water

MDPI

Article

Forecasting River Water Temperature Using Explainable Artificial Intelligence and Hybrid Machine Learning: Case Studies in Menindee Region in Australia

Leyde Briceno Medina [1], Klaus Joehnk [3], Ravinesh C. Deo [1,2,*], Mumtaz Ali [5], Salvin S. Prasad [4] and Nathan Downs [2,6]

1 Artificial Intelligence Applications Laboratory, School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia; u1136811@umail.usq.edu.au
2 Centre for Applied Climate Sciences, University of Southern Queensland, Springfield, QLD 4300, Australia; nathan.downs@usq.edu.au
3 Commonwealth Scientific and Industrial Research Organisation (CSIRO), Environment, Canberra, ACT 2601, Australia; klaus.joehnk@csiro.au
4 School of Sciences, College of Engineering, and Technical Vocational Education and Training, Fiji National University, Samabula, Suva P.O. Box 3722, Fiji; psdshalvin@gmail.com
5 UniSQ College, Springfield Campus, University of Southern Queensland, Springfield, QLD 4300, Australia; mumtaz.ali@unisq.edu.au
6 School of Mathematics, Physics and Computing, Toowoomba Campus, University of Southern Queensland, Toowoomba, QLD 4350, Australia
* Correspondence: ravinesh.deo@unisq.edu.au

Abstract: Water temperature (WT) is a crucial factor indicating the quality of water in the river system. Given the significant variability in water quality, it is vital to devise more precise methods to forecast temperature in river systems and assess the water quality. This study designs and evaluates a new explainable artificial intelligence and hybrid machine-learning framework tailored for hourly and daily surface WT predictions for case studies in the Menindee region, focusing on the Weir 32 site. The proposed hybrid framework was designed by coupling a nonstationary signal processing method of Multivariate Variational Mode Decomposition (MVMD) with a bidirectional long short-term memory network (BiLSTM). The study has also employed a combination of in situ measurements with gridded and simulation datasets in the testing phase to rigorously assess the predictive performance of the newly designed MVMD-BiLSTM alongside other benchmarked models. In accordance with the outcomes of the statistical score metrics and visual infographics of the predicted and observed WT, the objective model displayed superior predictive performance against other benchmarked models. For instance, the MVMD-BiLSTM model captured the lowest Root Mean Square Percentage Error (*RMSPE*) values of 9.70% and 6.34% for the hourly and daily forecasts, respectively, at Weir 32. Further application of this proposed model reproduced the overall dynamics of the daily WT in Burtundy (*RMSPE* = 7.88% and Mean Absolute Percentage Error (*MAPE*) = 5.78%) and Pooncarie (*RMSPE* = 8.39% and *MAPE* = 5.89%), confirming that the gridded data effectively capture the overall WT dynamics at these locations. The overall explainable artificial intelligence (*xAI*) results, based on Local Interpretable Model-Agnostic Explanations (LIME), indicate that air temperature (AT) was the most significant contributor towards predicting WT. The superior capabilities of the proposed MVMD-BiLSTM model through this case study consolidate its potential in forecasting WT.

Keywords: water temperature; hybrid modeling; MVMD; deep learning; Menindee

Citation: Briceno Medina, L.; Joehnk, K.; Deo, R.C.; Ali, M.; Prasad, S.S.; Downs, N. Forecasting River Water Temperature Using Explainable Artificial Intelligence and Hybrid Machine Learning: Case Studies in Menindee Region in Australia. *Water* **2024**, *16*, 3720. https://doi.org/10.3390/w16243720

Academic Editor: Junye Wang

Received: 21 October 2024
Revised: 12 December 2024
Accepted: 16 December 2024
Published: 23 December 2024

1. Introduction

Water temperature (WT) is a critical physical indicator of river water quality, with previous studies reporting a link between climate change and river WT [1,2]. Climate change and variations in land use have led to significant variability in the quantity and quality of

inland waters worldwide [3]. According to [4], the mean surface air temperature (AT) has increased by more than 1.4 °C since 1910, with 2019 recorded as the warmest year. Day and night-time temperatures have risen across Australia locations throughout the year. In addition, owing to a combination of factors, drier conditions have been observed across the south-east and south-west, with years of below-average rainfall occurring more frequently. Droughts and extreme temperatures in the Murray–Darling Basin are two critical factors affecting inflow volume, and these factors have led to fish kills [5–8]. Therefore, due to the significant water quality variability in the Murray–Darling Basin, accurate methods for predicting WT in the river system are crucial for assessing water quality.

Fish kill events occur frequently in the Murray–Darling Basin [6,9,10], and in recent years in the Menindee region, hundreds of thousands of large fish kills have occurred in the Darling Basin. Evidence shows that these fish killings were produced by low flows and low-oxygen conditions, high flows (also called blackwater events) and low-oxygen conditions, or the stranding of fish when there is insufficient water [5]. Fish kills are associated with poor water quality in river systems and affect local and regional communities; hence, WT prediction can inform management interventions.

WT models can be classified into physical-based (process-based, mechanistic, or deterministic) and statistical or stochastic models [11–13]. Physical-based models predict WT by quantifying energy fluxes at the air–water interface and have been applied with hourly and daily resolutions. Statistical or stochastic models primarily depend on data for predicting WT and have been used with daily, weekly, and monthly resolutions [11].

Physical-based models simulate the processes that control river temperature by considering hydrological, topographic, channel topologic, and morphologic factors [12]. These models require many inputs to assess flux components (advection, conduction, convection, and radiation), and the significance of energy exchanges at the riverbed can differ by location [14]. Most physical-based river temperature models, according to [15–18], are highly parameterized and require considerable data. Therefore, intermediate approaches have been proposed, such as the Air2Stream model, which is a hybrid statistical–physical model. This model combines a physical-based equation with a stochastic calibration of model parameters and is based on a given discharge scenario and AT [19]. Two mathematical models have been developed in Australia for river systems in south-eastern Australia [1], and for a shallow tidal wetland calibrated for hot, dry, and windy conditions in south-west Western Australia [20]. However, they are completely localized, and their applicability to other locations and conditions presents limitations due to site-specific temperature and water dynamics.

According to [13], statistical/stochastic models can be classified as linear regression, autoregressive, hybrid, or artificial intelligence (AI) models. AI models do not require data related to initial conditions that are necessary to simulate any given target variable in a physical-based model. Instead, an AI-based model prediction is largely based on a reasonably lengthy dataset that provides input features to construct a reliable model. Over the last few decades, a large and growing body of literature has explored the prediction of WT using AI-inspired models. These models can be divided into four main categories: artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems (ANFIS), gaussian process regression (GPR), and other AI models [13]. ANNs are widely used in water modeling. The multilayer perceptron (MLP) is the most basic ANN architecture. It has been used to forecast river WTs in various studies across different time horizons: 10-min [21], hourly [22], daily [23,24], and monthly [25]. ANFIS is a popular model, an adaptive network that integrates fuzzy if–then rules to enhance the ability to adapt to classical ANNs [26]. Even when ANFIS has been slightly more accurate in certain river stations than MLP, MLP neural network models have generally outperformed ANFIS [27]. The underlying premises of GPR are that the probability distribution is Gaussian, where prior knowledge is incorporated as kernels, and uncertainty measures for predictions are provided [28]. In reference [29] is proposed a daily stream WT prediction based on two GPR models describing long-term and short-term components for both periodic and non-periodic changes. GPR performs well

compared to traditional modeling approaches; however, ANNs surpass applications with complex relationships and little knowledge regarding the underlying process. Other AI models include the extreme learning machine [30], support vector machine [31], and hybrid models. Certain hybrid models were proposed by [32], combining the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise technique, random forest (RF), and extreme gradient boosting in their case study. Another hybrid model was proposed by [33], who found that coupling variational mode decomposition (VMD) with machine-learning (ML) methods could improve the accuracy of river WT prediction. However, the impact of other signal processing methods on the accuracy of WT predictions remains unclear.

Recently, investigators have used deep learning (DL) to predict WT and enhance performance. In reference [34] proposed a model using long short-term memory (LSTM) and found that LSTM outperformed several other models in forecasting daily river WT. Similarly, in reference [35] is developed a hybrid model based on an ANN to process missing values, discrete wavelet transforms to divide data into low- and high-frequency signals, and LSTM. This model outperformed the other models in terms of the different evaluation metrics. However, few studies have explored the impact of DL networks on river water temperature prediction; therefore, additional research on a DL method for forecasting river WTs is required.

This study aims to design and test a new model for predicting surface WT using a specific case study. The proposed model was improved by coupling the nonstationary signal processing method, such as the Multivariate Variational Mode Decomposition (MVMD), with DL predictive methods to achieve higher accuracy. Thus, this hybrid model was denoted as MVMD-bidirectional long short-term memory (BiLSTM). The hybrid model was developed to better understand WT changes in the Murray–Darling Basin River system, focusing on Weir 32. Its applicability in similar locations in the same region is explored, in situ measurements and gridded and simulation datasets are assessed, and explainable artificial intelligence (*xAI*) is used to explain the model's predictions.

The novelty and the scientific contributions of this study are summarized as follows:

(a) No prior study has developed a river water temperature using explainable artificial intelligence and hybrid machine learning for the Menindee region in Australia, which is an important ecological zone. Therefore, a primary contribution of our study is to adopt the MVMD method to first decompose the original time-series data that can enhance the predictive performance of the resulting MVMD-BiLSTM model for this region.

(b) To further improve the proposed model, we have designed and applied the hybridized model to predict the surface WT in the Menindee region that focuses on Weir 32. The proposed model, termed MVMD-BiLSTM, has improved the prediction capability and, therefore, makes a novel contribution to WT forecasting.

(c) To make the approach practical, we have evaluated the MVMD-BiLSTM model's effectiveness and accuracy with robust statistical score metrics and visual analysis of all tested data alongside other decomposition-based and standalone benchmarked models.

(d) As an added contribution, we have examined the behavior of the proposed MVMD-BiLSTM model outcomes at the local levels by incorporating for the first time the Local Interpretable Model-Agnostic Explanations (LIME) and therefore contributed to an *xAI* approach that has not been developed before for this specific purpose and study region. The application of these explanation metrics, especially for WT prediction in the present study region, has been largely absent in previous studies. Therefore, this study attempts to explain how the proposed model achieves a specific result driven by the model inputs across the different temporal ranges and for the seasons by testing its efficacy at Weir 32.

(e) Our study also makes an important scientific contribution through accurate and interpretable predictions of WT by the newly developed predictive system that can facilitate the end-users to deliver a more precise understanding of the WT and the

potential impact of its changes on aquatic organisms and ecological processes in the river.

The remainder of this paper is organized as follows: Section 2 provides a theoretical overview of the techniques used to construct the MVMD-BiLSTM model, including a description of the MVMD method and benchmark models. Section 3 includes the details of the study area, characteristics, data pre-processing, and prediction model development. Results are presented and discussed in Section 4. Section 5 summarizes the main findings of this study.

2. Theoretical Overview

In this study, we developed an MVMD-BiLSTM model for hourly and daily WT predictions at Weir 32. What follows next are the theoretical details of the developed model, including the objective (LSTM and Bidirectional LSTM) model, the benchmark AI models, the Air2Stream physical model, and the data decomposition method.

2.1. The Objective (LSTM and Bidirectional LSTM) Model

LSTM is a recurrent neural network (RNN) designed to address the vanishing gradient problem (difficult learning process), which impedes the learning process by preventing the adjustments of weights in earlier layers during backpropagation [36].

A set of connected memory cells constitutes an LSTM layer that can remember previous long-term time-series data. The content of the memory cells is modulated by the input, output, and forget gates using the hyperbolic tangent (tanh) activation function to update the cell. Gates control the flow of information internally, both within and outside the network. These gates are used to remember multiple data items. First, to protect the memory content from irrelevant perturbations, an input gate is introduced. This gate generates a fraction (between 0 and 1) that is multiplied by the output of the tanh block, therefore determining which content is allowed to enter. Second, the output gate protects other units from irrelevant content stored in the current unit, therefore generating another fraction. Third, the forget gate determines which content is relevant and which can be ignored by generating a fraction between 0 and 1 (0 indicates that everything should be forgotten, whereas 1.0 indicates that everything should be remembered). This is computed as follows:

The input gate:

$$i_t = \sigma(W_{i_h}[h_{t-1}], W_{i_x}[x_t], b_i) \quad (1)$$

$$\tilde{c}_t = \tanh(W_{c_h}[h_{t-1}], W_{c_x}[x_t], b_c) \quad (2)$$

The update for the LSTM memory:

$$c_t = f_t * c_{t-1} + i_t * (\tilde{c}_t) \quad (3)$$

The output gate:

$$o_t = \sigma(W_{o_h}[h_{t-1}], W_{o_x}[x_t], b_o) \quad (4)$$

$$h_t = o_t * \tanh(c_t) \quad (5)$$

The forget gate:

$$f_t = \sigma(W_{f_h}[h_{t-1}], W_{f_x}[x_t], b_f) \quad (6)$$

For the above equations, c_t indicates the value to be updated, o_t is the output value, and f_t is the output of the forget gate.

The decision is made based on the values of h_{t-1} and x_t, b_i, b_o and b_f are the bias values, $\tilde{c}_t$ is the vector of new candidate values, σ and tanh are activation functions, and W_i, W_c, W_o and W_f are learnable weight matrices.

In this study, we have adopted BiLSTM, which is an extension of the LSTM algorithm and is formed by two LSTM layers, forward and backward LSTM, without the limitation of

simply using input information in a present future frame [37]. In principle, one layer of the network processes inputs in the forward direction while another layer handles inputs in the backward direction. The final output is then based on the combined outputs of the two hidden layers. The accuracy of the model can then be improved as the LSTM algorithm is applied twice [38], leading to an overall efficacy of the proposed BiLSTM model.

2.2. The Benchmark Models

In this study, the objective (BiLSTM) model has been benchmarked against several competing models, such as the Gated Recurrent Network (GRU), Bidirectional GRU, and the Air2Stream physical model. Furthermore, the proposed BiLSTM and all benchmark models were also improved using the Multivariate Variational Mode Decomposition (MVMD) method to improve the accuracy of the proposed models. In this study, the MVMD approach has been utilized specifically to decompose all input and target features simultaneously, which provides the model greater clarity on the patterns, trends, and important features to create a reliable model [39]. Next, we discuss the overall theoretical details of these models.

2.2.1. Gated Recurrent Network and Bidirectional Gated Recurrent Network

The Gated Recurrent Network (GRU) is a variation of LSTM that reduces the number of gates [40]. It consists of two gates: update and reset gates, which control the information with activation functions. The update gate controls the extent to which new information is used, while the reset gate regulates the degree to which previous information is discarded. According to [41], this is computed as follows:

Activation of GRU using linear interpolation:

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t \tag{7}$$

The update gate:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{8}$$

The candidate activation:

$$\tilde{h}_t = \tanh(W_{x_t} + U(r_t \odot h_{t-1})) \tag{9}$$

The reset gate:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{10}$$

The decision is made based on the values of h_{t-1} and x_t, σ and tanh are activation functions, W_z, W_f, and W_r are the learnable weight matrices, and U_z, U and U_r are weights.

A Bidirectional Gated Recurrent Unit (BiGRU) is formed by two unidirectional GRU layers facing opposite directions [42] with one layer processing the input forward and the second layer processing it backward. The BiGRU is computed using these two GRU layers. Bidirectional neural networks can capture the long-term relationships between sequence pieces, enabling the handling of complex sequential dependencies.

2.2.2. Air2stream

Air2stream [19] is a physical process model that simulates WT using a minimal set of driving fields to describe the complex processes of heat exchange and changes in WT at a fixed point along a river. It is a hybrid model that combines heat transfer analysis with the stochastic calibration of model parameters to determine the daily average river WT. A reduction in the complexity of the process structure resulted in a set of eight parameters. The model could subsequently be simplified further under specific assumptions to yield even simpler models with seven, five, four, and three parameters. Further details can be found in the study conducted by [19,43]. The parameters were determined for each site using a known daily discharge time series and WT by applying a particle swarm optimization routine.

This model depends only on daily AT and discharge. It is recommended to consider discharge in the case of heavily regulated rivers, and it is not necessary to indicate the location of upstream boundary conditions. The equation for the eight-parameter form of the Air2Stream model is as follows:

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left(a_1 + a_2 T_a - a_3 T_w + \theta\left(a_5 + a_6 \cos\left(2\pi\left(\frac{t}{t_y} - a_7\right)\right)\right) - a_8 T_w\right) \tag{11}$$

where $\delta = \theta^{a_4}$ and $\theta = \frac{Q}{\overline{Q}}$, and where a_1–a_8 represents the eight model parameters estimated through calibration. This calibration captures features such as the characteristics of the river reach and local effects. T_a indicates the AT, T_w is the WT, t_y represents the duration of a year, and Q is discharge.

An application to approximately 60 river gauges in 5 catchments in South East Australia [1] showed that parameterization using 7 parameters delivered the best results under Australian conditions.

2.3. Multivariate Variational Mode Decomposition Method

In this study, we decomposed the original time series to develop an MVMD-BiLSTM model using a data pre-processing technique called MVMD [44]. MVMD extends the VMD algorithm, extracting multivariate modulated oscillations from input data and featuring a mode-alignment property across multiple data channels. It decomposes a signal into new components called intrinsic mode functions (IMFs). The sum of the IMFs and residuals exactly recovers the original signal, as follows:

$$x(t) = \Sigma_{k=1}^{K} IMF_K(t) + res(t) \tag{12}$$

The mode number (K) can be selected by minimizing the bandwidth summation of the modes, and each mode is derived by solving a variational problem as follows:

$$minimise\{u_{k,c}\}\{w_k\}\{\Sigma_k \Sigma_c || \partial_t [u_+^{k,c}(t) e^{-jw_k t}] ||_2^2\} \tag{13}$$

$$subject\ to\ \Sigma_k u_{k,c}(t) = x_c(t) \tag{14}$$

With $c = 1..C$, and where ∂_t is the partial derivative operation with respect to time, $u_+^{k,c}$ is the analytic signal corresponding to $u_{k,c}(t)$, $u_{k,c}$ denotes the k ($1 \leq k \leq K$) mode in the r data channel, K indicates the total number of modes, c is the data channel number, C is the set of AM–FM components, $e^{-jw_k t}$ is the complex exponential of frequency, w_k is the center frequencies of $u_{k,c}$, and $x(t)$ is the input data (original signal).

The optimization of the MVMD model aims to minimize the summation of the modes (Equations (12) and (13)), which is solved using an alternate direction method of multipliers (ADMM) approach until convergence. This can be described as follows:

1. Initialization

$$u_{k,c}^1, w_k^1, \lambda^1 \tag{15}$$

where λ is Lagrangian multipliers.

2. Mode update:

$$\hat{u}_{k,c}^{n+1}(w) = \frac{\hat{x}_c(w) - \Sigma_{i \neq k} \hat{u}_{i,c}(w) + \left(\frac{\hat{\lambda}_c(w)}{2}\right)}{1 + 2\alpha(w - w_k)^2} \tag{16}$$

with $c = 1..C$ and $k = 1 \ldots K$, and where α is the penalty factor and is calculated based on K.

3. Centre frequency update:

$$w_k^{n+1} = \frac{\Sigma_c \int_0^\infty w |\hat{u}_{k,c}(w)|^2 dw)}{\Sigma_c \int_0^\infty |\hat{u}_{k,c}(w)|^2 dw)} \tag{17}$$

with $k = 1 \ldots K$.

4. Lagrangian multiplier operators update:

$$\lambda_c^{n+1} = \lambda_c^n + \tau(x_c - \Sigma_k u_{k,c}^{n+1}) \quad (18)$$

with $c = 1 \ldots C$.

5. Repeat 3–5 until convergence.

Further information is provided on how the ADMM approach is applied to solve the MVMD optimization problem (see [44]).

3. Materials and Methods

This section details the study area, data, data pre-processing, development of the prediction model, and metrics to evaluate performance. Together, these elements provide an understanding of the study's methodology and evaluation.

3.1. Study Site and Data

The development and evaluation of the proposed MVMD-BiLSTM model for hourly and daily WT predictions were undertaken for a site immediately downstream of Menindee Lakes, Weir 32. It is in the lower Darling River, New South Wales, Australia, and is part of the Murray–Darling Basin River system. River flow exhibits significant variability, with prolonged droughts and sudden flood events in certain years linked to large-scale climatic patterns. This has led to problems related to blue-green algal blooms in the Darling River and other rivers in the Murray–Darling Basin [45], and in recent years (2018/19 and 2023), to massive fish kill events not previously observed [8]. Two other downstream sites (Pooncarie and Burtundy) were selected to test the daily MVMD-BiLSTM model. Figure 1 shows the geographic locations of these sites, indicating the weather and water monitoring stations that were considered for the provision of model input data for the proposed MVMD-BiLSTM model to predict WT.

Biogeochemical processes are strongly associated with WT and flow hydrology. Thus, the primary drivers of WT are meteorological and hydrological conditions. For this study, meteorological data for the two weather stations were obtained from the Bureau of Meteorology (BoM) [46]. Various WT drivers were available, including AT and relative humidity (RH), which were recorded daily at 9:00 am for Menindee and twice daily for Pooncarie (9:00 am and 3:00 pm), as well as daily minimum and maximum temperatures and daily solar exposure for both sites. Due to the unavailability of mean temperature data, only the daily minimum and maximum temperatures for these two locations were selected from this source. Hydrological data were obtained from Water New South Wales (WaterNSW) [47]. These data included WT, water level, and discharge, with annual, monthly, daily, hourly, and 15 min data intervals for these locations. Thus, hourly data were collected from WaterNSW for Weir 32, Pooncarie, and Burtundy.

We also used gridded climate data for locations without data records. Historical climate records from the Scientific Information for Land Owners (SILO) [48] are accessible for daily solar radiation and evaporation data for Australia, with a spatial resolution of approximately 5 km × 5 km and a temporal resolution of 1-d for each location (Weir 32, Pooncarie, and Burtundy). The data were used, selecting the nearest grid point from the SILO database for the three locations.

The amount of data required for ML models depends on different factors, such as the type of problem and model complexity. However, consistently measured data are not available everywhere. Therefore, for this study to incorporate the hourly AT, solar radiation, and RH into the features of the MVMD-BiLSTM model for Weir 32, we used simulated gridded data from Meteoblue [49]. These data had a spatial resolution of 30 km × 30 km and a temporal resolution of 1-h. To evaluate the performance of the MVMD-BiLSTM daily model in the absence of data, the daily average air temperature (AT) from Meteoblue was used for Pooncarie. For Burtundy, both the daily average and minimum ATs were utilized.

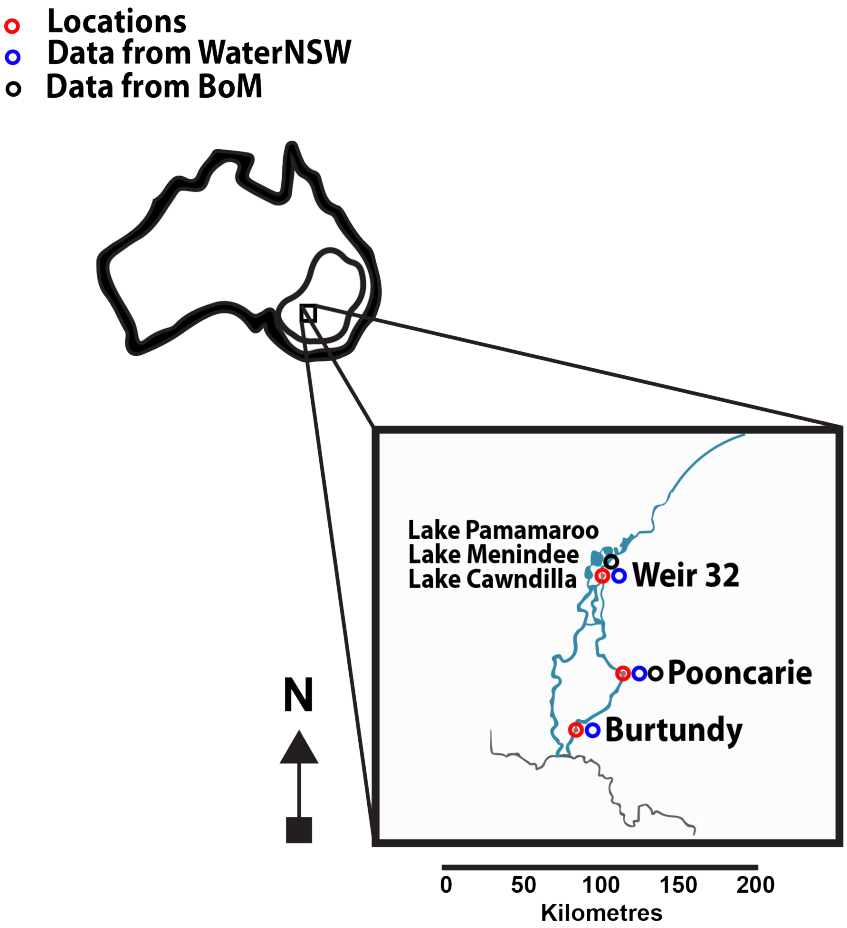


Figure 1. The study site (Weir 32) and another two sites close by (Pooncarie and Burtundy) located in the Murray–Darling Basin. The figure also includes the weather and water monitoring stations from where data were extracted to develop and evaluate the proposed Multivariate Variational Mode Decomposition (MVMD)-bidirectional long short-term memory (BiLSTM) model. WaterNSW: Water New South Wales; BoM: Bureau of Meteorology; SILO: Scientific Information for Land Owner.

The target (WT) and predictor variables, along with their daily descriptive statistics, are presented in Table 1. Notably, AT and RH data presented in this table were derived from Meteoblue datasets, WT from WaterNSW, and shortwave solar radiation (SR) and evaporation from SILO for these three locations.

Table 1. Descriptive statistics of the predictor (input) variables and the target variable (i.e., water temperature) for the present study site at Weir 32 and the two nearby locations in the Murray–Darling Basin from 1 January 2012 to 31 July 2023. The acronyms WT (°C), AT (°C), RH (%), SR (MJ/m^2), and EVA (mm) represent the predictors: water temperature, air temperature, relative humidity, shortwave solar radiation, and evaporation, respectively.

Predictor Variable	**Weir 32** **Lat: 32.44° S** **Lon: 142.38° E**	**Pooncarie** **Lat: 33.39° S** **Long: 142.57° E**	**Burtundy** **Lat: 33.75° S** **Long: 142.27° E**
Mean WT	19.288	20.435	18.814
Minimum WT	9.688	8.063	8.092
Maximum WT	32.621	36.966	32.192
STD WT	5.663	7.103	5.746

Table 1. *Cont.*

Predictor Variable	Weir 32 Lat: 32.44° S Lon: 142.38° E	Pooncarie Lat: 33.39° S Long: 142.57° E	Burtundy Lat: 33.75° S Long: 142.27° E
Mean AT	20.225	19.142	18.928
Minimum AT	5.600	5.419	4.947
Maximum AT	42.174	40.217	40.113
STD AT	7.437	7.184	7.107
Mean RH	43.078	47.044	47.900
Minimum RH	6.375	6.833	6.875
Maximum RH	93.458	92.792	92.667
STD RH	16.418	16.612	16.522
Mean SR	18.418	17.993	17.993
Minimum SR	2.950	2.850	2.850
Maximum SR	34.325	33.175	33.175
STD SR	7.370	7.559	7.559
Mean EVA	4.122	3.970	3.970
Minimum EVA	0.525	0.500	0.500
Maximum EVA	9.575	9.250	9.250
STD EVA	2.271	2.274	2.274

3.2. Development of the Proposed Predictive Model

The proposed hybrid MVMD-BiLSTM (and standalone benchmark models) were built using a Python Software environment 3.11.5, including packages for executing DL algorithms: the TensorFlow [50] and Keras [51] 2.13.1 frameworks. These models were executed using an Intel Core i7 @ 2.00 GHz processor and a 32.0 GB memory computer. The model development stages are shown in Figure 2 and were as follows:

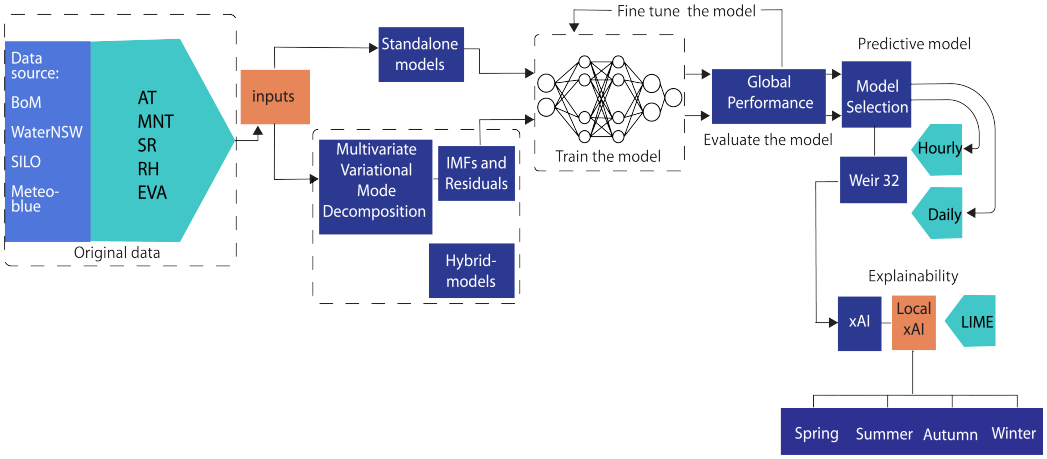


Figure 2. A schematic flowchart illustrating the steps used to develop the proposed MVMD-BiLSTM model for the daily (d) and hourly (h) water temperature prediction at Weir 32. Note that AT = air temperature, MNT = minimum air temperature, SR = shortwave solar radiation, RH = relative humidity, EVA = evaporation, IMF = intrinsic mode function, *xAI* = explainable artificial intelligence and LIME = Local Interpretable Model-Agnostic Explanations.

The model development phase involved proposing standalone (methods that function independently) and hybrid (methods that function together) models using RNNs (LSTM, BiLSTM, GRU, BiGRU) and inputs. For the hybrid modeling approach, input decomposition was performed using the MVMD method. These decomposed inputs were then fed directly into the RNN models. Next, the standalone and hybrid models were compared in terms of performance, and the models were selected based on the best global performance

for hourly and daily predictions at Weir 32. Subsequently, *xAI* was applied using LIME to elucidate individual predictions by season and generate a collection/family of local explanations to provide an approximated global explanation (Figure 2).

3.2.1. Data Pre-Processing

Missing Values

Time series data must be complete, gap-free, and regular to feed the models. If there are gaps (missing data), the impact must be evaluated to ensure that these gaps are insignificant, and imputation and interpolation methods must be considered. To perform the evaluation, the data were organized on an hourly and daily basis. The percentage of missing data in this dataset was determined to be less than 5%. Using Little's test [52] in SPSS 29.0, an appropriate imputation method was selected. This test provides evidence that the missing data of these features are not entirely random (MCAR). Therefore, at a significance level of 0.05, they are either Missing At Random (MAR) or non-ignorable. In principle, a MAR mechanism aims to describe systematic missingness of the data where missing data depends partly on other data in the dataset but not on any missing value [53]. Consequently, this dataset exhibits a MAR data pattern. Thus, three imputation methods and one interpolation method were explored to fill in the missing values of the features from WaterNSW (WT, water level, and discharge). These methods included forward and backward filling, mean imputation, and splines. The most appropriate imputation method for these features was selected based on the accuracy of the metrics, data gap structure, and the procedure proposed by [54].

Multivariate Variational Mode Decomposition Method

In this study, to develop the proposed MVMD-BiLSTM model, we applied the MVMD method [44] as the foremost pre-processing step using MATLAB R2023b (see method description in Section 2.3). MVMD was applied to the hybrid framework, where all input time series were decomposed into IMFs and residuals. This enabled the analysis of the data from a new perspective and yielded novel insights into inherent data that were not observable using the complete data. In the multivariate framework, our specific interest was in analyzing how different signals were related. Subsequently, the different signals were fed directly into the ML models.

The solution parameters were set to their default values: The parameter defining the bandwidth of extracted modes (α) was set to 2000, where α is calculated based on the number of modes to be recovered (K). The timestep of the dual accent (τ) was set to 0, all omega values start as uniformly distributed with ($init$) = 1, and the tolerance value for converge of the ADMM algorithm (tol) was set to 1×10^{-7}. In contrast, the number of modes to be recovered (K) varied between three and ten, with the best results obtained when $K = 7$. In this study, the number of IMFs was determined through a trial-and-error process, and the most informative IMFs were selected using a two-step selection method. First, the optimal IMF candidates were identified using a correlation matrix, and the IMFs were subsequently removed based on their impact on the actual prediction accuracy.

3.2.2. Model Design Procedure

The hourly and daily time series were divided into two groups: training (80%) and testing (20%). The training and testing sets were available for all seasons.

The following features were explored as inputs for the hourly and daily models: For the hourly models, mean AT, RH, shortwave radiation, and water level were used. For the daily models, the mean, minimum, and maximum ATs, mean shortwave radiation, evaporation, and water level were explored. The water level does not play a role because it was removed during the feature selection process. During this process, the impact of each feature on the performance degradation of the model was verified, and when the water level was removed, the accuracy of the model improved. The selected features are listed in Table 2.

Table 2. The set of input features used to construct the proposed hybrid MVMD-BiLSTM model for 1-h and 1-d horizons. The acronyms AT, RH, SR, and EVA represent the predictors of water, air temperature, shortwave solar radiation, relative humidity, and evaporation, respectively.

1-h Model Input Features	1-d Model Input Features
AT (°C)	AT (°C)
SR (MJ/m^2)	SR (MJ/m^2)
RH (%)	Minimum Air Temperature MNT (°C)
	Evaporation–Morton's shallow lake evaporation EVA (mm)

All input features were scaled using MinMaxScaler, normalizing them between zero and one, which is standard practice in developing predictive models.

The prediction model used zero lags as inputs and the highest correlated significant lag at 12 h for AT and RH in the 1-h time horizon model. For the 1-d time horizon model, inputs at 0 d for the mean and minimum ATs and 0 and 12 d for solar radiation and evaporation were used. However, another study could include lagged combinations of hourly and daily inputs.

3.2.3. Hyperparameter Tuning

The optimal hyperparameters for the MVMD-BiLSTM and benchmarked models were selected based on a grid search procedure, where, at every iteration, a model was trained and evaluated using a subset of values until all combinations were evaluated. This procedure is time-consuming and incurs a high computational cost; however, the computational time was reduced once the optimal parameters were identified (<4 min).

In addition, the MVMD-BiLSTM model utilizes dropout to control overfitting during the training period. The dropout layer randomly drops neurons by setting them to zero. The frequency was set to 0.1, which represents the fraction of the units set to zero.

Table 3 details the proposed model's optimal architecture. This includes the number of neurons or units, the layers that are formed by these neurons, the epoch or cycle where all the training dataset is trained in the neural network, the learning rate at which an algorithm updates to determine the minimum weight value, the batch size or number of samples used in one epoch, the dropout or fraction of units set to zero, the loss function that compares observed and predicted values, and the optimizer that adjusts the parameters of the neural network to minimize the loss function. It is important to note that these parameters were determined by trial-and-error testing to arrive at optimal model parameters for hourly and daily models.

Table 3. Parameter settings for 1-h and 1-d models including the datasets used for training and testing the models.

Site	Forecast Model	Time Horizon	Data Used	Train/Test (%)	Model Parameters
Weir 32	MVMD-BiLSTM	1-h	Train: 1 January 2012 –6 April 2021 Test: 7 April 2021–31 July 2023	Train 80% Test 20%	Layers = 2 Learning rate = 0.00975705 Neurons = 32 Epochs = 50 Batch_size = 512 Dropout = 0.1 Loss function = MSE Optimizer = Adam SGD
Weir 32	MVMD-BiLSTM	1-d	Train: 1 January 2012 –6 April 2021 Test: 7 April 2021–31 July 2023	Train 80% Test 20%	Layers = 2 Learning rate = 0.0095 Neurons = 32 Epochs = 250 Batch_size = 256 Dropout = 0.1 Loss function = MSE Optimizer = Adam SGD

3.3. Model Performance Metrics

The objective model, denoted as MVMD-BiLSTM, was evaluated against the benchmarked models applied in forecasting hourly (h) and daily (d) WTs at Weir 32 using different metrics widely used for assessing models: Pearson's correlation (r), Root Mean Square Error ($RMSE$) (°C), Mean Absolute Error (MAE) (°C), modified Willmott's index (WI), Nash Sutcliffe Coefficient (NSE), Legate-McCabe Efficiency Index (LME), Root Mean Square Percentage Error ($RMSPE$) (%), and Mean Absolute Percentage Error ($MAPE$) (%). These model evaluation metrics are given as follows:

$$r = \frac{\Sigma(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\Sigma(y_i - \bar{y})^2(x_i - \bar{x})^2}} \tag{19}$$

$$RMSE(^{\circ}\mathrm{C}) = \sqrt{\frac{(\Sigma(x_i - y_i)^2)}{n}} \tag{20}$$

$$MAE(^{\circ}\mathrm{C}) = \frac{|x_i - y_i|}{n} \tag{21}$$

$$WI = 1 - \frac{\Sigma_{i=1}^{n}|x_i - y_i|}{\Sigma_{i=1}^{n}(|y_i - \bar{x} + x_i - \bar{x}|)} \tag{22}$$

$$NSE = 1 - \frac{\Sigma_{i=1}^{n}(x_i - y_i)^2}{\Sigma_{i=1}^{n}(x_i - \bar{x})^2} \tag{23}$$

$$LME = 1 - \frac{\Sigma_{i=1}^{n}|x_i - y_i|}{\Sigma_{i=1}^{n}|x_i - \bar{x}|} \tag{24}$$

$$RMSPE(\%) = \sqrt{\left(\frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{x_i - y_i}{x_i}\right)^2\right)} \times 100 \tag{25}$$

$$MAPE(\%) = \frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{x_i - y_i}{x_i}\right) \times 100 \tag{26}$$

where x_i is the observed WT, y_i is the predicted WT, ($\bar{x}$, $\bar{y}$) are the average observed and predicted WTs, respectively, and n is the total number of data points. The values of r and NSE range from $-\infty$ to 1, where 1 is the ideal value for a perfect model. r records the level of agreement between observations and predictions, and NSE assesses the ability of the model to predict values different from the observed mean of the target variable. However, NSE often overestimates larger values and neglects smaller values [55]. The $RMSE$ and MAE range from 0 to ∞, where 0 is the ideal value; they measure the goodness of the model's fit; however, their sensitivity to outliers is a common concern. The WI and LME range from 0 to 1, where 1 is the ideal value. WI and LME have an advantage over r, $RMSE$, and MAE because they do not give outliers unnecessary weight. The $RMSPE$ and $MAPE$ range from 0 to 100%, where 0% is the ideal value, and they are measures of errors, expressed as a percentage in relation to observations, and it is used to compare models with different ranges; however if the observation is zero it is affected by a division error.

3.4. Explainability of Model Outcomes

In this study, we improved the practical utility of the proposed MVMD-BiLSTM model using *xAI* methods. In principle, *xAI* methods can provide both local and global explanations and interpretations of the model's behavior based on the feature inputs (or predictor variables). An explanation is an interpretable description of the model's behavior, where a local explanation model explains individual predictions, and a global model explanation explains the complete behavior of the model. In general, we have adopted *xAI* as explainability aims to be consistent with prior knowledge; if any explanation contradicts prior knowledge, there tends to be less trust in such explanations.

In this study, local explanations of the proposed MVMD-BiLSTM model were performed using LIME, which aimed to build the explanations of individual predictions by identifying an interpretable model, as presented by [56]. Three formats are supported by LIME: text, image, and tabular data [57]. Therefore, this study applied the recurrent tabular explainer format to produce local explainability. The following produced the LIME explanation ϵ:

$$\epsilon(x) = \frac{argmin}{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (27)$$

where g is the explanation model, G is the class of potential set of interpretable models, f is the original model, $\mathcal{L}$ (f,g,π_x) is the measurement of the unfaithfulness of model g in approximating the predictions of model f, π_x is the proximity measure, and $\Omega(g)$ is the measure of complexity of the explanation for all explanation models.

The LIME algorithm operates by taking a point (x_i), perturbing it several times by adding random Gaussian noise, and producing instances around x_i until a group of instances is generated in the local neighborhood. Subsequently, the predictions for each of these perturbations were determined using the underlying model. According to the distance to x_i, points closer to x_i have higher weights than those further away from x_i. Subsequently, a simple linear model was fitted to the weighted samples. This model provides a group of coefficients or weights associated with different features, which is the explanation. Essentially, LIME indicates the importance and direction of importance (positive or negative) for each feature.

4. Results and Discussion

In this section, we now describe the results obtained to forecast river water temperature using the proposed hybrid modeling approach for the case of the Menindee region in Australia. In doing so, this study evaluated the performance of the proposed hybrid predictive model developed by combining the MVMD and ML methods. The results are specifically evaluated for forecasting the one-hour and one-day-ahead WTs at Weir 32, located in the Menindee region in Australia.

4.1. Appraisal of Model Performance to Forecast One-Hour-Ahead Temperature

Based on the forecasted versus observed river water temperature, we now show in Table 4 the model performance metrics based on r, $RMSE$, and MAE, which are expected to register values close to 1 (for r) and 0 for $RMSE/MAE$, respectively. For the non-dimensional metrics WI, NSE, and LME, a perfect predictive model is expected to record a value close to 1, whereas the $RMSPE$ and $MAPE$ values are expected to be close to 0%. Table 4 shows the evaluation of the accuracy of the hourly model during the training period.

Table 4. Training performance (hybrid 1-h model) for Weir 32. Note: Pearson's correlation coefficient—r, Root Mean Square Error—$RMSE$ (°C), Mean Absolute Error—MAE (°C), Willmott's index—WI, Nash Sutcliffe Coefficient—NSE, Legate-McCabe Efficiency Index—LME, Root Mean Square Percentage Error—$RMSPE$ (%), Mean Absolute Percentage Error—$MAPE$ (%). The most accurate models in the training phase are boldfaced.

Forecast Model	r	$RMSE$	MAE	WI	NSE	LME	$RMSPE$	$MAPE$
MVMD-BiGRU	**0.953**	**1.714**	1.365	**0.862**	**0.909**	**0.730**	**9.840**	**7.500**
MVMD-BiLSTM	0.953	1.717	**1.344**	0.862	0.908	0.730	9.850	7.500
MVMD-GRU	0.951	1.758	1.370	0.857	0.904	0.723	10.230	7.770
MVMD-LSTM	0.950	1.781	1.366	0.854	0.901	0.717	10.500	8.000
BiGRU	0.917	2.277	1.784	0.812	0.839	0.644	13.690	10.120
BiLSTM	0.915	2.290	1.784	0.813	0.837	0.642	13.880	10.240
GRU	0.908	2.376	1.796	0.803	0.825	0.627	14.420	10.650
LSTM	0.907	2.388	1.791	0.801	0.823	0.625	14.510	10.740

The hybrid models MVMD-BiLSTM, MVMD-BiGRU, MVMD-GRU, and MVMD-LSTM have demonstrated the highest overall predictive accuracy compared with standalone models, with an average of r = 0.952, $RMSE$ = 1.743 °C, and MAE = 1.361 °C. In terms of the normalized metrics, we note a value of WI = 0.859, NSE = 0.906, and LME = 0.725, which also shows significantly accurate results. In terms of the relative errors, the hybrid models recorded an $RMSPE$ = 10.110%, and $MAPE$ = 7.690%, which is well below the 10% error recommended for excellent performance [58]. In general, our results indicate that the predictive capabilities of the MVMD-BiLSTM and MVMD-BiGRU models outperform those of other hybrid and standalone models when using the training dataset. These results suggest that the MVMD algorithm is significantly advantageous in increasing accuracy due to the mode-alignment property across all channels and excellent noise resistance [44].

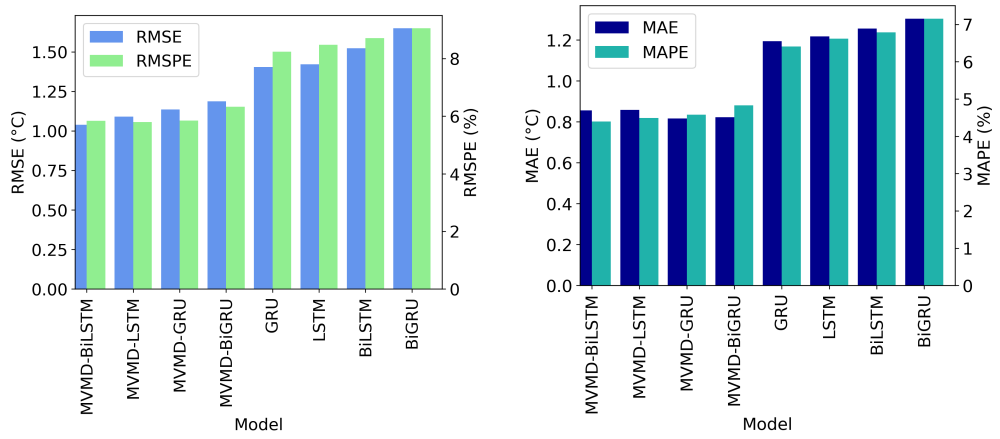
We now discuss the application of the proposed hybrid models (i.e., MVMD-BiLSTM, MVMD-BiGRU, MVMD-GRU, and MVMD-LSTM) with respect to standalone models for the testing phase. A comparison of the performance of all eight models adopted for predicting the 1-hourly WTs for the testing period showed that the proposed MVMD-BiLSTM hybrid model outperformed the other ML models. As listed in Table 5, the MVMD-BiLSTM hybrid model displayed the highest r value (0.955) (representing a high degree of agreement between the forecasted and the observed WT) and the lowest RMSE (1.690 °C) and MAE (1.347 °C) values for Weir 32 site. Indicating that predicted and observed hourly WT agree reasonably well and that the results of the testing phase are consistent with the results of the training phase.

Table 5. Testing performance (hybrid 1-h model) for Weir 32. The most accurate models in the testing phase are boldfaced.

Forecast Model	r	$RMSE$	MAE	WI	NSE	LME	$RMSPE$	$MAPE$
MVMD-BiLSTM	**0.955**	**1.690**	**1.347**	**0.860**	**0.907**	**0.724**	**9.700**	**7.600**
MVMD-BiGRU	0.955	1.702	1.365	0.859	0.906	0.723	9.800	7.600
MVMD-GRU	0.954	1.714	1.370	0.857	0.904	0.720	9.900	7.700
MVMD-LSTM	0.952	1.742	1.365	0.852	0.901	0.713	10.200	8.000
GRU	0.879	2.711	2.190	0.753	0.761	0.549	18.600	13.600
BiLSTM	0.877	2.722	2.187	0.753	0.759	0.547	18.600	13.700
BiGRU	0.878	2.708	2.187	0.751	0.761	0.547	18.400	13.600
LSTM	0.875	2.751	2.194	0.748	0.754	0.542	18.900	13.900

Similarly, the objective model (MVMD-BiLSTM) exhibited the highest WI (0.8602) and NSE values (0.907). The MVMD-BiLSTM model was assessed using the LME and outperformed the other models (i.e., LME = 0.722–0.724). All hybrid models combined with the MVMD method achieved LME values of 0.713–0.724 compared to 0.542–0.549 for the LSTM, GRU, BiLSTM, and BiGRU models. These results show that the inclusion of the MVMD contributes to the models' performances.

A comparison of the MVMD-BiLSTM model in terms of the $RMSE$, MAE, $RMSPE$, and $MAPE$ is shown in Table 5 and Figure 3. The results showed low values for the objective model: $RMSE$ = 1.690 °C, MAE = 1.347 °C, $RMSPE$ = 9.740%, and $MAPE$ = 7.560%, indicating good model performance. The model performance assessment revealed a significant improvement with the MVMD-BiLSTM model. These results demonstrate the influence of AT, SR, and HR as inputs on changes in hourly WT.



(**a**) (**b**)

Figure 3. (**a**) The *RMSE* and (**b**) *MAE* generated by the models during the testing phase for all 1-h horizon models.

Figure 4 shows a Taylor diagram, which determines the link between forecasted and observed WTs based on the correlation coefficient and standard deviation values. The diagram evaluates all predictive models: hybrid and standalone. The values predicted by the hybrid models were close to the observed WTs, making the MVMD-BiLSTM model the most accurate. In contrast, the standalone models showed the lowest accuracy, positioned further from the observed WT.

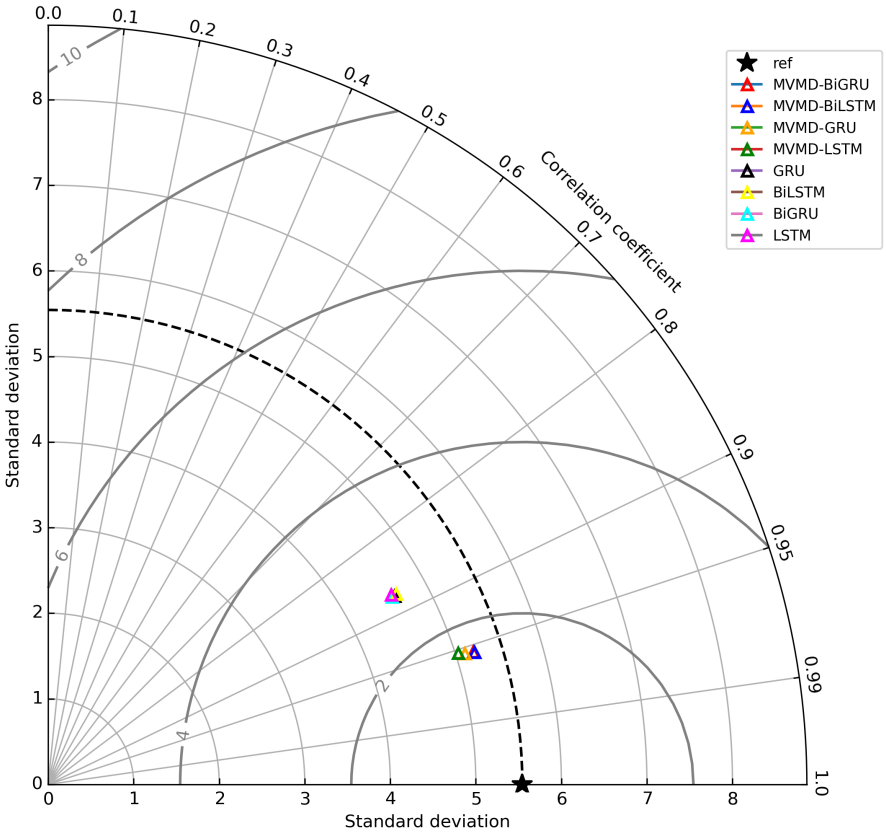


Figure 4. Taylor diagram depicting the observed and predicted WTs of all hourly models at Weir 32. (- -) The dashed line represents the correct standard deviation.

Consequently, the MVMD-BiLSTM model captured the highest *LME*, had low relative errors, and made predictions closest to the observed WT in the Taylor diagram, indicating that the objective model (MVMD-BiLSTM) displayed the lowest errors in predicting WT.

4.2. Appraisal of Model Performance to Forecast One-Day-Ahead Temperature

As listed in Table 6, during the training period, the one-day ahead hybrid model exhibited better performance than the standalone models, where the MVMD-BiLSTM presents a r = 0.984, and the $RMSE$ between predicted and observed values is 1.101 °C. This demonstrates that daily WT predictions correlate significantly with observed values, and the objective model (MVMD-BiLSTM) can replicate the WT well at Weir 32.

Further comparison of the models' performance was conducted based on LME values. The MVMD-BiLSTM, MVMD-LSTM, and MVMD-GRU models exhibited the highest LME values of 0.8266, 0.8337, and 0.8321, respectively, indicating that hybrid models demonstrate superior performance compared to standalone models of BiLSTM, LSTM, GRU, and BiGRU.

In terms of the $MAPE$, all hybrid models (MVMD-BiLSTM, MVMD-LSTM, MVMD-GRU, and MVDM-BiGRU) exhibit values lower than standalone models (BiLSTM, LSTM, GRU, BiGRU), which are lowers than 10% [58], consequently, hybrid models can achieve better predictions.

We now discuss the proposed hybrid models and standalone models for the testing phase. For all experiments with the testing datasets, the proposed MVMD-BiLSTM model for the 1-d horizon achieved the highest r value to 3 decimal places (i.e., 0.984) with actual values between 0.9836–0.9835 for MVMD-LSTM, MVMD-GRU, and MVMD-Bi-GRU, respectively. This represents the forecasted and observed WTs being closest to unity for the proposed MVMD-BiLSTM model for the 1-d horizon and the lowest $RMSE$ of 1.039 (Table 7). Hence, daily WT appeared to be well reproduced by AT, SR, MNT, and EVA at Weir 32.

Table 6. Training performance (hybrid 1-d model) for Weir 32. The most accurate models in the training phase are boldfaced.

Forecast Model	r	$RMSE$	MAE	WI	NSE	LME	$RMSPE$	$MAPE$
MVMD-BiLSTM	**0.984**	**1.101**	**0.862**	0.910	**0.962**	0.827	6.340	4.900
MVMD-LSTM	0.981	1.110	0.861	0.915	0.962	**0.834**	**5.620**	4.380
MVMD-GRU	0.980	1.131	0.841	**0.916**	0.960	0.832	5.660	**4.360**
MVMD-BiGRU	0.982	1.196	0.867	0.903	0.956	0.811	7.130	5.440
GRU	0.974	1.316	1.073	0.893	0.946	0.792	7.400	5.740
BiLSTM	0.975	1.335	1.076	0.891	0.945	0.788	7.100	5.690
BiGRU	0.976	1.369	1.088	0.887	0.942	0.783	7.000	5.670
LSTM	0.971	1.375	1.082	0.888	0.941	0.781	8.030	6.170

Table 7. Testing performance (hybrid 1-d model) for Weir 32. The most accurate models in the testing phase are boldfaced.

Forecast Model	r	$RMSE$	MAE	WI	NSE	LME	$RMSPE$	$MAPE$
MVMD-BiLSTM	**0.984**	**1.039**	0.856	**0.917**	**0.965**	**0.839**	5.840	**4.400**
MVMD-LSTM	0.984	1.091	0.859	0.913	0.961	0.831	**5.800**	4.490
MVMD-GRU	0.984	1.136	**0.817**	0.910	0.958	0.826	5.850	4.580
MVMD-BiGRU	0.984	1.187	0.822	0.904	0.954	0.818	6.330	4.830
GRU	0.970	1.404	1.194	0.878	0.936	0.767	8.240	6.410
LSTM	0.970	1.421	1.218	0.876	0.934	0.763	8.480	6.620
BiLSTM	0.966	1.522	1.256	0.870	0.924	0.749	8.710	6.790
BiGRU	0.964	1.649	1.304	0.858	0.911	0.730	9.050	7.150

The objective model (MVMD-BiLSTM) showed the highest WI (0.9172) and NSE (0.965) and the highest LME (i.e., LME = 0.820–0.840). Again, the hybrid models exhibited higher WI, NSE, and LME values than the standalone models. Results demonstrate that including MVMD greatly improves the model performances; however, this effect is more pronounced for the MVMD-BiLSTM model.

The MVMD-BiLSTM model also performed better in terms of the MAPE, achieving a value of 4.400%, which is below 10% [58]. Table 7 and Figure 5 show the magnitudes of

RMSPE and *MAPE* for all tested models. The objective model (MVMD-BiLSTM), when compared with the other models, exhibited low relative errors. Therefore, the efficacy of the MVMD-BiLSTM was verified, including relative predictive errors, and it can be clearly seen that these features are major factors contributing to the daily WT prediction at Weir 32.

Earlier results in Tables 4–7 showed that there was a certain degree of variability in the training and testing performance among the prescribed models developed using the MVMD approach. Therefore, we have summarized the performance of the MVMD-based model as follows:

- For the 1-h forecast horizon: The proposed MVMD-BiLSTM model performed slightly better in terms of the training phase based on the *MAE* value, whereas the MVMD-BiGRU model was the best based on all the other metrics. In the testing phase, however, the proposed MVMD-BiLSTM was the best model if we ignored the slightly larger value of *r* for the MVMD-BiGRU model.
- For the 1-d forecast horizon: There has been a larger variability in the training phase for this forecast horizon. For example, the proposed MVMD-BiLSTM model was better than other models based on *r*, *RMSE*, *NSE*, whereas based on *MAE*, the proposed MVMD-LSTM was superior based on the larger magnitude of *WI*, the MVMD-GRU was superior whereas based on *MAPE*, the MVMD-BiGRU was the best model. For the testing phase, however, the proposed MVMD-BiLSTM attained the best performance based on *r*, *RMSE*, *WI*, *NSE*, *LME*, and *MAPE*.

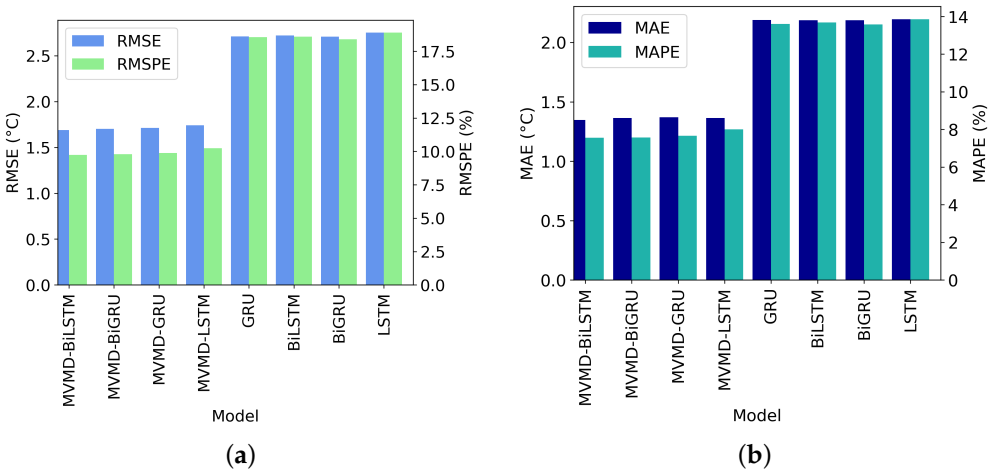


Figure 5. (**a**) The *RMSE* and (**b**) *MAE* generated by the model during the testing phase for all 1-d horizon models.

Despite the subtle variability in performance among the trained and the tested models, it is noteworthy that the hybrid models were always better than the standalone models, and the proposed MVMD-BiLSTM was the best model for 1-h and the MVMD-BiLSTM was the best model for the 1-d forecast horizon if a majority of the model performance metrics were considered.

The Taylor diagram (Figure 6) shows that the result of the MVMD-BiLSTM model is positioned close to the observed WT (reference data point); therefore, the objective model (MVMD-BiLSTM) was the most accurate. Besides the hourly horizon, in the daily forecast horizon, the standalone models exhibited the largest discrepancy between the testing and observed data, located farther from the reference point. Consequently, it was evident that the proposed MVMD-BiLSTM model for the 1-d time horizon outperformed all comparable models during testing.

A study [34] assessed daily WT forecasting for seven rivers in China, the USA, and Switzerland and found that an LSTM model performed better than RF and Back Propagation Neural Networks. However, the overall trend in this study was consistent with previous findings from another time-series study. Researchers in reference [38] found that BiLSTMs outperformed unidirectional LSTMs for daily, weekly, and monthly forecasting.

They stated that BiLSTMs could capture the underlying context by learning from past and future data; however, they observed that the training was slower than that for LSTMs.

In summary, for the 1-h and 1-d time horizons, the hybrid approach exhibited the highest model performance. In contrast, the standalone models showed poor accuracy. In particular, the MVMD-BiLSTM model exhibited superior performance compared to the MVMD-LSTM, MVMD-GRU, MVMD-BiGRU, LSTM, GRU, BiLSTM, and BiGRU models. One implication of this study is the potential to estimate hourly and daily river WTs at Weir 32 using MVMD with ML models.

Another point to note is that an increase in the forecast time horizon leads to a decrease in the RMSPE and MAPE. An increase in the time horizon from 1-h to 1-d leads to an improvement in model performance, and this result was similar to that reported by [59]. One reason for this could be the addition of minimum AT and the use of evaporation instead of HR in the daily model.

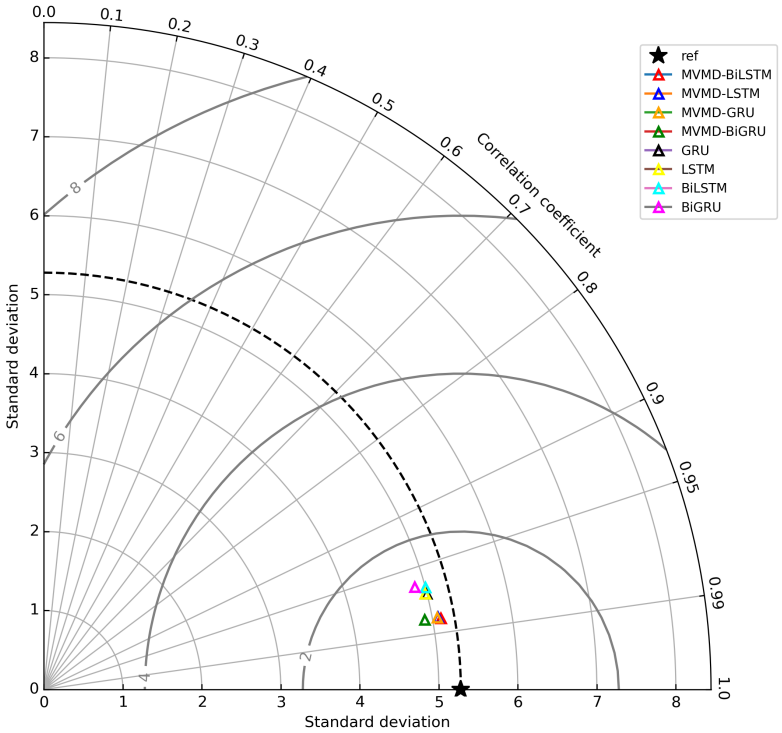


Figure 6. Taylor diagram illustrating the observed and predicted WTs for all daily models at Weir 32. (- -) The dashed line represents the correct standard deviation.

Figures 7 and 8 show visual comparisons of the interannual variability in WT for hourly and daily time horizons. Overall, the models had good dynamics to reproduce WT, and in general, the daily model presented a better response across all seasons compared to the response of the hourly model. They followed annual oscillations, in which the seasonal variation in WT is represented by the annual component.

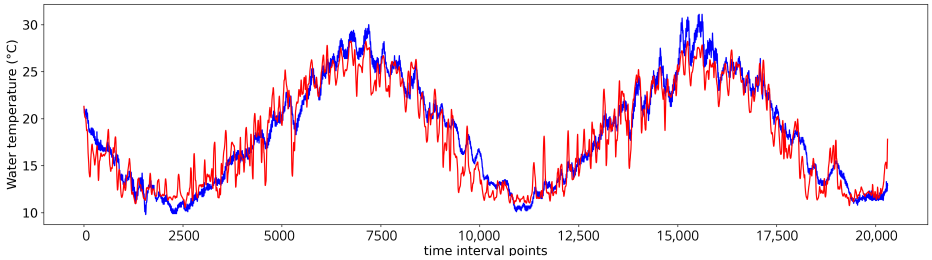


Figure 7. Evaluation of the hourly MVMD-BiLSTM model, illustrating the observed and predicted WTs at Weir 32 over a two-year horizon. Blue line: observed. Red line: predicted.

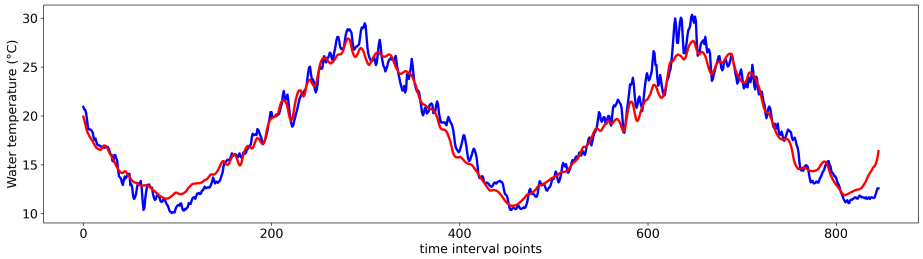


Figure 8. Evaluation of the daily MVMD-BiLSTM model, illustrating the observed and predicted WTs at Weir 32 over a two-year horizon. Blue line: observed. Red line: predicted.

These figures reproduce this for two years, with the second being the warmer year. The models did not reproduce these peaks during warmer summers. The response of the MVMD-BiLSTM 1-h time horizon to changes during summer and autumn was underestimated (Figure 7), whereas the MVMD-BiLSTM 1-d time horizon underestimated changes during spring and summer, particularly during the second year (Figure 8).

When we used these features, the WT was mostly lower, particularly during daytime and across different seasons. The biases decreased at lower temperatures, and notable differences were observed in the maximum temperatures during spring for the daily model. These differences could be associated with not considering the maximum temperature as a feature in the model because it presents a high correlation with the mean AT. However, when considering the maximum temperature, the large differences during spring did not change, and the model's accuracy decreased; for example, the *LME* decreased from 0.840 to 0.800. It is likely that factors that were not considered may have affected this result. Therefore, events during spring were poorly predicted by the model, with hourly and daily models underestimating peaks during summer. One factor to be considered is that upstream of Weir 32, the river forms a weir pool, which is strongly affected by the river input and the inputs from the Menindee Lakes. Another factor is the use of gridded simulated data to incorporate hourly and daily mean ATs into the models, which could introduce bias.

We evaluated the MVMD-BiLSTM model for *n*-hour and *n*-day ahead forecasts, and we observed that when the number of days increased, the indices were impacted (Table 8). When the model behavior of MVMD-BiLSTM was compared with the forecast, the error bias was higher when the number of hours increased from 1-h to 24-h, and the same pattern was found when the number of days increased from 1-d to 7-d. On the one hand, when focusing on the *RMSE* and *MAE*, if the number of days increased, then the indices increased; for instance, the *RMSE* increased by 38.660% and the *MAE* by 33.480% for the hourly model, while the *RMSE* increased by 62.980% and the *MAE* by 64.470% for the daily model. On the other hand, focusing on *WI*, the *NSE*, and the *LME*, when the number of days was increased, these indices were reduced. For the hourly model, the indices were reduced between 1-h and 24-h: *WI*, 10.150%; *NSE*, 12.050%; and *LME*, 19.710%. For the daily model, these indices were reduced from 1-d to 7-d: *WI*, 16.370%; *NSE*, 30.230%; and *LME*, 58.280%. This gradual decline in these metrics indicates that the performance of the model is degrading and that the model's performance degrades more quickly in the daily model than in the hourly model. Poorly performing models can emerge over time because of changes in the properties of the input data, shifts in the underlying relationships between the target variable and input features, changes in underlying data distribution, changes in the importance of features, or alterations in the model's underlying assumptions.

Table 8. The model performance for *n*-hour and *n*-day ahead forecasts in terms of *r*, *RMSE* (°C), *MAE* (°C), *WI*, *NSE*, *LME*, *RMSPE* (%) and *MAPE* (%). Note: 'h' stands for hourly, and 'd' stands for daily forecast model.

Model and Forecast Horizon	*r*	*RMSE*	*MAE*	*WI*	*NSE*	*LME*	*RMSPE*	*MAPE*
Hourly Horizon								
MVMD-BiLSTM: 1-h	0.980	1.369	1.190	0.868	0.939	0.753	9.400	7.400
MVMD-BiLSTM: 6-h	0.979	1.803	1.566	0.819	0.894	0.676	11.100	9.100
MVMD-BiLSTM: 12-h	0.988	1.920	1.609	0.812	0.880	0.667	9.900	8.600
MVMD-BiLSTM: 24-h	0.985	2.232	1.789	0.788	0.838	0.629	10.400	9.100
Daily Horizon								
MVDM-BiLSTM: 1-d	0.986	1.046	0.820	0.917	0.965	0.831	5.800	4.600
MVDM-BiLSTM: 3-d	0.985	1.994	1.639	0.851	0.872	0.663	14.700	11.000
MVDM-BiLSTM: 5-d	0.982	2.194	1.764	0.838	0.844	0.637	16.400	12.000
MVDM-BiLSTM: 7-d	0.974	2.826	2.308	0.788	0.741	0.525	20.900	15.400

4.3. Explainability

The bar graphs (Figure 9) show the local explainability of instances for different seasons using LIME. In this study, the number of instances was equal to the number of time points in the testing phase (instances for the hourly model range from instance 0 to 846, and the daily model range from 0 to 20,304). These instances were selected randomly for each season. Instances illustrate individual contributions toward predicting WT during the testing period. Thus, the features can positively or negatively influence the predicted value. Positive contributions are in green, negative contributions are in red, and the most important features with the highest coefficients are at the top of the bar graphics.

LIME is an approximation method because a simple linear model approximates another non-linear model in the neighborhood of an instance. It selects an arbitrary model, giving more weight to samples near the point being explained and less weight to those far away. Therefore, the weights are not necessarily the same. Moreover, it is based on sampling, and true complexity cannot be captured. Even the random nature of perturbation methods, such as LIME, can lead to inconsistent or unstable results [60]. According to [61], the disadvantages of LIME include its assumption of independence among features, as well as sampling issues and uncertainty in defining the correct neighborhood. These factors decrease accuracy and lead to problems with consistency, confidence, and stability in the interpretation of tabular data. However, certain patterns can be observed within these instances. Therefore, by using average patterns, these approximations can provide an idea of how the model learns during training.

Figure 9 shows the comparison results for the hourly and daily models at Weir 32. Five instances per season were randomly selected for the 1-h and 1-d models, and certain patterns were observed between the instances across different seasons.

For the 1-h MVMD-BiLSTM model in winter, at instance 4379, solar radiation had a positive effect, while AT at the 12-h lag (AT12) had a negative effect. The AT at the 12-h lag was the most important feature (Figure 9a). At instance 5122, Solar radiation and AT12 push positively during spring, with AT12 being the most crucial feature (Figure 9b). At instance 8302, AT12 was the most influential feature during summer, with a positive effect along with solar radiation and temperature at a 0-h lag (AT0) (Figure 9c). During autumn, at instance 12,602, AT12 (the most important feature) and AT0 had a negative effect, while solar radiation and RH had a positive effect (Figure 9d). In summary, AT, with the highest lag correlation, was the most important feature, and SR had a positive influence across all seasons. AT12 was the most influential feature during winter and autumn, but it also had a negative influence due to the decrease in AT. In contrast, during spring and summer, the most crucial feature (AT12) had a positive influence due to the increase in AT.

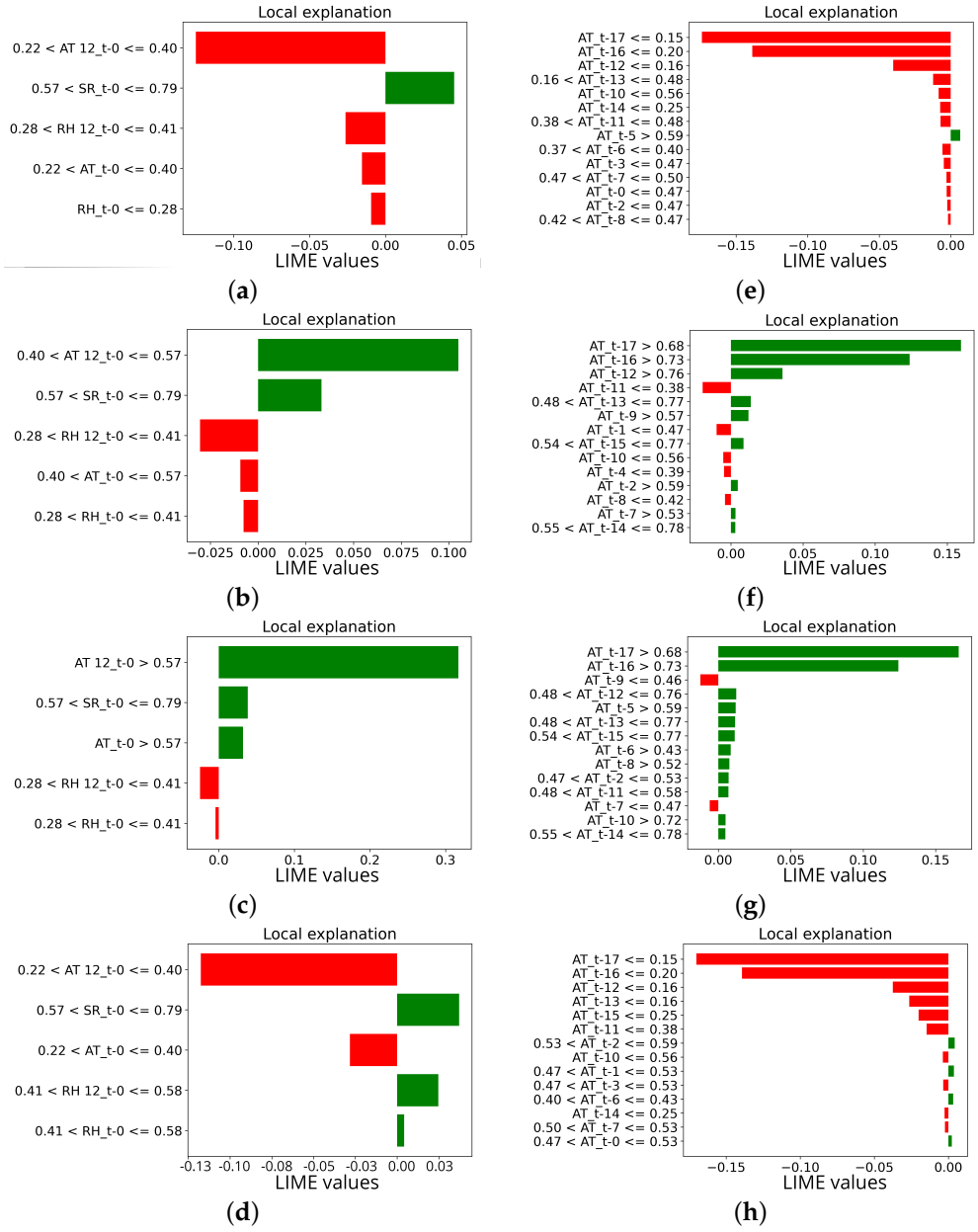


Figure 9. LIME explanation bar plots for the hourly (**a**–**d**) and daily (**e**–**h**) MVMD-BiLSTM models at Weir 32: (**a**) 4379 and (**e**) 136 random instances during winter, (**b**) 5122 and (**f**) 306 random instances during spring, (**c**) 4379 and (**g**) 327 random instances during summer, and (**d**) 12,602 and (**h**) 455 random instances during autumn. Positive contributions are in green and negative contributions are in red color.

For the 1-d MVMD-BiLSTM model, the most significant features had a negative effect in winter (instance 136) and autumn (instance 455) but a positive effect in spring (instance 306). Summer (instance 327) did not present a specific pattern; it featured a weight in each direction. These figures can be read as feature importance plots, with the most important features having the highest coefficients. Therefore, AT was the most important feature across all seasons, which was discretized, with either positive or negative effects on the final result depending on the weather conditions. Other features also have an impact but are not sufficient to significantly influence the WT prediction.

Consequently, when modeling hourly data, it is important to take heat fluxes into account because of the significant daily variations and heat exchange in WT over the day. The 1-d component can be modeled using AT data. Thus, these overall LIME results can

enhance the trustworthiness of the hybrid MVMD-BiLSTM model. Applying this approach offers explanations of the proposed black-box' model, addressing a gap in the literature.

The aforementioned findings indicate that the demand for explainability in AI models is increasingly critical across various sectors due to regulatory requirements for features like interpretability, transparency, and traceability [62,63]. To cultivate trustworthy AI in the realm of WT prediction, it is essential that AI models adhere to legal and ethical standards [63]. Additionally, robust ML systems based on AI are very sensitive to small perturbations in practice, complicating the interpretation of predicted results. Variations in input data can greatly impact outcomes, particularly in critical areas that face challenges due to low-quality datasets that do not conform to independent and identically distributed data standards [62]. Thus, our study seeks to tackle these vital issues by utilizing an advanced *xAI*-based LIME tool to ensure trustworthy AI in the WT predictive framework.

Figure 10 shows boxplots for the 1-h (a) and 1-d (b) time horizon models. These figures were obtained by randomly selecting five instances for each season and then calculating the absolute average difference between the predicted values and local predictions. Thus, the average difference between the predicted values per model and local predictions was calculated. As demonstrated by the boxplots, the hourly model presented more considerable differences across all seasons compared to those of the daily model, with poorer performance during summer (with a median of 1.700 °C) and optimal performance during autumn (with a median of 1.000 °C). For the daily model, spring and summer exhibited higher differences (medians of 0.850 °C and 0.300 °C, respectively), and winter exhibited the lowest differences (median of 0.350 °C). These results confirm the visual evaluation in Figures 7 and 8, which show considerable differences during summer and spring for the hourly and daily models, respectively.

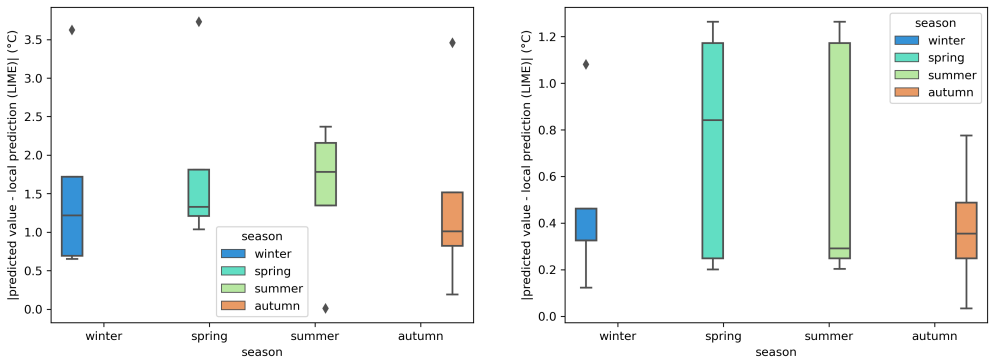


Figure 10. Boxplots of the absolute average difference between predicted values and local predictions by season using LIME for (**a**) hourly and (**b**) daily MVMD-BiLSTM models at Weir 32

4.4. Comparison to Air2Stream

The results of the MVMD-BiLSTM 1-d time horizon and Air2Stream models showed good performance with respect to the observations.

The best Air2Stream model was achieved using the parameterization $p = 5$, which assumes a constant discharge. When the model with eight parameters was executed, similar results were obtained. However, discharge is generally a minor factor in Air2Stream, particularly in our rivers, where low flow or discharge is insufficient. In low-rate rivers with slow flow and due to the local climate, there is minimal influence on discharge.

The objective model, the MVMD-BiLSTM 1-d time horizon model, was compared to the five-parameter Air2Stream parameters version (Air2Stream-5). The objective model performed better than Air2Stream-5 at Weir 32, Pooncarie, and Burtundy, with lower $RMSE$, MAE, and relative error values and higher r, WI, NSE, and LME indices (Table 9).

Table 9. Comparison of the proposed MVMD-BiLSTM objective model against the five-parameter Air2Stream physical model evaluated in respect to the *r*, *RMSE* (°C), *MAE* (°C), *WI*, *NSE*, *LME*, *RMSPE* (%) and *MAPE* (%) derived from forecasted and observed WT.

Model and Study Site	*r*	*RMSE*	*MAE*	*WI*	*NSE*	*LME*	*RMSPE*	*MAPE*
MVMD-BiLSTM | Menindee	0.984	1.039	0.856	0.917	0.965	0.839	5.840	4.400
Air2stream-5 | Menindee	0.963	1.511	1.002	0.876	0.926	0.753	8.230	6.610
MVMD-BiLSTM | Pooncarie	0.983	1.369	1.017	0.913	0.962	0.833	8.390	5.890
Air2stream-5 | Pooncarie	0.966	2.423	1.463	0.826	0.879	0.686	12.430	10.120
MVMD-BiLSTM | Burtundy	0.982	1.274	0.953	0.902	0.952	0.812	7.880	5.780
Air2stream-5 | Burtundy	0.968	1.749	1.973	0.865	0.910	0.729	11.320	8.750

The Air2Stream-5 model performed better at Menindee (LME = 0.753), followed by Burtundy (LME = 0.7290) and Pooncarie (LME = 0.686). The WI, NSE, and LME seemed sensitive to over- and underestimations, presenting significant differences in the Air2Stream-5 model at Pooncarie during the summer season, underestimating low and high temperatures during this period. However, Air2Stream is a local method that requires input from local stations and needs WT for calibration, which is typically unavailable.

Prior studies have predicted the mean daily river WT using ML methods and concluded that these models perform well and even more accurately than the Air2Stream model [34]. They conducted a comparison between an LSTM model integrated with AT and discharge as the primary predictors and the Air2Stream model that incorporated eight parameters for predicting the mean daily WT. The LSTM model achieved superior performance compared to the Air2Stream model. This was attributed to the model considering the input and output as two time-series sequences, allowing it to learn to forget unimportant information and when to retain information.

4.5. Performance of the MVMD-BiLSTM Model in Other Locations

The proposed model was trained at Weir 32 and tested in Pooncarie (approximately 150 km from Weir 32) and Burtundy (approximately 100 km from Pooncarie) to assess how well the model predicted within other datasets using gridded data.

The overall predictive skill of the MVMD-BiLSTM daily model was obtained by comparing the relative model error. The best performance was obtained at Menindee ($RMSPE$ = 5.84% and $MAPE$ = 4.4%), followed by Burtundy ($RMSPE$ = 7.88% and $MAPE$ = 5.78%) and Pooncarie ($RMSPE$ = 8.39% and $MAPE$ = 5.89%). These results exhibit the same trend as those obtained using the Air2Stream-5 model.

The WTs at Weir 32 and Burtundy have comparable mean (19.29 and 18.81 °C), maximum (32.62 and 32.19 °C), and STD (5.66 and 5.74 °C) values. In contrast, at Pooncarie, the mean WT difference was higher (Δ = 1.1473 °C), and the mean AT difference was lower ($\Delta = -1.0838$ °C). Although both locations exhibited good performances, the MVMD-BiLSTM model provided overall lower relative errors in Burtundy than in Pooncarie (Table 9).

Certain studies have compared gridded and observational data. A study in the Apulia region (southern Italy) compared the use of gridded data with observed data, focusing on AT and precipitation data. They revealed that gridded data can reproduce the observed trend patterns with some deviations, indicating the general spatiotemporal evolution of these variables in the region [64]. Another study assessed how well gridded data replicated observed data for precipitation and temperature and observed that the coarse spatial resolution of the gridded data could misrepresent complex topographies, such as the Chilean topography [65]. We do not have implications regarding the topography of this region, but they demonstrated differences between gridded data and observations.

In this study, despite merging observation (MNT) and gridded data (AT, RH, SR, and EVA), the data with a coarse spatial resolution (SILO 5 km × 5 km data and Meteoblue 30 km × 30 km data) could explain the underestimation of WT during the spring and summer seasons. It is also possible that these data cannot capture certain underlying

processes or rare events, ignoring their existence. It is important to balance the data to increase the weight of these rare events and to incorporate physics at various stages in the ML process such that the dynamics can be captured. However, performance is not always good or bad when the grid is coarser, as the MVMD-BiLSTM model can reproduce the overall dynamics of the WT in these locations.

5. Conclusions and Future Work

This study aimed to design an MVMD method combined with a BiLSTM model (MVMD-BiLSTM) to forecast WT. The results confirmed that the MVMD-based hybrid models are considerably better for forecasting the hourly and daily WTs at Weir 32 than the standalone model versions of the LSTM, GRU, BiLSTM, and BiGRU models. In particular, among the hybrid and standalone models considered, the results showed that the proposed MVMD-based BiLSTM model was significantly superior to the BiLSTM (and other standalone) model, which was demonstrated from Tables 4–7. The MVMD method offers significant benefits in inaccurate model creation, which leads to accurate performance [39]. First, the MVMD approach splits the complex dataset into variational mode signals. This ability to decompose and analyze multiple correlated variables simultaneously leverages cross-variable relationships among predictor and target variables that provide improved insights and predictive performance due to the integration of dependencies among variables. This is carried out using variational principles to decompose signals into intrinsic mode functions (IMFs) or components that are smoother and more interpretable. The primary advantage here is a better separation of mixed signals or patterns, leading to more robust feature extraction. Second, with respect to noise reduction and signal clarity for the BiLSTM predictive model, the MVMD effectively separates noise from true signal components in multivariate datasets. Therefore, this capability of the MVMD method provides enhanced signal-to-noise ratio detection and improves the accuracy and reliability of the BiLSTM model. Thirdly, compared to other decomposition tasks, the MVMD method aims to mitigate the mode-mixing issues by separating distinct frequency bands and patterns more effectively, which leads to improved performance of the resulting BiLSTM model. A higher accuracy of the Bi-LSTM is therefore achieved by the provision of cleaner, non-overlapping signal components, which is critical for accurate analysis and predictions generated by the BiLSTM model, as noted in Tables 4–7. The distinct advantage of MVMD-based models over standalone models was clearly demonstrated using several performance metrics (r, $RMSE$, MAE, WI, NSE, LME, $RMSPE$, and $MAPE$). Overall, the performance of the MVMD-BiLSTM model was much better compared with all the other comparison models.

Overall, the MVMD-BiLSTM model demonstrated good dynamics in reproducing WT, and the 1-d model generally shows improved performance compared to the 1-h model. However, models performing poorly over time were observed when the time horizon increased from 1-h to 24-h and from 1-d to 7-d. This was due to model degradation, which led to less accurate predictions and caused the model to potentially miss new trends.

The model-agnostic LIME tool delivers local model explainability, and this study attempted to explain the MVMD-BiLSTM model's predictions for WT. Therefore, in both time horizon models (1-h and 1-d), AT was the most significant predictor variable for WT at Weir 32, which was consistent with highly associated variables. However, in the 1-h horizon model, other features (SR and RH) significantly contributed and changed depending on the season.

In addition, this study demonstrated the ability of gridded data to reproduce the overall dynamics of WT in these locations. However, deviations were observed owing to the inability of gridded data to capture certain underlying processes or rare events, as it can only explain variability to a certain extent. We also acknowledge that in the present study, no lagged combinations of predictor datasets were used for the AI models. Therefore, we recommend that future studies address this limitation and benefit by testing lagged features, which might capture temporal dependencies more effectively. Furthermore, we note that the proposed MVMD-BiLSTM model was developed specifically for Weir 32, but

the performance has also been assessed for two other locations with similar datasets. These locations include Pooncarie and Burtundy sites, which demonstrated reliable and accurate performance. However, it would be useful to discuss the generalizability of the proposed MVMD-BiLSTM Model to other regions with similar datasets in future work.

This case study supports the merits of the MVMD-BiLSTM predictive model for attaining accuracy in forecasting WT. However, the reproducibility of this research is a potential limitation due to the lack of open data available for these locations. Even when in situ measurements and gridded data were used in this study, additional simulated data for these locations was required from Meteoblue because of this lack of data. Another limitation in this study includes the use of a unique *xAI* method (LIME) to generate local explanations, which present accuracy issues for tabular data. For future tests, the WT can be forecast at other points where we currently lack sufficient data combining in situ measurements gridded and simulated data. Nevertheless, to determine if the approach is transferable to other locations, future research should focus on further testing DL methods and incorporating physics at various stages in the ML process to capture the underlying dynamics, as well as other *xAI* methods should be used to produce explanations in deep neural networks.

In the Murray–Darling Basin, it is imperative to develop tools that rely on a few input variables, primarily meteorological drivers, due to the scarcity of data in many locations. This case study demonstrates the merits of the MVMD-BiLSTM predictive model in achieving accuracy for forecasting WT. Therefore, one application of the case study is predicting hourly and daily WTs when observed data are limited. This prediction helps in understanding the WT and the potential impact of WT changes on aquatic organisms and ecological processes in the river. Also, an accurate WT forecast can be used for management decisions in the case of water quality concerns, and if we have an accurate temperature forecast, we can better forecast other water quality parameters that depend on WT, like dissolved oxygen, and therefore predict fish kills.

Author Contributions: L.B.M.: Writing—original draft, Writing—review and editing, Methodology, Data curation, Conceptualization, Model Development, Visualization. K.J.: Writing—review and editing, Methodology, Data curation, Conceptualization, Supervision, Funding Acquisition. R.C.D.: Writing—review and editing, Methodology, Data curation, Conceptualization, Supervision, Funding Acquisition, Project administration. M.A.: Writing—review and editing, Supervision. S.S.P.: Writing—review and editing. N.D.: Writing—review and editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Southern Queensland Domestic PhD Scholarship and a Top-Up Scholarship by the CSIRO.

Data Availability Statement: The data presented in this study are available on request from the first author, excluding some data protected by copyright.

Acknowledgments: The first author (LBM) is supported by a UniSQ Domestic Stipend, Australian Government Research and Training (RTP) and CSIRO Environment Top-up Scholarship (2022–2025) funding.

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

References

1. Joehnk, K.D.; Graham, K.; Sengupta, A.; Chen, Y.; Aryal, S.K.; Merrin, L.; Durr, P.A. The Role of Water Temperature Modelling in the Development of a Release Strategy for Cyprinid Herpesvirus 3 (CyHV-3) for Common Carp Control in Southeastern Australia. *Water* **2020**, *12*, 3217. [CrossRef]
2. Simpson, H.; Cane, M.; Herczeg, A.; Zebiak, S.; Simpson, J. Annual river discharge in southeastern Australia related to El Nino? Southern Oscillation forecasts of sea surface temperatures. *Water Resour. Res.* **1993**, *29*, 3671–3680. [CrossRef]
3. Kaushal, S.S.; Gold, A.J.; Mayer, P.M. Land Use, Climate, and Water Resources-Global Stages of Interaction. *Water* **2017**, *9*, 815. [CrossRef] [PubMed]
4. Climate Change in Australia. *Australian Climate Trends*; CSIRO: Canberra, Australia, 2024.

5. Australian Academy of Science. *Investigation of the Causes of Mass Fish Kills in the Menindee Region NSW over the Summer of 2018–2019*; Australian Academy of Science: Canberra, Australia, 2019.
6. Murray-Darling Basin Authority. *Climate Change and the Murray-Darling Basin Plan*; Murray-Darling Basin Authority: Canberra, Australia, 2019.
7. Murray-Darling Basin Authority. *Response Fish Death Events Recommended Action Plan*; Murray-Darling Basin Authority: Canberra, Australia, 2019.
8. Vertessy, R.; Barma, D.; Baumgartner, L.; Mitrovic, S.; Sheldon, F.; Bond, N. *Independent Assessment of the 2018–19 Fish Deaths in the Lower Darling*; Australian Government: Canberra, Australia, 2019.
9. Ellis, I.; Meredith, S. *An Independent Review of the February 2004 Lower Darling River Fish Deaths: Guidelines for Future Release Effects on Lower Darling River Fish Populations*; La Trobe: Victoria, Australia, 2004.
10. Scholz, O.; Gawne, B.; Ebner, B.; Ellis, L.; Betts, F.; Meredith, S. *The Impact of Drying on the Ecology of the Menindee Lakes*; La Trobe: Victoria, Australia, 1999.
11. Caissie, D. The thermal regime of rivers: A review. *Freshw. Biol.* **2006**, *51*, 1389–1406. [CrossRef]
12. Dugdale, S.J.; Hannah, D.M.; Malcolm, I.A. River temperature modelling: A review of process-based approaches and future directions. *Earth-Sci. Rev.* **2017**, *175*, 97–113. [CrossRef]
13. Zhu, S.; Piotrowski, A.P. River/stream water temperature forecasting using artificial intelligence models: A systematic review. *Acta Geophys.* **2020**, *68*, 1433–1442. [CrossRef]
14. Webb, B.W.; Hannah, D.M.; Moore, R.D.; Brown, L.E.; Nobilis, F. Recent advances in stream and river temperature research. *Hydrol. Process.* **2008**, *22*, 902–918. [CrossRef]
15. Cole, T.; Buchack, E. *CE-QUAL-W2, a Two Dimensional, Laterally Averaged Hydrodynamic and Water Quality Model*; US Army Engineer Waterways Experiment Station: Vicksburg, MS, USA, 1995; Version 2.
16. King, I.P. *RMA-11—A Three Dimensional Finite Element Model for Water Quality in Estuaries and Streams*; Resource Modelling Associates: Sydney, Australia, 2003.
17. Wool, T.; Ambrose, R.; Martin, J. *WASP7 Temperature and Fecal Coliform Model Theory and User's Guide*; US Environmental Protection Agency, EPA: Washington, DC, USA, 2008; Volume 600.
18. Boyd, M.; Kasper, B. *Analytical Methods for Dynamic Open Channel Heat and Mass Transfer: Methodology for Heat Source Model Version 7.0*; Watershed Sciences Inc.: Portland, OR, USA, 2003. Available online: https://www.oregon.gov/deq/FilterDocs/heatsourcemanual.pdf (accessed on 12 December 2024).
19. Toffolon, M.; Piccolroaz, S. A hybrid model for river water temperature as a function of air temperature and discharge. *Environ. Res. Lett.* **2015**, *10*, 114011. [CrossRef]
20. Staples, K.; Richardson, S.; Neville, P.J.; Oosthuizen, J. An Improved Shallow Water Temperature Model for An Australian Tidal Wetland Environment Using Publicly Available Data. *Water* **2023**, *15*, 2221. [CrossRef]
21. Hong, Y.S.T. Dynamic nonlinear state-space model with a neural network via improved sequential learning algorithm for an online real-time hydrological modeling. *J. Hydrol.* **2012**, *468*, 11–21. [CrossRef]
22. Hebert, C.; Caissie, D.; Satish, M.G.; El-Jabi, N. Modeling of hourly river water temperatures using artificial neural networks. *Water Qual. Res. J. Can.* **2014**, *49*, 144–162. [CrossRef]
23. Piotrowski, A.P.; Osuch, M.; Napiorkowski, M.J.; Rowinski, P.M.; Napiorkowski, J.J. Comparing large number of metaheuristics for artificial neural networks training to predict water temperature in a natural river. *Comput. Geosci.* **2014**, *64*, 136–151. [CrossRef]
24. Zhu, S.; Nyarko, E.K.; Hadzima-Nyarko, M. Modelling daily water temperature from air temperature for the Missouri River. *PeerJ* **2018**, *6*, e4894. [CrossRef] [PubMed]
25. Temizyurek, M.; Dadaser-Celik, F. Modelling the effects of meteorological parameters on water temperature using artificial neural networks. *Water Sci. Technol.* **2018**, *77*, 1724–1733. [CrossRef]
26. Jang, J.S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [CrossRef]
27. Zhu, S.; Heddam, S.; Nyarko, E.K.; Hadzima-Nyarko, M.; Piccolroaz, S.; Wu, S. Modeling daily water temperature for rivers: Comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models. *Environ. Sci. Pollut. Res.* **2019**, *26*, 402–420. [CrossRef] [PubMed]
28. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.
29. Grbić, R.; Kurtagić, D.; Slišković, D. Stream water temperature prediction based on Gaussian process regression. *Expert Syst. Appl.* **2013**, *40*, 7407–7414. [CrossRef]
30. Zhu, S.; Heddam, S.; Wu, S.; Dai, J.; Jia, B.; Dai, J.; Jia, B. Extreme learning machine-based prediction of daily water temperature for rivers. *Environ. Earth Sci.* **2019**, *78*, 202. [CrossRef]
31. Rehana, S. River water temperature modelling under climate change using support vector regression. In *Hydrology in a Changing World: Challenges in Modeling*; Springer: Gewerbestrasse/Cham, Switzerland, 2019; pp. 171–183.
32. Lu, H.; Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **2020**, *249*, 126169. [CrossRef]
33. Heddam, S.; Ptak, M.; Sojka, M.; Kim, S.; Malik, A.; Kisi, O.; Zounemat-Kermani, M. Least square support vector machine-based variational mode decomposition: A new hybrid model for daily river water temperature modeling. *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 71555–71582. [CrossRef] [PubMed]

34. Qiu, R.; Wang, Y.; Rhoads, B.; Wang, D.; Qiu, W.; Tao, Y.; Wu, J. River water temperature forecasting using a deep learning method. *J. Hydrol.* **2021**, *595*, 126016. [CrossRef]
35. Wu, J.; Wang, Z. A Hybrid Model for Water Quality Prediction Based on an Artificial Neural Network, Wavelet Transform, and Long Short-Term Memory. *Water* **2022**, *14*, 610. [CrossRef]
36. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
37. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
38. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019.
39. Ali, M.; Prasad, R.; Jamei, M.; Malik, A.; Xiang, Y.; Abdulla, S.; Deo, R.C.; Farooque, A.A.; Labban, A.H. Short-term wave power forecasting with hybrid multivariate variational mode decomposition model integrated with cascaded feedforward neural networks. *Renew. Energy* **2024**, *221*, 119773. [CrossRef]
40. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.
41. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.
42. Xiong, C.; Merity, S.; Socher, R. Dynamic memory networks for visual and textual question answering. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.
43. Piccolroaz, S.; Calamita, E.; Majone, B.; Gallice, A.; Siviglia, A.; Toffolon, M. Prediction of river water temperature: A comparison between a new family of hybrid models and statistical approaches. *Hydrol. Process.* **2016**, *30*, 3901–3917. [CrossRef]
44. Rehman, N.U.; Aftab, H. Multivariate Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2019**, *67*, 6039–6052. [CrossRef]
45. Bowling, L.; Baldwin, D.; Merrick, C.; Brayan, J.; Panther, J. Possible drivers of a Chrysosporum ovalisporum bloom in the Murray River, Australia, in 2016. *Mar. Freshw. Res.* **2018**, *69*, 1649–1662. [CrossRef]
46. Bureau of Meteorology. *Climate Data Online*; Bureau of Meteorology: Sydney, NSW, Australia, 2024.
47. WaterNSW. *Continuous Water Monitoring Network*; WaterNSW: Parramatta, NSW, Australia, 2024.
48. Jeffrey, S.J.; Carter, J.O.; Moodie, K.B.; Beswick, A.R. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ. Model. Softw.* **2001**, *16*, 309–330. [CrossRef]
49. Meteoblue, A.G. *Weather Simulation Data*; University of Basel: Basel, Switzerland, 2024.
50. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for Large-Scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016.
51. Ketkar, N. Introduction to Keras. In *Deep Learning with Python: A Hands-On Introduction*; Apress: New York, NY, USA, 2017; pp. 97–111. [CrossRef]
52. Li, C.; Il, E. Little's test of missing completely at random. *Stata J.* **2013**, *13*, 795–809. [CrossRef]
53. Enders, C.K. *Applied Missing Data Analysis*; Guilford Publications: New York, NY, USA, 2022.
54. Muzellec, B.; Josse, J.; Boyer, C.; Cuturi, M. Missing data imputation using optimal transport. In Proceedings of the International Conference on Machine Learning, Online, 12–18 July 2020.
55. Legates, D.R.; McCabe, G.J., Jr. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [CrossRef]
56. Ribeiro, M.T.; Singh, S.; Guestrin, C. Why should i trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
57. Garreau, D.; von Luxburg, U. Looking deeper into tabular LIME. *arXiv* **2020**, arXiv:2008.11092.
58. Lewis, C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*; Butterworth Scientific: Oxford, UK, 1982.
59. Deo, R.C.; Wen, X.; Qi, F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl. Energy* **2016**, *168*, 568–593. [CrossRef]
60. Ghorbani, A.; Abid, A.; Zou, J. Interpretation of neural networks is fragile. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019.
61. An, J.; Zhang, Y.; Joe, I. Specific-Input LIME Explanations for Tabular Data Based on Deep Learning Models. *Appl. Sci.* **2023**, *13*, 8782. [CrossRef]
62. Holzinger, A. The Next Frontier: AI We Can Really Trust. In *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*; Springer Nature: Gewerbestrasse/Cham, Switzerland, 2021; pp. 427–440. [CrossRef]
63. Holzinger, A.; Dehmer, M.; Emmert-Streib, F.; Cucchiara, R.; Augenstein, I.; Ser, J.D.; Samek, W.; Jurisica, I.; Díaz-Rodríguez, N. Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence. *Inf. Fusion* **2022**, *79*, 263–278. [CrossRef]

64. My, L.; Di Bacco, M.; Scorzini, A.R. On the use of gridded data products for trend assessment and aridity classification in a Mediterranean context: The case of the Apulia Region. *Water* **2022**, *14*, 2203. [CrossRef]
65. Schumacher, V.; Justino, F.; Fernández, A.; Meseguer Ruiz, O.; Sarricolea, P.; Comin, A.; Peroni Venancio, L.; Althoff, D. Comparison between observations and gridded data sets over complex terrain in the Chilean Andes: Precipitation and temperature. *Int. J. Climatol.* **2020**, *40*, 5266–5288. [CrossRef]