.91	1.29	0.78	1.63	2.54
	MAPE (%)	8.99	15.61	22.43	11.88	23.53	33.46	9.68	19.39	32.01	13.19	30.60	55.13

the forecasts made by AR-RBF and MLP-RF-PR considering lagged flow rate, precipitation, and temperature values (referred to as Model A) with those obtained with only lagged flow rate (referred to as Model B).

3. Results

Tables 4 and 5 provide the evaluation metrics for the training and testing stages, respectively, including all models and forecast horizons. A representation of the predictions made for the three rivers with both AR-RBF and MLP-RF-PR with Model A and during the testing stage is provided in Figs. 6–8.

The Elbe at Wittenberge exhibited the highest mean streamflow rate among the three rivers. During the testing stage, with Model A, the AR-RBF model outperformed the MLP-RF-PR model for all forecast horizons (Fig. 6). The most accurate predictions were observed for the shortest forecast horizon (1 day) with the stacked AR-RBF model (R^2 = 0.998, KGE = 0.989, RMSE = 21.98 m^3/s, MAPE = 2.02 %) slightly outperforming the MLP-RF-PR model (R^2 = 0.996, KGE = 0.983, RMSE = 26.53 m^3/s, MAPE = 2.14 %). A performance reduction as the forecasting horizon increases was observed. However, for a forecast horizon of 3 days, both AR-RBF (R^2 = 0.971, KGE = 0.939, RMSE = 73.18 m^3/s, MAPE = 5.77 %) and MLP-RF-PR (R^2 = 0.958, KGE = 0.931, RMSE = 87.67 m^3/s, MAPE = 6.35 %) models showed a good accuracy. Moving from a 3-day forecast horizon to a 7-day forecast horizon, both models showed a more pronounced worsening of forecasts than moving from 1 to 3 days, with, however, AR-RFB (R^2 = 0.871, KGE = 0.806, RMSE = 134.91 m^3/s, MAPE = 10.64 %) that still outperforming MLP-RF-PR (R^2 = 0.840, KGE = 0.840, RMSE = 157.22 m^3/s, MAPE = 12.51 %) for all metrics with the exception of the only KGE.

Overall, the decrease in performance of both models with increasing forecast horizons was primarily linked to a higher tendency to underestimate streamflow rates, especially during flood peaks. This can be observed in Fig. 6 by the significant disparity between the measured and predicted flow values in the time series plots (left side) and the notable scattering of data points away from the 1:1 line in the scatter plots (right side).

For the second river investigated, the Leine at Herrenhausen, the performance of the AR-RBF and MLP-RF-PR models during the testing stage (Model A) was quite similar to each other (Fig. 7). In particular, the AR-RBF model showed better values of RMSE (AR – RBF – RMSE = 5.98 m^3/s – 11.84 m^3/s, MLP-RF-PR – RMSE = 6.36 m^3/s – 13.89 m^3/s) for all forecasting horizon, while the MLP-RF-PR highlighted better KGE (AR – RBF – KGE = 0.940–0.652, MLP-RF-PR – KGE = 0.963–0.842) and MAE (AR – RBF – MAE = 2.93 m^3/s – 9.05 m^3/s, MLP-RF-PR – MAE = 2.87 m^3/s – 7.91 m^3/s). Similar to the Elbe at Wittenberge, both models exhibited underestimation of peak streamflows when predicting 7 days ahead.

The third case study, the Saale at Hof, exhibited the lowest mean streamflow rate among the three rivers. As for the Elbe River at Wittenberge, also in this case, during the testing stage (Model A), the AR-RBF model outperformed the MLP-RF-PR model for all forecast horizons (Fig. 8). However, as observed for the other two rivers, the underestimation of the streamflow rate became increasingly pronounced as the forecast horizon increased, with R^2 values decreasing from 0.939 for AR-RBF and 0.930 for MLP-RF-PR, for 1-day ahead predictions, to 0.772 for AR-RBF and 0.703 for MLP-RF-PR, for 7-days ahead predictions.

Regarding Model B, which includes only lagged flow rates as input, it was consistently outperformed by Model A across all forecast horizons, both for AR-RBF and MLP-RF-PR. Furthermore, as the forecast horizon increased, Model B exhibited a significantly sharper decline in performance compared to Model A. Specifically, for the Elbe at Wittenberge, both AR-RBF and MLP-RF-PR yielded R^2 values lower than 0.8. For the

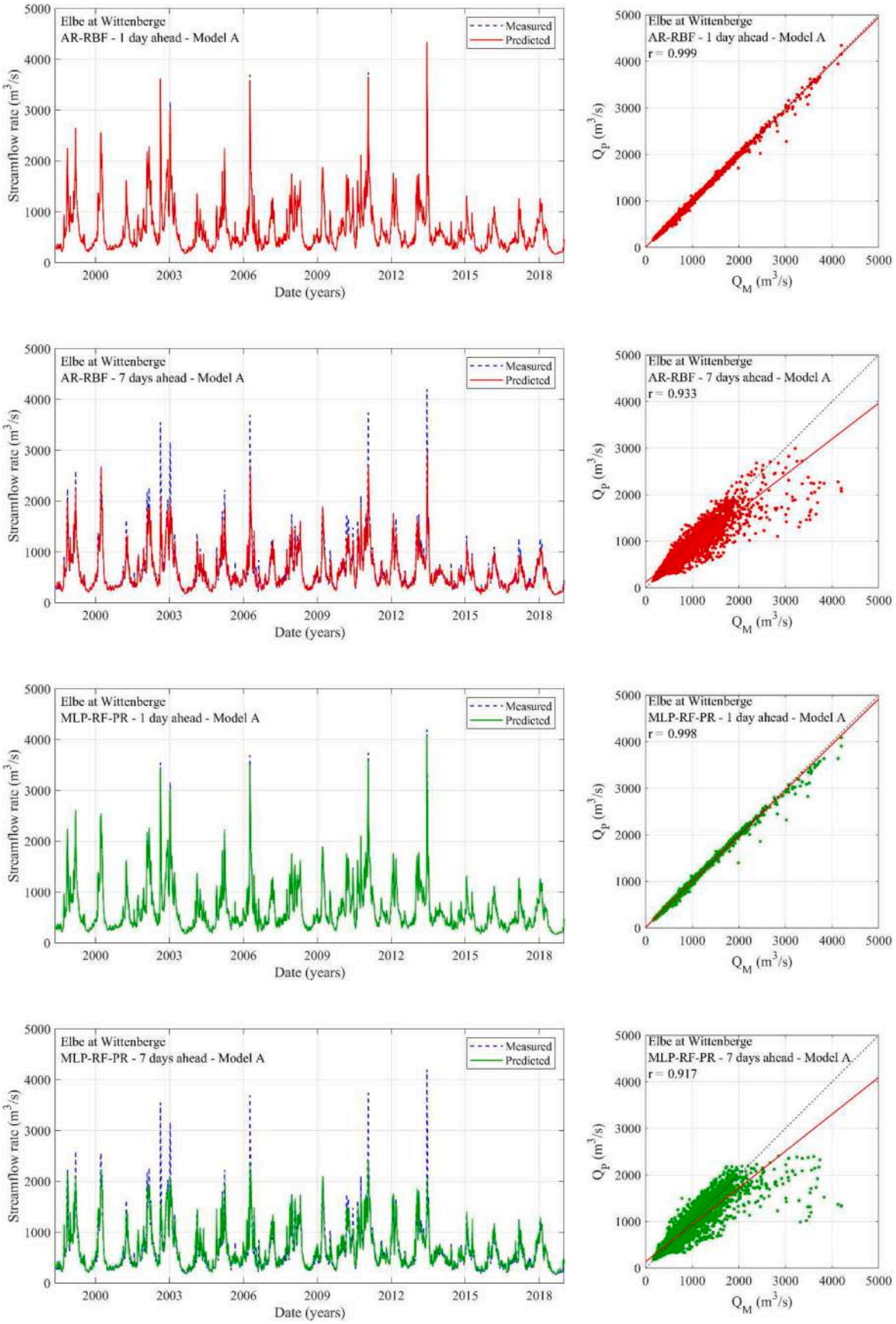

Fig. 6. Streamflow rate prediction for Elbe at Wittenberge – Model A – Testing stage.

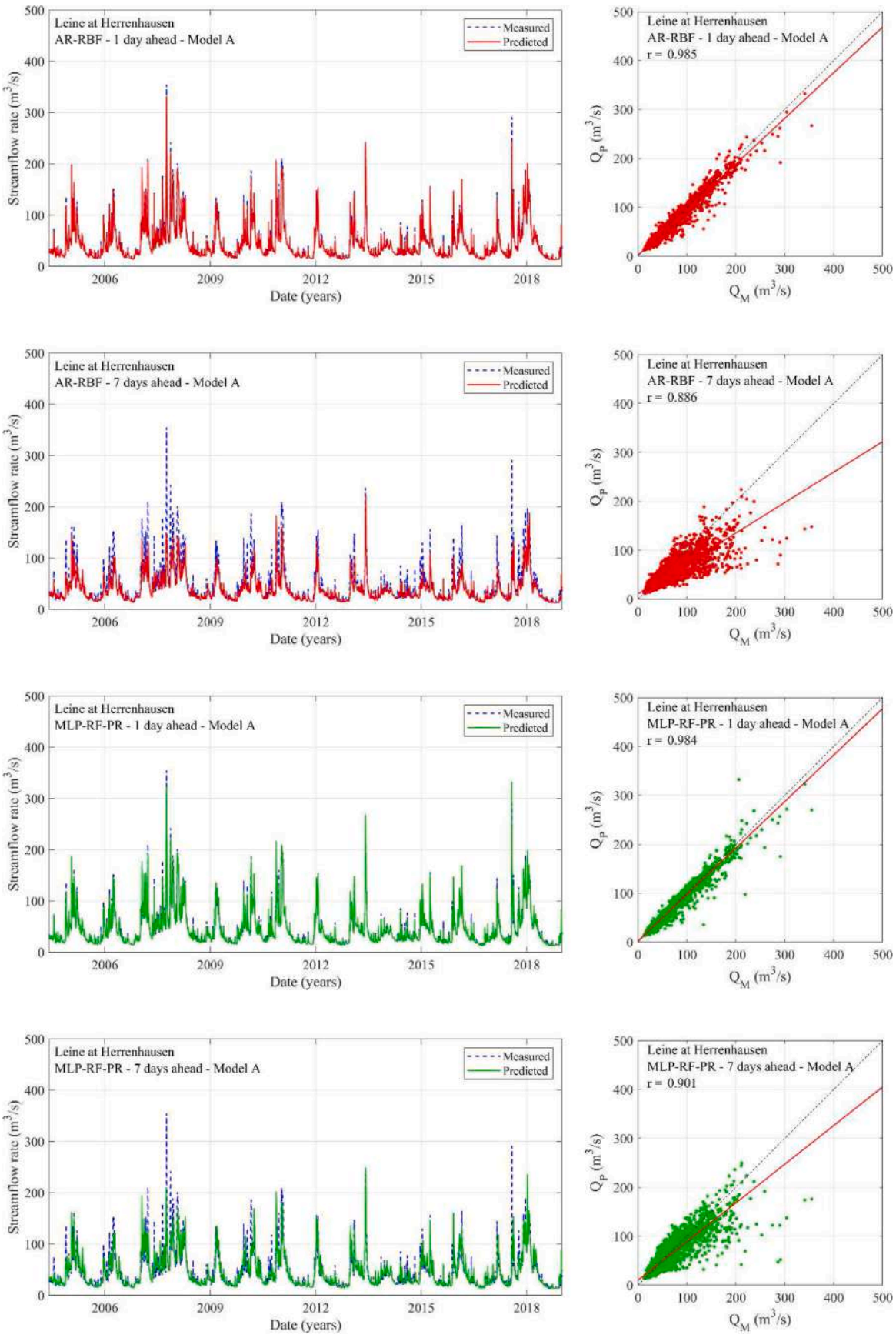

Fig. 7. Streamflow rate prediction for Leine at Herrenhausen – Model A – Testing stage.

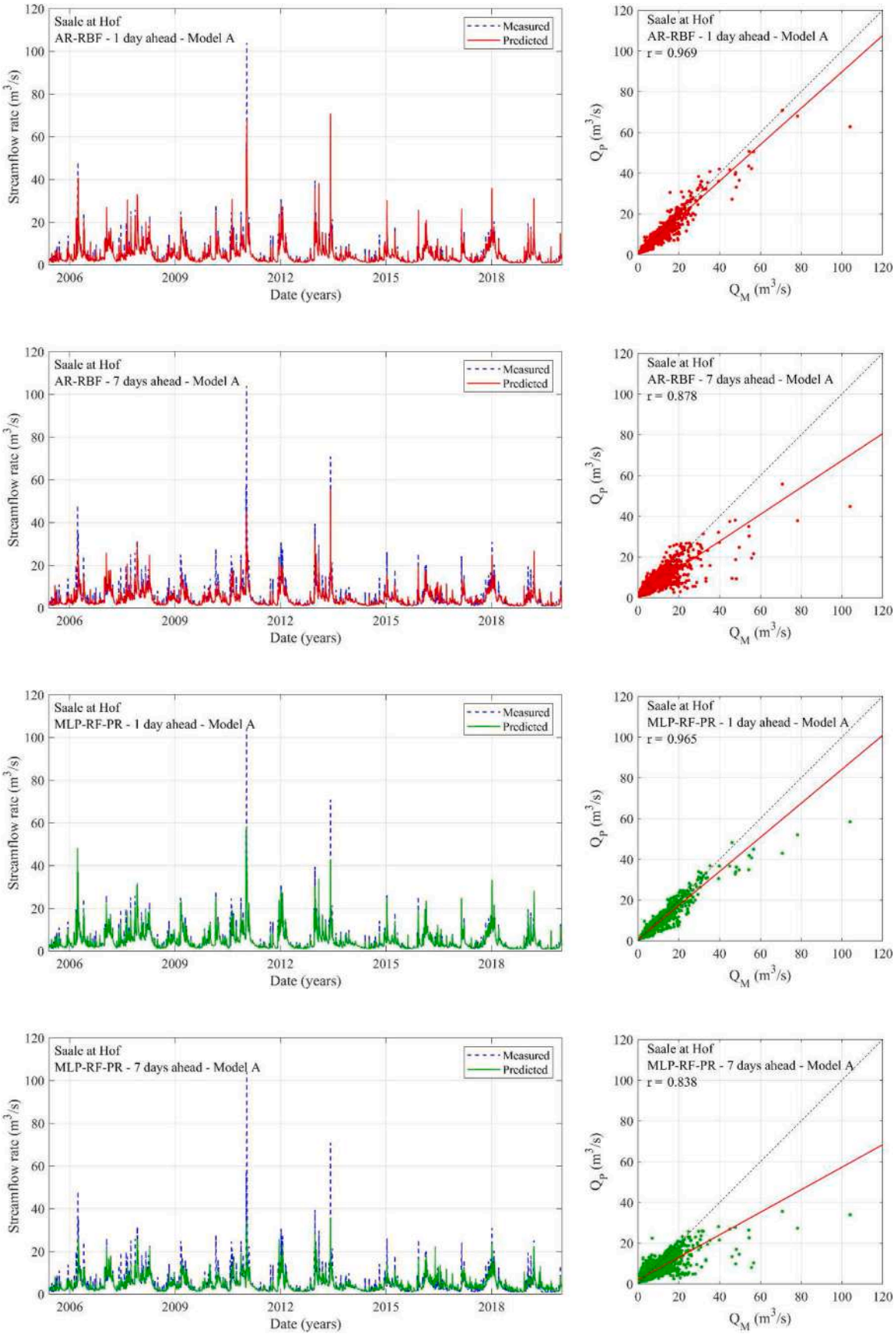

Fig. 8. Streamflow rate prediction for Saale at Hof – Model A – Testing stage.

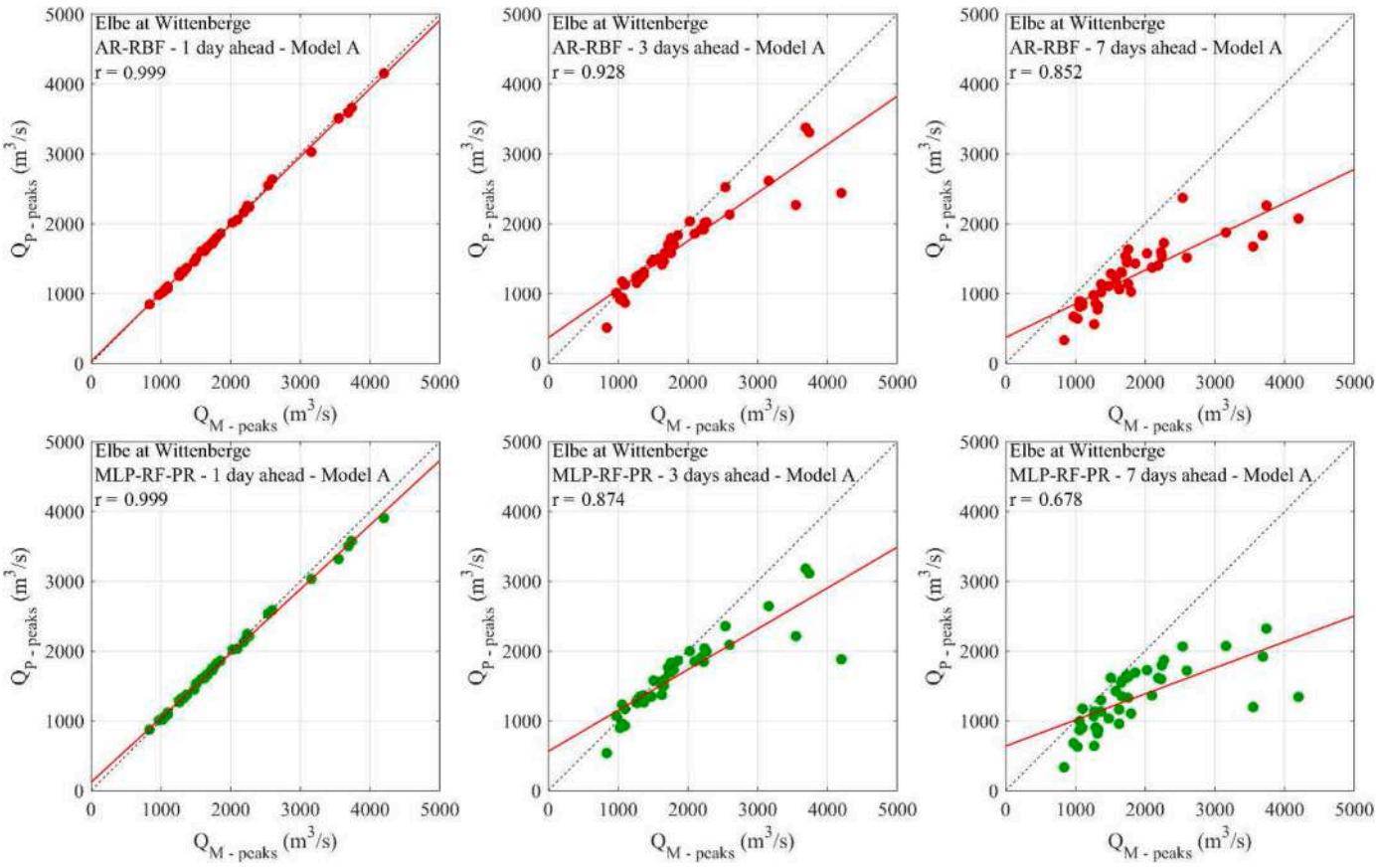

Fig. 9. Measured versus predicted streamflow rate peaks – Elbe at Wittenberge – Model A – Testing stage.

other two rivers, Leine at Herrenhausen and Saale at Hof, R^2 values below 0.5 and 0.4 were obtained, respectively. This indicates a much more pronounced underestimation of the streamflow rate as the forecast horizon extends when compared to Model A.

Additionally, Figs. 9–11 illustrate the comparison between the measured and predicted streamflow rate peaks during the testing stage, with Model A, for the three rivers. In the case of the Elbe at Wittenberge, both models effectively reproduced peak values in the 1-day-ahead prediction, with r values up to 0.999. Nevertheless, their performance gradually deteriorated as the forecasting horizon extended to 3- days- (AR-RBF – r = 0.928, MLP-RF-PR – r = 0.874) and 7-days-ahead predictions (AR-RBF – r = 0.852, MLP-RF-PR – r = 0.678), with a tendency to increasingly underestimate peaks (as illustrated in Fig. 9). The description above is similarly applicable to the case study at the Leine at Herrenhausen (Fig. 10) and Saale at Hof (Fig. 11). In particular, for Leine at Herrenhausen, the r values ranged between 0.867 (1 day ahead) and 0.502 (7 days ahead) for AR-RBF and between 0.845 (1 day ahead) and 0.372 (7 days ahead) for MLP-RF-PR. For Saale at Hof, a less marked reduction in r was observed compared to Leine at Herrenhausen, with r that ranged between 0.877 (1 day ahead) and 0.746 (7 days ahead) for AR-RBF and between 0.869 (1 day ahead) and 0.651 (7 days ahead) for MLP-RF-PR. However, also in this case, a tendency to increasingly underestimate peaks was observed.

Overall, the performance of the two models in these two case studies was poorer compared to the Elbe at Wittenberge case study, especially for Leine at Herrenhausen. Moreover, the AR-RBF model exhibited superior performance compared to the MLP-RF-PR model in predicting streamflow rate peak for all three rivers.

Fig. 12 shows the R^2 and MAPE values computed for the streamflow rate peaks, for both AR-RBF and MLP-RF-PR models and for all forecasting horizons. A decrease in R^2 and the increase in MAPE across all cases as the forecasting horizon increases was observed, indicating the challenging nature of longer forecasting horizons predictions for peak streamflow. The outcomes of the Elbe at Wittenberge case study revealed the highest R^2 and lowest MAPE among the three rivers. In particular, the AR-RBF model led to R^2 values between 0.999 (1 day ahead) and 0.727 (7 days ahead) and MAPE values between 1.09 % (1 day ahead) and 30.05 % (7 days ahead), while the MLP-RF-PR model exhibited R^2 values between 0.998 (1 day ahead) and 0.459 (7 days ahead) and MAPE values between 1.58 % (1 day ahead) and 25.45 % (7 days ahead). This suggests that both models performed exceptionally well for this specific river, which also showed the highest values of mean streamflow rate and the lowest CV and Skew (see also Table 1). Discrepancies between both R^2 and MAPE were observed for the Leine at Herrenhausen and Saale at Hof. For Saale at Hof (AR-RBF – R^2 between 0.556 and 0.769, MLP-RF-PR – R^2 between 0.424 and 0.756), most of the R^2 values were better than those computed for Leine at Herrenhausen (AR-RBF – R^2 between 0.252 and 0.751, MLP-RF-PR – R^2 between 0.139 and 0.715). Simultaneously, the MAPE values were lower for Leine at Herrenhausen (AR-RBF – MAPE between 12.49 % and 43.29 %, MLP-RF-PR – MAPE between 12.39 % and 27.71 %) compared to Saale at Hof (AR-RBF – MAPE between 22.24 % and 40.72 %, MLP-RF-PR – MAPE between 24.67 % and 45.85 %). This complexity in model evaluation prompts a consideration of the modeler's objectives when selecting an appropriate evaluation metric that aligns with their intended purposes.

A comprehensive assessment of the streamflow rate forecasting accuracy of AR-RBF and MLP-RF-PR was carried out using three Taylor diagrams, as depicted in Fig. 13, for the three rivers and for both models A and B. Taylor diagrams were constructed based on the geometric relationship between the correlation coefficient r, standard deviation, and RMSE. The Taylor representation emphasizes both the accuracy and efficiency of the tested models concerning the observed values, following the approach proposed by Kim et al. [33]. Due to its effectiveness, the Taylor diagram has gained widespread use in hydrological

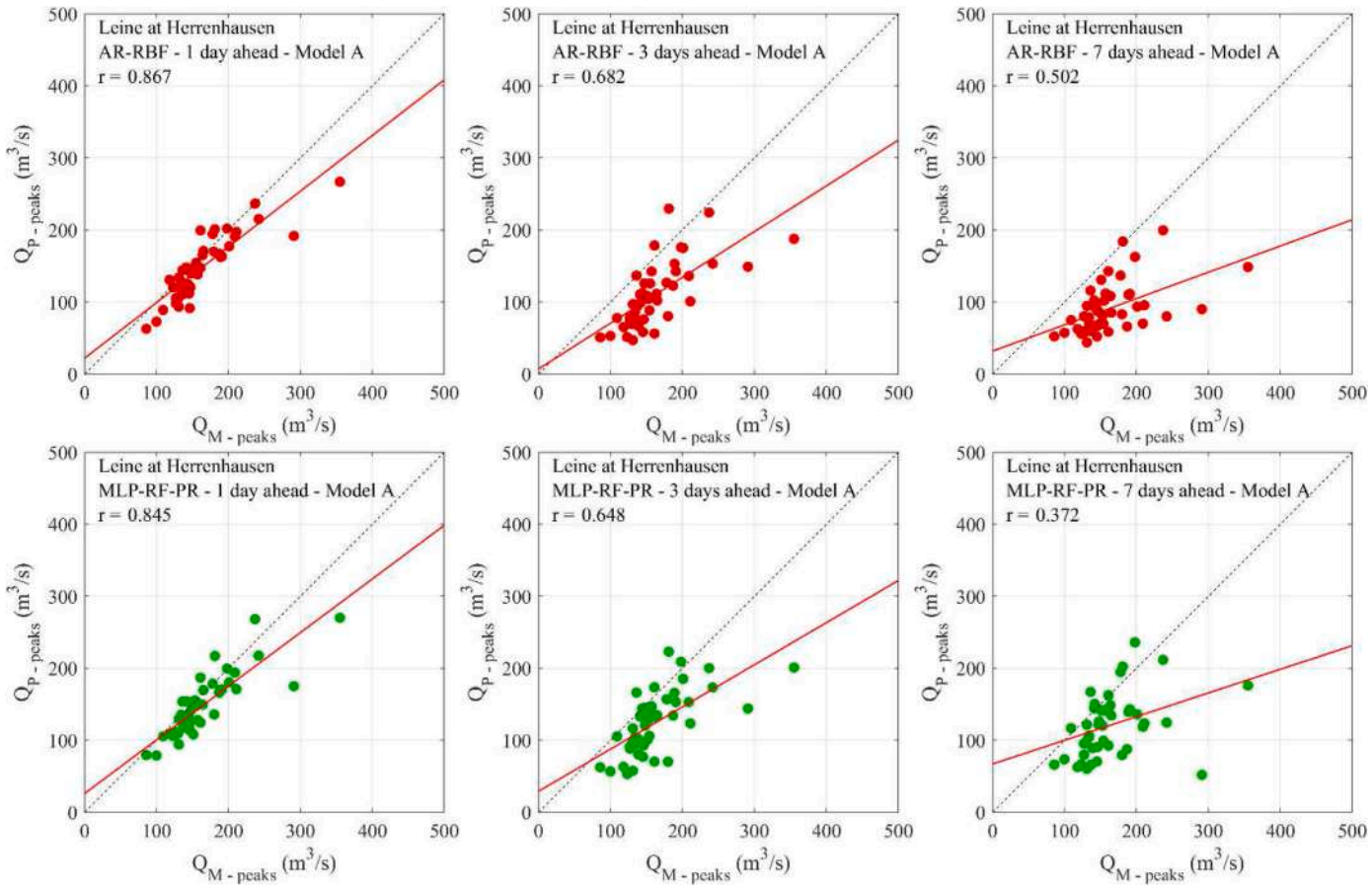

Fig. 10. Measured versus predicted streamflow rate peaks – Leine at Herrenhausen – Model A – Testing stage.

studies [34,35].

As for the Elbe at Wittenberge, it is remarkably evident to observe the distinct separation and gradual divergence of the measured points within the three clusters: circles, squares, and triangles, corresponding to 1-, 3-, and 7-day-ahead predictions. This increasing divergence of the measured points reaffirms the accuracy of the statement that model performance deteriorates with the increase in lead time across all cases and models. Each shape (circle, square, and triangle) is associated with four points representing both AR-RBF and MLP-RF-PR and the two combinations of predictors (Model A and Model B). All the models within Model A demonstrate superior performance over those in Model B, a point previously discussed in the above paragraphs. Within Model A, AR-RBF (orange) consistently surpasses MLP-RF-PR (green) across all lead times of 1-, 3-, and 7-day-ahead predictions (Fig. 13). Conversely, in Model B, MLP-RF-PR (purple) slightly outperforms AR-RBF (red), suggesting that MLP-RF-PR might be better suited for time series forecasting without the further predictors.

As for the Leine at Herrenhausen, the clustering of square and triangle shapes is not as distinct as observed in the case study at the Elbe at Wittenberge. Evidence of this lies in a triangle (green) being closer to the measured point in comparison to two squares (red and purple). This indicates that MLP-RF-PR-7-day-ahead-Model A performed even better than AR-RBF-3-day-ahead-Model B. There is only one exception, other statements related to the performance of models and scenarios at the Elbe at Wittenberge also apply to the Leine at Herrenhausen.

At the Saale at Hof, the clusters are also not distinctly separated. There are two triangle shapes (orange and green) that are closer to the measured point in comparison to two squares (red and purple). This indicates that the 7-day-ahead predictions by Model A are even superior to those of the 3-day-ahead predictions by Model B. This suggests that streamflow forecasting for this river is more intricate, with exogenous variables (precipitation and temperature) playing a significant role.

Moreover, a combined box and violin plots representation [36] of the relative error evaluated for the testing stage, Model A, and for both AR-RBF and MLP-RF-PR, is provided in Fig. 14. For the Elbe at Wittenberge, the median values of the relative error were close to 0 for the AR-RBF model for all forecasting horizons while ranged between 0 and 0.05, increasing with the forecasting horizon, for the MLP-RF-PR model. A positive median indicates a slight overestimation of the streamflow rate. Moreover, for both models, an increase of the interquartile range (IQR) as the forecasting horizon increases was observed, ranging from 0.03 to 0.16 for AR-RBF and between 0.03 and 0.18 for MLP-RF-PR. For the Leine at Herrenhausen, the median values of the relative error were negative for AR-RBF, ranging between −0.01 (forecasting horizon equal to 1 day) and −0.05 (7 days), indicating a slight underestimation of the streamflow rate, while were positive for MLP-RF-PR, ranging between 0.01 (1 day) and 0.05 (7 days). However, IQRs were wider than those estimated for Elbe at Wittenberge, ranging between 0.06 (1 day) and 0.20 (7 days) for both models. For the Saale at Hof, AR-RBF model showed a median close to 0 for both 1-day-ahead and 3-days-ahead forecasting horizon while showed a slight reduction, with a median equal to −0.02, for a 7-days-ahead forecasting horizon, indicating a slight tendency to underestimate the streamflow rate. Conversely, MLP-RF-PR model showed positive median values, ranging between 0.02 (1 day) and 0.07 (7 days). The IQRs were wider than those computed for both Elbe at Wittenberge and Leine at Herrenhausen, ranging between 0.11 (1 day) and 0.30 (7 days) for AR-RBF and between 0.10 (1 day) and 0.35 (7 days) for MLP-RF-PR. In addition, the number of outliers also increased from a few outliers in the Elbe at Wittenberge case study to many outliers in the Saale at Hof case study, with the Leine at Herrenhausen that represents an intermediate situation.

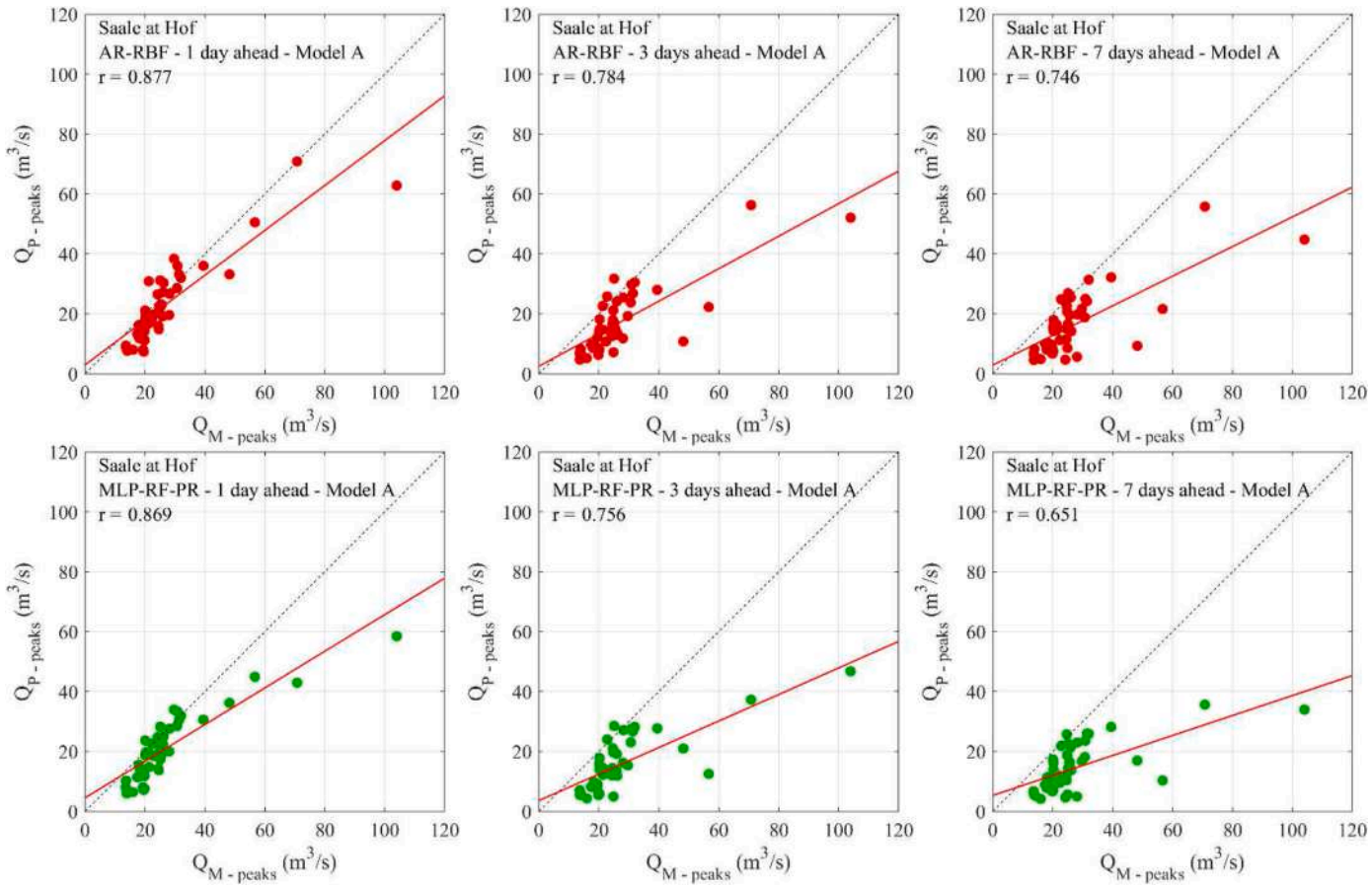

Fig. 11. Measured versus predicted streamflow rate peaks – Saale at Hof – Model A – Testing stage.

4. Discussion

This study proposed a comparison of two ensemble ML models, AR-RBF and MLP-RF-PR, in providing short-term (1–3 days) and medium-term (7 days) streamflow rate predictions, for three rivers located in Germany. Across all three rivers, forecasts with a 1-day-ahead horizon exhibited highly favorable outcomes. The predictive performance of the models showed a gradual decline as the forecasting horizon. This pattern is commonly observed and has been mentioned in some previously published research papers [12,15]. This aspect can be correlated with the inherent challenges associated with making predictions over significant time horizons for complex hydrological systems. Nevertheless, the results for 3-day-ahead and 7-day-ahead predictions remain within an acceptable range in this study.

Among the three rivers investigated, the most accurate predictions were obtained for the Elbe at Wittenberge. This result can be attributed mainly to the lower flow variability that characterizes the Elbe river compared to the other two, with lower CV and skew values. This promotes more accurate prediction of flow rates. For the Leine at Herrenhausen the second-best predictions were obtained, while for the Saale at Hof, the least accurate results were obtained. The latter river also showed the lowest mean discharge and highest CV, indicating significant variability in streamflow, which makes forecasting a more challenging task. In the context of predicting peak flood values, which represents a complex issue of considerable practical interest [37], AR-RBF consistently outperformed MLP-RF-PR across all rivers and scenarios examined in this study. As a result, for applications that prioritize the precision of high-flow predictions, AR-RBF emerges as a more viable choice and should be taken into consideration.

The selection of lagged values holds paramount importance and significantly influences model performance. While the process may be time-intensive, it is imperative to dedicate effort to identifying suitable predictive models. Within this study, the optimal lagged values for streamflow were determined as 20, 30, and 12 for Elbe at Wittenberge, Leine at Herrenhausen, and Saale at Hof, respectively. Similarly, those values are 7, 7, and 5 for temperature and precipitation. These chosen lagged values demonstrated the prolonged influence of antecedent streamflow compared to precipitation and temperature. This finding underscores that historical streamflow carries a more enduring impact, emphasizing the necessity of capturing long-term patterns when constructing predictive models. Some previous studies have used only three lagged values [38].

Regarding the selection of the exogenous inputs, in the present study, only precipitation and temperature were used as predictors (Model A). The predictions made with Model A were also compared with those obtained including only the lagged values of streamflow rate (Model B), with Model A outperforming Model B for both AR-RBF and MLP-RF-PR and all forecasting horizons. Moreover, the performance gap between models A and B became increasingly pronounced as the forecasting horizon extended, growing significantly from 1-day-ahead to 7-day-ahead predictions.

In the context of applying hybrid ML models for streamflow forecasting, Li et al. [39] developed hybrid models specifically tailored for the Yuetan Basin in China. Among their models, the one built upon Particle Swarm Optimization and Support Vector Regression (PSO-SVR) yielded the most accurate results, showcasing a Nash-Sutcliffe Efficiency score (which has a mathematical expression almost identical to R^2) of 0.82, which is lower than the R^2 values computed, for the same forecasting horizon equal to 1 day, in the present study. The super ensemble model proposed by Tyralis et al. [14], based on 10 different machine learning algorithms, for a one-step-ahead daily streamflow forecasting in several USA basins led to R^2 values in the range of 0.60–0.65. Lee and

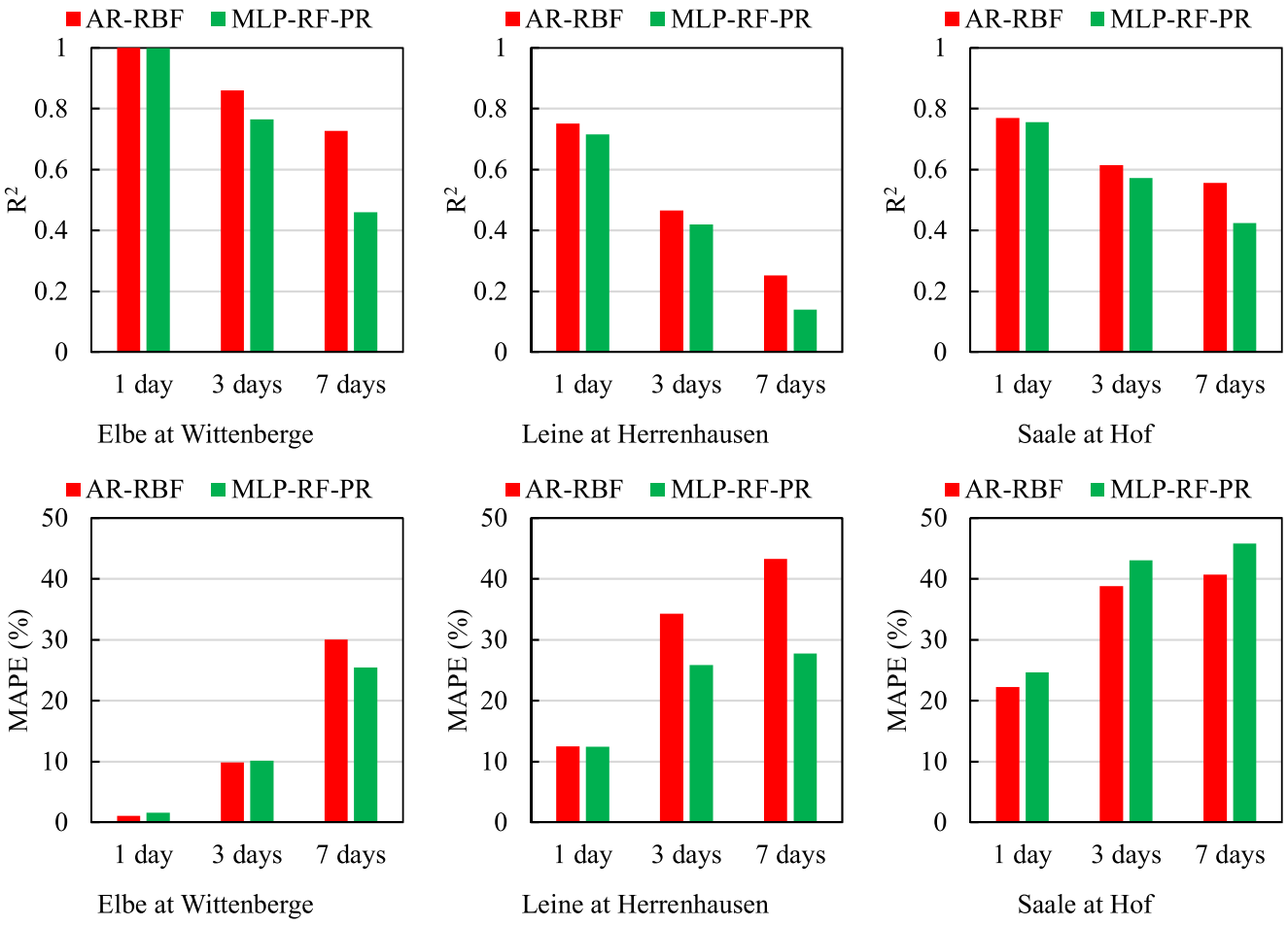

Fig. 12. R^2 and MAPE values computed for the streamflow rate peaks – Model A.

Ahn [40] also proposed a stacking model based on four ML algorithms: Support Vector Machine (SVM), Gradient boosting machine (GBM), Cubist, and Bayesian Regularized Neural Networks (BRNN), for the short and medium-term (1–7 days) streamflow rate prediction in South Korea. The authors computed values of NSE up to 0.48, showing a performance reduction as the forecast horizon increased, as observed in the present study. Further comparison can be drawn with the previously mentioned study by Granata et al. [15]. In this study, for 1-day-ahead streamflow rate prediction, R^2 values between 0.80 and 0.92 with both Stacked RF-MLP and Bi-LSTM models. However, they also showed a reduction in prediction accuracy as the forecast horizon increased, already for the 3-days forecast, with R^2 values between 0.53 and 0.76 with both models. Kilinc et al. [41] proposed a hybrid model based on PSO algorithm and extreme gradient boosting (XGBoost) for short-term streamflow forecasting in Turkey's Meriç basin. The authors showed how the PSO-XGBoost outperformed different ML/DL models (e.g., Linear Regression, LSTM) in streamflow prediction, with R^2 values between 0.7460 and 0.9582. These values are in line and in some cases lower than those achieved in the present study for the short-term forecasting horizon.

Overall, the literature studies demonstrate the complexity involved in selecting the proper streamflow rate forecasting model, which is contingent upon the particular hydrological features of each basin, available data, and the desired level of accuracy. Each type of model has its own set of advantages and limitations, leading researchers and practitioners to often utilize a combination of models to address complex challenges in discharge prediction. This can help explain the variety in findings across literature studies regarding the effectiveness of machine learning models, such as those evaluated in the current study, for predicting streamflow rates. From this perspective, in the future, further ML or DL algorithms, coupled with different combinations of exogenous inputs, could be integrated into the forecasting process to enhance the prediction reliability. This can lead to higher forecast accuracy for medium and long-term horizons. Furthermore, a limitation of the current study is that the model comparison was centered on three rivers. While these rivers exhibit distinct characteristics in terms of streamflow rate, precipitation, and temperature, they represent a limited number of cases. In the future, it would be intriguing to assess these models for rivers situated in regions with diverse climatic conditions, where the patterns of streamflow rate, precipitation, and temperature may differ from those examined in the present study.

Furthermore, accurate streamflow prediction models like AR-RBF and MLP-RF with PR as meta-learner are crucial for fostering novel environmental remediation projects from a sustainability perspective [42]. These models enable proactive planning and management of water resources, aiding in flood mitigation and ensuring the sustainability of activities like agriculture and hydroelectric power generation. By providing reliable short- and medium-term predictions, they assist in identifying and addressing environmental challenges efficiently. Their ease of implementation and minimal parameter optimization requirements make them valuable tools for environmental practitioners and policymakers, facilitating agile decision-making and project implementation for sustainable water management.

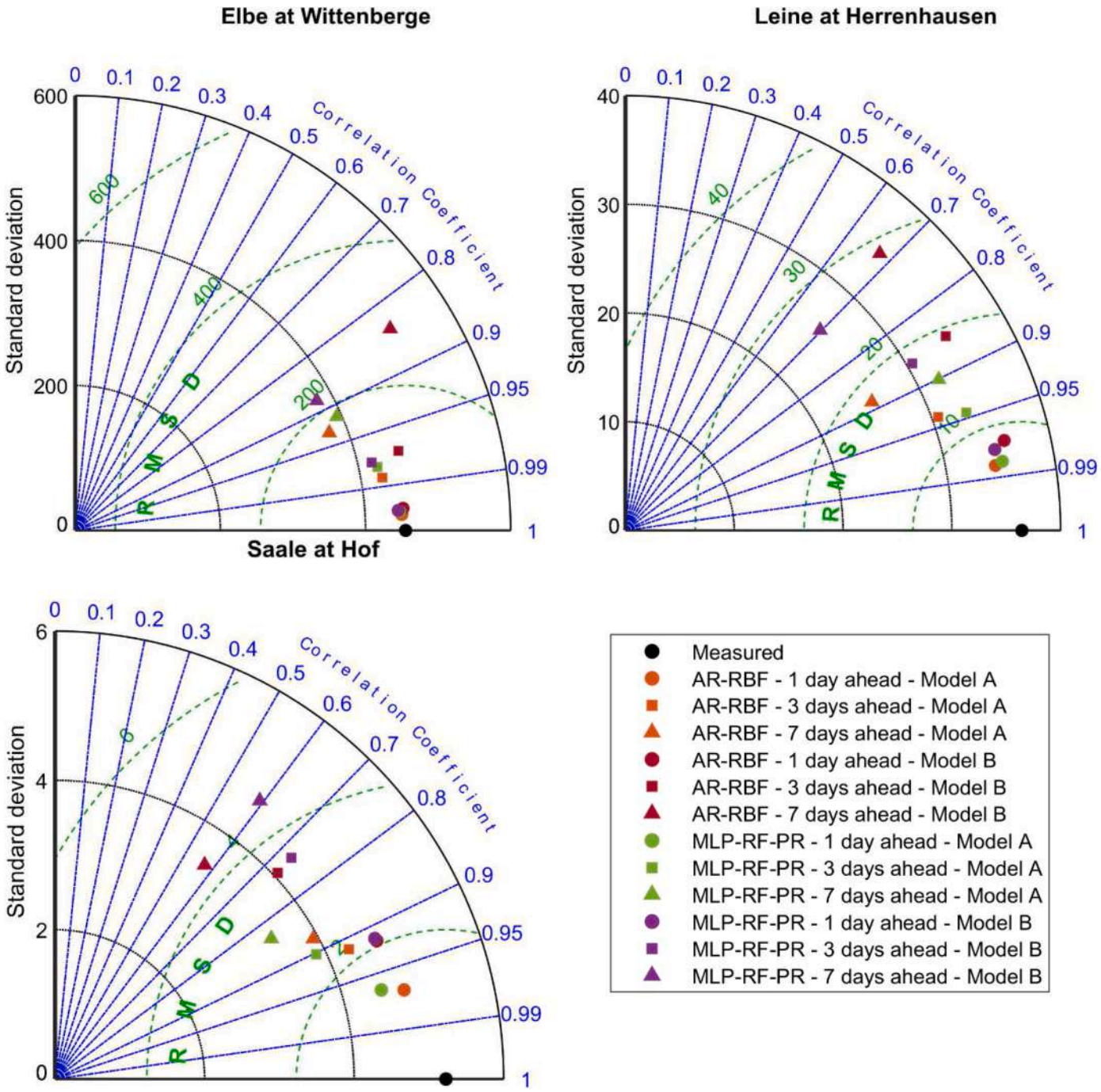

Fig. 13. Taylor diagrams for the three rivers – Testing stage.

5. Conclusions

This study compared two different models, AR-RBF and MLP-RF-PR, in short- and medium-term (up to 7 days) flow forecasting of three different rivers in Germany. The input variables consisted of the lagged values of flow rates, precipitation and temperature. The optimal number of lagged values for each input variable and the hyperparameters of the various algorithms were chosen through the BO algorithm. Regarding short-term forecasts (1 day ahead), both AR-RBF and MLP-RF-PR showed accurate predictions, with a slight tendency to underestimate streamflow rate peaks. As the forecast horizon increased, a reduction in forecast accuracy was observed, although both AR-RBF and MLP-RF-PR were still able to provide a reliable prediction of the overall flow trend, even for 7-day-ahead forecasting horizon. Specifically, regarding streamflow rate peaks, AR-RBF consistently exhibited superior performance compared to MLP-RF-PR across all rivers and scenarios investigated in this study. Furthermore, overall performance appears to be better for Elbe at Wittenberge, which showed the highest average streamflow rate along with the lowest CV value among the three rivers. In contrast, Saale at Hof, characterized by the lowest average streamflow rate and the highest CV, exhibited lower performance. This evidence confirms the greater complexity of making predictions for watercourses characterized by low and highly variable discharge rates.

This study has demonstrated that both AR-RBF and MLP-RF-PR are powerful tools for generating predictions in both the short- and medium-term. Moreover, they offer the significant advantage of requiring only a limited number of parameters to be optimized, making them easy to implement and relatively time-efficient.

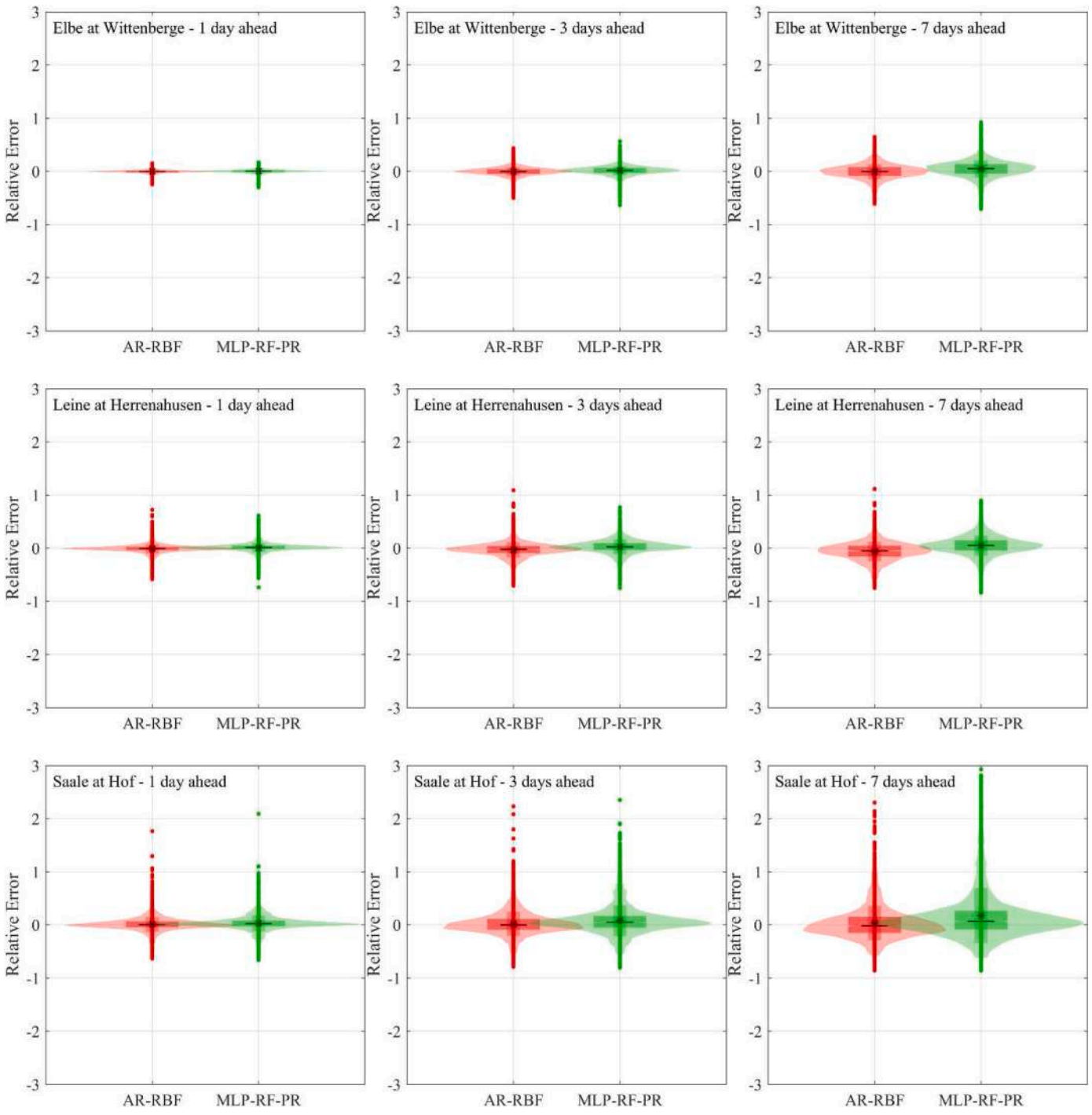

Fig. 14. Combined box and violin plots for the three rivers – Testing stage, Model A.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Funding

No external funding

CRediT authorship contribution statement

Francesco Granata: Writing – original draft, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Fabio Di Nunno:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Quoc Bao Pham:** Visualization, Validation, Data curation, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] P. Dobriyal, R. Badola, C. Tuboi, S.A. Hussain, A review of methods for monitoring streamflow for sustainable water resource management, Appl. Water Sci. 7 (6) (2017) 2617–2628.

[2] E.T. Alemu, R.N. Palmer, A. Polebitski, B. Meaker, Decision support system for optimizing Reservoir operations using ensemble streamflow predictions, J. Water Resour. Plann. Manag. 137 (1) (2011) 72–82.

[3] J.G. Arnold, R. Srinivasan, R.S. Muttiah, J.R. Williams, Large area hydrologic modeling and assessment. Part I. model development, J. Am. Water Resour. Assoc. 34 (1998) 73–89.

[4] C. Shen, M.S. Phanikumar, A process-based, distributed hydrologic model based on a large-scale method for surface– subsurface coupling, Adv. Water Resour. 33 (12) (2010) 1524–1541.

[5] S. Kostić, M. Stojković, S. Prohaska, N. Vasović, Modeling of river flow rate as a function of rainfall and temperature using response surface methodology based on historical time series, J. Hydroinf. 18 (4) (2016) 651–665.

[6] P. Sharma, D. Machiwal, Chapter 1 - streamflow forecasting: overview of advances in data-driven techniques, Advances in Streamflow Forecasting (2021) 1–50.

[7] Z.M. Yaseen, A. El-Shafie, O. Jaafar, H.A. Afan, K.N. Sayl, Artificial intelligence based models for stream-flow forecasting: 2000–2015, J. Hydrol. 530 (2015) 829–844.

[8] G. Papacharalampous, H. Tyralis, D. Koutsoyiannis, Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes, Stoch. Environ. Res. Risk Assess. 33 (2) (2019) 481–514.

[9] R.M. Adnan, Z. Liang, S. Heddam, M. Zounemat-Kermani, O. Kişi, B. Li, Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs, J. Hydrol. 586 (2020) 124371.

[10] F. Di Nunno, F. Granata, R. Gargano, G. de Marinis, Prediction of spring flows using nonlinear autoregressive exogenous (NARX) neural network models, Environ. Monit. Assess. 193 (2021) 350.

[11] A. Elbeltagi, F. Di Nunno, N.L. Kushwaha, G. de Marinis, F. Granata, River flow rate prediction in the Des Moines watershed (Iowa, USA): a machine learning approach, Stoch. Environ. Res. Risk Assess. (2022) 1–21.

[12] F. Granata, F. Di Nunno, Neuroforecasting of daily streamflows in the UK for short- and medium-term horizons: a novel insight, J. Hydrol. 624 (2023) 129888.

[13] J.F. Ruma, M.S.G. Adnan, A. Dewan, R.M. Rahman, Particle swarm optimization based LSTM networks for water level forecasting: a case study on Bangladesh river network, Results in Engineering 17 (2023) 100951.

[14] H. Tyralis, G. Papacharalampous, A. Langousis, Super ensemble learning for daily streamflow forecasting: large-scale demonstration and comparison with multiple machine learning algorithms, Neural Comput. Appl. 33 (8) (2021) 3053–3068.

[15] F. Granata, F. Di Nunno, G. de Marinis, Stacked machine learning algorithms and bidirectional long short-term memory networks for multi-step ahead streamflow forecasting: a comparative study, J. Hydrol. 613 (2022) 128431.

[16] K.S.M.H. Ibrahim, Y.F. Huang, A.N. Ahmed, C.H. Koo, A. El-Shafie, A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting, Alex. Eng. J. 61 (1) (2022) 279–303.

[17] F. Di Nunno, G. de Marinis, F. Granata, Short-term forecasts of streamflow in the UK based on a novel hybrid artificial intelligence algorithm, Sci. Rep. 13 (2023) 7036.

[18] K.W. Ng, Y.F. Huang, C.H. Koo, K.L. Chong, A. El-Shafie, A.N. Ahmed, A review of hybrid deep learning applications for streamflow forecasting, J. Hydrol. 625 (Part B) (2023) 130141.

[19] Y. Lin, D. Wang, G. Wang, et al., A hybrid deep learning algorithm and its application to streamflow prediction, J. Hydrol. 601 (2021) 126636.

[20] P.S.D.M. Neto, P.R.A. Firmino, H. Siqueira, Y.D.S. Tadano, T.A. Alves, J.F.L. De Oliveira, F. Madeiro, Neural-based ensembles for particulate matter forecasting, IEEE Access 9 (2021) 14470–14490.

[21] F. Granata, R. Gargano, G. de Marinis, Artificial intelligence based approaches to evaluate actual evapotranspiration in wetlands, Sci. Total Environ. 703 (2020) 135653.

[22] J. Park, I.W. Sandberg, Universal approximation using radial-basis-function networks, Neural Comput. 3 (2) (1991) 246–257.

[23] W. Li, Y. Shi, F. Huang, H. Hong, G. Song, Uncertainties of collapse susceptibility prediction based on remote sensing and GIS: effects of different machine learning models, Front. Earth Sci. 9 (2021) 731058.

[24] F. Murtagh, Multilayer perceptrons for classification and regression, Neurocomputing 2 (5–6) (1991) 183–197.

[25] M.M. Hasan, M.S.M. Nilay, N.H. Jibon, R.M. Rahman, LULC changes to riverine flooding: a case study on the Jamuna River, Bangladesh using the multilayer perceptron model, Results in Engineering 18 (2023) 101079.

[26] J. Sarafaraz, F.A. Kaleybar, J.M. Karamjavan, N. Habibzadeh, Predicting river water quality: an imposing engagement between machine learning and the QUAL2Kw models (case study: aji-Chai, river, Iran), Results in Engineering 21 (2024) 101921.

[27] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[28] Y. Wang, I.H. Witten, Pace Regression, University of Waikato, Department of Computer Science, Hamilton, New Zealand, 1999. Working paper 99/12).

[29] A. Wilhelms, N. Börsig, J. Yang, A. Holbach, S. Norra, Insights into phytoplankton dynamics and water quality monitoring with the BIOFISH at the Elbe River, Germany, Water 14 (13) (2022) 2078, https://doi.org/10.3390/w14132078.

[30] K. Nödler, T. Licha, S. Fischer, B. Wagner, M. Sauter, A case study on the correlation of micro-contaminants and potassium in the Leine River (Germany), Appl. Geochem. 26 (12) (2011) 2172–2180, https://doi.org/10.1016/j.apgeochem.2011.08.001.

[31] C. Scherer, A. Weber, F. Stock, et al., Comparative assessment of microplastics in water and sediment of a large European river, Sci. Total Environ. 738 (2020) 139866, https://doi.org/10.1016/j.scitotenv.2020.139866.

[32] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, Adv. Neural Inf. Process. Syst. 25 (2012).

[33] S. Kim, Y. Seo, M. Rezaie-Balf, et al., Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: peninsula vs continent, Theor. Appl. Climatol. 137 (2019) 693–712, https://doi.org/10.1007/s00704-018-2627-x.

[34] M.A. Ghorbani, R.C. Deo, Z.M. Yaseen, et al., Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran, Theor. Appl. Climatol. 133 (2018) 1119–1131, https://doi.org/10.1007/s00704-017-2244-0.

[35] Q.B. Pham, M. Kumar, F. Di Nunno, et al., Groundwater level prediction using machine learning algorithms in a drought-prone area, Neural Comput. Appl. 34 (2022) 10751–10773, https://doi.org/10.1007/s00521-022-07009-7.

[36] A. Legouhy, MATLAB central file exchange. al_goodplot - Boxblot and Violin Plot, 2023. https://www.mathworks.com/matlabcentral/fileexchange/91790-al_goodplot-boxblot-violin-plot. (Accessed 3 August 2023).

[37] J. Li, X. Yuan, P. Ji, Long-lead daily streamflow forecasting using Long Short-Term Memory model with different predictors, J. Hydrol.: Reg. Stud. 48 (2023) 101471.

[38] G. Liu, S. Ouyang, H. Qin, et al., Assessing spatial connectivity effects on daily streamflow forecasting using Bayesian-based graph neural network, Sci. Total Environ. 855 (2023) 158968.

[39] X. Li, J. Sha, Y. Li, Z.L. Wang, Comparison of hybrid models for daily streamflow prediction in a forested basin, J. Hydroinf. 20 (2018) 191–205.

[40] D.G. Lee, K.H. Ahn, A stacking ensemble model for hydrological postprocessing to improve streamflow forecasts at medium-range timescales over South Korea, J. Hydrol. 600 (2021) 126681.

[41] H.C. Kilinc, B. Haznedar, F. Ozkan, et al., An evolutionary hybrid method based on particle swarm optimization algorithm and extreme gradient boosting for short-term streamflow forecasting, Acta Geophys. (2024).

[42] S. Koley, Contemporary practices in groundwater arsenic remediation and wastewater management in West Bengal, India: a systematic review, International Journal of Advanced Technology and Engineering Exploration 8 (80) (2021) 797.