

River Water Temperature Prediction Based on Boosting Algorithms and Ensemble Empirical Mode Decomposition: CatBoost, XGBoost, LightGBM versus AdaBoost

Khaled Merabet¹, Fabio Di Nunno^{2,3}, Francesco Granata^{2,3}, Ozgur Kisi^{4,5,6,*}, Salim Heddami¹, Sungwon Kim⁷, Mariusz Ptak⁸, Mariusz Sojka⁹, Christoph Kulls^{4,*} and Mohammad Zounemat-Kermani¹⁰

¹ Faculty of Science, Agronomy Department, Hydraulics Division, University 20 Août 1955, Route El Hadaik, BP 26, Skikda, Algeria

² Department of Civil and Mechanical Engineering (DICEM), University of Cassino and Southern Lazio, via Di Biasio, 43, Cassino, Frosinone, Italy

³ European University of Technology, European Union, Cassino, Italy

⁴ Department of Civil Engineering, Technische Hochschule Lübeck, Lübeck, Germany

⁵ Department of Civil Engineering, School of Technology, Ilia State University, Tbilisi, Georgia

⁶ School of Civil, Environmental and Architectural Engineering, Korea University, Seoul, Republic of Korea

⁷ Department of Railroad Construction and Safety Engineering, Dongyang University, Yeongju, Republic of Korea

⁸ Department of Hydrology and Water Management, Adam Mickiewicz University, Krygowskiego 10, Poznań, Poland

⁹ Department of Land Improvement, Environmental Development and Spatial Management, Poznań University of Life Sciences, Piątkowska 94E, Poznań, Poland

¹⁰ Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

INFORMATION

Keywords:

River water temperature
boosting algorithms
machine learning
ensemble empirical mode
decomposition (EEMD)
CEEMDAN
hybrid models

DOI: 10.23967/j.rimni.2026.10.78336



River Water Temperature Prediction Based on Boosting Algorithms and Ensemble Empirical Mode Decomposition: CatBoost, XGBoost, LightGBM versus AdaBoost

Khaled Merabet¹, Fabio Di Nunno^{2,3}, Francesco Granata^{2,3}, Ozgur Kisi^{4,5,6,*}, Salim Heddami¹, Sungwon Kim⁷, Mariusz Ptak⁸, Mariusz Sojka⁹, Christoph Kulls^{4,*} and Mohammad Zounemat-Kermani¹⁰

¹Faculty of Science, Agronomy Department, Hydraulics Division, University 20 Août 1955, Route El Hadaik, BP 26, Skikda, Algeria

²Department of Civil and Mechanical Engineering (DICEM), University of Cassino and Southern Lazio, via Di Biasio, 43, Cassino, Frosinone, Italy

³European University of Technology, European Union, Cassino, Italy

⁴Department of Civil Engineering, Technische Hochschule Lübeck, Lübeck, Germany

⁵Department of Civil Engineering, School of Technology, Ilia State University, Tbilisi, Georgia

⁶School of Civil, Environmental and Architectural Engineering, Korea University, Seoul, Republic of Korea

⁷Department of Railroad Construction and Safety Engineering, Dongyang University, Yeongju, Republic of Korea

⁸Department of Hydrology and Water Management, Adam Mickiewicz University, Krygowskiego 10, Poznań, Poland

⁹Department of Land Improvement, Environmental Development and Spatial Management, Poznań University of Life Sciences, Piątkowska 94E, Poznań, Poland

¹⁰Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

ABSTRACT

This study presents a *hybrid modelling framework* for predicting river water temperature (T_w) by combining advanced boosting algorithms with signal decomposition techniques. Four models were compared: (i) categorical boosting (CatBoost), (ii) extreme gradient boosting (XGBoost), (iii) adaptive boosting (AdaBoost), and (iv) light gradient boosting machine (LightGBM). Each model was developed using air temperature (T_a) as input, both with and without signal decomposition. Three algorithms, empirical mode decomposition (EMD), ensemble EMD (EEMD), and complete EEMD with adaptive noise (CEEMDAN), were applied to decompose T_a into intrinsic mode functions (IMFs), which served as additional predictors. The models were trained and validated using data from four Polish river stations, and their performance was evaluated using root mean squared error (RMSE), mean absolute error (MAE), correlation coefficient (R), and Nash–Sutcliffe efficiency (NSE). T_w was modelled as an instantaneous function of T_a , without incorporating a forecasting horizon, to assess the ability of decomposition-boosting hybrids to reproduce observed thermal dynamics at the same time step rather than perform multi-step predictions. Unlike previous single-hybrid studies, this work systematically compares multiple

OPEN ACCESS

Received: 29/12/2025

Accepted: 12/02/2026

Published: 07/08/2026

DOI

10.23967/j.rimni.2026.10.78336

Keywords:

River water temperature
boosting algorithms
machine learning
ensemble empirical mode
decomposition (EEMD)
CEEMDAN
hybrid models

boosting-decomposition combinations under identical settings, isolating the decomposition effect and assessing accuracy, efficiency, and robustness. Results show that integrating signal decomposition substantially improves prediction accuracy. The best performance was achieved by XGBoost_EEMD (RMSE = 0.804, MAE = 0.598, R = 0.993, NSE = 0.987), reflecting a $\sim 64\%$ improvement over the standalone model. The proposed *hybrid modelling framework* offers methodological innovation and practical potential for real-time, data-driven river temperature prediction in support of climate-resilient water management.

1 Introduction

River water temperature (Tw) is a key environmental variable controlling a wide range of physical, chemical, and biological processes in freshwater ecosystems. It directly affects metabolic rates, biogeochemical cycling, dissolved oxygen solubility, and species distributions, thereby exerting a strong influence on water quality, ecosystem functioning, and biodiversity [1,2]. Many aquatic organisms exhibit narrow thermal tolerance ranges, and even moderate deviations from optimal conditions may induce physiological stress or mortality. Consequently, the accurate characterization and prediction of river water temperature dynamics have become increasingly important, particularly under the combined effects of climate change and growing anthropogenic pressures.

River Tw variability is driven by a combination of atmospheric, hydrological, and catchment-related factors. Among these, air temperature (Ta) plays a dominant role, as it largely governs the thermal equilibrium between the river and the surrounding environment, especially at daily and seasonal time scales [3]. Additional influences include solar radiation, streamflow variability, precipitation, snowmelt, and groundwater contributions, the latter often exerting a buffering effect due to its relatively stable thermal regime. Human interventions, such as riparian vegetation removal, urbanization, irrigation practices, and reservoir operations, further modify natural thermal regimes by altering shading, flow paths, and heat exchanges, frequently amplifying temperature variability and ecological stress.

Given its ecological and operational relevance, Tw prediction has attracted considerable attention in water resources management, fisheries protection, hydropower operation, and climate adaptation planning. Over the past two decades, a wide spectrum of modelling approaches has been developed, ranging from physically based heat-budget models to empirical and data-driven formulations. While process-based models provide mechanistic insight into energy exchanges, their practical applicability is often constrained by extensive data requirements and computational complexity, particularly in data-scarce regions or real-time applications. In contrast, empirical and machine learning (ML) approaches offer a flexible alternative, capable of capturing nonlinear and nonstationary relationships between Tw and its dominant drivers using relatively limited input information.

Recent studies have highlighted the strong potential of ML and deep learning (DL) approaches for water temperature prediction across a wide range of aquatic environments. Neural networks, ensemble tree-based models, and hybrid architectures have been successfully applied to lakes, rivers, reservoirs, and coastal systems, often achieving high predictive accuracy when driven by meteorological and hydrological inputs [4–6]. In particular, ensemble learning techniques, including boosting algorithms

such as CatBoost, XGBoost, AdaBoost, and LightGBM, have attracted increasing attention due to their robustness, ability to capture nonlinear interactions, and strong generalization capabilities.

Despite these advances, river water temperature time series are inherently nonstationary and characterized by multiscale variability, reflecting the superposition of short-term atmospheric forcing, seasonal cycles, and longer-term hydrological and climatic influences. When applied directly to raw input signals, even advanced machine learning models may struggle to fully exploit such complex temporal structures, potentially leading to reduced stability and suboptimal generalization performance.

To overcome these limitations, signal decomposition techniques—including Empirical Mode Decomposition (EMD), Ensemble EMD (EEMD), and Complete Ensemble EMD with Adaptive Noise (CEEMDAN)—have been increasingly adopted in hydrological and environmental modelling as effective preprocessing tools for handling nonstationary and multiscale time series. These data-driven methods decompose complex signals into a finite set of intrinsic mode functions (IMFs), each representing oscillatory components associated with distinct temporal scales, together with a residual term capturing long-term trends [7–9]. By separating high-frequency variability from medium- and low-frequency components, decomposition techniques can mitigate noise contamination, reduce mode mixing, and enhance the representation of scale-dependent dynamics.

When used in conjunction with ML models, the resulting multiscale components provide enriched feature sets that facilitate improved learning of nonlinear relationships and temporal dependencies that may be obscured in the original signal. Accordingly, several hybrid decomposition–ML frameworks have been proposed in the literature, combining EMD-based techniques with regression, tree-based, or deep learning models. Representative examples include EEMD-SVR, EEMD-LSTM, and CEEMDAN-RF architectures, which have reported notable performance gains in hydrological prediction tasks, including streamflow, groundwater level, and water temperature modelling [10–12].

However, the majority of existing studies have focused on single hybrid configurations or model-specific optimizations, often tailored to a particular algorithm or case study. As a consequence, the relative contribution of signal decomposition is frequently entangled with that of the learning algorithm itself, making it difficult to isolate and quantify the actual effect of decomposition across different modelling paradigms. This lack of systematic, cross-model benchmarking under comparable experimental conditions hampers a comprehensive understanding of when and why decomposition-based preprocessing yields genuine predictive benefits.

More recently, alternative hybridization strategies integrating ensemble learning with metaheuristic optimization or physics-informed components have been explored to enhance parameter tuning, feature selection, and model interpretability [13,14]. While these approaches have demonstrated promising results in specific applications, they address a different research objective, primarily related to model optimization rather than to the controlled evaluation of signal preprocessing strategies. Consequently, their direct comparison with decomposition-based hybrids remains challenging under heterogeneous experimental settings.

In this context, the present study aims to provide a clear and systematic assessment of decomposition–boosting synergies for river water temperature prediction. The main contribution of this work is methodological rather than algorithmic. Specifically, a unified benchmarking framework is established to compare four widely used boosting algorithms—XGBoost, LightGBM, CatBoost, and AdaBoost—combined with three signal decomposition techniques (EMD, EEMD, and CEEMDAN) under strictly identical data, preprocessing, and evaluation conditions. To isolate the effect of decomposition and minimize confounding influences, river water temperature is modelled exclusively as a function of air temperature, which represents the dominant and most widely available predictor.

By simultaneously evaluating predictive accuracy, robustness, and computational efficiency across multiple stations and model configurations, this study clarifies when and to what extent signal decomposition provides tangible benefits in river temperature modelling. Beyond performance comparison, the proposed framework offers practical insights for real-time, data-driven water temperature monitoring and management, supporting climate-resilient decision-making in freshwater systems.

2 Materials and Methods

2.1 Study Area and Available Data

The present paper analyses four hydrological stations located in Poland, with two rivers situated in the northern and two in the southern parts of the country (Fig. 1). The rivers analyzed in the article are different in terms of hydrological characteristics and catchment parameters. This approach allows the proposed methodology to be used in different environmental conditions. Detailed data for individual rivers are presented below, in Tables 1 and 2. The Łyna River is the largest tributary of the Pregolya, which flows into the Vistula Lagoon (Baltic Sea). Hydrologically, it is characterized by a moderately developed nival regime, where the average spring discharge ranges from 130% to 180% of the average annual discharge [15]. The Drwęca River, also located in the northern part of the country, flows into the Vistula (the longest river in Poland). Like the Łyna, it features a moderately developed nival regime. The sources of the San River are in Ukraine, and its upper section flows through the Bieszczady Mountains, a range constituting part of the Carpathians. It exhibits a moderately developed pluvial-nival regime, meaning that the average summer discharge is higher or nearly equal to the average spring discharge, amounting to 130% to 180% of the average annual discharge. Like the Drwęca, it flows into the Vistula. Lastly, the Dunajec River drains the highest mountain range in Poland (the Tatra Mountains), which results in significant water level fluctuations. Its hydrological regime is typically mountainous, characterized by a moderately developed pluvial-nival regime. Water temperature measurements were conducted by the Institute of Meteorology and Water Management at a depth of 0.4 m below the surface, at the same point near the gauge. The dataset covers the period 1987–2014 and consists of 10,220 daily observations for each station. In addition, air temperature data from four corresponding stations were used as the sole predictor for water temperature estimation. To enhance the input information, the air temperature time series was decomposed using three decomposition techniques, namely EMD, EEMD, and CEEMDAN. The decomposition resulted in 9 IMFs plus one residual for EMD, 12 IMFs plus one residual for EEMD, and 13 IMFs plus one residual for CEEMDAN, which were considered as input features for the prediction models. The results of the air temperature decomposition for the Łyna station are provided in the Supplementary Materials, including the IMFs obtained using EMD, EEMD, and CEEMDAN (Figs. S1–S3).

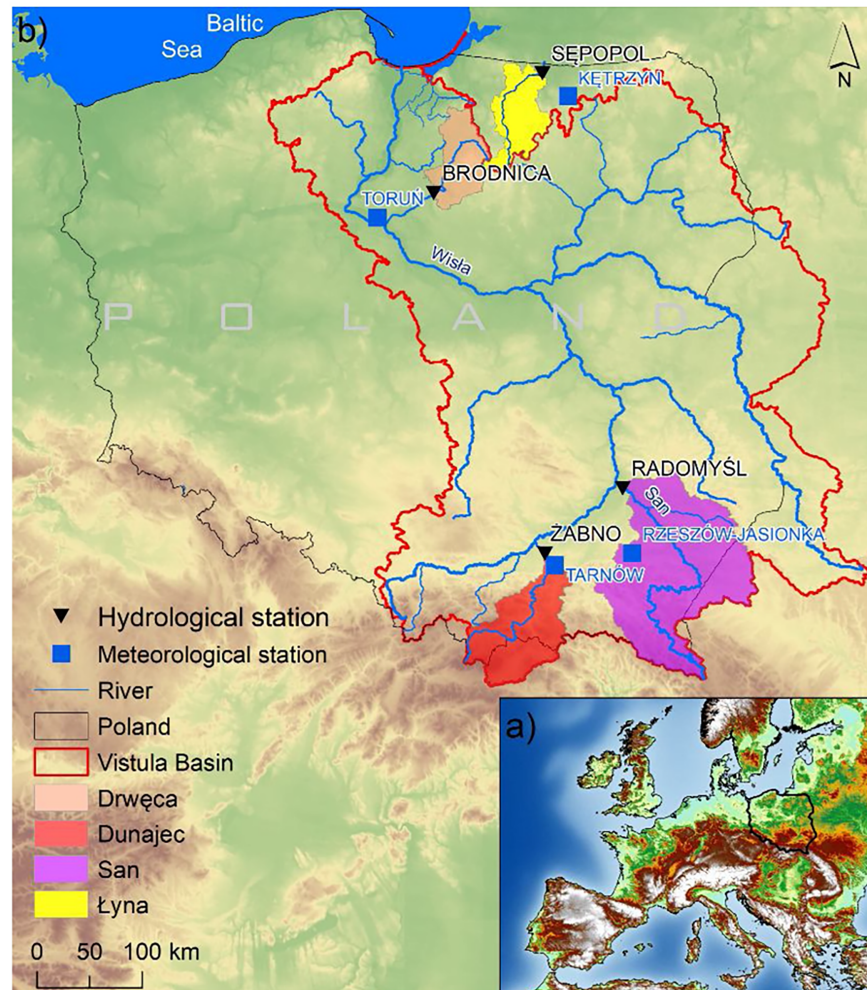


Figure 1: Study site location. (a) Location of the Vistula River Basin in Europe. (b) Detailed map of the Vistula Basin in Poland showing the selected sub-basins (Drwęca, Łyna, Dunajec, and San), river network, and the locations of hydrological (black triangles) and meteorological stations (blue squares). The figure was generated using ArcGIS Pro v3.1.0 (<https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>) based on the Copernicus Global Digital Elevation Model developed by the European Space Agency

Table 1: Characteristics of rivers and catchments along with the locations of meteorological stations

River/Stations	Unit	<i>Łyna</i> <i>Sępól</i>	<i>Drwęca</i> <i>Brodnica</i>	<i>San</i> <i>Radomyśl</i>	<i>Dunajec</i> <i>Żabno</i>
Location	(Decimal degree)	21.0059 E 54.2662 N	19.4009 E 53.2554 N	21.9307 E 50.6735 N	20.8620 E 50.1318 N
Location	(m asl)	26.22	67.72	138.75	173.43

(Continued)

Table 1 (continued)

River/Stations	Unit	<i>Lyna Sępapol</i>	<i>Drweca Brodnica</i>	<i>San Radomyśl</i>	<i>Dunajec Żabno</i>
Catchment area	km ²	5973.2	5698.2	16866.8	6787.2
Total river length	km	210.3	240.8	459.4	250.0
Sub catchment area	(km ²)	3640.1	3539.8	16837.9	6738.5
Percentage of total catchment area	(%)	60.9	62.1	99.8	99.3
River length	(km)	191.6	145.19	449.68	232.74
Subcatchment circulatory ratio	(-)	0.12	0.21	0.26	0.29
Mean elevation ¹	(m asl)	127.4	132.8	309.5	651.4
Mean slope ¹	(%)	3.6	3.5	7.2	18.2
Drainage density	(km·km ⁻²)	1.5	1.3	2.0	2.2
Artificial surfaces ²	(%)	4.0	4.3	6.4	5.8
Agricultural areas ²	(%)	57.0	57.3	46.6	46.3
Forest and seminatural areas ²	(%)	35.1	33.4	46.1	47.0
Wetlands ²	(%)	0.4	0.5	0.1	0.1
Water bodies ²	(%)	3.5	4.5	0.8	0.8
Meteorological station	/	Kętrzyn	Toruń	Rzeszów- Jesionka	Tarnów
Mean annual precipitation	mm	594	540	665	726

Note: ¹European Space Agency (2024). Copernicus Global Digital Elevation Model. Distributed by Open Topography. <https://doi.org/10.5069/G9028PQB> (accessed on 08 October 2024). ²European Environment Agency (2020). CORINE Land Cover 2018 (vector), Europe, 6-yearly-version 2020_20u1, May 2020. <https://doi.org/10.2909/71c95a07-e296-44fc-b22b-415f42acfd0> (accessed on 08 October 2024).

Table 2: Brief statistics for water and air temperatures recorded at the four stations

Variables	Subset	Unit	X_{mean}	X_{max}	X_{min}	S_x	C_v	R
<i>Lyna Station</i>								
T_w	Training	°C	9.341	25.700	0.100	7.050	0.755	1.000
	Validation	°C	9.317	24.600	0.100	6.990	0.750	1.000
	All data	°C	9.334	25.700	0.100	7.031	0.753	1.000
T_a	Training	°C	7.816	28.500	−26.000	8.534	1.092	0.919
	Validation	°C	7.777	28.000	−21.200	8.461	1.088	0.919
	All data	°C	7.804	28.500	−26.000	8.512	1.091	0.919
<i>Drweca Station</i>								
T_w	Training	°C	9.611	24.400	0.100	7.123	0.741	1.000
	Validation	°C	9.791	24.000	0.100	7.066	0.722	1.000
	All data	°C	9.665	24.400	0.100	7.106	0.735	1.000
T_a	Training	°C	8.606	30.100	−24.800	8.361	0.971	0.916
	Validation	°C	8.884	28.700	−23.600	8.419	0.948	0.916
	All data	°C	8.690	30.100	−24.800	8.379	0.964	0.916
<i>Sun Station</i>								
T_w	Training	°C	10.331	27.400	0.100	7.812	0.756	1.000
	Validation	°C	10.326	26.500	0.100	7.775	0.753	1.000
	All data	°C	10.330	27.400	0.100	7.801	0.755	1.000
T_a	Training	°C	8.627	28.000	−25.300	8.666	1.005	0.918
	Validation	°C	8.524	27.700	−22.800	8.639	1.014	0.917
	All data	°C	8.596	28.000	−25.300	8.658	1.007	0.917
<i>Dunajec Station</i>								
T_w	Training	°C	9.691	24.400	0.100	6.871	0.709	1.000
	Validation	°C	9.773	24.600	0.100	6.997	0.716	1.000
	All data	°C	9.716	24.600	0.100	6.909	0.711	1.000
T_a	Training	°C	9.110	28.400	−24.400	8.463	0.929	0.893
	Validation	°C	9.221	29.500	−22.300	8.656	0.939	0.897
	All data	°C	9.143	29.500	−24.400	8.521	0.932	0.894

Note: [Abbreviations: X_{mean} , mean; X_{max} , maximum; X_{min} , minimum; S_x , standard deviation; C_v , coefficient of variation; R , coefficient of correlation between T_w and T_a , T_w : river water temperature, T_a : air temperature].

2.2 Accuracy Evaluation Criteria

Five performance metrics, namely root-mean-square error (RMSE), mean absolute error (MAE), Nash-Sutcliffe efficiency (NSE), Kling-Gupta efficiency (KGE) and correlation coefficient R , were used for the models' performance evaluation.

$$MAE = \frac{\sum_{i=1}^N |Tw_{pre,i} - Tw_{obs,i}|}{N} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Tw_{obs,i} - Tw_{pre,i})^2}{N}} \quad (2)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^N (Tw_{obs,i} - Tw_{pre,i})^2}{\sum_{i=1}^N (Tw_{obs,i} - \overline{Tw_{obs}})^2} \right] \quad (3)$$

$$R = \left(\frac{\sum_{i=1}^N (Tw_{obs,i} - \overline{Tw_{obs}}) (Tw_{pre,i} - \overline{Tw_{pre}})}{\sqrt{\sum_{i=1}^N (Tw_{obs,i} - \overline{Tw_{obs}})^2 \sum_{i=1}^N (Tw_{pre,i} - \overline{Tw_{pre}})^2}} \right) \quad (4)$$

$$KGE = 1 - \sqrt{(R - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (5)$$

Here $Tw_{obs,i}$ and $Tw_{pre,i}$ are measured and calculated daily T_w values at an i th time step, N is the quantity of observations. $\overline{Tw_{obs}}$ and $\overline{Tw_{pre}}$ are the means of measured and calculated daily T_w . α measures the variability ratio (standard deviation of calculated/ standard deviation of measured), whereas β indicates the ratio of average predicted results to average experimental data.

3 Machine Learning Models

3.1 Categorical Boosting (CatBoost)

The CatBoost model, introduced by [16], represents a state-of-the-art gradient-boosting decision tree framework designed to optimize the treatment of categorical features during the modeling process. This innovative approach enhances the model's generalization capabilities and performance by effectively addressing issues related to low bias and high variance. To further mitigate the risks of overfitting and underfitting, CatBoost incorporates a modified implementation of ordered boosting, a refined variant of conventional gradient boosting methods [17–19]. The core mechanism of CatBoost relies on the sequential arrangement of variables according to a predefined order, enabling the creation of multiple models through systematic sampling of training data [20]. The model employs a randomized organization of all sampling data, ensuring consistency across feature classes. Additionally, prior to initiating the sampling procedure, CatBoost calculates the target value of each data point and assigns corresponding weights and priorities, enhancing its ability to handle categorical data with precision [16,21]. CatBoost assists the task of converting categorical features into numerical values utilizing a reliable approach (i.e., ordered target metrics), which can be expressed as follows [16]:

$$X_{\sigma p,k} = \frac{\sum_{j=1}^{p-1} [X_{\sigma j,k} = X_{\sigma p,k}] Y_{\sigma s} + \beta \cdot P}{\sum_{j=1}^{p-1} [X_{\sigma j,k} = X_{\sigma p,k}] + \beta} \quad (6)$$

where P and β are prior values and the smoothing parameter.

CatBoost also integrates a greedy optimization algorithm that iteratively refines sequential estimates to minimize a specified loss function [16,22]. This iterative approach enables the model to achieve superior predictive accuracy while maintaining computational efficiency. For a deeper exploration of CatBoost's applications in solving predictive problems, comprehensive discussions can be found in [16,18], among others.

3.2 Extreme Gradient Boosting (XGBoost)

The XGBoost model, developed by [23], stands as one of the most prominent algorithms for supervised learning tasks, particularly in addressing classification and regression challenges. Grounded in the gradient boosting framework, XGBoost has achieved widespread recognition for its ability to enhance gradient optimization and deliver superior predictive performance [24,25]. By employing a robust implementation of the tree-based boosting paradigm, the model strategically combines multiple weak learners to form a strong and highly accurate ensemble [26]. Boosting, the foundational principle of XGBoost, operates by iteratively improving upon weak models, each slightly better than random chance, to construct an ensemble with markedly enhanced predictive capabilities [27]. This iterative process weights input features and outputs strategically, allowing the model to refine predictions by reducing residual errors over successive iterations [28]. Within this framework, XGBoost employs a sophisticated tree ensemble approach, utilizing classification and regression trees (CART) as its base learners. Each individual tree contributes to the final prediction, with results aggregated to optimize overall accuracy and robustness [23,29]. Therefore, XGBoost established itself as a powerful tool that can effectively handle a wide range of regression and classification tasks (Chen and Guestrin, 2016). Given that the approach consists of k decision trees, the integrated strategy can be expressed as follows [23]:

$$\hat{y}_i = \sum_1^k f_k(x_i), f_k \in F \quad (7)$$

where $\hat{y}_i$ = the calculated T_w at four stations, f_k = the k -th tree. Moreover, the variable k = the sample size, x_i = input parameter, and F = all possible CARTs.

A key strength of XGBoost lies in its ability to handle complex and large datasets with efficiency, leveraging advanced features such as regularization, sparsity awareness, and parallel processing to mitigate overfitting and improve scalability. These attributes make it a highly versatile tool, adaptable to a broad spectrum of predictive modeling applications. For comprehensive insights into the development and applications of XGBoost, readers are directed to foundational works by [23,25,26], among others.

3.3 Adaptive Boosting (AdaBoost)

The AdaBoost model, introduced by [30], is a pioneering ensemble learning method designed to improve the predictive accuracy of weak classifiers. By iteratively combining multiple machine learning models, AdaBoost constructs a unified framework that excels in addressing classification problems. The model operates by sequentially training classifiers, where each iteration focuses on the misclassified instances from the previous classifier, assigning them higher weights to emphasize their importance in the subsequent iteration [31]. A distinctive feature of AdaBoost is its ability to transform weak classifiers models with a classification error slightly better than random guessing into a robust ensemble capable of achieving high accuracy. This is accomplished through a weighted voting mechanism, where each weak classifier contributes to the final decision proportionally to its performance. The cumulative effect of these weighted contributions results in a strong classifier that minimizes overall prediction error [32,33]. The general structure of AdaBoost can be expressed as follows [30]:

$$F_n(x) = F_{n-1}(x) + \arg \min \sum_{i=1}^n L(y_i, F_{n-1}(x_i) + h(x_i)) \quad (8)$$

where y_i = output of i -th tree, $h(x_i)$ = the tree, $F_{n-1}(x_i)$ = the value that calculated in previous step, and $F_n(x)$ = the final model.

The model's iterative nature ensures continuous refinement, progressively reducing the classification error. By dynamically adjusting the weights of training samples, AdaBoost prioritizes instances that are challenging to classify, enabling the ensemble to focus on the most complex aspects of the data. For correctly classified samples, the assigned weights are reduced, while the probability of selecting challenging samples in subsequent iterations increases [30,34]. AdaBoost's resilience against overfitting and underfitting further enhances its versatility, making it suitable for a wide range of classification tasks. Weak classifiers that exhibit significant training error are iteratively refined or replaced, ensuring that only the most robust models are retained within the ensemble [35]. For an in-depth understanding of AdaBoost's theoretical foundations and applications, readers are referred to the seminal work of [30], as well as more recent studies by [33,34].

3.4 *Light Gradient Boosting Machine (LightGBM)*

The LightGBM model, developed as an advanced gradient boosting framework, represents a breakthrough in adaptive boosting techniques by addressing key limitations of traditional approaches, such as computational complexity, memory usage, and processing time. Built upon the principles of gradient boosting decision trees (GBDT), LightGBM efficiently handles both classification and regression tasks, offering superior scalability and performance [36–38]. Two innovative strategies underpin the model's efficiency: gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). GOSS reduces computational overhead by focusing on instances with larger gradients, which carry more information for model optimization, while EFB compresses sparse features into bundled groups to further streamline computation. These advancements make LightGBM particularly effective for high-dimensional datasets, achieving remarkable speed and precision without sacrificing accuracy [39,40]. LightGBM also introduces a histogram-based algorithm and a depth-limiting strategy for leaf-wise growth, significantly reducing memory consumption while maintaining high predictive accuracy. By constraining tree depth, the model mitigates the risks of overfitting and underfitting, delivering robust performance across diverse datasets [41,42]. Additionally, the model incorporates a maximum depth constraint and an innovative design feature to enhance computational efficiency during training [39,40].

Considering the mathematical formation, the objective function utilized in the LightGBM can be expressed as follows [39,40]:

$$OF_{LGBM} = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{j=1}^t \Omega(f_j) \quad (9)$$

where $\sum_{j=1}^t \Omega(f_j)$ = A regularization step to assist the choice of a simple prediction function and reduce the risk of overfitting in the model and $l(y_i, \hat{y}_i)$ = difference between the predicted and observed values.

A distinguishing feature of LightGBM is its ability to handle categorical data directly, bypassing the need for one-hot encoding, a common but computationally intensive preprocessing step in many machine learning models. This direct approach not only simplifies the handling of categorical features but also enhances computational efficiency and model accuracy [42,43]. For a detailed exploration of LightGBM's architecture and applications in addressing predictive challenges, readers are referred to seminal works by [39,40,42,43]. These studies provide valuable insights into the model's innovations and its transformative impact on gradient boosting methodologies.

3.5 *Signal Decomposition Methods*

In this study, we have employed three signal decomposition techniques to preprocess the air temperature data as a nonlinear and complicated time series at each of the four stations in Poland prior

to entering the observed dataset into the four gradient boosting models. The applied techniques include (i) Empirical Mode Decomposition (EMD), (ii) Ensemble Empirical Mode Decomposition (EEMD), and (iii) Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). In these techniques, the observed time series data are broken into simpler oscillatory components known as Intrinsic Mode Functions (IMFs). Below are the detailed descriptions and mathematical formulations for each technique.

3.5.1 Empirical Mode Decomposition (EMD)

As a data-driven method, the EMD decomposes a signal $x(t)$ (e.g., a meteorological time series) into a finite number of IMFs. Each IMF represents a simple oscillatory mode embedded within the data. It is worth mentioning that the EMD is an iterative process that does not require any predetermined basis functions [7]. The decomposition process for decomposing air temperature data to generate IMFs can be summarized as follows:

- (1) Identify extrema: Identify the local extrema of the air temperature time series signal $x(t)$ (both maxima and minima).
- (2) Create and adjust envelopes: Generate upper and lower envelopes by interpolating between the local maxima and minima. Afterwards, the mean envelope is calculated:

$$m(t) = (U(t) + L(t)) / 2 \quad (10)$$

Then, the mean of the upper and lower envelopes is subtracted from the original signal to obtain a candidate IMF:

$$h(t) = x(t) - m(t) \quad (11)$$

At this stage, the $h(t)$ signal should satisfy the conditions of IMFS (i.e., the average of the $U(t)$ and $L(t)$ must be zero).

- (3) The iteration process: there are two iterative loops in the EMD method. The first loop will be continued until it identifies the IMFs. If $h(t)$ meets the IMF criteria, the confirmed IMF would be labelled as the first IMF₁. Otherwise, $h(t)$ would be considered as the new $x(t)$ signal and the former steps are repeated. Once the IMF1 is identified, it is subtracted from the original $x(t)$ time series signal:

$$x(t) = x(t) - \text{IMF}_1 \quad (12)$$

In the second loop, the process is repeated iteratively to extract successive IMFs (e.g., IMF₂, IMF₃, ..., IMF_n) until the monotonic residual signal $r(t)$ shows no oscillatory behavior.

- (4) Final output: At the end, the original time series is decomposed and reconstructed as

$$x(t) = \sum_{i=1}^n \text{IMF}_i(t) + r(t) \quad (13)$$

The resulting IMFs represent different low- to high-frequency components of the time series signal, which are used as features for the model.

3.5.2 Ensemble Empirical Mode Decomposition (EEMD)

EEMD is an improvement over the traditional EMD method. It is an extension of EMD that addresses mode mixing by adding white noise to the original time series [8]. The added noise helps to

separate the IMFs more effectively, and an ensemble average of multiple decompositions is taken to obtain the final IMFs. The key steps in EEMD are:

- (1) Adding white noise: Add white noise $n(t)$ using an amplitude coefficient α to the original time series signal and perform EMD on the noisy signal:

$$x_n(t) = x(t) + \alpha n(t) \quad (14)$$

- (2) Performing the EMD: Repeat the process multiple times with different noise realizations $x_n(t)$ to extract the IMFs:

$$IMF_{i,n}(t) = EMD(x_n(t)) \quad (15)$$

Next, the obtained IMFs are averaged to compute an ensemble average of $IMF_i(t)$. By the averaging process, the noises are mitigated and a more stable decomposition signal is generated.

- (3) The iteration process: The EMD process is iterated until the desired number of IMFs and their related residuals $r(t)$ are extracted:

$$r(t) = x(t) - \sum_{i=1}^n IMF_i(t) \quad (16)$$

- (4) Final output: Finally, the new time series is reconstructed as:

$$x(t) = \sum_{i=1}^n IMF_i(t) + r(t) \quad (17)$$

3.5.3 Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

CEEMDAN is a developed version of the EEMD that modifies the EEMD process by adding adaptive noise at each stage of the decomposition rather than only at the beginning. This technique was proposed by [9] to eliminate the residual white noise in the decomposition process of EEMD. In this regard, the CEEMDAN can extract more accurate IMFs and, consequently, reduce the risk of residual noise affecting the decomposition. The general procedure of the CEEMDAN is similar to the explained stages in the last section for the EEMD technique. However, there are some differences. Below the main steps are expressed:

- (1) Initial noise addition: this stage is applied similarly to the EEMD technique. Thereafter, we have $x_n(t) = x(t) + \alpha n(t)$. It should be noted that unlike the EEMD technique, the CEEMDAN inserts adaptive noise that is derived from the previous IMFs. Also, the CEEMDAN adapted the amplitude α based on the current stage of decomposition.
- (2) Performing the EMD: apply the EMD process to each realization $x_{n1}(t)$ to extract the IMFs. Unlike the EEMD technique, where IMFs occur for each realization of the signal independently, in the CEEMDAN, each IMF is extracted iteratively, and the residual is updated after extracting each IMF. Next, the extracted IMFs are averaged, and the noises are removed.
- (3) The iteration process: a new adaptive noise addition $n_1(t)$ is generated based on the $r_1(t)$ residual signal and then added to the original signal.

$$x_{n2}(t) = r_1(t) + n_1(t) \quad (18)$$

Then the residual signal is updated as:

$$r_2(t) = x_{n2}(t) - \sum_{i=1}^n \text{IMF}_i(t) \quad (19)$$

Next, the second adaptive noise $n_2(t)$ is added to calculate $x_{n3}(t) = r_2(t) + n_2(t)$. This process is repeated until a stopping criterion is met.

(4) Final output: the final formulation is

$$x(t) = \sum_{i=1}^n \text{IMF}_i(t) + r(t) \quad (20)$$

4 Results and Discussion

In this study, we explored the application of ensemble machine learning methods, namely, CatBoost, XGBoost, AdaBoost, and LightGBM, to estimate water temperature (T_w) based on air temperature (T_a) data collected from four stations: Lyna, Drweca, San, and Dunajec. To enhance the predictive accuracy, we also applied three signal decomposition techniques, Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), to the air temperature data. These decompositions generated intrinsic mode functions (IMFs), which served as additional inputs for the models. The models were then evaluated for training and validation datasets using performance metrics such as the correlation coefficient (R), Nash-Sutcliffe Efficiency (NSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The flowchart of the overall modelling framework is depicted in Fig. 2.

4.1 Models Structures

In this study, water temperature (T_w) was modelled based on three different scenarios namely:

1. Scenario 1: $T_w = f(T_a)$
2. Scenario 2: $T_w = f(T_a \text{ with decomposition EMD})$
3. Scenario 3: $T_w = f(T_a \text{ with decomposition EEMD})$
4. Scenario 4: $T_w = f(T_a \text{ with decomposition CEEMDAN})$

Thus, we apply the models as; AdaBoost, AdaBoost_EMD, AdaBoost_EEMD, and AdaBoost_CEEMDAN, and so on for the others models.

4.2 Results at Lyna Station

Table 3 summarizes the performance of AdaBoost, CatBoost, LightGBM, and XGBoost models at the Lyna station, both with and without signal decomposition techniques (EMD, EEMD, CEEMDAN). AdaBoost shows consistent but moderate performance across both stages, with correlation coefficients (R) around 0.945 and NSE values near 0.89. RMSE and MAE are slightly higher in the validation stage, indicating that the model maintains relatively stable but not exceptional predictive power without decomposition. Decomposition methods improve AdaBoost's accuracy considerably. EEMD provides the best enhancement, achieving $R = 0.981$ in training and 0.980 in validation. However, there is a slight drop in performance from training to validation (NSE decreases from 0.957 to 0.953, RMSE increases from 1.459 to 1.508), suggesting some sensitivity to the training data.

CatBoost performs well on its own, achieving R values around 0.95 and NSE near 0.91 in both training and validation stages. It outperforms AdaBoost but still shows room for improvement, as indicated by RMSE values above 2.

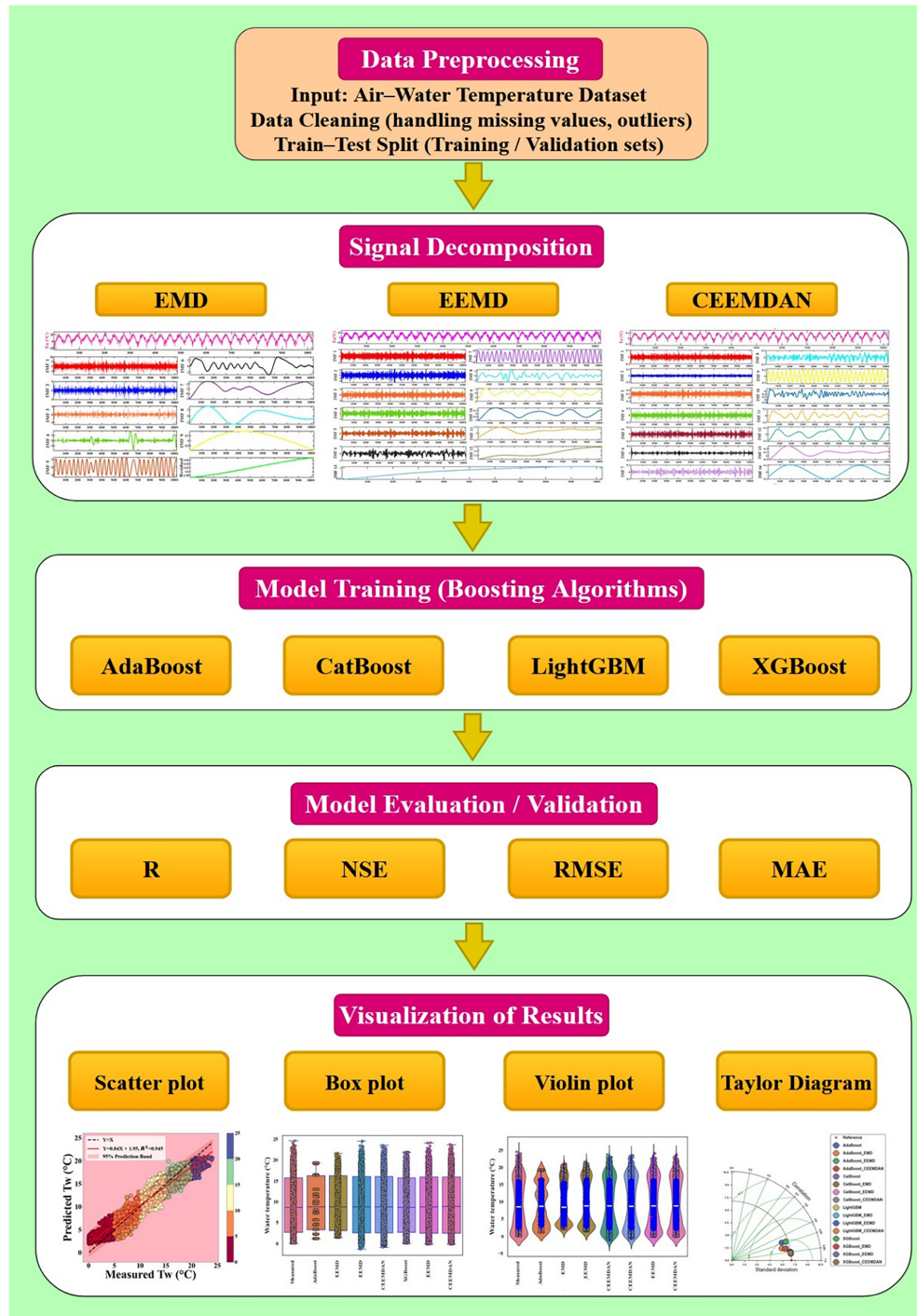


Figure 2: Flowchart of modelling water temperature-based signal decomposition

Table 3: Performances of different models at the *Lyna* station

Models	Training					Validation				
	R	NSE	RMSE	MAE	KGE	R	NSE	RMSE	MAE	KGE
AdaBoost	0.948	0.894	2.290	1.827	0.873	0.945	0.888	2.334	1.844	0.868
AdaBoost_EMD	0.973	0.931	1.852	1.538	0.852	0.973	0.930	1.848	1.532	0.850
AdaBoost_EEMD	0.981	0.957	1.459	1.226	0.909	0.980	0.953	1.508	1.259	0.907
AdaBoost_CEEMDAN	0.979	0.955	1.502	1.261	0.916	0.978	0.952	1.532	1.278	0.914
CatBoost	0.953	0.909	2.129	1.614	0.934	0.950	0.902	2.192	1.650	0.931
CatBoost_EMD	0.991	0.981	0.965	0.753	0.984	0.989	0.978	1.046	0.805	0.982
CatBoost_EEMD	0.995	0.991	0.682	0.530	0.992	0.990	0.981	0.966	0.734	0.989
CatBoost_CEEMDAN	0.996	0.991	0.656	0.507	0.993	0.991	0.981	0.951	0.729	0.988
LightGBM	0.954	0.911	2.107	1.595	0.935	0.948	0.898	2.234	1.688	0.931
LightGBM_EMD	0.998	0.996	0.427	0.324	0.995	0.993	0.986	0.828	0.617	0.987
LightGBM_EEMD	0.999	0.997	0.369	0.279	0.995	0.993	0.987	0.800	0.596	0.990
LightGBM_CEEMDAN	0.999	0.997	0.353	0.267	0.996	0.993	0.987	0.804	0.598	0.989
XGBoost	0.954	0.911	2.107	1.595	0.935	0.948	0.898	2.234	1.688	0.931
XGBoost_EMD	0.998	0.996	0.427	0.324	0.995	0.993	0.986	0.831	0.618	0.988
XGBoost_EEMD	0.999	0.997	0.369	0.278	0.995	0.993	0.987	0.804	0.598	0.990
XGBoost_CEEMDAN	0.999	0.997	0.355	0.268	0.995	0.993	0.986	0.818	0.610	0.988

Decomposition, especially CEEMDAN and EEMD, significantly enhances CatBoost's performance. In training, CEEMDAN boosts R to 0.996, NSE to 0.991, with an RMSE of 0.656, while validation metrics are similarly strong with R = 0.991, NSE = 0.981, and RMSE = 0.951. The small differences between training and validation suggest robust generalization and an effective fit to the data when using decomposed signals. LightGBM shows very similar performance to CatBoost without decomposition, with R = 0.954 and NSE = 0.911 in training, and slightly lower metrics in validation (R = 0.948, NSE = 0.898). The RMSE and MAE indicate that the model performs reasonably well but could benefit from additional signal enhancement. Like CatBoost, LightGBM also sees marked improvements with CEEMDAN and EEMD, which raise its R to 0.996 in training and maintain it at 0.993 in validation. The model's RMSE and MAE decrease significantly, demonstrating that LightGBM, when combined with decomposition, offers strong predictive accuracy and is well-suited to generalization across both datasets. XGBoost slightly outperforms other models in its raw form, with R values around 0.95 and NSE slightly above 0.91 across both training and validation, with lower RMSE and MAE values than the others. In addition to conventional error-based metrics, the Kling–Gupta Efficiency (KGE) values further confirm the superiority of decomposition-based hybrid models. For the *Lyna* station, models combining EEMD or CEEMDAN with XGBoost and LightGBM consistently achieved KGE values exceeding 0.98 during the validation stage, indicating improved balance between bias, variability, and correlation. This demonstrates that performance gains are not limited to reduced prediction errors, but also reflect enhanced hydrological consistency and stability of the simulated water temperature dynamics.

This indicates that XGBoost benefits the most from the decomposition techniques, as its performance was significantly improved when the input signal is decomposed. In contrast, without

decomposition, XGBoost ranks behind LightGBM and CatBoost. XGBoost sees the most dramatic improvement with decomposition, particularly EEMD and CEEMDAN. In training, XGBoost with EEMD achieves an R of 0.999 and an NSE of 0.997, and this high performance is mirrored in validation with $R = 0.993$ and $NSE = 0.987$. The low RMSE (0.804 in validation) and MAE (0.598 in validation) suggest that XGBoost combined with EEMD maintains exceptional accuracy and robustness across both stages. Overall, XGBoost with EEMD or CEEMDAN decomposition stands out as the top performer, providing consistent and superior accuracy and error minimization across both training and validation stages. CatBoost and LightGBM with CEEMDAN or EEMD also achieve very high performance, indicating that these models, when enhanced by decomposition techniques, are also reliable and robust. AdaBoost, while improved by decomposition, does not quite reach the same level of performance, particularly in the validation stage. This suggests that while all models benefit from decomposition, XGBoost paired with EEMD or CEEMDAN offers the best overall approach for generalizing water temperature predictions from air temperature data in Lyna station.

Fig. 3 gives scatter plots for the Lyna station, showing the relationship between predicted and measured water temperatures across various models (AdaBoost, CatBoost, LightGBM, and XGBoost) with and without signal decomposition techniques (EMD, EEMD, and CEEMDAN). The base models (especially AdaBoost) show more scatter and deviation from the 1:1 line, indicating higher prediction errors. Although XGBoost and CatBoost perform slightly better, there is still visible dispersion, especially at higher temperatures. CEEMDAN delivers the most accurate predictions, with data points closely following the 1:1 line across all models, especially LightGBM and XGBoost. These models achieve near-perfect R^2 values, indicating that CEEMDAN is particularly effective at reducing prediction errors and enhancing model fit. Taylor diagram given in Fig. 4 visualizes the performance of various models for the Lyna station in terms of correlation, standard deviation, and how closely each model's predictions align with the reference data (indicated by the red star). Models closer to the top of the diagram (near 1.0 on the correlation axis) indicate a higher correlation with the reference data, suggesting stronger predictive accuracy. Models with decomposition methods, particularly CEEMDAN, generally achieve higher correlation coefficients, indicating improved alignment with observed data. The models are plotted relative to the reference's standard deviation. Those closer to the reference point (star) have a standard deviation more similar to the observed data. XGBoost_CEEMDAN and LightGBM_CEEMDAN are positioned closest to the reference point, indicating that they not only have high correlations but also closely match the variability in the observed data. AdaBoost models tend to cluster with higher standard deviations and lower correlation coefficients compared to CatBoost, LightGBM, and XGBoost, indicating relatively lower accuracy.

Fig. 5 displays both a violin plot (top) and a box plot (bottom) for the Lyna station, comparing the distribution of water temperatures across different models (AdaBoost, CatBoost, LightGBM, and XGBoost) with and without decomposition methods (EMD, EEMD, and CEEMDAN) to the measured data. The measured data shows a broad distribution with a noticeable concentration around the median, indicating natural variability in water temperature. Models with EEMD and CEEMDAN decomposition tend to show distributions that more closely resemble the measured data in terms of shape, particularly for LightGBM and XGBoost. AdaBoost models, especially without decomposition, exhibit narrower distributions, suggesting that these models may not capture the full variability of the measured water temperatures. According to the box plots, the measured data has a median around the center of the interquartile range, reflecting a balanced distribution. Models using EEMD and CEEMDAN have box plots with medians and ranges that closely match the measured data, particularly for XGBoost and LightGBM. These models capture both the central tendency and

variability more accurately. While all models generally follow a similar range, AdaBoost tends to have a more compressed range, especially without decomposition, indicating less alignment with the natural variability of the measured temperatures.

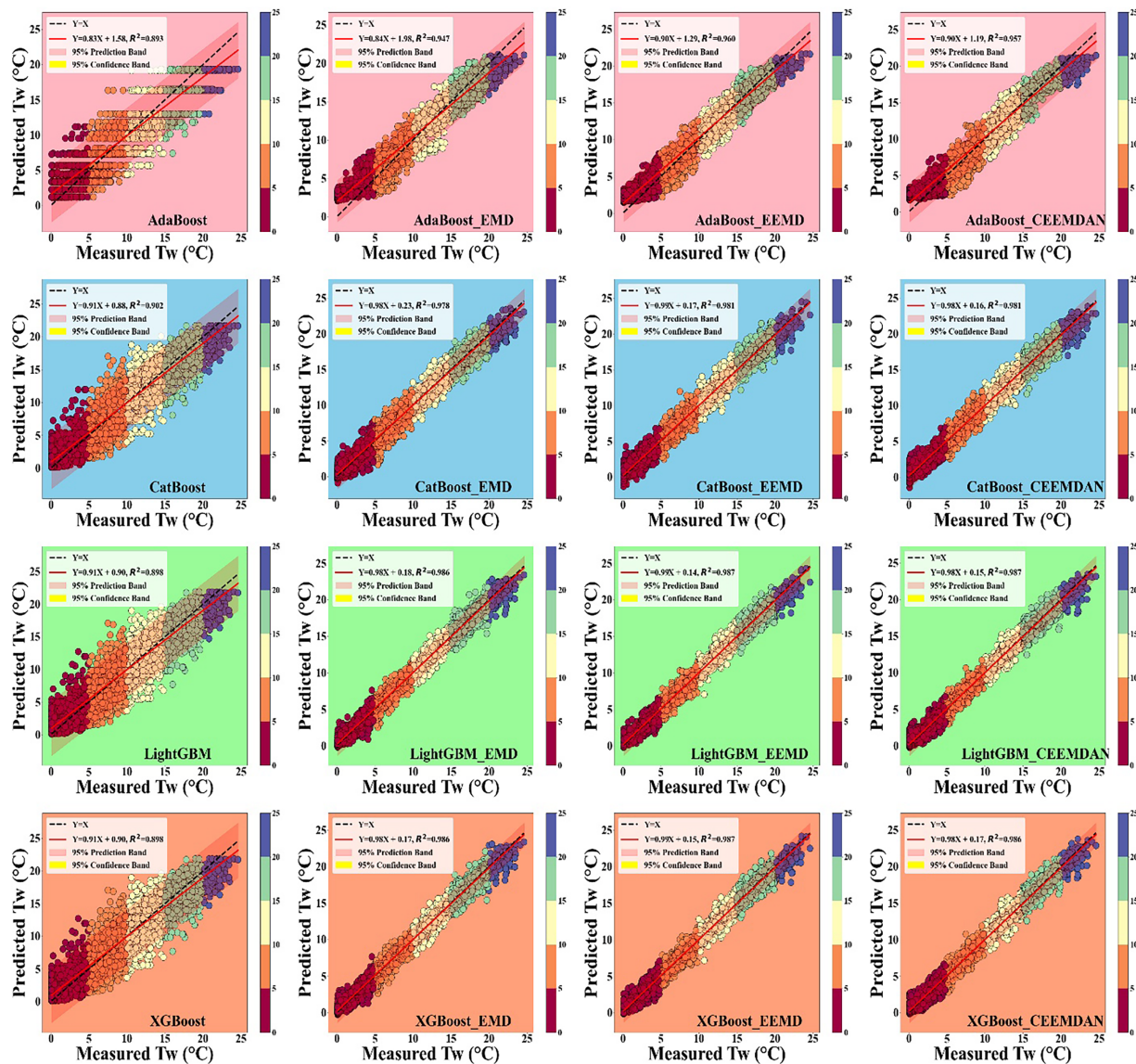


Figure 3: Scatterplots of measured against calculated water temperature (T_w) for the validation dataset: Lyna Station

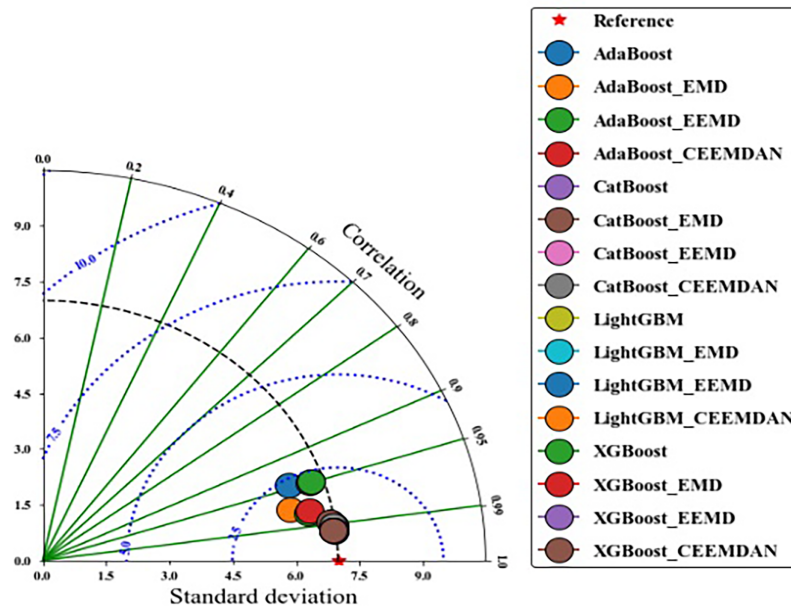


Figure 4: Taylor Diagram for models' comparison between measured and predicted water temperature for the validation dataset: *Lyna Station*

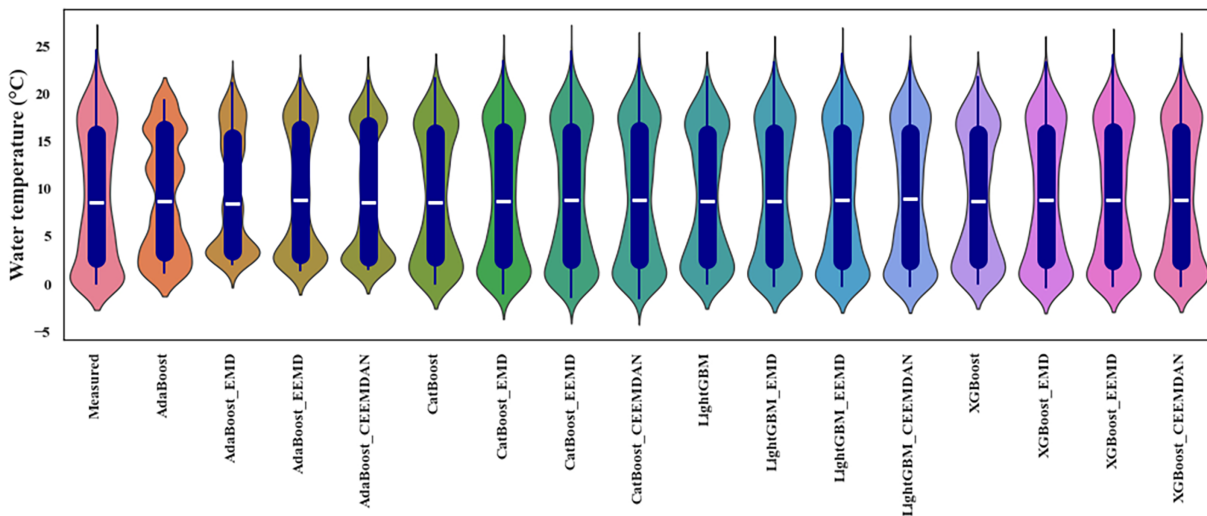


Figure 5: (Continued)

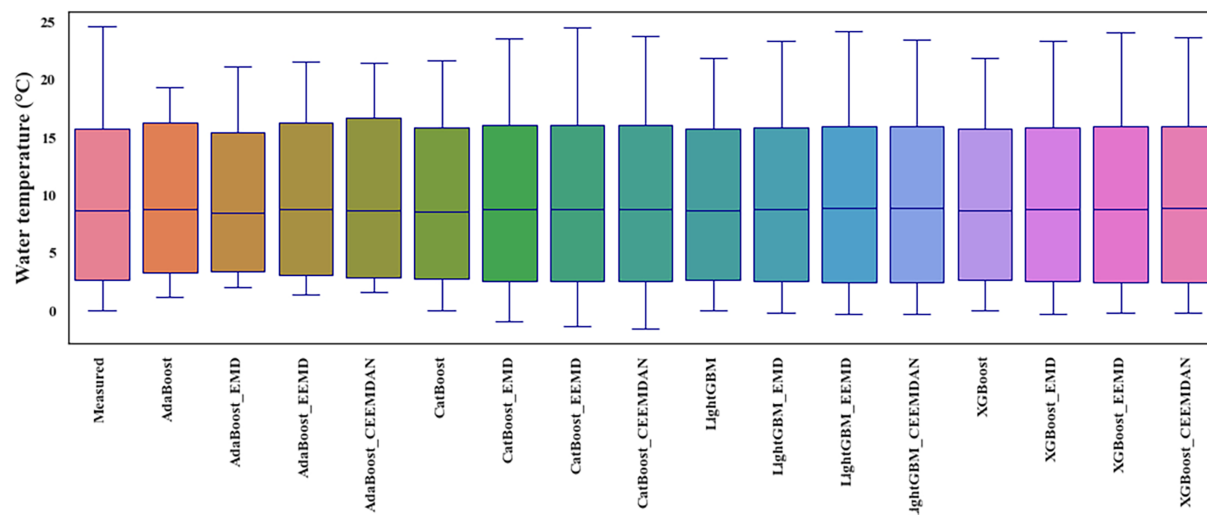


Figure 5: Violinplot upper panel and boxplot lower panel for models' comparison between measured and predicted water temperature for the validation dataset: *Lyna Station*

4.3 Results at Drweca Station

Table 4 presents the performance metrics of different machine learning models in estimating the water temperature of the Drweca station. AdaBoost shows moderate performance in both training and validation stages, with R values around 0.94 and NSE around 0.88. RMSE and MAE are relatively high (over 2.4 and 1.9, respectively), indicating that AdaBoost struggles to provide highly accurate predictions without decomposition. Decomposition techniques (especially EEMD) significantly improve AdaBoost's performance. EEMD achieves R of 0.979 in training and 0.977 in validation, with corresponding NSE values of approximately 0.95. However, there is a slight increase in RMSE from training to validation (1.555 to 1.582), suggesting some sensitivity to unseen data, though the model's accuracy is enhanced considerably with decomposition. CatBoost's base performance is consistent across stages, with R values around 0.945 and NSE near 0.89. RMSE and MAE are around 2.3 and 1.7, slightly outperforming AdaBoost without decomposition. CatBoost shows substantial gains with signal decomposition, particularly with EEMD and CEEMDAN. In training, CatBoost achieves an R of 0.995 and NSE of 0.990, which are mirrored in validation with similar high values (R $\approx$ 0.990 and NSE $\approx$ 0.980). RMSE remains below 1.0 in both stages, indicating consistent accuracy and error minimization. This suggests that CatBoost is robust and generalizes well when enhanced with decomposed signals. LightGBM shows slightly better base performance compared to AdaBoost and is on par with CatBoost, with R $\approx$ 0.947 and NSE $\approx$ 0.89 across both stages. However, its error metrics (RMSE around 2.3 and MAE around 1.7) indicate it could benefit from additional refinement. LightGBM achieves near-perfect performance with decomposition, particularly with CEEMDAN. The KGE metric further supports these findings for the Drweca station. Hybrid models based on EEMD and CEEMDAN, particularly when coupled with LightGBM and XGBoost, achieved consistently high KGE values, indicating improved agreement between observed and simulated mean values, variability, and temporal structure. This confirms that the decomposition-based framework enhances not only statistical accuracy but also the overall hydrological coherence of the predictions.

Table 4: Performances of different models at the *Drweca* station

Models	Training					Validation				
	R	NSE	RMSE	MAE	KGE	R	NSE	RMSE	MAE	KGE
AdaBoost	0.942	0.884	2.423	1.924	0.872	0.941	0.883	2.417	1.905	0.876
AdaBoost_EMD	0.975	0.932	1.856	1.568	0.846	0.972	0.930	1.865	1.559	0.853
AdaBoost_EEMD	0.979	0.952	1.555	1.301	0.908	0.977	0.950	1.582	1.314	0.918
AdaBoost_CEEMDAN	0.979	0.949	1.602	1.334	0.895	0.976	0.947	1.634	1.356	0.905
CatBoost	0.946	0.895	2.304	1.737	0.924	0.944	0.890	2.340	1.750	0.929
CatBoost_EMD	0.994	0.988	0.767	0.594	0.990	0.989	0.979	1.027	0.790	0.988
CatBoost_EEMD	0.995	0.990	0.697	0.535	0.992	0.990	0.981	0.984	0.744	0.990
CatBoost_CEEMDAN	0.995	0.990	0.694	0.537	0.993	0.990	0.980	0.999	0.761	0.990
LightGBM	0.947	0.897	2.284	1.720	0.925	0.942	0.888	2.367	1.768	0.929
LightGBM_EMD	0.998	0.997	0.419	0.313	0.995	0.993	0.986	0.834	0.613	0.990
LightGBM_EEMD	0.999	0.997	0.380	0.283	0.995	0.993	0.986	0.823	0.598	0.991
LightGBM_CEEMDAN	0.999	0.998	0.355	0.265	0.995	0.993	0.986	0.823	0.603	0.989
XGBoost	0.947	0.897	2.287	1.723	0.925	0.943	0.888	2.364	1.766	0.929
XGBoost_EMD	0.996	0.991	0.660	0.503	0.990	0.991	0.982	0.935	0.712	0.987
XGBoost_EEMD	0.996	0.992	0.624	0.473	0.991	0.991	0.983	0.922	0.696	0.990
XGBoost_CEEMDAN	0.996	0.993	0.610	0.464	0.991	0.991	0.983	0.925	0.692	0.988

In the training stage, CEEMDAN boosts R to 0.999 and NSE to 0.998, and these metrics are maintained in validation with only slight increases in RMSE and MAE (from 0.355 to 0.823 and 0.265 to 0.603). LightGBM's consistent metrics across training and validation reflect strong predictive accuracy and robustness, suggesting it handles decomposition-enhanced data exceptionally well. XGBoost's base performance is similar to LightGBM and CatBoost, with $R \approx 0.947$ and $NSE \approx 0.89$ across both stages. Its error metrics are slightly lower than those of LightGBM and CatBoost, indicating a better fit to the data in its base form. XGBoost also benefits significantly from decomposition, particularly with EEMD and CEEMDAN. In training, XGBoost with CEEMDAN achieves $R = 0.996$ and $NSE = 0.993$, with similarly high performance in validation ($R \approx 0.991$ and $NSE \approx 0.983$). Its RMSE and MAE remain low across both stages, with minimal discrepancies between training and validation. This demonstrates that XGBoost, especially with CEEMDAN or EEMD, is highly effective and maintains robust generalization across datasets. Overall, LightGBM and XGBoost, particularly with CEEMDAN or EEMD, provide the best balance of accuracy and consistency at the Drweca station. CatBoost with CEEMDAN or EEMD is also a strong option, while AdaBoost, though improved by decomposition, remains less competitive in comparison.

Fig. 6 illustrates the scatter plots for the Drweca station. The base models show moderate accuracy, with visible scatter around the 1:1 line, indicating prediction errors. AdaBoost, in particular, shows more deviation compared to the other models, with data points scattered more widely. XGBoost and CatBoost perform slightly better than AdaBoost and LightGBM in terms of alignment with the 1:1 line, but overall, the base models without decomposition indicate room for improvement. CEEMDAN yields the most accurate predictions for the Drweca station, with the highest alignment along the 1:1 line. XGBoost_CEEMDAN and LightGBM_CEEMDAN demonstrate near-perfect

R^2 values, indicating excellent predictive power and minimized errors. Even AdaBoost benefits significantly from CEEMDAN, with a marked reduction in scatter compared to its base form. However, XGBoost and LightGBM achieve the best overall accuracy with CEEMDAN, showcasing its effectiveness.

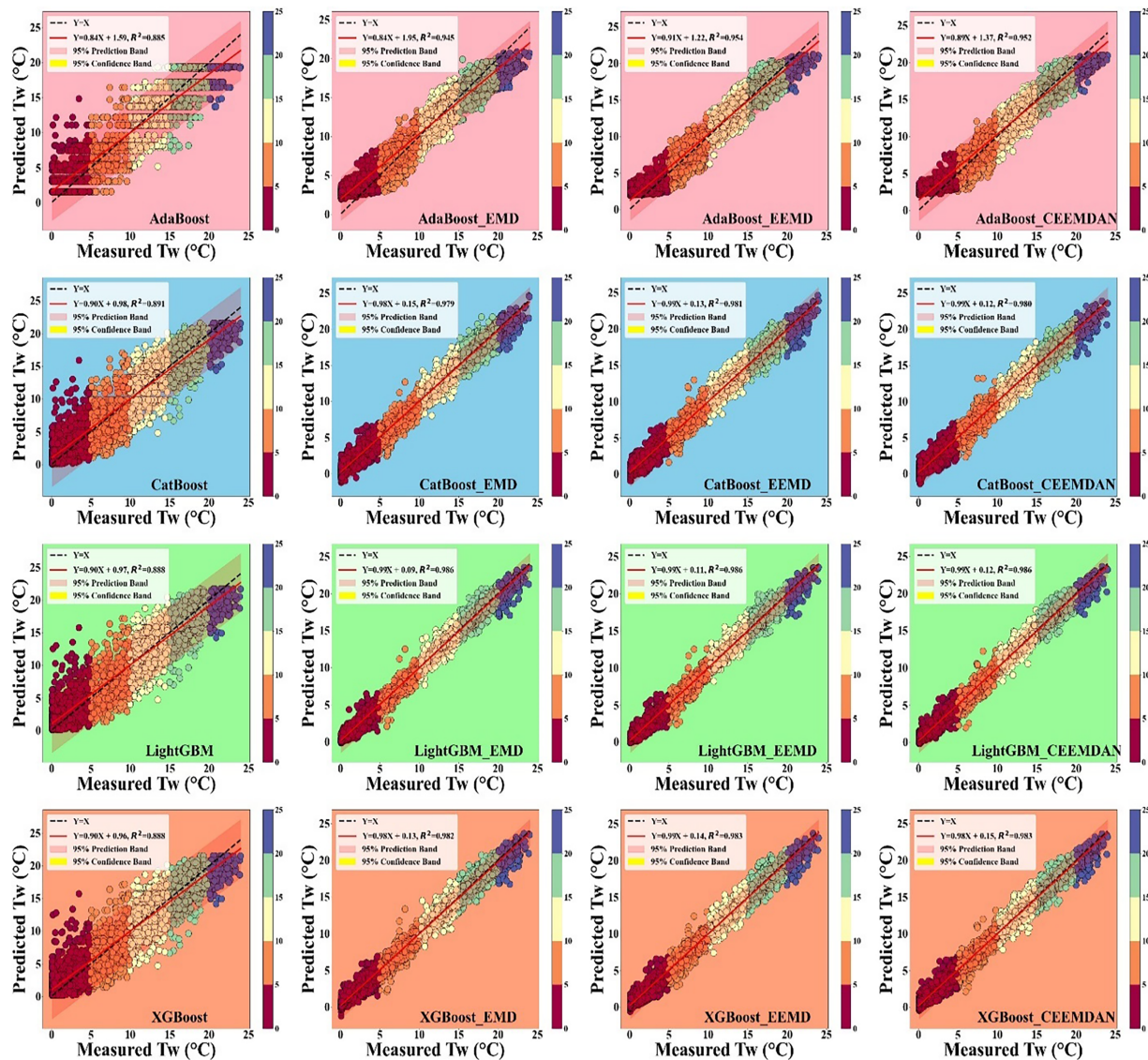


Figure 6: Scatterplots of measured against calculated water temperature (T_w) for the validation dataset: Drweca Station

Taylor diagram given in Fig. 7 represents the performance of different models for the Drweca station. The CEEMDAN enhanced models, particularly XGBoost_CEEMDAN and LightGBM_CEEMDAN, show very high correlations, indicating strong agreement with observed data. XGBoost_CEEMDAN and LightGBM_CEEMDAN are closest to the reference point, suggesting that these models not only have a high correlation but also closely match the observed standard deviation, making them highly reliable. AdaBoost models generally have a lower correlation and

exhibit greater distance from the reference point compared to other models, reflecting relatively lower accuracy. CatBoost, LightGBM, and XGBoost models enhanced by EEMD and CEEMDAN are positioned near the reference, showing higher accuracy and better representation of the observed data's variability. EMD models exhibit improvement over base models but generally lag behind EEMD and CEEMDAN in both correlation and standard deviation alignment. Fig. 8 shows both a violin plot (top) and a box plot (bottom) for the Drweca station. Models using EEMD and CEEMDAN, particularly LightGBM and XGBoost, show distributions that closely resemble the measured data in terms of spread and density, suggesting that these decompositions capture variability more effectively. AdaBoost, especially without decomposition, exhibits a more compressed and narrower distribution, indicating a less accurate capture of the temperature variability compared to other models. The box plot of the measured data shows a balanced distribution around the median. LightGBM and XGBoost models with EEMD and CEEMDAN exhibit medians and quartile ranges closely matching the measured data, indicating these models are capturing both central tendency and variability accurately. AdaBoost, especially without decomposition, has a narrower range and a different median alignment, reflecting less accuracy in capturing the full range of the measured data.

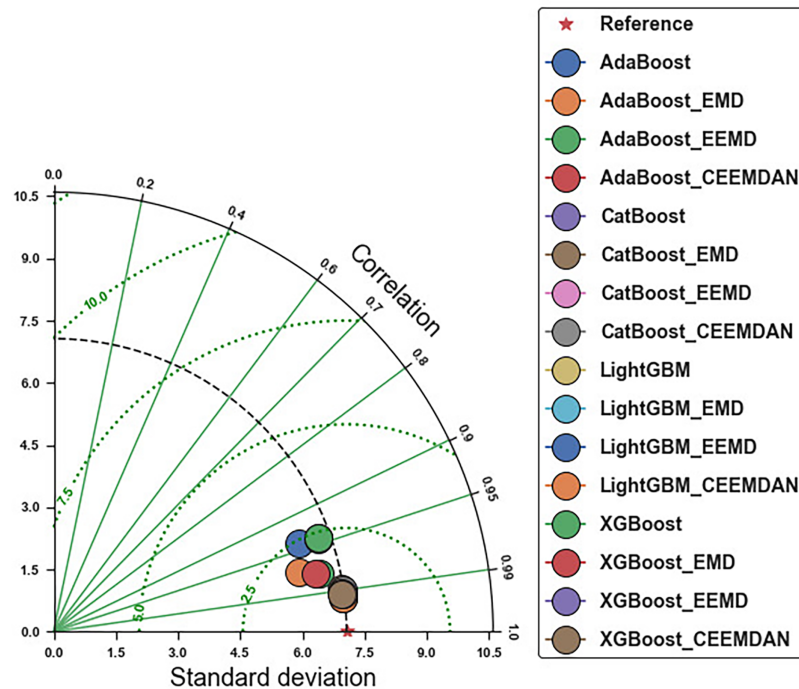


Figure 7: Taylor diagram for models' comparison between measured and predicted water temperature for the validation dataset: *Drweca* station

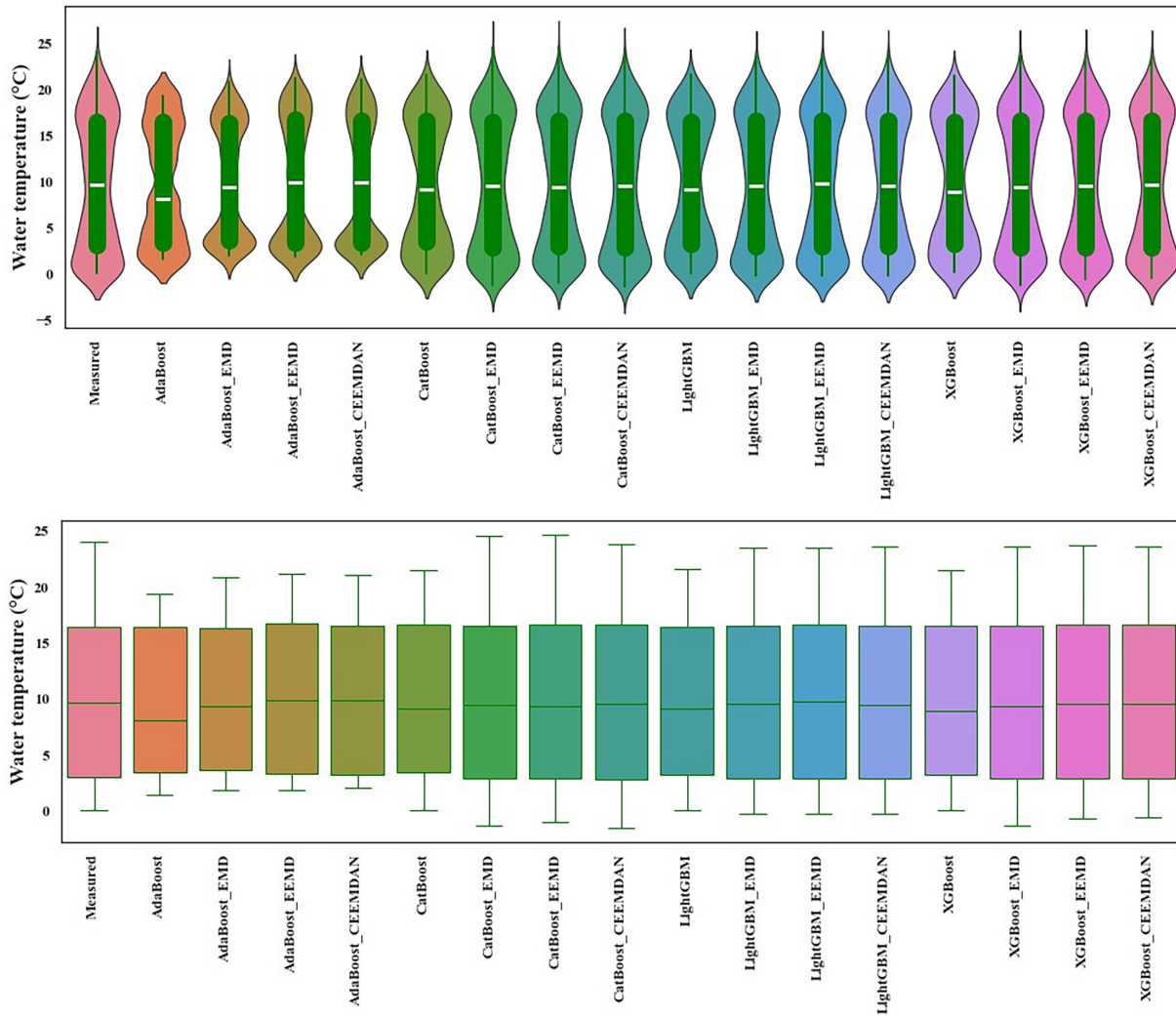


Figure 8: Violinplot upper panel and boxplot lower panel for models' comparison between measured and predicted water temperature for the validation dataset: *Drweca* station

4.4 Results at San Station

Table 5 presents the performance metrics of different machine learning models in estimating the water temperature of San station.

In both stages, AdaBoost exhibits moderate performance, with R values around 0.94 and NSE near 0.88. However, RMSE and MAE are high, exceeding 2.6 in training and 2.7 in validation, indicating that while consistent, AdaBoost struggles with accuracy. Decomposition techniques, particularly CEEMDAN, significantly enhance AdaBoost's performance. In training, CEEMDAN raises R to 0.976 and NSE to 0.943, with RMSE and MAE dropping to 1.868 and 1.576. In validation, this improvement is maintained, with R = 0.974 and NSE = 0.939, and RMSE and MAE also reduced (1.923 and 1.612). Although decompositions help, AdaBoost still falls short compared to other models with decomposition. CatBoost's base performance is slightly better than AdaBoost,

with R values around 0.95 and NSE around 0.90 across both stages. However, RMSE and MAE, above 2.4 and close to 1.9, suggest limited accuracy without decomposition. For the San station, the inclusion of KGE provides additional insight into model performance. Decomposition-enhanced models, especially XGBoost_EEMD and LightGBM_CEEMDAN, exhibited high KGE values during validation, reflecting a well-balanced representation of bias, variability, and correlation. These results indicate that the observed improvements are not solely driven by reduced errors, but also by improved physical consistency of the modeled water temperature series.

Table 5: Performances of different models at the *San* station

Models	Training					Validation				
	R	NSE	RMSE	MAE	KGE	R	NSE	RMSE	MAE	KGE
AdaBoost	0.943	0.884	2.656	2.107	0.865	0.939	0.877	2.728	2.185	0.856
AdaBoost_EMD	0.962	0.898	2.492	2.110	0.800	0.962	0.896	2.505	2.103	0.799
AdaBoost_EEMD	0.975	0.940	1.910	1.615	0.877	0.973	0.936	1.964	1.661	0.875
AdaBoost_CEEMDAN	0.976	0.943	1.868	1.576	0.883	0.974	0.939	1.923	1.612	0.882
CatBoost	0.949	0.901	2.457	1.847	0.928	0.945	0.893	2.542	1.915	0.923
CatBoost_EMD	0.993	0.985	0.951	0.727	0.989	0.987	0.974	1.250	0.954	0.982
CatBoost_EEMD	0.994	0.988	0.853	0.657	0.990	0.990	0.979	1.114	0.870	0.985
CatBoost_CEEMDAN	0.994	0.988	0.846	0.646	0.990	0.989	0.977	1.172	0.901	0.982
LightGBM	0.949	0.901	2.457	1.847	0.928	0.945	0.892	2.555	1.925	0.923
LightGBM_EMD	0.998	0.995	0.533	0.393	0.993	0.992	0.984	0.994	0.730	0.983
LightGBM_EEMD	0.998	0.996	0.467	0.346	0.994	0.992	0.984	0.982	0.726	0.984
LightGBM_CEEMDAN	0.998	0.997	0.440	0.325	0.994	0.993	0.985	0.950	0.699	0.983
XGBoost	0.949	0.901	2.461	1.850	0.928	0.945	0.892	2.551	1.924	0.923
XGBoost_EMD	0.995	0.989	0.818	0.618	0.988	0.990	0.979	1.118	0.838	0.979
XGBoost_EEMD	0.995	0.991	0.754	0.573	0.990	0.991	0.981	1.062	0.807	0.981
XGBoost_CEEMDAN	0.995	0.991	0.744	0.555	0.990	0.991	0.981	1.064	0.803	0.981

Decomposition considerably enhances CatBoost's accuracy and consistency. In training, CEEMDAN raises R to 0.994, NSE to 0.988, with RMSE at 0.846 and MAE at 0.646. These high metrics are well maintained in validation, where CEEMDAN gives an R of 0.989, NSE of 0.977, RMSE of 1.172, and MAE of 0.901. This indicates strong generalization and robustness, as the model handles decomposed signals well. LightGBM's performance is comparable to CatBoost, with $R \approx 0.949$ and $NSE \approx 0.90$, as well as RMSE and MAE in the 2.4–2.5 and 1.8–1.9 ranges. This base performance is solid but leaves room for improvement. LightGBM shows exceptional gains with decomposition, particularly with CEEMDAN. In training, it achieves $R = 0.998$ and $NSE = 0.997$, with very low RMSE and MAE (0.440 and 0.325). This accuracy translates well into validation, where CEEMDAN maintains $R = 0.993$, $NSE = 0.985$, and keeps RMSE and MAE low at 0.950 and 0.699. This suggests LightGBM with CEEMDAN is highly effective, capturing detailed patterns and maintaining predictive accuracy across datasets. XGBoost's base form mirrors LightGBM and CatBoost, with $R \approx 0.949$ and NSE around 0.90, along with RMSE and MAE around 2.4 and 1.9. These metrics are stable but not particularly high in terms of predictive accuracy. XGBoost achieves substantial improvement with decomposition, especially CEEMDAN. In training, CEEMDAN boosts R to 0.995

and NSE to 0.991, with RMSE and MAE reduced to 0.744 and 0.555. Validation results also reflect this enhancement, with $R = 0.991$, $NSE = 0.981$, RMSE at 1.064, and MAE at 0.803, indicating strong robustness and accuracy. Overall, LightGBM and XGBoost with CEEMDAN or EEMD provide the best balance of accuracy, consistency, and generalization for the San station, while CatBoost with decomposition is also a strong contender. AdaBoost, despite improvements, shows more variability and is less competitive, suggesting it is less suitable for this particular predictive task compared to the other models.

Fig. 9 provides the scatter plots for the San station. The base models exhibit considerable scatter around the 1:1 line, particularly AdaBoost, which shows noticeable deviation and higher prediction errors. XGBoost and CatBoost perform slightly better, with more data points closer to the 1:1 line, but they still have some dispersion. Overall, these base models display moderate accuracy, suggesting the need for decomposition methods to improve predictions. CEEMDAN yields the highest accuracy for the San station, with data points closely following the 1:1 line, particularly for LightGBM and XGBoost. These models achieve near-perfect R^2 values, indicating minimized errors and excellent predictive performance. Even AdaBoost shows marked improvement with CEEMDAN, although XGBoost and LightGBM stand out with the best overall fit and highest accuracy. Taylor diagram depicted in Fig. 10 represents the performance of various models at the San station. LightGBM_CEEMDAN and XGBoost_CEEMDAN show very high correlation values, suggesting a strong match with the observed water temperatures. XGBoost_CEEMDAN and LightGBM_CEEMDAN again are closest to the reference point, indicating they not only have a strong correlation but also capture the observed variability accurately. AdaBoost models generally have lower correlation and higher standard deviation compared to CatBoost, LightGBM, and XGBoost, reflecting comparatively lower accuracy. CEEMDAN-decomposed models, especially with XGBoost and LightGBM, achieve both high correlation and alignment with the standard deviation of the reference data, highlighting their accuracy and reliability.

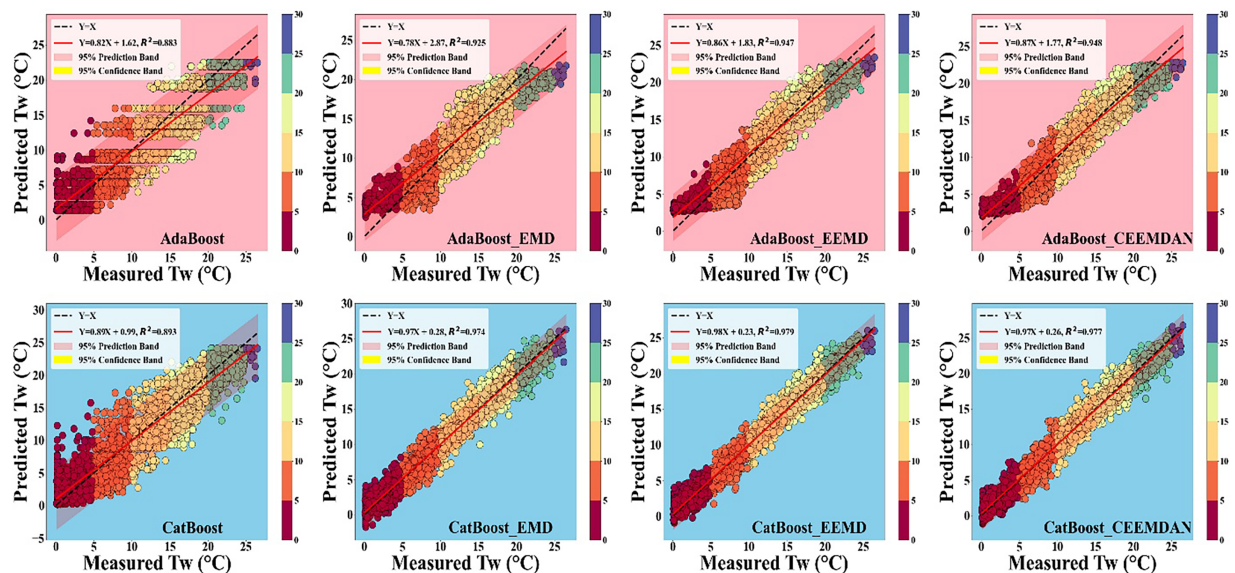


Figure 9: (Continued)

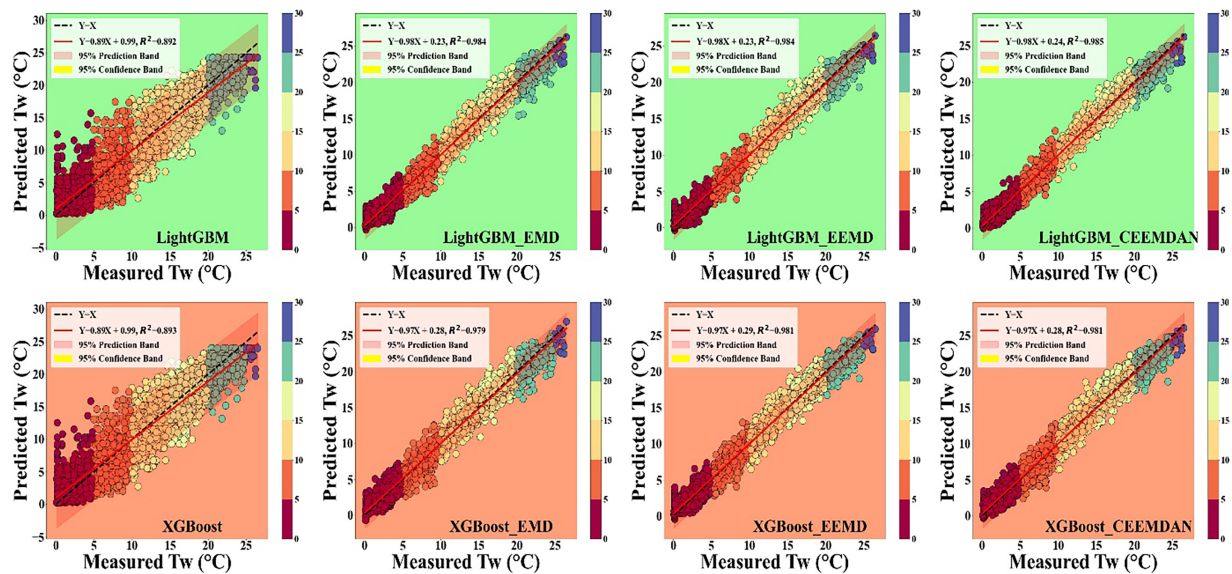


Figure 9: Scatterplots of measured against calculated water temperature (T_w) for the validation dataset: *San Station*

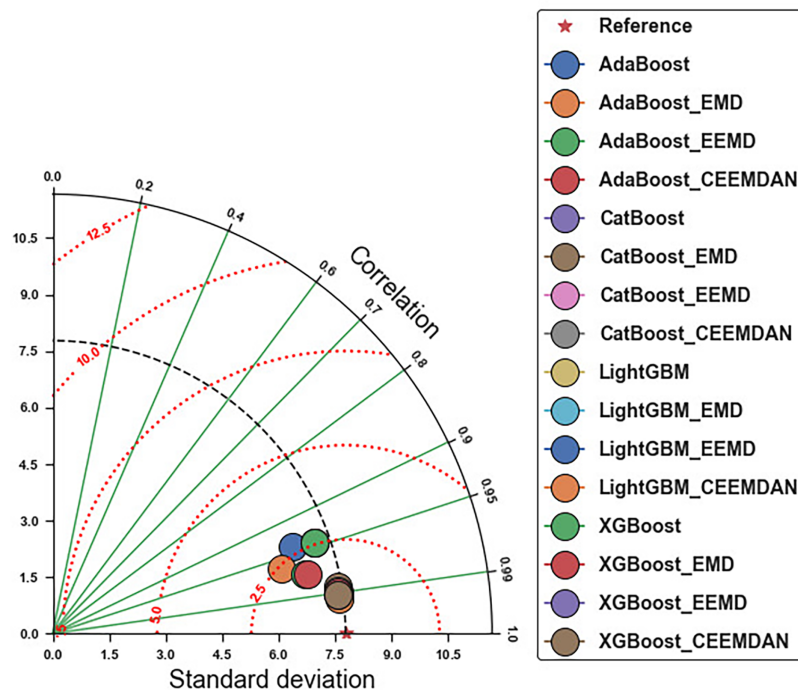


Figure 10: Taylor diagram for models' comparison between measured and predicted water temperature for the validation dataset: *San station*

Fig. 11 presents violin (top) and box plots (bottom) for the San station. Models with EEMD and CEEMDAN (especially for LightGBM and XGBoost) have distributions that closely resemble the measured data, with a similar spread and concentration around the median. This suggests these decompositions capture temperature variability effectively. AdaBoost models, particularly without decomposition, exhibit narrower distributions with less variability, indicating a less accurate fit to the observed data compared to models with decomposition. LightGBM and XGBoost models with EEMD and CEEMDAN decompositions display quartile ranges and medians that align closely with the measured data, indicating they capture both central tendency and variability well. AdaBoost models, especially without decomposition, have a more compressed range, suggesting they may not capture the observed data's full variability.

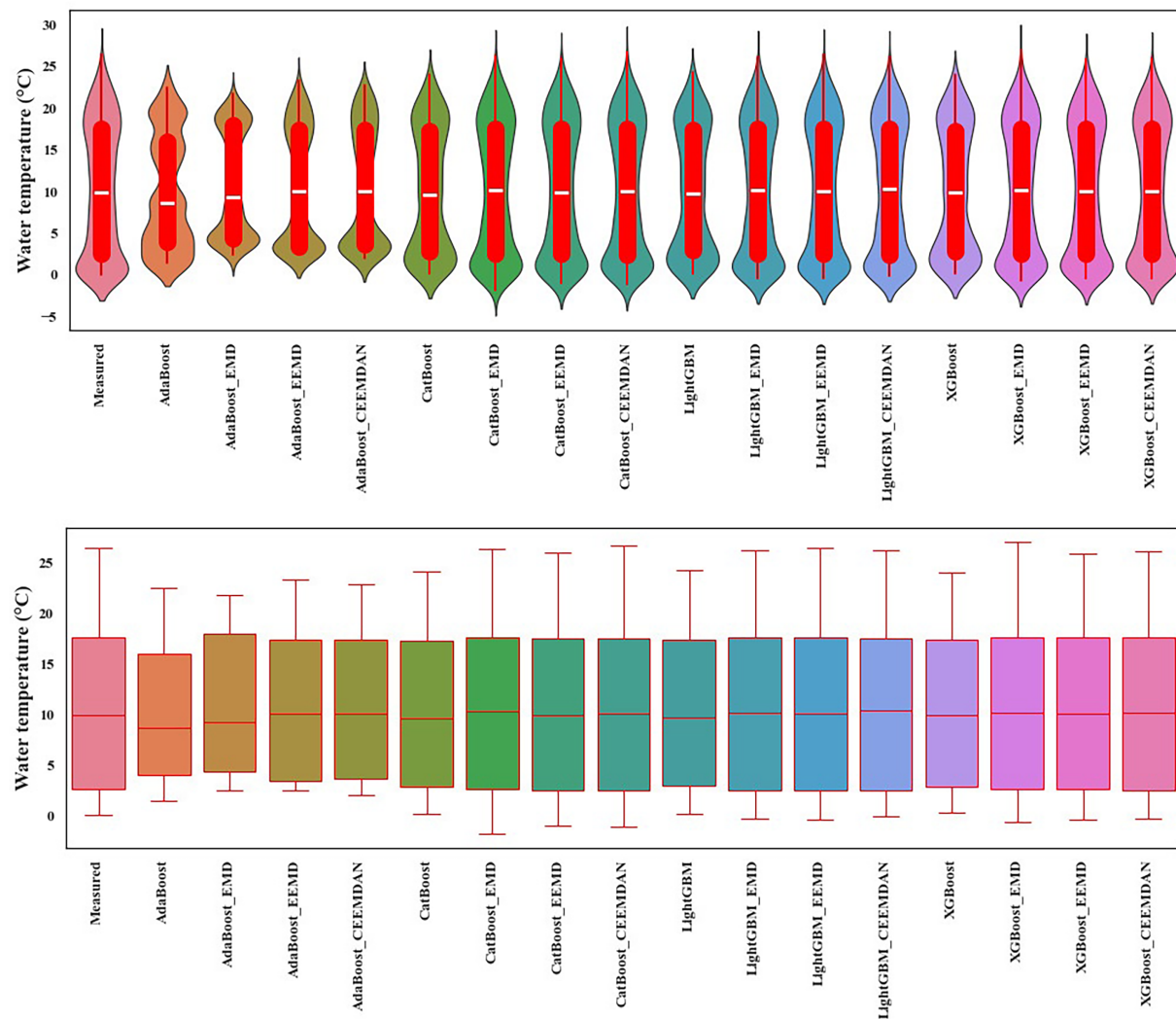


Figure 11: Violinplot upper panel and boxplot lower panel for models' comparison between measured and predicted water temperature for the validation dataset: *San* station

4.5 Results at Dunajec Station

The performance metrics of the machine learning models in estimating water temperature are shown in Table 6 for the Dunajec station. AdaBoost displays moderate performance with R values around 0.92 and NSE around 0.83–0.84 in both stages. However, RMSE and MAE are relatively high (above 2.8 and around 2.2, respectively), indicating limited accuracy and a tendency to produce larger errors in both training and validation. Decomposition methods enhance AdaBoost’s performance, especially EEMD, which raises R to 0.964 in training and 0.964 in validation. NSE improves to around 0.91, while RMSE and MAE reduce significantly in both stages (RMSE drops to around 2.1 and MAE to 1.8). Although decomposition helps AdaBoost achieve better predictive accuracy, it still falls behind other models with decomposition. CatBoost’s baseline performance is slightly better than AdaBoost’s, with R values near 0.92 and NSE around 0.85. RMSE and MAE are also slightly lower, suggesting a marginally better fit, though still limited without decomposition. Signal decomposition markedly improves CatBoost’s performance, with EEMD yielding the best results. During training, EEMD boosts R to 0.994 and NSE to 0.989, with reduced RMSE and MAE (around 0.72 and 0.57). In validation, CatBoost with EEMD maintains high accuracy with R = 0.989, NSE = 0.979, and RMSE and MAE around 1.01 and 0.78, respectively. These results indicate CatBoost’s ability to generalize well when trained with decomposed data. LightGBM shows similar baseline performance to CatBoost, with R ≈ 0.92 and NSE around 0.84–0.85 across both stages. However, it doesn’t outperform CatBoost without decomposition. Decomposition dramatically boosts LightGBM’s accuracy. CEEMDAN provides the best results, achieving R = 0.998 and NSE = 0.996 in training, with extremely low RMSE (0.409) and MAE (0.314). In validation, CEEMDAN maintains high metrics with R = 0.992, NSE = 0.983, and low RMSE and MAE (0.913 and 0.692). This indicates that LightGBM with CEEMDAN can capture complex patterns while minimizing error across both datasets. XGBoost performs on par with LightGBM and CatBoost in its baseline form, with R around 0.92 and NSE around 0.84 in both stages. RMSE and MAE indicate stable, though moderate, predictive accuracy. XGBoost’s performance improves notably with decomposition, particularly with EEMD. In training, EEMD raises R to 0.995 and NSE to 0.991, while keeping RMSE and MAE low (0.659 and 0.509). In validation, XGBoost maintains high performance with R = 0.991, NSE = 0.982, and RMSE and MAE around 0.937 and 0.724. This shows that XGBoost, especially with EEMD or CEEMDAN, offers robust generalization and handles complex data well. Overall, LightGBM and XGBoost with CEEMDAN or EEMD offer the best combination of accuracy and robustness at the Dunajec station, followed closely by CatBoost with EEMD. AdaBoost shows some benefits from decomposition but remains less competitive. A similar pattern is also observed when considering KGE. The highest-performing hybrid models based on EEMD and CEEMDAN achieved KGE values close to unity in the validation phase, demonstrating that multi-scale decomposition improves the balance between simulated and observed mean, variability, and temporal dynamics. This further confirms the robustness of the proposed hybrid framework under contrasting hydrological conditions.

Table 6: Performances of different models at the *Dunajec* station

Models	Training					Validation				
	R	NSE	RMSE	MAE	KGE	R	NSE	RMSE	MAE	KGE
AdaBoost	0.915	0.831	2.820	2.248	0.826	0.918	0.836	2.834	2.265	0.828
AdaBoost_EMD	0.956	0.868	2.492	2.119	0.751	0.958	0.870	2.520	2.141	0.746

(Continued)

Table 6 (continued)

Models	Training					Validation				
	R	NSE	RMSE	MAE	KGE	R	NSE	RMSE	MAE	KGE
AdaBoost_EEMD	0.964	0.908	2.089	1.793	0.824	0.964	0.905	2.157	1.852	0.813
AdaBoost_CEEMDAN	0.963	0.909	2.069	1.766	0.841	0.962	0.907	2.130	1.815	0.835
CatBoost	0.919	0.845	2.705	2.093	0.885	0.920	0.847	2.736	2.112	0.887
CatBoost_EMD	0.993	0.986	0.811	0.637	0.989	0.989	0.977	1.055	0.818	0.980
CatBoost_EEMD	0.994	0.989	0.720	0.566	0.991	0.989	0.979	1.012	0.785	0.984
CatBoost_CEEMDAN	0.994	0.988	0.739	0.574	0.991	0.989	0.978	1.035	0.806	0.981
LightGBM	0.921	0.848	2.679	2.071	0.888	0.918	0.843	2.775	2.141	0.887
LightGBM_EMD	0.998	0.995	0.473	0.362	0.993	0.992	0.984	0.880	0.666	0.979
LightGBM_EEMD	0.998	0.996	0.427	0.329	0.994	0.992	0.984	0.896	0.683	0.980
LightGBM_CEEMDAN	0.998	0.996	0.409	0.314	0.994	0.992	0.983	0.913	0.692	0.976
XGBoost	0.921	0.847	2.684	2.076	0.888	0.918	0.843	2.770	2.137	0.887
XGBoost_EMD	0.995	0.990	0.702	0.542	0.988	0.991	0.982	0.951	0.741	0.976
XGBoost_EEMD	0.995	0.991	0.659	0.509	0.989	0.991	0.982	0.937	0.724	0.977
XGBoost_CEEMDAN	0.995	0.990	0.685	0.524	0.988	0.990	0.980	0.982	0.764	0.975

Fig. 12 shows the scatter plots for the Dunajec station. The base models, particularly AdaBoost, display considerable scatter around the 1:1 line, indicating higher prediction errors. CatBoost and LightGBM perform slightly better, with reduced scatter and more points closer to the 1:1 line, suggesting better predictive accuracy. XGBoost shows moderate accuracy, but like the other base models, it would benefit from decomposition techniques to reduce errors further. CEEMDAN yields the highest accuracy, with data points closely following the 1:1 line, particularly for XGBoost and LightGBM. These models achieve R^2 values above 0.98, indicating near-perfect predictions with minimal error. Even AdaBoost sees significant improvement with CEEMDAN, although XGBoost and LightGBM demonstrate the best overall accuracy, with high alignment with the 1:1 line and reduced errors.

Taylor diagram illustrated in Fig. 13 represents the performance of various models at the Dunajec station. XGBoost_CEEMDAN and LightGBM_CEEMDAN exhibit very high correlations, indicating a strong alignment with the observed water temperature data. XGBoost_CEEMDAN and LightGBM_CEEMDAN are closest to the reference point, demonstrating that these models capture both the correlation and variability of the observed data accurately, which makes them highly reliable. AdaBoost models generally have lower correlation and are positioned farther from the reference compared to CatBoost, LightGBM, and XGBoost, indicating relatively lower accuracy. EMD improves model performance over the base, but EEMD and CEEMDAN further enhance performance. The EEMD and CEEMDAN versions of CatBoost, LightGBM, and XGBoost are consistently positioned closer to the reference, indicating higher predictive power. CEEMDAN-decomposed models, particularly XGBoost and LightGBM, demonstrate optimal performance with both high correlation and standard deviation close to the observed data, showcasing their superior accuracy.

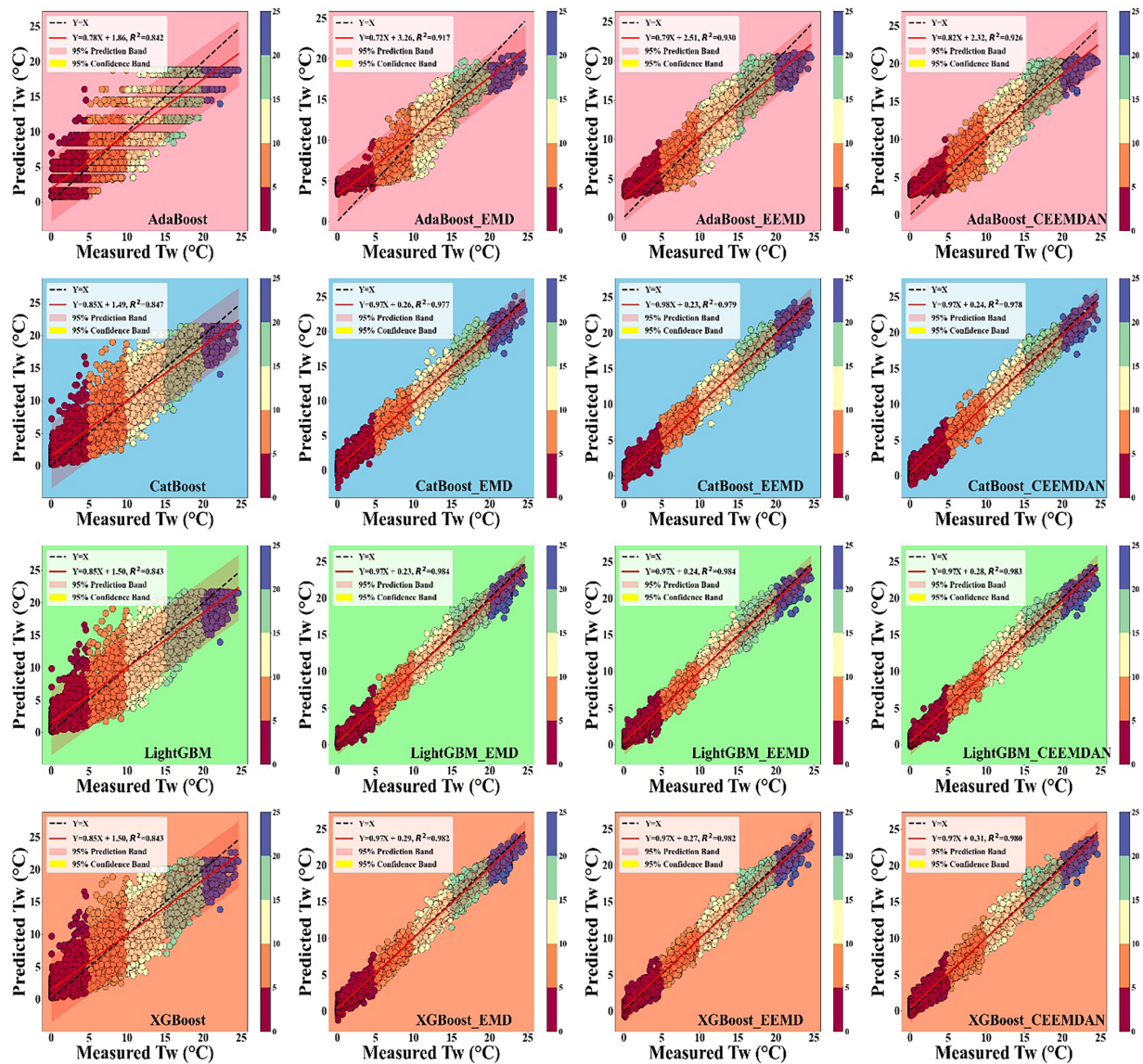


Figure 12: Scatterplots of measured against calculated water temperature (T_w) for the validation dataset: Dunajec station

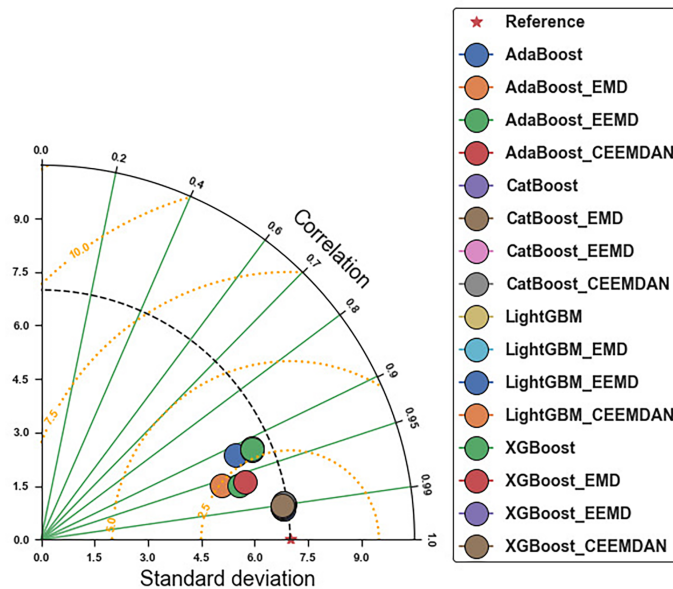


Figure 13: Taylor diagram for models' comparison between measured and predicted water temperature for the validation dataset: *Dunajec* station

Fig. 14 displays violin (top) and box plots (bottom) for the *Dunajec* station. Models with EEMD and CEEMDAN decompositions, especially for LightGBM and XGBoost, have distributions that closely match the measured data, suggesting that these decompositions are effective in capturing temperature variability. AdaBoost models, particularly without decomposition, exhibit narrower distributions, reflecting a more limited ability to capture the full range of temperature variability compared to the decomposed versions of CatBoost, LightGBM, and XGBoost. LightGBM and XGBoost with EEMD and CEEMDAN exhibit medians and interquartile ranges closely aligned with the measured data, indicating they capture both the central tendency and variability accurately. AdaBoost, especially without decomposition, shows a narrower interquartile range, indicating it may not fully capture the natural variability present in the measured data.

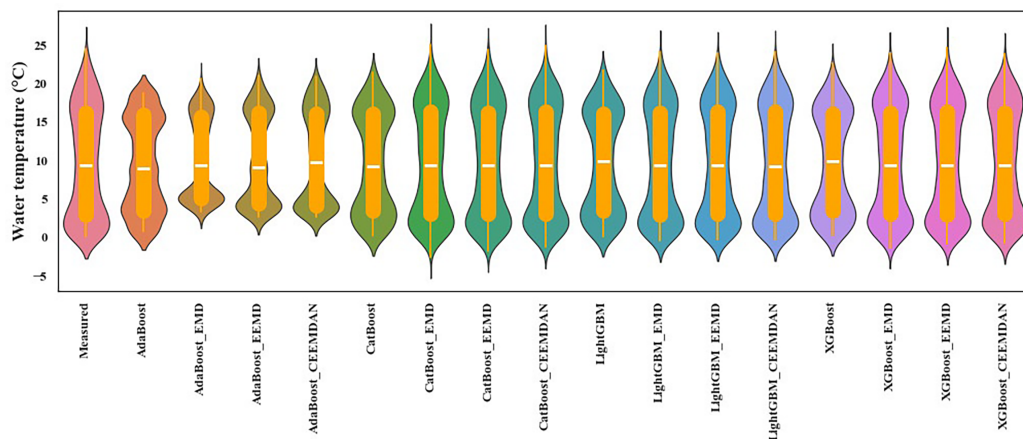


Figure 14: (Continued)

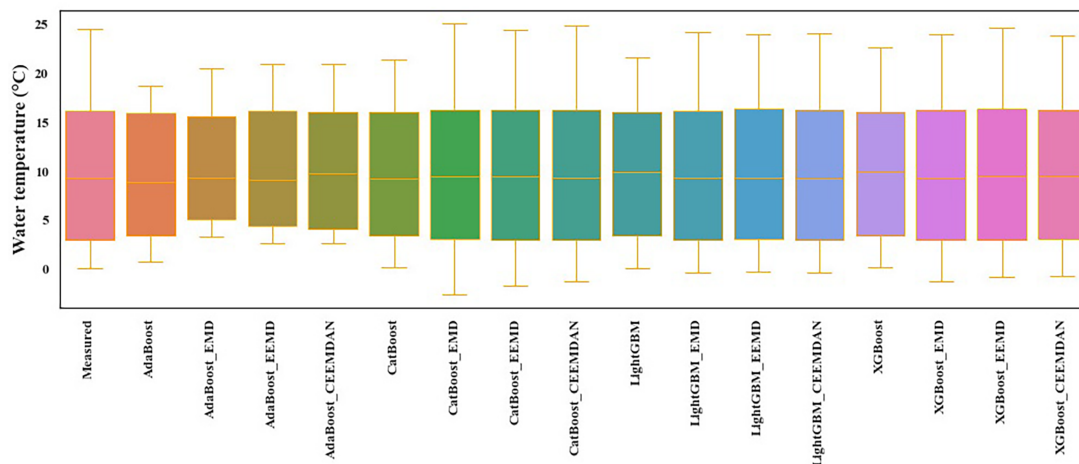


Figure 14: Violinplot upper panel and boxplot lower panel for models' comparison between measured and predicted water temperature for the validation dataset: *Dunajec* station

4.6 Discussion and Overall Comparison

AdaBoost generally has the lowest baseline performance across stations, with R values around 0.92, NSE between 0.83 and 0.89, and relatively high RMSE and MAE. This indicates that, without decomposition, AdaBoost struggles to capture the relationships between air and water temperature accurately. When EEMD or CEEMDAN is applied, AdaBoost sees improvements across all metrics. R values typically increase by about 5%–8%, and NSE improves by 10%–12%, indicating a stronger fit. The RMSE and MAE also decrease by approximately 20%–25% and 18%–22%, respectively. While these improvements are substantial, AdaBoost still does not match the enhanced accuracy of the other models after decomposition. CatBoost performs better than AdaBoost without decomposition, with R values around 0.94–0.95 and NSE ranging from 0.85 to 0.90 across stations. RMSE and MAE are moderately lower, indicating a better baseline fit than AdaBoost. CatBoost experiences significant improvements when EEMD or CEEMDAN is applied. R values typically increase by 4%–6%, and NSE improves by 8%–10%, reaching close to 0.99 across stations. RMSE decreases by 60%–70%, while MAE reduces by around 65%–70%, showing that CatBoost can greatly benefit from decomposition. These improvements indicate that CatBoost is highly capable of capturing complex data patterns when paired with decomposed signals. LightGBM performs similarly to CatBoost in its baseline form, with R values near 0.94–0.95 and NSE around 0.90. RMSE and MAE are relatively moderate, suggesting a solid starting accuracy. With EEMD or CEEMDAN, LightGBM achieves some of the highest accuracy improvements. R values typically increase by about 5%–7%, with NSE improving by 10%–12%, reaching nearly perfect levels around 0.99 in both training and validation. RMSE reductions are particularly notable, often improving by over 80%, while MAE decreases by about 70%–75%. This model consistently delivers excellent predictive accuracy and robustness, with decomposition making it one of the most reliable choices across all stations. XGBoost shows a strong baseline performance similar to LightGBM, with R around 0.94–0.95 and NSE slightly above 0.90. This indicates a solid ability to generalize even without decomposition. XGBoost sees dramatic improvements with decomposition techniques, especially EEMD. R values increase by 4%–6%, and NSE improves by about 10%–12%, reaching around 0.99. RMSE typically decreases by over 70%, and MAE improves by about 65%–70%, reflecting XGBoost's ability to leverage decomposed data for exceptional accuracy. Across stations, XGBoost with EEMD consistently delivers some of the

best generalization and error minimization, similar to LightGBM. LightGBM and XGBoost with CEEMDAN or EEMD emerge as the most effective models, showing consistent improvements across all stations, with near-perfect accuracy in both training and validation. Both models typically achieve 5%–7% improvements in R, 10%–12% in NSE, and substantial reductions in RMSE and MAE, often exceeding 70%. CatBoost also performs well with decomposition, though not quite matching the optimal accuracy of LightGBM and XGBoost. Its improvements are significant, particularly with a 65%–70% reduction in RMSE and MAE, making it a solid choice. AdaBoost, while improved with decomposition, still lags in comparison. Although it sees 20%–25% reductions in RMSE and MAE, its overall accuracy is limited, making it less suitable for this predictive task relative to the other models. In summary, LightGBM and XGBoost consistently provide the most reliable and accurate predictions across all stations when enhanced with decomposition techniques, particularly EEMD and CEEMDAN. This underscores the potential of advanced decomposition techniques in improving environmental temperature modeling.

Signal decomposition algorithms are very popular and robust approaches for improving the performances of single and standalone machine learning models. However, there are few similar studies related to modelling of water temperature based on signal decomposition. In the literature, we found that all reported studies that predict water temperature usually adopt the decomposition of air temperature or at least fewer other meteorological variables into various sub-signals included as inputs to the machine learning models, and the obtained results were compared with models without decomposition. Always with this approach, it was demonstrated that a hybrid model based on signal decomposition often guaranteed high improvement in terms of the model's performance or predictive accuracy. For example, Ref. [44] used the multivariate variational mode decomposition (MVMD) combined with various deep learning models, i.e., bidirectional long short-term memory network (BiLSTM), the LSTM, the gated recurrent network (GRU), the bidirectional GRU, and the Air2Stream physical model, for modelling hourly and daily water temperature measured at various stations located at the Menindee Lakes, lower Darling River, New South Wales, Australia. They used various climatic variables as input, i.e., T_a , relative humidity (RH%), solar radiation (SR), and evaporation (E_p). From the obtained results, it was found that, at the hourly time step, R and NSE values ranging from 0.952 to 0.955 and from 0.901 to 0.907 were obtained by the MVMD, while at the daily time step, the performances were superior exhibiting the values of $R \approx 0.984$ and $NSE \approx 0.965$, however, none of the proposed models by [44] was able to provide a value of R and NSE nearly equal to 0.993 and 0.987, and the best RMSE value obtained by the XGBoost_EEMD at Lyna station. In a previously published paper, Ref. [45] were among the first authors in the literature that have tried to use the EMD algorithms for T_w prediction measured at prawn culture ponds. They combine the multilayer perceptron neural network (MLPNN) with the EMD. Excellent results were obtained with RMSE value of approximately 0.484, without reporting the values of R and NSE. Ref. [46] used the variational mode decomposition (VMD) to improve the performances of the extreme learning machine models applied for predicting water T_w measured at two stations in Missouri River, USA. They compared ELM, outlier robust extreme learning machine (ORELM), regularized extreme learning machine (RELM), and the weighted regularized extreme learning machine (WRELM), with and without signal decomposition. The ORELM model have provided the best numerical performances with NSE and MAE values of approximately ≈ 0.924 and ≈ 2.645 , respectively, significantly lower than the value obtained in our present study. Ref. [47] used two signal decompositions, namely, the empirical wavelet transform (EWT) and maximum overlap discrete wavelet transform (MODWT) combined with the MLPNN, the optimally pruned extreme learning machine (OPELM), the adaptive boosting (AdaBoost), and the bootstrap aggregating (Bagging).

Excellent numerical performances were obtained with R, NSE, RMSE and MAE of approximately 0.99, 0.98, 0.718 and 0.599, respectively. All models were developed using data from five Polish rivers, making the comparison with the obtained results in the present study very attractive. Finally, Ref. [48] used the wavelet threshold denoising (WTD) coupled with the long short-term memory (LSTM) and transformer encoder for modelling T_w . They reported that, the hybrid LSTM-TE-WTD exhibited an NSE value of approximately 0.714, significantly lower than the value obtained in the present study.

Several studies have reported the use of deep learning models and their successful applications to both river and lake water temperature prediction, with varying levels of performance depending on the dataset characteristics and model configuration. Ref. [49] graph convolutional network-long short-term memory (GCN-LSTM) model for modelling water temperature calculated as the amplitude ratio (A_r) between air and water temperature. The model was developed based on hydroclimate features, i.e., air temperature, precipitation, shortwave radiation, and streamflow, combined with watershed physical features (e.g., slope). From the obtained results, excellent prediction results were obtained with R^2 and NSE values of 0.86 and 0.73 for one-day-ahead, significantly lower than the values obtained in our present study. Ref. [50] introduced a new hybrid deep learning models called CNN-LSTM-PID, i.e., the convolutional neural network-long short-term memory proportional integral-derivative (PID) error corrector for modelling water temperature measured at an aquatic pond. The hybrid CNN-LSTM-PID was compared with temporal convolutional network (TCN), TCN-LSTM, and CNN-LSTM, using various climatic variables namely, relative humidity, solar radiation, atmospheric pressure, wind speed, wind direction, rainfall, air temperature. However, based on correlation analysis only the air temperature and relative humidity were selected as relevant features. Results obtained revealed that the CNN-LSTM-PID was the most accurate having an RMSE and MAE of approximately 0.185 and 0.133, respectively. The significant contribution of the signal decomposition in improving the performances of the machine learning and deep learning for water temperature prediction was further highlighted by [51]. By combining wavelet noise reduction-deep learning algorithms, they obtained excellent prediction accuracy of water temperature with R reaching the values of 0.982 and 0.996, and RMSE values of 0.099 and 0.223, respectively, showing that denoising the air temperature data using wavelet decomposition, helped significantly in improving the performances of the deep learning models. Ref. [52] used the EEMD combined with boosting algorithms, i.e., the AdaBoost, and two deep learning models, i.e., LSTM and gated recurrent unit (GRU). Yet, three deep learning models were compared namely, EEMD-AdaBoost, EEMD-LSTM, and EEMD-GRU with their corresponding single models, i.e., the AdaBoost, LSTM, and GRU, using data collected from two stations in USA. Excellent results were obtained with RMSE = 0.192°C, MAE = 0.148°C, and $R^2 = 0.998$. In another study, Ref. [53] applied the LSTM deep learning model integrated with the ERA5-Land reanalysis using the Alaska water temperature database (AKTEMP) data set for modelling arctic river water temperature predictions. By integrating various features, i.e., the soil temperature, solar radiation, and thermal radiation associated with energy balance, excellent performances were achieved by the LSTM with NSE and RMSE values of approximately 0.95 and 1.0, respectively.

The findings of this study have significant implications for addressing the challenges posed by climate change and ensuring sustainable water resource management. Rising global temperatures and increasing climatic variability are expected to exacerbate the thermal regimes of river systems, with profound effects on aquatic ecosystems, water quality, and the availability of freshwater resources. By leveraging advanced hybrid models that integrate signal decomposition techniques, this work provides a robust and scalable framework for accurately predicting river water temperature under evolving climatic conditions. Such predictive capabilities are critical for informing adaptive management strategies, including the optimization of hydropower operations, the protection of biodiversity in

thermal-sensitive habitats, and the development of early warning systems to mitigate thermal stress in aquatic environments. Furthermore, the methodological advancements presented here can support decision-makers in devising data-driven policies aimed at balancing ecological conservation with human water needs, thereby contributing to a more resilient and sustainable management of freshwater systems in a warming world.

Although the proposed hybrid decomposition and boosting framework shows consistently high predictive performance across all the river stations studied, it is crucial to acknowledge some limitations regarding spatial and climatic generalization. All the river stations analyzed in this study are in Poland, so they share temperate climate characteristics with relatively similar atmospheric conditions. While the selected rivers vary significantly in their hydrological patterns, with some being lowland and others mountainous, applying the results to regions with different climates, such as arid, tropical, or monsoon-dominated areas, is uncertain. In arid and semi-arid regions, river water temperature dynamics are often greatly affected by solar radiation, evaporation, groundwater inflows, and changing flow conditions. This can weaken the strong connection between air and water temperature assumed in this study. Likewise, in tropical river systems, less seasonal temperature variation and stronger hydrological controls might change how well decomposition-based predictors work.

For rivers affected by human activities, like those controlled by dams or reservoirs, sudden thermal changes, managed flow, and altered residence times may not be fully represented by air temperature alone, even with multi-scale signal decomposition. In these cases, it might be necessary to include more hydrological or operational variables to maintain model reliability. Still, the proposed framework is particularly appealing for areas with limited data since it relies on a single, widely accessible predictor. Predictive uncertainty may rise with scant data or shorter records, but the decomposition and boosting strategy strikes a practical balance between model complexity and data needs. This makes it suitable for operational uses in stable climatic conditions.

5 Conclusion and Future Research

This study proposed a hybrid modeling framework that combines boosting algorithms with signal decomposition techniques to predict river water temperature using air temperature as the only predictor. The results clearly show that using multi-scale decomposition methods, especially EEMD and CEEMDAN, significantly improves model accuracy, reliability, and hydrological consistency across all tested stations.

Among the configurations tested, XGBoost with EEMD and LightGBM with CEEMDAN consistently delivered the best performance. They achieved notable reductions in RMSE and MAE, along with near-optimal values for R, NSE, and KGE in both training and validation phases. These findings confirm that decomposition-based preprocessing helps boosting algorithms better use the natural timing patterns of air temperature signals.

While the framework indicates strong potential for temperate river systems and situations with limited data, its application to rivers with different climates or those affected by human activities needs further study. Future research should aim to expand the framework by including more hydroclimatic predictors, exploring IMF selection strategies to enhance understanding, and evaluating performance in non-stationary and climate change scenarios.

Acknowledgement: During the preparation of this work the authors used generative AI in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization: Salim Heddam. Data curation: Mariusz Ptak and Mariusz Sojka. Formal analysis: Sungwon Kim, Mohammad Zounemat-Kermani, Fabio Di Nunno. Validation: Ozgur Kisi, Mohammad Zounemat-Kermani, Francesco Granata, Christoph Kulls. Supervision: Salim Heddam, Khaled Merabet, Francesco Granata, Ozgur Kisi. Writing original draft: All authors. Visualization: Mohammad Zounemat-Kermani, Sungwon Kim, Khaled Merabet. Investigation: All authors. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data presented in this study will be available on interested request from the corresponding author.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at <https://doi.org/10.23967/j.rimni.2026.10.78336>.

References

1. Caissie D. The thermal regime of rivers: A review. *Freshw Biol.* 2006;51(8):1389–406. doi:10.1111/j.1365-2427.2006.01597.x.
2. Webb BW, Hannah DM, Moore RD, Brown LE, Nobilis F. Recent advances in stream and river temperature research. *Hydrol Process.* 2008;22(7):902–18. doi:10.1002/hyp.6994.
3. Benyahya L, Caissie D, Satish MG, El-Jabi N. Long-wave radiation and heat flux estimates within a small tributary in Catamaran Brook (New Brunswick, Canada). *Hydrol Process.* 2012;26(4):475–84. doi:10.1002/hyp.8141.
4. Chen C, Xue X. Multi-step ahead prediction of lake water temperature using neural network and physically-based model. *J Hydraul Res.* 2024;62(4):370–82. doi:10.1080/00221686.2024.2377132.
5. Kim HI, Kim D, Salamattalab MM, Mahdian M, Bateni SM, Noori R. Machine learning-based modeling of surface water temperature dynamics in arctic lakes. *Environ Sci Pollut Res.* 2024;31(49):59642–55. doi:10.1007/s11356-024-35173-x.
6. Bonaglia A, Shen C, Padrón RS, Bogner K, Fopp F, Rubin A, et al. Sub-seasonal forecasting of thermal stress for Swiss river fishes during heatwaves. *Ecol Model.* 2025;507:111171. doi:10.1016/j.ecolmodel.2025.111171.
7. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond A.* 1998;454(1971):903–95. doi:10.1098/rspa.1998.0193.
8. Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal.* 2009;01(01):1–41. doi:10.1142/S1793536909000047.
9. Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2011 May 22–27. Prague, Czech Republic. p. 4144–7. doi:10.1109/icassp.2011.5947265.
10. Ouyang Q, Lu W, Xin X, Zhang Y, Cheng W, Yu T. Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. *Water Resour Manage.* 2016;30(7):2311–25. doi:10.1007/s11269-016-1288-8.

11. Zhang X, Zhang Q, Zhang G, Nie Z, Gui Z, Que H. A novel hybrid data-driven model for daily land surface temperature forecasting using long short-term memory neural network based on ensemble empirical mode decomposition. *Int J Environ Res Public Health*. 2018;15(5):1032. doi:10.3390/ijerph15051032.
12. Bera K, Raj P, Selvaraju N. Forecasting financial volatility: a hybrid approach integrating ICEEMDAN and asymmetric volatility models. *J Financ Data Sci*. 2025;7(1):54–75. doi:10.3905/jfds.2024.1.174.
13. Ouadi B, Khatir A, Magagnini E, Mokadem M, Abualigah L, Smerat A. Optimizing silt density index prediction in water treatment systems using pressure-based gradient boosting hybridized with Salp Swarm Algorithm. *J Water Process Eng*. 2024;68:106479. doi:10.1016/j.jwpe.2024.106479.
14. Bouhacina O, Khatir A, Matoug SA, Tamine T. A novel hybrid TOARS-optimized ensemble of tree-based models for predicting soil temperature at shallow depths. *Sustainability*. 2025;18(1):294. doi:10.3390/su18010294.
15. Wrzesiński D. Entropia odpływu rzek w Polsce. In: *Studia i Prace z Geografii i Geologii*. Vol. 33. Poznań, Poland: Bogucki Wydawnictwo Naukowe; 2013.
16. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. arXiv:1706.09516. 2018. doi:10.48550/arXiv.1706.09516.
17. Huang G, Wu L, Ma X, Zhang W, Fan J, Yu X, et al. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J Hydrol*. 2019;574:1029–41. doi:10.1016/j.jhydrol.2019.04.085.
18. Dong L, Zeng W, Wu L, Lei G, Chen H, Srivastava AK, et al. Estimating the pan evaporation in northwest China by coupling CatBoost with bat algorithm. *Water*. 2021;13(3):256. doi:10.3390/w13030256.
19. Jabeur SB, Gharib C, Mefteh-Wali S, Arfi WB. CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technol Forecast Soc Change*. 2021;166:120658. doi:10.1016/j.techfore.2021.120658.
20. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. *J Big Data*. 2020;7(1):94. doi:10.1186/s40537-020-00369-8.
21. Zhou F, Pan H, Gao Z, Huang X, Qian G, Zhu Y, et al. Fire prediction based on CatBoost algorithm. *Math Probl Eng*. 2021;2021:1929137. doi:10.1155/2021/1929137.
22. Bo Y, Liu Q, Huang X, Pan Y. Real-time hard-rock tunnel prediction model for rock mass classification using CatBoost integrated with Sequential Model-Based Optimization. *Tunn Undergr Space Technol*. 2022;124:104448. doi:10.1016/j.tust.2022.104448.
23. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016 Aug 13–17. San Francisco, CA, USA. p. 785–94. doi:10.1145/2939672.2939785.
24. Meddage DPP, Ekanayake IU, Weerasuriya AU, Lewangamage CS, Tse KT, Miyanawala TP, et al. Explainable machine learning (XML) to predict external wind pressure of a low-rise building in Urban-like settings. *J Wind Eng Ind Aerodyn*. 2022;226:105027. doi:10.1016/j.jweia.2022.105027.
25. Hosseini S, Pourmirzaee R, Armaghani DJ, Sabri Sabri MM. Prediction of ground vibration due to mine blasting in a surface lead-zinc mine using machine learning ensemble techniques. *Sci Rep*. 2023;13(1):6591. doi:10.1038/s41598-023-33796-7.
26. Liu Y, Liu L, Yang L, Hao L, Bao Y. Measuring distance using ultra-wideband radio technology enhanced by extreme gradient boosting decision tree (XGBoost). *Autom Constr*. 2021;126:103678. doi:10.1016/j.autcon.2021.103678.
27. Amjad M, Ahmad I, Ahmad M, Wróblewski P, Kamiński P, Amjad U. Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation. *Appl Sci*. 2022;12(4):2126. doi:10.3390/app12042126.
28. Sagi O, Rokach L. Approximating XGBoost with an interpretable decision tree. *Inf Sci*. 2021;572:522–42. doi:10.1016/j.ins.2021.05.055.

29. Yun KK, Yoon SW, Won D. Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Syst Appl.* 2021;186:115716. doi:10.1016/j.eswa.2021.115716.
30. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci.* 1997;55(1):119–39. doi:10.1006/jcss.1997.1504.
31. Liu H, Tian HQ, Li YF, Zhang L. Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. *Energy Convers Manag.* 2015;92:67–81. doi:10.1016/j.enconman.2014.12.053.
32. Tsai JK, Hung CH. Improving AdaBoost classifier to predict enterprise performance after COVID-19. *Mathematics.* 2021;9(18):2215. doi:10.3390/math9182215.
33. Demir S, Sahin EK. An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost. *Neural Comput Applic.* 2023;35(4):3173–90. doi:10.1007/s00521-022-07856-4.
34. Busari GA, Lim DH. Crude oil price prediction: a comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. *Comput Chem Eng.* 2021;155:107513. doi:10.1016/j.compchemeng.2021.107513.
35. Li W, Suhayb MK, Thangavelu L, Abdulameer Marhoon H, Pustokhina I, Alqsair UF, et al. Implementation of AdaBoost and genetic algorithm machine learning models in prediction of adsorption capacity of nanocomposite materials. *J Mol Liq.* 2022;350:118527. doi:10.1016/j.molliq.2022.118527.
36. Chakraborty D, Elhegazy H, Elzarka H, Gutierrez L. A novel construction cost prediction model using hybrid natural and light gradient boosting. *Adv Eng Inform.* 2020;46:101201. doi:10.1016/j.aei.2020.101201.
37. Sun Y, Zhang F, Lin H, Xu S. A forest fire susceptibility modeling approach based on light gradient boosting machine algorithm. *Remote Sens.* 2022;14(17):4362. doi:10.3390/rs14174362.
38. Guo J, Yun S, Meng Y, He N, Ye D, Zhao Z, et al. Prediction of heating and cooling loads based on light gradient boosting machine algorithms. *Build Environ.* 2023;236:110252. doi:10.1016/j.buildenv.2023.110252.
39. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst.* 2017;30:3149–57.
40. Li L, Liu Z, Shen J, Wang F, Qi W, Jeon S. A LightGBM-based strategy to predict tunnel rock-mass class from TBM construction data for building control. *Adv Eng Inform.* 2023;58:102130. doi:10.1016/j.aei.2023.102130.
41. Fan J, Ma X, Wu L, Zhang F, Yu X, Zeng W. Light Gradient Boosting Machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric Water Manage.* 2019;225:105758. doi:10.1016/j.agwat.2019.105758.
42. Shehadeh A, Alshboul O, Al Mamlook RE, Hamedat O. Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression. *Autom Constr.* 2021;129:103827. doi:10.1016/j.autcon.2021.103827.
43. Xu K, Han Z, Xu H, Bin L. Rapid prediction model for urban floods based on a light gradient boosting machine approach and hydrological-hydraulic model. *Int J Disaster Risk Sci.* 2023;171:81. doi:10.1007/s13753-023-00465-2.
44. Briceno Medina L, Joehnk K, Deo RC, Ali M, Prasad SS, Downs N. Forecasting river water temperature using explainable artificial intelligence and hybrid machine learning: case studies in menindee region in Australia. *Water.* 2024;16(24):3720. doi:10.3390/w16243720.
45. Liu S, Xu L, Li D. Multi-scale prediction of water temperature using empirical mode decomposition with back-propagation neural networks. *Comput Electr Eng.* 2016;49:1–8. doi:10.1016/j.compeleceng.2015.10.003.
46. Mirzania E, Roshni T, Ghorbani MA, Heddami S. River water temperature prediction using a hybrid model based on variational mode decomposition (VMD) and outlier robust extreme learning machine. *Environ Process.* 2024;11(3):39. doi:10.1007/s40710-024-00716-4.

47. Heddami S, Merabet K, Difi S, Kim S, Ptak M, Sojka M, et al. River water temperature prediction using hybrid machine learning coupled signal decomposition: EWT versus MODWT. *Ecol Inform.* 2023;78:102376. doi:10.1016/j.ecoinf.2023.102376.
48. Chen J, Xue X. A transfer learning-based long short-term memory model for the prediction of river water temperature. *Eng Appl Artif Intell.* 2024;133:108605. doi:10.1016/j.engappai.2024.108605.
49. Behbahani MRM, Rey DM, Briggs MA, Bagtzoglou AC. A spatiotemporal deep learning approach for predicting daily air-water temperature signal coupling and identification of key watershed physical parameters in a montane watershed. *J Hydrol.* 2025;663:134139. doi:10.1016/j.jhydrol.2025.134139.
50. Zhou X, Hao Y, Liu Y, Dang L, Qiao B, Zuo X. Short-term prediction of dissolved oxygen and water temperature using deep learning with dual proportional-integral-derivative error corrector in pond culture. *Eng Appl Artif Intell.* 2025;142:109964. doi:10.1016/j.engappai.2024.109964.
51. Cheng T, Wang J, Sui J, Song F, Fu H, Wang T, et al. Simulation and prediction of water temperature in a water transfer channel during winter periods using a new approach based on the wavelet noise reduction-deep learning method. *J Hydrol Hydromech.* 2024;72(1):49–63. doi:10.2478/johh-2023-0036.
52. Gheisari M, Shafi J, Kosari S, Amanabadi S, Mehdizadeh S, Fernandez Campusano C, et al. Development of improved deep learning models for multi-step ahead forecasting of daily river water temperature. *Eng Appl Comput Fluid Mech.* 2025;19(1):2450477. doi:10.1080/19942060.2025.2450477.
53. Chang SY, Schwenk J, Solander KC. Deep learning advances arctic river water temperature predictions. *Water Resour Res.* 2025;61(6):e2024WR039053. doi:10.1029/2024WR039053.