

# Groundwater level predictions in the Thames Basin, London over extended horizons using Transformers and advanced machine learning models

Ali J. Ali<sup>a,\*</sup>, Ashraf A. Ahmed<sup>a,\*\*</sup>, Maysam F. Abbod<sup>b</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, Brunel University London, Uxbridge, UB8 3PH, UK

<sup>b</sup> Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge, UB8 3PH, UK

## ARTICLE INFO

Handling Editor: Jin-Kuk Kim

### Keywords:

Transformer model  
LSTM  
Fractured porous media  
Long-term time-series predictions  
AI  
Deep learning

## ABSTRACT

This study breaks new ground by using the Temporal Fusion Transformer (TFT) method for groundwater level prediction, addressing the complex dynamics of the Thames Basin aquifer in England. Our research combines extensive hydrological data collected from the Thames Basin with advanced machine learning, where a complex network of rivers and streams substantially affects groundwater dynamics. Unlike previous studies, this research focuses on long-term forecasting with deep learning, offering, for the first time, a 60-day prediction horizon based on daily data. To rigorously examine the model performance and robustness on new, unseen data, we applied the walk-forward validation method and other matrices such as RMSE and  $R^2$  coupled with the Holdout technique. The models used were Long Short-Term Memory (LSTM), Attention-based LSTM, LSTM with Bayesian optimisation, Attention-based LSTM with Bayesian optimisation and TFT. They were used on the basin's Chalk, Jurassic Limestone, and Lower greensand aquifers. Whilst both LSTM models were optimised using the Bayesian technique, TFT was applied for its inherent capability in complex time series. Our methodology processed historical groundwater and rainfall data from 2001 to 2023, accounting for the potential lag in aquifer response to the proximity of the river system. The dataset served as training, validation, and holdout for each model, focusing on capturing the dynamic temporal fluctuation. The results clearly showed the superiority of the TFT model in all aquifer types compared to other models across all horizons. The Limestone had the greatest result in the 7-day projections, with an RMSE of 0.02 and  $R^2$  of 0.98; Whilst the Chalk and Lower greensand, had RMSEs of 0.03 with  $R^2$  values of 0.75 and 0.95, respectively. The Limestone aquifer performed best for the 30-day horizon again (RMSE = 0.06,  $R^2$  = 0.85), with the Chalk and Lower greensand aquifer yielding RMSE of 0.04 and 0.12 and  $R^2$  values of 0.64 and 0.74, respectively. In the 60 days predictions, the best results were observed in the limestone aquifer with RMSE of 0.09 and  $R^2$  of 0.65 in holdout validation. However, in chalk and lower greensand aquifers, the TFT showed RMSEs of 0.05 and 0.15 and  $R^2$ s of 0.45 and 0.58, respectively. Traditional LSTM models demonstrated limited predictive power compared to the main model TFT, while the attention mechanism slightly improved the accuracy. This study not only sets a new benchmark in hydrological modelling accuracy but also highlights the potential of advanced machine learning in managing complex aquifers and predicting the water table.

## 1. Introduction

In England, 27% of groundwater is used for drinking water supply, which requires blending and treatment to meet national standards (United Kingdom Water Industry Research, 2004). However, this vital source of groundwater faces challenges in achieving satisfactory quantitative status, as assessed by the Environment Agency (2022). Therefore, a comprehensive understanding of the historical, present, and

future conditions of groundwater empowers water sector practitioners to develop strategies for sustainable socio-economic development through improving water resources planning and management (Wada et al., 2010). These sectors aligned with the goals of Agenda (2030) introduced by the United Nations (United Nations, 2015), particularly goal 6, which addresses different Sustainable development goals slated for completion by 2030.

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [ali.ali@brunel.ac.uk](mailto:ali.ali@brunel.ac.uk) (A.J. Ali), [ashraf.ahmed@brunel.ac.uk](mailto:ashraf.ahmed@brunel.ac.uk) (A.A. Ahmed).

<https://doi.org/10.1016/j.jclepro.2024.144300>

Received 3 March 2024; Received in revised form 10 October 2024; Accepted 22 November 2024

Available online 28 November 2024

0959-6526/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### 1.1. Limitation of classical models

Groundwater Level (GWL) simulation presents a challenging task due to several climatic, hydrogeological and topographic factors (Sadeghi et al., 2017; Afzaal et al., 2019). Historically, groundwater modelling has relied on numerous simulation techniques, including physically rooted conceptual models (Kennedy et al., 2003; Saberian et al., 2017; Gupta et al., 2019; Liu et al., 2023), as well as numerical models, such as the meshless method (Mohtashami et al., 2017), element free (Pathania et al., 2019), and boundary element method (El-Harrouni et al., 1996). Despite the robustness and reliability classical models have demonstrated, these traditional modelling methods have shown limitations in precision and accuracy. These methods are constrained by various factors, such as aquifer properties, porous media geology and basement topography (Buczko et al., 2010; Barnett et al., 2013). Moreover, the physical-based model comes with large data requirements, reliance on simplifying assumptions and complex mathematical tools (Duan et al., 1992; Beven, 2011; Mehr et al., 2013); these challenges restrict the validation accuracy in the model (Condon et al., 2021).

Recognising these complex challenges in predicting GWL, specifically due to climatic, hydrogeological and topographic variabilities, Artificial Intelligence (AI) methods have overcome some limitations of traditional models and are increasingly used in groundwater quality and quantity (Pannu, 2015; Khaki et al., 2015; Sakizadeh, 2016; Bagheri et al., 2017; Muñoz et al., 2023; Ahmed et al., 2024). This advancement has paved the way for more sophisticated approaches, such as Long short-term memory (LSTM) introduced by Hochreiter and Schmidhuber (1997), which are particularly suited for analysing time series data. The integration of Bayesian optimisation (BO) further enhances the predictive capabilities of these models by searching and choosing the best hyperparameters, allowing better performance and more precise groundwater predictions (Sameen et al., 2020).

### 1.2. LSTM and attention mechanism in hydrology

In the realm of hydrological and meteorological forecasting, LSTM is recognised as one of the most effective methods for predicting hydrological variables (Sherstinsky, 2020); its strength lies in its ability to understand long-term dependencies in sequential data. To further enhance the predictive accuracy of the LSTM model, researchers have put an effort into identifying and emphasising the most instructive time points whilst downplaying less relevant ones (Jozefowicz et al., 2016; Ding et al., 2020). The attention mechanism, a relatively new and dynamic feature inspired by human attention, has been deeply introduced in several papers (Vaswani et al., 2017). Its application across various fields has demonstrated improved efficiency and precision (Ghaffarian et al., 2021; Lieskovská et al., 2021; Li et al., 2023). While its integration in hydrological forecasting is still emerging, early implementations have shown promising results (Ding et al., 2019; Chen et al., 2020, 2021; Ehteram, 2023). A comprehensive review of the different AI models for GW predictions is provided by Ahmed et al. (2024).

### 1.3. Temporal Fusion Transformer

Building upon the foundational strengths of LSTM models and the enhanced focus provided by attention mechanisms, this study also incorporates the innovative Temporal Fusion Transformer (TFT) to leverage groundwater level prediction. The TFT, a cutting-edge time series model, excels in handling and integrating several types of inputs, such as static variables, known future inputs and historical time-series data (Ali and Ahmed, 2024; Lim et al., 2021). This paves the way for a comprehensive understanding of temporal patterns and relationships, which is crucial in hydrological studies.

Unlike traditional recurrent models, TFT employs a novel architecture that combines convolutional components for local processing with

attention mechanisms for multi-horizon forecasting (Lim et al., 2021). This model offers a unique balance between flexibility and interpretability in handling complex data, making it privileged in forecasting hydrological aspects (Fayer et al., 2023). In the context of groundwater forecasting, the ability of TFT to process and integrate diverse data sources, including hydrological, climatic and anthropogenic factors, surpasses other methods. A noteworthy aspect of the TFT is its inherent proficiency in handling such complex scenarios, even without applying hyperparameter tuning such as BO (Lim et al., 2021). The significance of this advantageous capacity lies in the model's ability to effectively learn and adapt to the data's temporal dynamics, making it suited for hydrological studies where the variability and interaction of several factors over time are complex and subtle.

### 1.4. Model performance metrics and validation techniques

In terms of evaluations, it is crucial to address the approaches used to assess their success in time-series forecasting. Rolling window analysis has traditionally been used as such evaluations, providing insight into a model's stability over time (Kombo et al., 2020; Hussein et al., 2020). This method computes parameter estimates over a fixed-size window throughout the sample, which provides a gauge for parameter constant, which is crucial in dynamic situations. However, applying this technique in hydrological forecasting requires careful consideration, particularly for GWL. Because of their complex systems and temporal dynamics, groundwater systems may not comply with the assumption of constant parameters, a limitation addressed in financial time series (Zivot et al., 2003). To complement rolling window analysis, our research also integrates walk-forward validation and holdout technique (Roelofs et al., 2019; Cerqueira et al., 2020; Guo et al., 2021). Walk-forward validation enables ongoing adaption of the training dataset by testing the model against previously unseen data points, thus mimicking real-world forecasting scenarios under changeable conditions. The holdout technique boosts our strategy by retaining a subset of the dataset for a final single test, ensuring that the performance of the model is tested against entirely new data.

### 1.5. Challenges of Thames Basin

The river basin management plans by the Environment Agency (2022) published some pressing challenges affecting the current and potential future of the water environment. These challenges represent the key issues of managing and sustaining these environments. Many water bodies, including rivers, lakes, coastal areas, streams and aquifers, suffered harm due to factors like urban development and industrial activities (Miller and Hutchins, 2017), agriculture (Taylor et al., 2016) and changes made to provide flood protection (Rubinato et al., 2019). The additional pressures of climate change and population growth add to these issues (Frederick and Major, 1997). In addition, excessive groundwater abstraction can directly impact the river ecosystem, leading to unsustainable water bodies (Zektser et al., 2005). According to the Environment Agency (2022), if no action is taken, the problem will increase, and many areas of England will face water shortages by 2050. The complexity of the Thames Basin, characterised by its three principal aquifers—the Chalk, the Jurassic limestones, and the Lower greensand—can pose challenges even for the most advanced AI methods (Mathers et al., 2014).

### 1.6. Research gap and novelty

Our work contributes greatly to the knowledge and implementation of machine learning in hydrological forecasting by offering numerous unique techniques and methodologies. Key contributions include:

- Most previous research, such as Cheng et al. (2020), focuses on predicting for up to 20 days. We extend the forecasting period to 30

and 60 days, filling the significant gap in long-term GWL projection that has received less attention in the literature.

- We present TFT model for capturing complex temporal correlation in hydrology data, establishing a new standard in the field. This study differs from prior research in that it employs classic and advanced ML approaches to improve forecast accuracy and give in-depth insight into water resources management in the Thames Basin.
- For assessing model performance and identifying overfitting, the majority of earlier research used fundamental measures including RMSE, MAE, and MSPE solely (May-Lagunes et al., 2023; Chen et al., 2020). Even though these measures are important, it's possible that they may not accurately represent model accuracy or regression performance quality, especially when used with fresh, untested data across longer predicting horizons (Chicco et al., 2021). In order to close this gap, our work uses walk-forward validation approaches and holdout methods in addition to these fundamental measures. These extra methods offer a more thorough evaluation, guaranteeing that the model is resilient across various data subsets and periods and that it is constantly checked for indications of overfitting.
- Our study focuses on the 7–30 – 60 days prediction period based on daily data, which is significant for water resources management in the Thames Basin and practical given the available data and environmental conditions.

In addition to our comprehensive methodology, this study addresses a critical challenge of model overfitting in GWL prediction particularly for extended forecasting periods incorporating fractured and granular porous medium, in addition to our thorough approach. While traditional methods, such as rolling window analysis, are helpful in assessing the stability of the model, they frequently fall short in capturing the seasonal fluctuations and long-term hydrological patterns that are crucial in these intricate systems. Our methodology incorporates walk-forward validation, the holdout technique, and delayed and rolling window features to reduce the danger of overfitting and provide a thorough evaluation across various aquifer types. This method improves our models' predictability and accuracy and provides insightful information for hydrological modelling in the region's varied aquifer systems.

## 2. Materials and methods

In this section, we embark on a comprehensive investigation of GWL prediction in the Thames Basin using advanced machine learning. Our methodology integrates a blend of deep learning models – the standard

LSTM and attention-based LSTM, in addition to their variant, which is fine-tuned through Bayesian optimisation and TFT. First, the data collection and pre-processing will be explained; then, the ML models will be explored. Finally, the evaluation matrices will be explained.

### 2.1. Data collection and pre-processing

We compiled a large historical groundwater and rainfall database from the Environment Agency's [Hydrology Data Explorer \(Hydrology Data Explorer - Explore\)](#). This resource provides extensive spatial and temporal hydrological data across the United Kingdom. Thames Basin district covers over 16 000 km<sup>2</sup> (Fig. 1) located in London, the UK's capital, and it is considered one of the most densely populated river basins serving over 15 million people (Bearcock and Smedley, 2010; Environment Agency, 2022). This district encompasses not only the entirety of Greater London but also extends from north Oxfordshire southwards to Surrey and Gloucester in the west to the Thames Estuary, including parts of Kent in the east. The Thames Basin area is among the driest in the country, getting, on average, around 690 mm of rainfall. This is less than the national average of 897 mm, in addition to the increasing population, which places significant stress on the available water supplies which are used to the fullest (Bearcock and Smedley, 2010).

The Thames basin is a region critical due to its dense population and significant hydrological challenges due to its complex hydraulic properties (Neal et al., 2006; Roysse et al., 2009). The site selection criteria were severe; we selected representative point in each aquifer – chalk, limestone, and lower greensand, focusing on data continuity from 2001 to 2023 and the geographical pairing of GWL monitoring locations with nearby rainfall stations. This strategic choice was made to capture the interplay between atmospheric precipitation and subsurface water levels, which is critical to the aquifer dynamics of the Thames Basin.

The initial data was recorded hourly, presenting some temporal coverage inconsistencies. Therefore, the normalisation technique was conducted to uniform daily time series (Kang and Tian, 2018), where the mean value of available data was computed for each day. This was critical to offset the impact of missing data records and provide a daily temporal resolution that is more congruent with hydrological trends and water resources management practices. A linear interpolation approach was used, ensuring temporal continuity since it preserves the integrity of the time series (Huang, 2021). This is critical for modelling purposes as the interruption may lead to interpretations of the underlying water dynamics.

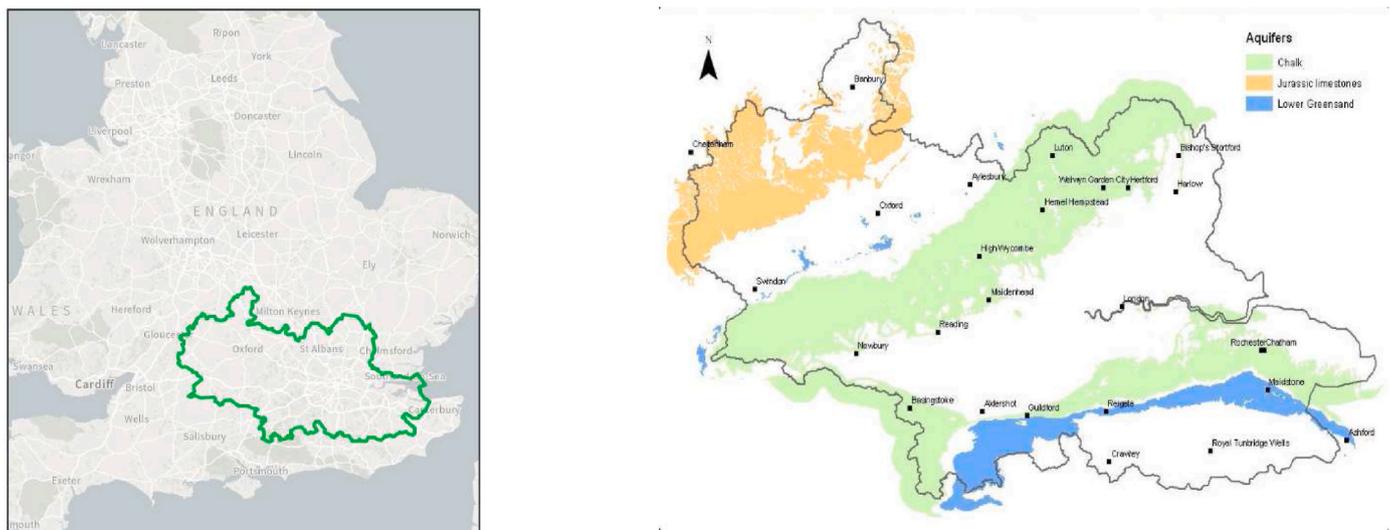


Fig. 1. Thames basin (Environment Agency, 2022).

We revised this approach to rainfall data processing in light of the proximate river's potential of minimising the direct effect of rainfall on the groundwater recharge. Recognising that the existence of the river may introduce a lag in the aquifer's response to precipitation (Randall and Albany, 1978), we thoroughly standardised the rainfall data, treating it as an important rather than a direct, impactful factor in our hydrological modelling framework. This advanced understanding guided our data integration method, where we aligned the rainfall and groundwater datasets under a single temporal framework.

Recognising the deep link between rainfall and groundwater levels (Fetter, 2001), we methodically integrated the rainfall and groundwater levels based on their corresponding dates. Combining these two datasets allows for a thorough investigation of how groundwater levels fluctuate over time in response to rainfall. While also understanding that the river's proximity may cause a lag rather than an immediate shift in groundwater level.

Following that, we concentrated on splitting the combined groundwater and rainfall data, and the dataset was divided into training, validation, and holdout purposes. This division was crucial as it is not only for the effective functioning of the ML models (Ransom et al., 2017), but also for assuring their relevance and application in hydrological forecasting within Thames Basin. The quality of these datasets determines the dependability of walk-forward validation findings. The ability of a model to perform effectively in walk-forward validation reflects not only on its design but also on the capacity of the historical data to include the complete range of environmental conditions affecting the aquifer (Kaastra and Boyd, 1996). The training set received 70% of our dataset. The requirement is to offer the model with a full understanding of the hydrological patterns and relationships between GWL and rainfall. Given the complex and dynamic nature of the hydrological processes, such a significant volume of training data was required for the models to learn and duplicate these intricate patterns effectively. The remaining 30% sample was divided into validation and holdout sets of 15% each. The validation set plays a critical role in the model development phase, primarily used for fine-tuning the models so that they can adapt to and accurately predict hydrological trends without overfitting to training data. The 15% holdout set was held aside for a final single evaluation to assess the robustness of the model in handling unseen data. The holdout sets are the best practice in ML that ensures the final model evaluation is rigorous and trustworthy (Dwork et al., 2015).

The data was further processed to suit the needs of the models; this included bending the data into specified forms that reflected the time-step and feature considerations required for these advanced modelling approaches. The dimensions of the reshaped data set for each geological formatting-Chalk, Limestone, and Lower greensand-were as follows:

- Chalk Formation: Training set (5494), Validation set (1107), Holdout set (1108)
- Limestone Formation: Training set (5718), Validation set (1155), Holdout set (1156).
- Lower greensand Formation: Training set (5562), Validation set (1122), Holdout set (1122).

## 2.2. Machine learning models

### 2.2.1. Bayesian optimisation

Bayesian optimisation (BO) is a powerful method for hyperparameter tuning in machine learning models, and it is particularly useful in time series forecasting. By using the surrogate function, the BO predict the conditional likelihood of validation set performance given hyperparameters (Alizadeh et al., 2021). Unlike random or grid search, BO keeps records of previous historical evaluations to prevent wasting computation on poor hyperparameters. When compared to standard approaches, the algorithm uses an acquisition function to find potential hyperparameters for further evaluation, optimising the search with fewer evaluations (Du et al., 2022). In our research, we combined BO

with LSTM models for fine-tuning the hyperparameters, where:

- **Model configuration:** LSTM models were built with varied numbers of units and learning rates. The architecture was designed to handle the temporal dynamics of groundwater and rainfall data effectively.
- **Parameter space:** The learning rate, number of LSTM units, and dropout rates were all placed within certain ranges to ensure robust parameter space exploration.
- **Optimisation process:** The optimisation procedure was executed over several iterations, with each iteration narrowing the search depending on the model's performance on the validation set. The goal was to reduce the mean square error (MSE), which was a significant parameter for model accuracy in our study.

Through BO, we effectively determined the ideal set of hyperparameters for our LSTM models, improving their capacity to capture the complicated relationship between GWL and rainfall data.

### 2.2.2. Long Short-Term Memory

Long short-term memory (LSTM) networks use memory blocks that consist of gates and cells, where the cell works as a conveyor belt which carries information and runs through the entire chain, whilst the gates choose what information to add and remove. Each gate is expressed by its own sets of weights and biases. The forget gate, which decides which information to erase from the memory, can be expressed as follows (Khozani et al., 2022):

$$f_t = \sigma(\omega_f \cdot [h_{t-1}, x_t], b_f)$$

where  $f_t$  is the forget gate output,  $\sigma$  is the sigmoid function,  $h_{t-1}$  is the concatenation of the previous hidden state at  $t - 1$  time,  $x_t$  is the current input,  $\omega_f$  is the weight matrix related to the forget gate,  $b_f$  is the bias associated with the forget process. So, the sigmoid function is utilised to the total of the weighted input and bias to compute the output of the forgot gate (Kong et al., 2021). Moving on, deciding what information to be stored in the memory cell is controlled by the input gate. The input gate controls the flow of new information into the memory cell by using the tanh function, which creates a new cell to save new information (Gundu and Simon, 2021). The architectures are defined as follows:

1. **Model definition:** a sequential model with multiple LSTM model is used, where the first layer is configured to return sequences, followed by another layer of LSTM. The LSTM units in each layer and the learning rate are optimised through BO.
2. **Dropout and Regularisation:** Dropout layers were employed after each LSTM layer to randomly drop out neurons during the training process (Wei et al., 2020).
3. **Output layer:** The final layer in the model is a dense layer with a (linear) activation function responsible for predicting GWL (Liu et al., 2020).
4. **Model compilation and optimisation:** The model is then complied with an MSE loss function and optimised with the (Adam) optimisation. Based on the optimised hyperparameters, the learning rate is determined.
5. **Bayesian Optimisation:** We use BO to systematically explore the hyperparameter space for hyperparameter tuning, achieving the most efficient model training.
6. **Training and validation:** Following, custom Keras Regressor Wrapper with Early Stopping and call-backs (Gulli and Pal, 2017) were used in the training process to ensure that our model is neither underfit nor overstrained.
7. **Final model training:** After hyperparameter optimisation, we build the final LSTM model with our chosen parameters. The model is trained on the full dataset with a validation split to monitor performance.

Alongside the BO LSTM models, we used a normal LSTM model with a fixed hyperparameter that was extracted from the median values of the BO search space to guarantee fairness and comparability. Reflecting the median values from the BO search space, this model was constructed using a fixed configuration of 50 LSTM units in two layers, with a 0.3 dropout rate and 0.001 learning rate. This simplified technique aims to strike a balance between efficiency and performance, using dropout layers to reduce overfitting and a dense linear activation for GWL forecasting. It is important to note that this architecture was also used for attention-based LSTM.

### 2.2.3. Attention-based LSTM cell

A way to enhance LSTM to capture high non-linearity is to incorporate a self-attention mechanism into its architecture, which applies attention scores to each observed value. This method has shown effective results in various fields in handling complex dependencies in sequential data (Pei et al., 2017). It allows the model to weigh the importance of different inputs at various time steps. This approach is particularly useful in hydrology, where the interdependency of GWL sequences is often complex and non-linear. In the used model, the self-attention mechanism is formulated using the scaled dot-product attention approach (Alizadeh et al., 2021). This approach enables the model to prioritise inputs at different time steps, improving its focus on significant elements in the data. The attention layer is strategically placed after the first LSTM layer, allowing the network to focus on the most relevant element before passing them through another LSTM layer. This can dynamically allow the attention mechanism to adjust the focus of the network, enhancing its ability to recognise significant temporal patterns. Following that is a dropout layer to reduce overfitting and a dense output layer to predict the GWL. Adding this attention mechanism to our model will leverage it by a sophisticated capacity to recognise and focus on relevant temporal patterns, which is critical for accurate long-term GWL forecasting. It represents a significant advancement in dealing with the complex dynamic and nonlinear nature of hydrological data (Vaswani et al., 2017).

### 2.2.4. Temporal Fusion Transformer model architecture

The novel attention-based architecture in TFT model offers a breakthrough in interpretability for deep learning methods, particularly fitted for complex hydrological forecasting in the Thames Basin (Junankar et al., 2023). The model employs a number of specialist components that pick key features and a series of gating layers to filter out unnecessary elements, allowing it to provide impressive performance across a range of real-world datasets (Wu et al., 2022). It is uniquely designed for multi-horizon forecasting and is capable of processing a complex mix of inputs, including static, known future and exogenous time-series data, all of which are critical in hydrological forecasting (Marcellino et al., 2006; Li et al., 2019). The TFT architecture consists of many components, each of which contributes to the ability of the model to collect the complex temporal dynamics in hydrological data (Lim et al., 2021):

1. **Gating mechanisms blocks** are used for selective non-linear processing data to adjust to changing hydrological variables, such as rainfall intensity and duration thereby enhancing model accuracy and efficiency. The formula for Gated Residual Network (GRN) has a primary input  $a$ , an optional context vector  $c$ , and uses exponential linear unit (ELU) activation (Clevert et al., 2015) and Gated Linear Units (GLUs):

$$GRN_{\omega}(a, c) = LayerNorm(a + GLU_{\omega}(\eta_1))$$

$$\eta_1 = W_1\omega\eta_2 + b_1\omega$$

$$\eta_2 = ELU(W_2\omega a + W_3\omega c + b_2\omega)$$

2. **Variable selection networks** to focus on the most important variables at each time step, limiting the influence of noisy data, which is critical in hydrology where factors like river stream and rainfall intensity can affect GWL (Appels et al., 2015; Dauphin et al., 2017). Creating variable selecting weights involves transforming input variables into a  $d_{model}$  dimensional vector, then applying the GRN and Softmax layer (Lim et al., 2021):

$$v_{x_t} = Softmax(GRN_{v_x}(\Xi t, cs))$$

3. **Static covariate encoders and temporal processing** with a sequence layer for short-term and a unique interpretable multi-head attention block for long-term dependencies, which is vital for understanding the interaction between geological features and rainfall patterns in the Thames Basin.
4. **Interpretable multi-head attention** is employed by TFT, which modifies the multi-head attention mechanism (Vaswani et al., 2017), allowing it to learn and highlight long-term relationships across different time steps, improving the detection of key occurrences such as continuous rainfall events or dry seasons.
5. **Temporal fusion decoder** to identify and incorporate significant hydrological events. Enhancing short-term interpretation and prediction accuracy.

TFT contains extra layers to handle temporal data and interpretive multi-head attention. To process input, the sequence-to-sequence layer is often comprised of LSTM units for short-term dependencies, complementing the long-term focus of the multi-head attention mechanism. Context vectors are employed to process contextual information, which is crucial for adjusting predictions on static characteristics such as soil type or land use. The model uses suitable loss functions, mean square error (MSE) or mean absolute error (MAE), and an Adam optimiser for training. The model is structured in sequence with an embedding layer for the input transformations, a multi-head self-attention layer for capturing long-term dependencies, an LSTM layer to capture complicated temporal patterns, and a dense output layer for predicting GWL. An early-stopping call-back is added to the training process to promote quick learning and prevent overfitting.

## 2.3. Evaluation methods

### 2.3.1. Rolling window analysis

Each model was evaluated using a similar assessment process. We used a rolling window approach for post-data pre-processing, which was crucial in assessing model performance. This method generates lagged and rolling window features and simulates the model's operation in real-world scenarios to forecast future GWLs. The rolling window approach estimates the model at time instance  $i$  using a sample of  $i - 1$  to  $i - N$  prior observations, providing  $h$ -step forward projections (Amor et al., 2016). The procedure is repeated, with the window moving forward one period at a time, updating the model constantly and providing fresh projection.

### 2.3.2. Holdout method

The holdout method is for estimating the performance of time series forecasting models. It is performed by splitting the data into two parts: a training and testing set. Where the model is trained on the first portion of the data and tested on the second, ensuring that the testing occurs on unseen data. It is particularly suitable for non-stationary time series since it produces more validation compared to other methods like cross-validation (Cerqueira et al., 2020). It is significant since it aids in assessing the ability of the model to generate new, unseen data, which is crucial for reliable prediction in real-world applications.

### 2.3.3. Walk-forward technique

The walk-forward validations approach is crucial for assessing time series models since it allows the model a chance to make predictions at each time step (Guo et al., 2021). This method involves training the model on historical data and then testing it on a smaller, more recent sample. The data window is pushed back, and the training and testing procedure is repeated (Wan et al., 2019). This strategy allows the model to continuously adjust to recent data, improving the prediction accuracy. It is especially useful for models that need to adjust to conditions that are changing over time, ensuring that they stay relevant and accurate.

### 2.3.4. Performance metrics (RMSE, $R^2$ , and MAE)

We evaluated the performance using three essential metrics: Root Mean Square Error (RMSE), R-squared ( $R^2$ ), and Mean Absolute Error (MAE). These metrics are chosen to provide a thorough understanding of the model predictive accuracy and reliability (Chicco et al., 2021).

- RMSE measures the predictive errors by calculating the squared root of the average of squared differences between the actual and predicted values, indicating the model's accuracy. RMSE can be calculated as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2}; \text{ (best value} = 0, \text{ worst value} = +\infty)$$

- MAE calculates the average magnitude of errors in a set of forecasts without considering their direction. It is a simple measure of prediction accuracy, where lower values indicate better performance. It can be calculated as follows:

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|; \text{ (best value} = 0, \text{ worst value} = +\infty)$$

- $R^2$  measures the proportion of variation in the dependent variable that can be redacted by the independent variables. It indicates the goodness of the fit, providing us an insight into how well our model captures the variability in GWL influences by factors like rainfall. It can be calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2}; \text{ (best value} = +1, \text{ worst value} = -\infty)$$

Calculations were made for different datasets (training, validation and holdout); these predictions, along with the actual values, are inverted using a scalar. These matrices assess and evaluate the model to capture complicated interactions between GWL and rainfall, offering a thorough understanding of how capable the model predictive abilities are in a hydrological context.

### 2.3.5. Look-back window selection

The look-back window defines a range of previous time steps that the model can observe to recognise the patterns and relationships within the sequence data. A fully connected neural network with weight sharing separates the temporal and trend aspects (Zhao et al., 2023). For the 7-day prediction horizon, a 7-day lookback window is ideal because it allows the model to capture short-term temporal trends that are relevant to the following week without overfitting. This brief timeframe gives an accurate view of the most recent conditions, which is critical for making short-term predictions.

To cover a 30-day prediction, a lookback window of 30 days is chosen to catch any major trends in GWL and rainfall. This time frame is especially useful for spotting monthly weather fluctuations and seasonal changes that may impact GWL in the near term. A longer look-back

period provides a more diversified dataset for the model to learn from, which is especially important for long-term forecasts. It helps the model grasp more complex dependencies that may not be visible in a shorter time. When the forecast horizon is extended to 60 days, a 30-day look-back time compromises between capturing longer-term patterns and being attentive to recent alterations that may affect near-future GWL. Therefore, the lookback window is set to 30 days to mitigate the risk of the model overemphasising earlier data, which may not be as indicative of upcoming situations, whilst avoiding overfitting as well.

## 3. Results and discussion

We processed a rich dataset encapsulating the complex relationship between groundwater levels and rainfall over three unique aquifer systems, chalk, limestone, and lower greensand, using a suite of advanced ML. The results (see Table 1) are systematically analysed across different temporal forecasting horizons of 7–30–60 days, emphasising hydrological consistency and predictive accuracy as measured by robust metrics (i.e. RMSE, MAE and  $R^2$ ).

The study used advanced ML techniques to address overfitting, a challenge often noted in many studies (Tao et al., 2022; Pandya et al., 2024). Using walk-forward validations, holdout sets, and rolling windows ensured the reliability and generalisability of the models used, particularly the TFT. Furthermore, few studies have put effort into predicting extended horizons, like Cheng et al. (2020), which focused on the 20-day lead periods using the LSTM model. Our research extended the forecasting to 30 and 60 days using both LSTM and TFT models. Our findings (Table 1) reflect the effectiveness of these models in the Thames Basin's complex hydrological context.

By reducing overfitting, the technique used improves forecast accuracy and generalisability in the Thames basin complex system. While the walk-forward validation approach has shown positive results in Tables 1 and it is essential to recognise the foundational role played by the prior training, testing, and holdout procedures. The robustness of the predictive variables derived from the walk-forward validation depends on the depth and breadth of the data used in these initial phases (Kaastra and Boyd, 1996). The TFT model's outstanding performance is based on the high quality of these datasets, which capture the complexities of the Thames Basin's hydrological profile.

Leveraging the information on aquifer properties such as the dual porosity of chalk (Brouyère, 2006), limestone's strong yet low permeability (Selvadurai, 2019), and the unique flow process of lower greensand-demanded the models to capture these complex dynamics. BO was used to fine-tune LSTM models, with the goal of leveraging the unique aquifer properties to increase forecasting accuracy and enhance the model performance by searching wide parameter space. However, contrary to expectations, the small difference in RMSE and  $R^2$  between conventional LSTM and LSTM with BO, in addition to the attention-based LSTM, demonstrated that extensive parameter search does not always lead to considerably enhanced predicting accuracy, as Alizadeh et al. (2021) have conducted in their study.

This minor difference emphasises an important issue of model selection and optimisation in hydrology forecasting. It suggests that the choice between a standard LSTM and a BO variant should consider not only accuracy metrics but also aspects like computational efficiency and the target aquifer properties. It also emphasises the need for careful initial parameter selection in LSTM models, which in certain cases may be as effective or even superior to a more computationally demanding optimisation approach.

However, the TFT model, which was not subjected to Bayesian optimisation, produced good results by including the different hydrological features of the Thames basin. The model excelled across all aquifer types, achieving the lowest RMSE and the best  $R^2$ , notably in the limestone aquifer (Table 1). This result demonstrates the capacity of the TFT model to capture the rapid reaction of limestone precipitation and the other groundwater behaviour of chalk and lower greensand,

**Table 1**  
Results of each model in all the three aquifers.

a) The prediction 7 days ahead							
Aquifer Type	Model	Holdout			Walk-forward		
		RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
Chalk	LSTM	0.08	0.06	0.52	0.08	0.06	0.51
	ATTETNION-LSTM	0.06	0.05	0.69	0.08	0.06	0.41
	LSTM + BO	0.06	0.05	0.68	0.08	0.06	0.49
	ATTETNION-LSTM + BO	0.05	0.04	0.73	0.09	0.07	0.37
	TFT	0.03	0.02	0.75	0.03	0.02	0.40
Limestone	LSTM	0.42	0.36	0.96	0.29	0.22	0.98
	ATTETNION-LSTM	0.28	0.20	0.98	0.32	0.25	0.98
	LSTM + BO	0.31	0.24	0.98	0.24	0.17	0.99
	ATTETNION-LSTM + BO	0.28	0.20	0.98	0.24	0.17	0.98
	TFT	0.02	0.01	0.98	0.02	0.01	0.97
Lower Green-sand	LSTM	0.41	0.23	0.93	0.51	0.32	0.89
	ATTETNION-LSTM	0.40	0.22	0.93	0.65	0.41	0.83
	LSTM + BO	0.39	0.21	0.94	0.61	0.37	0.85
	ATTETNION-LSTM + BO	0.41	0.21	0.93	0.64	0.43	0.83
	TFT	0.06	0.03	0.95	0.05	0.03	0.91
b) The prediction 30 days ahead							
Chalk	LSTM	0.08	0.07	0.44	0.08	0.06	0.53
	ATTETNION-LSTM	0.08	0.06	0.45	0.07	0.06	0.53
	LSTM + BO	0.07	0.05	0.51	0.07	0.06	0.50
	ATTETNION-LSTM + BO	0.10	0.08	0.15	0.07	0.06	0.56
	TFT	0.04	0.03	0.64	0.04	0.03	0.85
Limestone	LSTM	0.78	0.55	0.87	0.30	0.23	0.98
	ATTETNION-LSTM	0.73	0.52	0.88	0.29	0.23	0.98
	LSTM + BO	0.71	0.48	0.89	0.27	0.21	0.98
	ATTETNION-LSTM + BO	0.73	0.49	0.88	0.24	0.18	0.98
	TFT	0.06	0.04	0.85	0.04	0.03	0.87
Lower Green-sand	LSTM	0.86	0.58	0.70	0.64	0.38	0.83
	ATTETNION-LSTM	0.83	0.57	0.72	0.78	0.47	0.75
	LSTM + BO	0.81	0.54	0.73	0.55	0.33	0.88
	ATTETNION-LSTM + BO	0.83	0.56	0.72	0.73	0.45	0.78
	TFT	0.12	0.08	0.74	0.10	0.06	0.68
c) The prediction 60 days ahead							
Chalk	LSTM	0.11	0.09	0.03	0.07	0.06	0.57
	ATTETNION-LSTM	0.09	0.07	0.24	0.08	0.06	0.53
	LSTM + BO	0.09	0.07	0.33	0.08	0.06	0.52
	ATTETNION-LSTM + BO	0.10	0.08	0.24	0.07	0.06	0.56
	TFT	0.05	0.04	0.45	0.03	0.02	0.99
Limestone	LSTM	1.24	0.81	0.67	0.28	0.22	0.98
	ATTETNION-LSTM	1.55	0.90	0.49	0.32	0.24	0.98
	LSTM + BO	1.37	0.82	0.60	0.29	0.22	0.98
	ATTETNION-LSTM + BO	1.33	0.86	0.62	0.26	0.19	0.99
	TFT	0.09	0.07	0.65	0.06	0.05	0.61
Lower Green-sand	LSTM	1.10	0.75	0.49	0.58	0.37	0.86
	ATTETNION-LSTM	1.04	0.74	0.54	0.71	0.43	0.79
	LSTM + BO	1.01	0.74	0.57	0.59	0.38	0.85
	ATTETNION-LSTM + BO	1.07	0.77	0.52	0.68	0.44	0.80
	TFT	0.15	0.11	0.58	0.12	0.08	0.49

demonstrating its advanced analytical ability in hydrological forecasting.

Although the TFT excelled across all aquifers and horizons, its performance varied slightly in certain cases. In predicting 60 days, the TFT yielded the highest R<sup>2</sup> for the limestone aquifer but did not get the lowest RSME. These results emphasise the inherent complexity and unpredictability of groundwater systems rather than highlighting the model's limitations. In hydrological forecasting, the trade-off between R<sup>2</sup> and RMSE is a well-documented phenomenon that reflects the difficulties faced by various error measures in capturing all facets of model performance.

The best models were chosen based on the lowest RMSE values and the best R<sup>2</sup> obtained throughout the training, validation and holdout set (Chicco et al., 2021). The TFT model performed exceptionally well, demonstrating consistency across all horizons in the chalk aquifer (Table 1) and demonstrating its resilience in capturing the hydrological response of the aquifer to environmental variables. This superior performance demonstrates the TFT capacity to integrate complicated data relationships while accurately reflecting the actual groundwater fluctuations. Further insight into the model performance, including the attention-based LSTM and the standard LSTM, reveals that the TFT advanced temporal understanding gives a significant advantage in predicting accuracy.

The TFT model provided low RMSE and high R<sup>2</sup> values throughout all horizons and aquifer types. In chalk aquifer, for example, the TFT model produced a 7-day forecast with an RMSE of 0.03 and R<sup>2</sup> of 0.98, whereas the other models produced greater RSME values and Worse R<sup>2</sup> scores. The limestone and lower greensand showed similar patterns, with TFT retaining lower RMSE values and demonstrating greater accuracy and resilience. This shows that TFT is quite effective at modelling short-term groundwater changes in aquifers that respond quickly to rainfall. However, in more geologically complicated systems, such as the limestone aquifer, RMSE values were greater across all models, indicating the difficulty of forecasting groundwater levels in aquifers with lower permeability and more irregular water flow. However, the high RMSE values in the limestone aquifer, even for the TFT model, can be related to karst formations and fault networks, which cause fast, unexpected water level variations (Abesser et al., 2005).

A measure of the average size of error (MAE), provides information about the model's overall accuracy without considering the direction of the mistakes. Interestingly, while TFT consistently exceeds the other models in terms of RMSE, its MAE was lower, in lower greensand aquifer, for example, over the 60-day horizon, where it attained an MAE of 0.11. This suggests that, while the model made significant mistakes (as seen by the greater RMSE), the overall trend of its predictions was closer to the observed values (Chai and Draxler, 2014).

When comparing TFT and attention-LSTM models in terms of MAE, we notice that TFT had superior R<sup>2</sup> values in the chalk and lower greensand aquifers, showing it accurately recorded groundwater changes. However, in limestone aquifer, Attention-LSTM had greater MAE values than TFT, although having stronger R<sup>2</sup> in some cases (e.g., Attention LSTM R<sup>2</sup> = 0.88 vs TFT R<sup>2</sup> = 0.85) in Table 1(B). Attention-LSTM had better R<sup>2</sup> but was less accurate at predicting individual data points. This contrast shows that, whilst TFT had fewer overall prediction errors, Attention LSTM was better at capturing the overall form of the time series over longer durations.

The small R<sup>2</sup> values for TFT in the limestone aquifer may be related to the model's sensitivity to quick, high-frequency fluctuations in GWL, which are more typical of karstic systems like limestone. This emphasises the necessity of matching model capabilities to specific aquifer features since models that succeed in one type may fail in another due to differences in geological processes.

The detailed results below will show each model's quantitative performance measure across the different aquifers and prediction horizons. The RMSE of the holdout sets is used as a measurement, as it is one of the most reliable indicators that clearly explains how well the model

is performing in real-world scenarios (Cerqueira et al., 2020). This is because the holdout sets are completely unseen datasets during the training and validation processes.

### 3.1. The chalk aquifer

This group is the primary aquifer in the Thames basin, and it is crucial for a public water supply and river flow. It features dual porosity storage within its matrix and fractures, significantly affecting pollutant behaviour such as nitrates. These features allow it to quickly respond to rainfall, as evidenced by the sharp rises in GWL following precipitation events, as shown in Fig. 3, to transmit groundwater. Most fractures within this aquifer run parallel to the bedding plane within the upper part, exhibiting a subsequent reduction in permeability with increasing depth. Furthermore, transmissivity is influenced by topographical variations, with valleys displaying higher transmissivity values decreases towards the interfluvies (Smedley et al., 2003; Shand et al., 2003a; Neal et al., 2006).

The TFT model showed remarkable performance in this aquifer. Although this aquifer responds rapidly to changes due to precipitation, TFT consistently had the lowest RMSE scores over the 7, 30, and 60-day prediction horizons, as shown in Fig. 2, indicating its better capabilities in capturing the aquifer response to varying hydrological circumstances.

The TFT model outperformed the other models with an RMSE of 0.03 for the 7-day horizon. This demonstrates the usefulness of TFT for short-term forecasts where quick response to precipitation is essential. Even after being adjusted by BO, the attention-based LSTM and LSTM models still performed admirably, but the TFT surpassed both in terms of accuracy. This shows the advanced design of TFT, which efficiently combines temporal static and factors, was especially advantageous in modelling the hydrodynamic properties of the chalk aquifer (Lim et al., 2021).

Notably, LSTM and attention-based LSTM models also showed capabilities in capturing the hydrological dynamics. The  $R^2$  values shown in Table 1 in the holdout set over all horizons were moderate, which indicates that the model did somehow well in capturing the trends and was not overfit to the training set, which allowed them to generalise effectively to new data (Chicco et al., 2021). On the other hand, the TFT model provided high  $R^2$  values, showing a strong fit and good data generalisation. It is important to note that, while the LSTM with BO model did not have the lowest RMSE and best  $R^2$  values, overall, it was

the second in terms of performance, especially for the intermediate 30 and 60-day horizons. This might indicate that the models make excellent use of previous data to capture temporal relationships. It is worth mentioning that the attention mechanism in the standard LSTM also aided in capturing the important trends of GWL by using temporal dependencies and prioritising relevant information over time.

### 3.2. The limestone aquifer

The limestone aquifer had consistently higher RMSE in the holdout set compared to chalk and lower greensand aquifer. Due to various geological factors, including valleys and faults, water flow is affected, which results in significant water level fluctuations throughout the year. The limestone intricate and tough structure is distinguished by lithological heterogeneity and severe karst system that involves rapid underground flow through massive conduits (Abesser et al., 2005). Furthermore, this aquifer presents a unique hydrology since it has limited storage due to low porosity and is prone to rain runoff due to its unconfined nature, resulting in rapid change in GWL, as seen in Fig. 5. As a result, the upper parts of many rivers dry out annually, which leads to frequent drought conditions in this region. It provides a stable base-flow and consists of contributing river flow and springs and consists of two aquifers separated by Fuller Earth clay (Oubagaranadin et al., 2007), with the possibility of vertical water movement between them through fault systems (Neumann et al., 2003; Maurice et al., 2008). This results in continuously low GWL and lower river flows over the summer and autumn months, which are common and expected in this area (Neumann et al., 2003).

The chemical composition of groundwater in the aquifer varies greatly due to the natural geotechnical processes and occasional hydrothermal mineralisation, which contributes to its unpredictable behaviour, affecting the model's ability to effectively capture the slow hydrological response of this aquifer (Maurice et al., 2008). Despite the advanced capabilities of TFT illustrated in Fig. 4 in capturing such complexities, the underlying geological and chemical variety of the limestone aquifer poses significant modelling challenges, resulting in higher error matrices.

For the first prediction horizon (7 days), the TFT model achieved an RMSE of 0.02, significantly lower than other models, which were LSTM (0.42), attention LSTM and its variant with BO (0.28) and LSTM with BO (0.31). This implies that despite the complicated hydrogeological

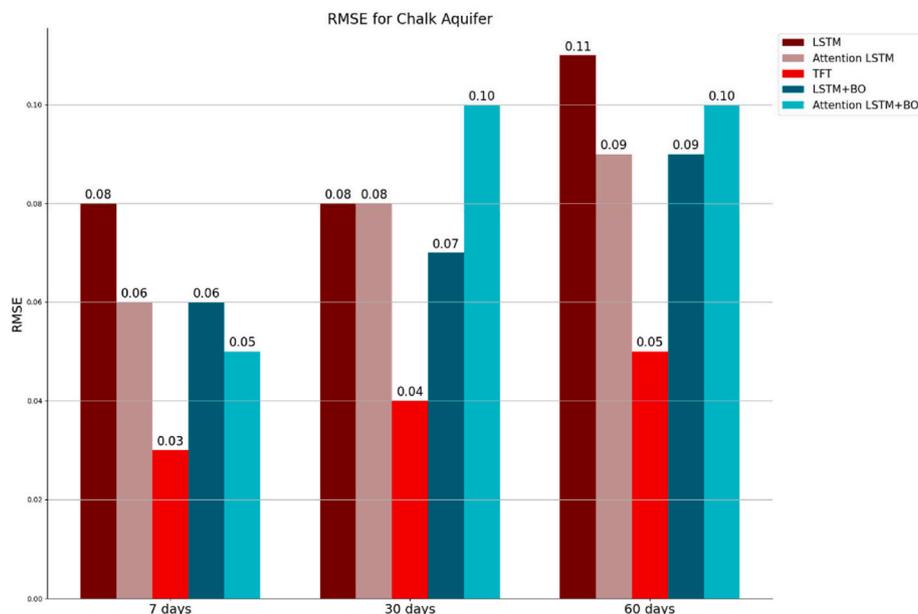


Fig. 2. The holdout RMSE across all horizons for the chalk.

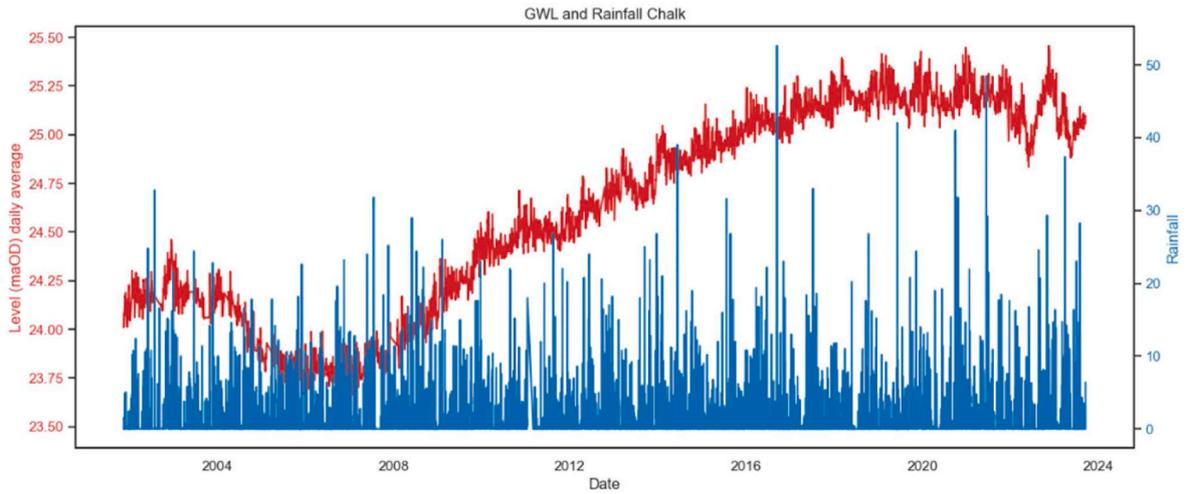


Fig. 3. Groundwater levels and rainfall interplay from year 2001–2023 of chalk aquifer.

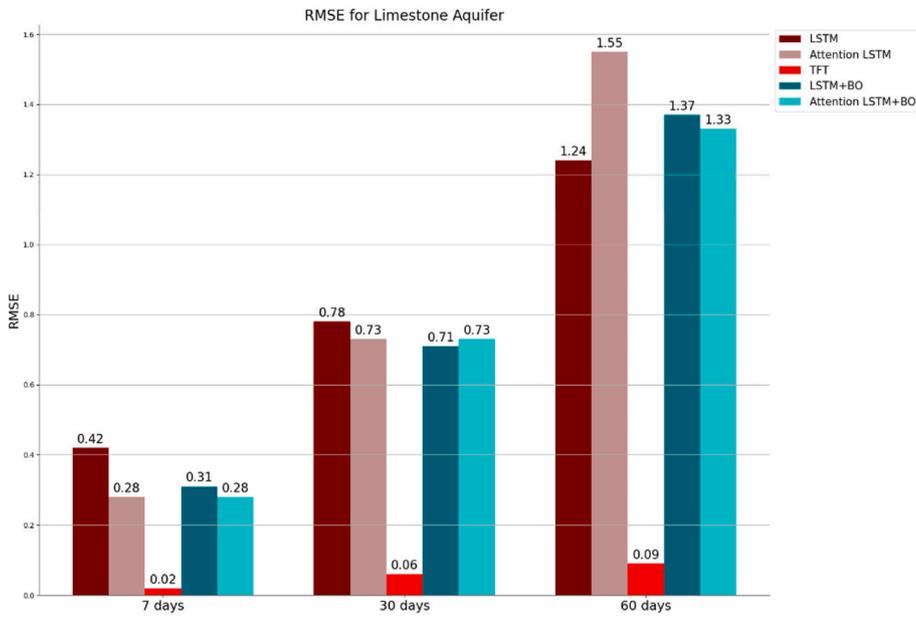


Fig. 4. The holdout RMSE across all horizons for the limestone aquifer.

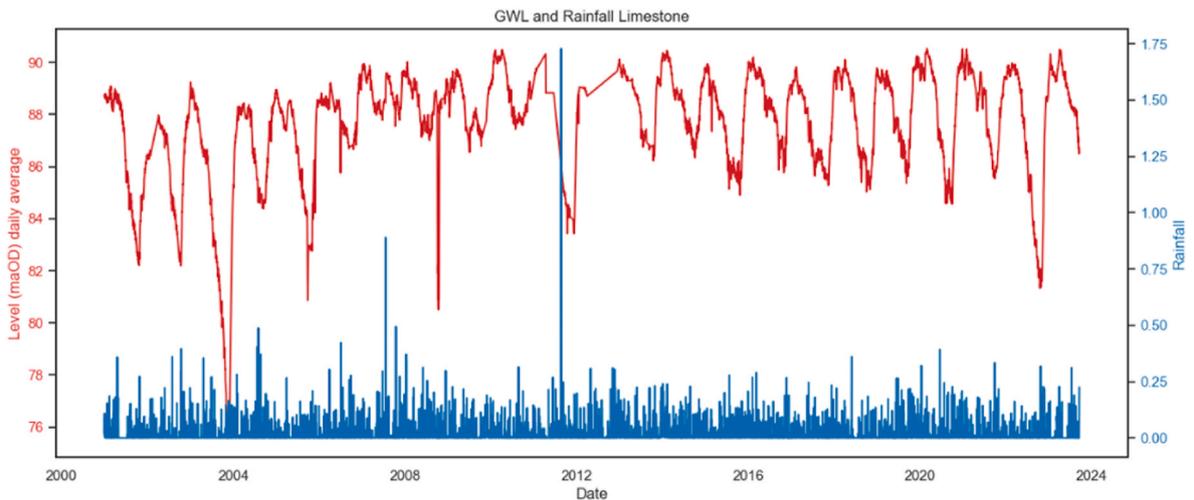


Fig. 5. Groundwater levels and rainfall interplay from year 2001–2023 of limestone aquifer.

features of the limestone aquifer, TFT is quite successful in making short-term projections. Additionally, the results showed that the TFT model has a superior capability for predicting long-term hydrological responses (30 and 60 days horizons).

While the TFT model performed admirably in difficult settings, its RMSE scores for all horizons still highlight the inherent modelling problems associated with the limestone aquifer. Although LSTM and attention-LSTM models were fine-tuned using BO, they did not outperform TFT. Nevertheless, they present a noteworthy attempt to deal with the aquifer’s hard hydrological properties. However, it is important to note that the Attention-LSTM model has emerged as a strong contender, surpassing both the standard LSTM and its BO variant. These results highlight the ability of Attention-LSTM to leverage temporal dependencies and prioritise relevant information over time, which is critical for modelling the limestone aquifer slow and complex response patterns.

### 3.3. The lower greensand aquifer

The lower greensand aquifer levels shown in Fig. 7, is known by its unique hydraulically independent Hythe and Folkestone formations, which function hydraulically independently. In the Hythe formation, water flows through both fractures and intergranular flow. The Folkestone formation is homogeneous, characterised by intergranular flow only, resulting in low transmissivity. Piezometer readings and groundwater dates show that each aquifer has independent water flow behaviour (Shand et al., 2003b). and the presence of both fracture and intergranular flow. It should be noted the period from 2008 to 2010 contained zero values from the original source. It is important to note that, unlike other cases of missing data that were addressed using methods such as rolling window averages, the zeros were kept as is due to the extended period of missing information as has been done in Somasundaram and Nedunchezian (2011).

However, the model outputs accurately reflected the aquifer’s unique geological characteristics, such as its variable composition and the presence of the Sandgate formation (Shand et al., 2003b), which allows for vertical leakage and poses a challenging environment for predictive modelling. The Folkestone formation, with high storage, ensures a consistent baseflow to nearby rivers. As one moves further from the outcrop, the lower greensand becomes confined, reducing permeability and storage. In this confined area, the Hythe and the Folkestone

formations display more uniform water flow behaviour, restricting the water extraction capabilities compared to the unconfined region.

As illustrated in Fig. 6, the TFT model excelled in prediction accuracy. These findings reflect the geological complexities of the aquifer, effectively capturing the nuanced processes of fracture and intergranular flow that distinguish the various formations within the aquifer. Overall, the findings demonstrate that the models were consistently accurate over the course of all forecast timeframes. It also shows that the use of attention mechanism allowed to effectively analyse and prioritise important hydrological patterns.

The variation in predicted accuracy amongst aquifers demonstrates the importance of both hydrological and geological dynamics on machine learning algorithms. In addition to performance indicators like RMSE and  $R^2$ , the aquifers in the Thames Basin show the models’ ability to capture the nuances of water flow dynamics as well as their limits.

The lower greensand’s dual nature, with both confined and unconfined zones, made it a fascinating topic for model performance analysis. Confined places tend to hinder vertical water flow, whereas unconfined parts have better permeability. The TFT demonstrated an improved ability to adjust to this variation by focusing on the important temporal features, making it more resilient to differences in water flow between these two zones.

The flexibility of TFT contrasts with the rigid static structure of LSTM models used. These models performed well in settings with more consistent hydrological responses, such as the Chalk aquifer, but struggled in locations with more variable flow patterns. This finding implies that model flexibility is critical for effectively estimating GWL in aquifers with complicated flow regimes. The TFT’s capacity to integrate long-term dependencies enabled it to better portray the lower greensand aquifers’ complex hydrological response, particularly in constrained locations with limited water extraction capability.

The substantial time lag between groundwater response and rainfall, notably in the Chalk and limestone aquifers, presented a difficulty for the models, especially for long-term projections of more than 30-day horizon. With the dual porosity in the Chalk aquifer, it responds quickly to rainfall, followed by a longer, more gradual recharging (Smedley et al., 2003). This dual-phase response posed challenges for LSTM-based models, which are designed to prioritise short-term temporal dynamics (Khozani et al., 2022). TFT, on the other hand, was a more reliable option for forecasting long-term water level variations due to its capacity to mimic both immediate and delayed reactions to

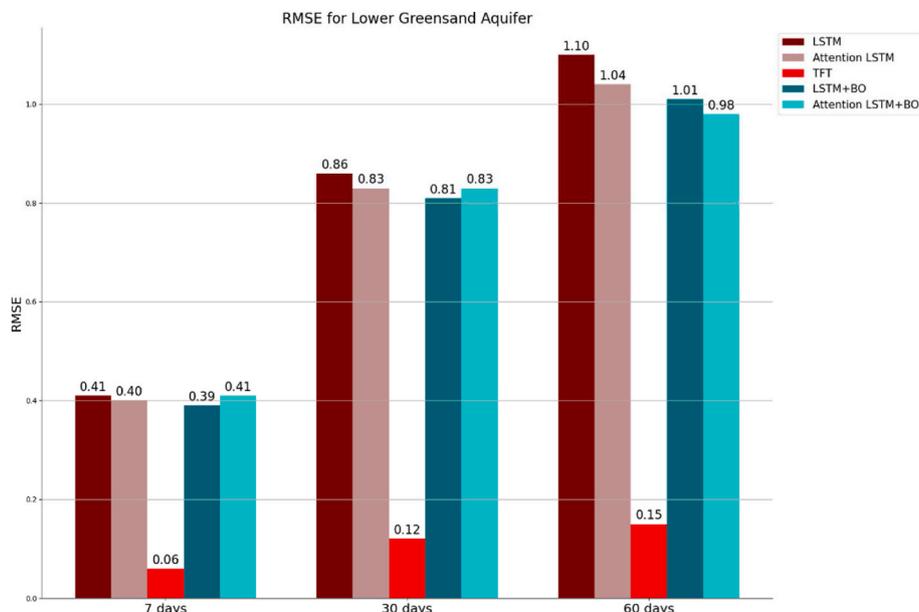


Fig. 6. The holdout RMSE across all horizons for the lower greensand aquifer.

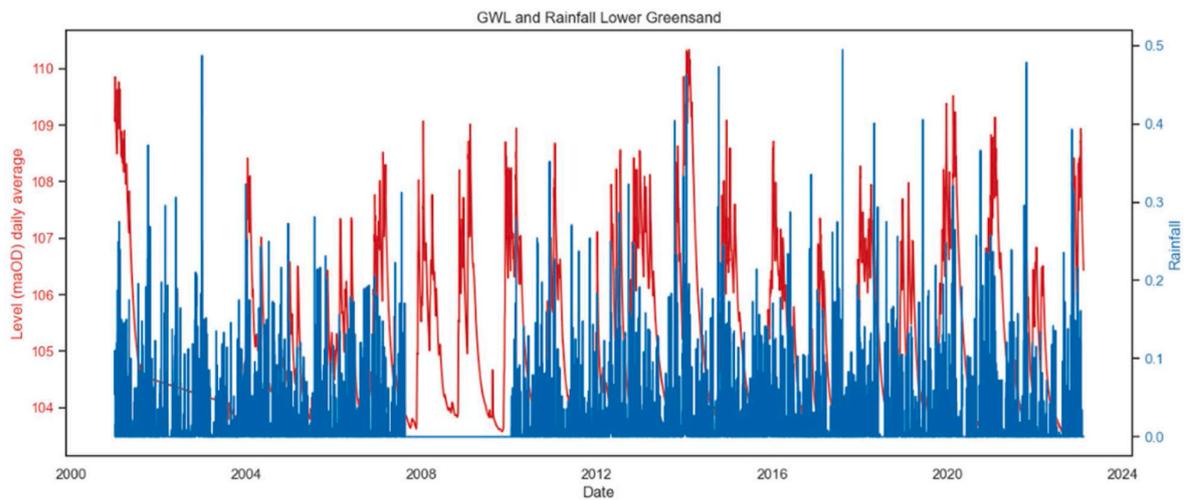


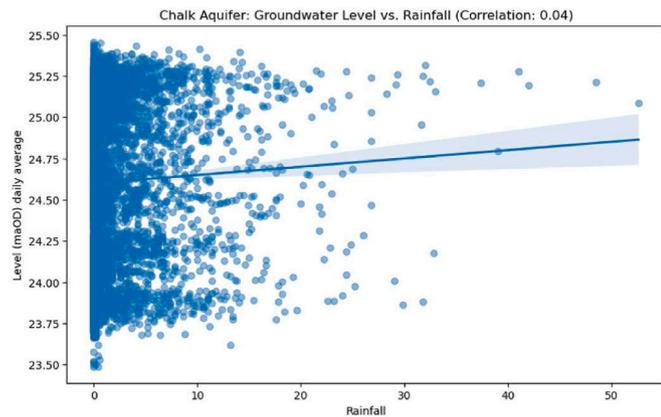
Fig. 7. Groundwater levels and rainfall interplay from year 2001–2023 of lower greensand aquifer.

rainfall.

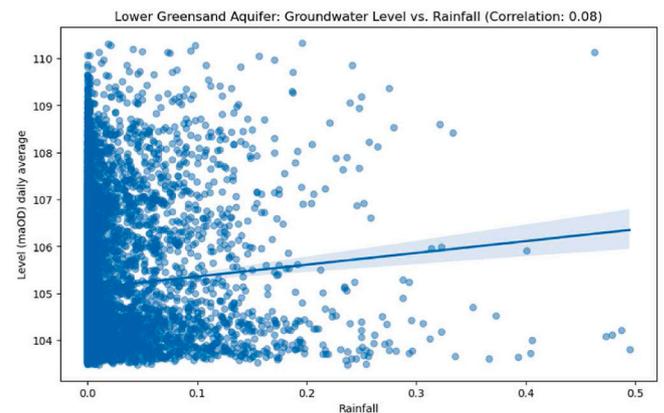
In the limestone aquifer, where river systems are more important in controlling GWL (Oubagaranadin et al., 2007), the time lag between GW, rainfall, and river level fluctuation recharge added to the complexity (Polomčić et al., 2013). Whilst LSTM models caught some of the short-term variations, TFT outperformed them because of its ability to detect both rapid precipitation responses and delayed, river-mediated recharging. This emphasises the necessity of taking into consideration

both direct and indirect hydrological causes when designing models.

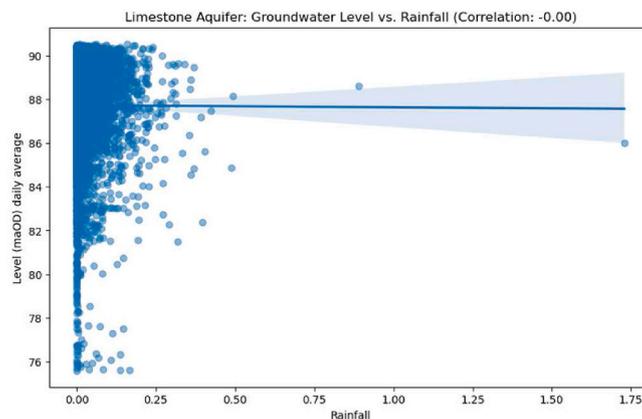
Given that the TFT model has been shown to be extremely successful over a variety of forecasting horizons and aquifer types, the present results are adequate for application in water resources management. The combination of rolling window, holdout set and walk-forward validation approach assures that the model is robust and can handle real-world data. The model’s ability to capture groundwater oscillations makes it a useful tool for long-term water management plans, particularly in



Chalk aquifer



Lower greensand aquifer



Limestone aquifer

Fig. 8. Scatter plots that present the correlation between groundwater levels and rainfall infiltration.

complicated systems such as the Thames Basin.

The linearity between groundwater and rainfall was measured using Pearson correlation (Jebli et al., 2021). The correlation between rainfall and GWL across the three aquifers is explored using scatter plots, as presented in Fig. 8. In the Chalk aquifer, the correlation value of 0.04 indicates a very weak positive relationship, implying that variables other than rainfall might significantly influence GWL. Similarly, the lower greensand aquifer has a correlation coefficient of 0.08, reinforcing the claim that rainfall has only a minimal influence on GWL in this aquifer.

The limestone displays the most notable result of 0.00, contrary to expectations given its hydrological responsiveness (Bricker et al., 2014). Despite the sensitivity of limestone to climatic factors, it seems that neighbouring river systems may have an impact that could overshadow the effect of rainfall through processes such as greater riverbank infiltration or aquifer water drainage into the river system (Polomčić et al., 2013). This concludes the particular intriguing since it shows that rainfall has no direct or major impact on GWL. Given the limestone’s tough structure and lower intergranular permeability, it might be hypothesised that the proximity of river systems could influence the aquifer’s response to rainfall. Rivers in close proximity to aquifers may promote quick runoff or absorption of rainfall, thereby limiting the potential recharge that would otherwise be reflected in the GWL rise.

To further understand the reasons contributing to the weak correlation between GWL and rainfall in some aquifers, it is necessary to investigate the water infiltration. Fig. 9 illustrates the correlation between the river level and rainfall. Chalk and Lower greensand aquifers show a moderate correlation coefficient of 0.32 and 0.48, respectively, indicating that river levels respond considerably to rainfall events. This supports the hypothesis that river levels might serve as mediators in the hydrological cycle, affecting the GWL through interconnected surface and subsurface water processes.

River proximity is especially important in the limestone aquifer, where the correlation between GWL and rainfall is insignificant. The river system nearby this aquifer likely leads to quick runoff or quick overland flow of rainfall, thereby restricting the recharge capacity (Fetter, 2001). This phenomenon aligns with the aquifer known geological characteristics as its strong structure and limited intergranular permeability, which might lead to a slower reaction to rainfall.

The response of an aquifer to precipitation events is not always direct and is influenced by various intervening factors. In particular, the existence of rivers can substantially impact this hydrological interaction. During rainfall events, river levels may rise quickly, and the increased flow can help recharge surrounding aquifers (Fetter, 2001). However, such recharging may not be directly detectable in groundwater level measurements, particularly in the short term. This lag is due to the time it takes for the water to percolate through the stream bank and into the aquifer from the river system, which is well-documented but often overlooked in casual analysis (Randall and Albany, 1978). This temporal difference between precipitation and apparent changes in GWL emphasises the importance of a comprehensive approach to hydrological investigation that considers the dynamic and frequently delayed interactions between surface bodies and groundwater aquifers. This method is critical for creating accurate groundwater models and effective water resources management strategies. The findings support the TFT model’s ability to adjust to the Thames basin complex hydrological conditions. Similarly, the findings support the efficacy of the Attention based-LSTM model, which overall produced better results than standard LSTM, particularly for longer-term predictions, demonstrating its potential to capture the dynamic behaviour of this system.

The results show that the TFT has exceptional ability in predicting GWL in the Thames Basin, outperforming LSTM and attention-based LSTM models in forecasting an extended horizon period in GWL prediction. The use of advanced validation techniques confirmed the ability

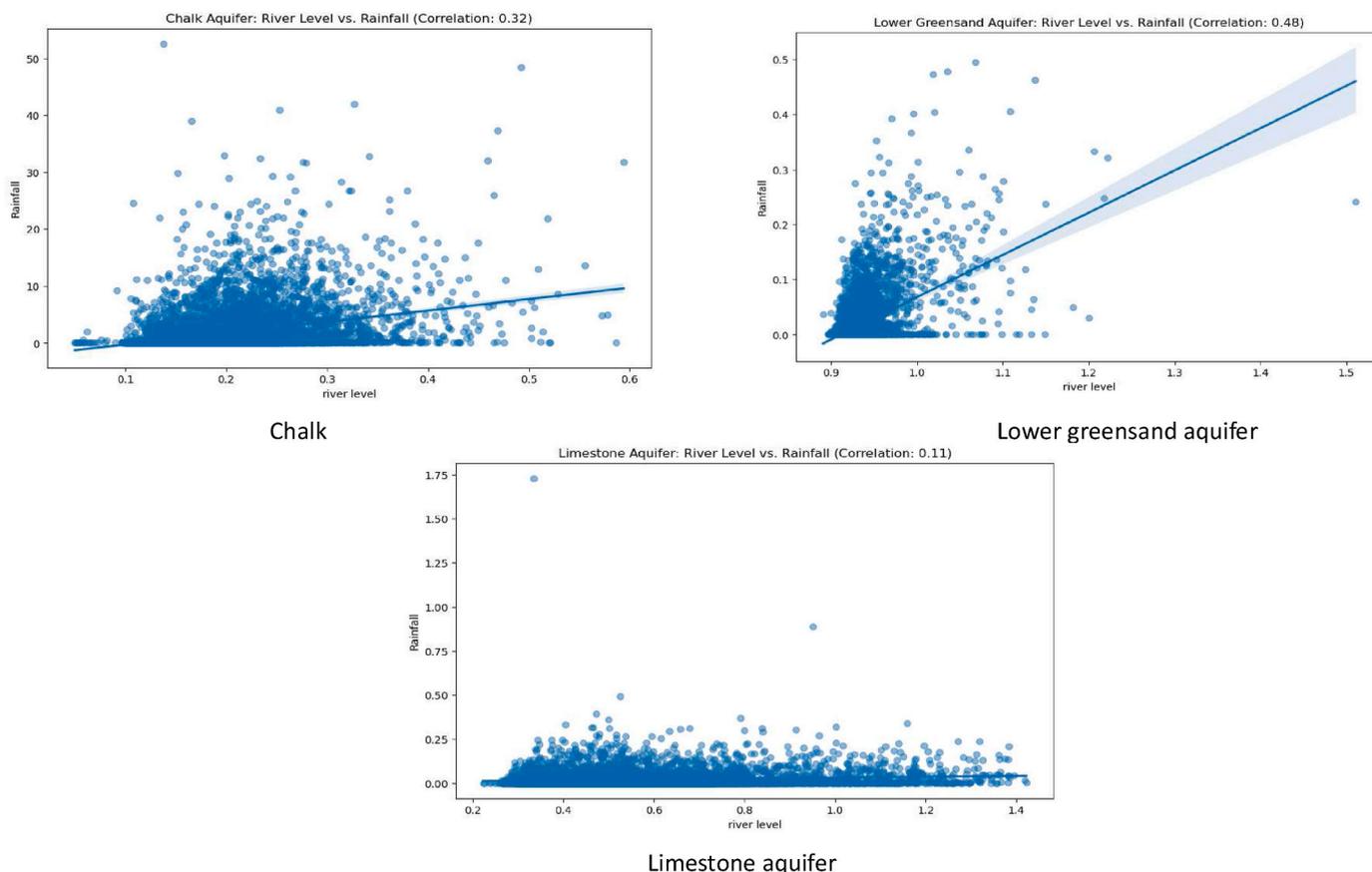


Fig. 9. The correlation between river level and rainfall.

of TFT to produce robust and reliable predictions, especially in the Chalk aquifer and highlighted potential overfitting that is often an issue in advanced machine learning like LSTM models, as evidenced by higher  $R^2$ . This demonstrates the value of using several types of hydrological data to improve groundwater level forecasting. By leveraging the capability of advanced machine learning algorithms, this study marks a considerable advancement in GWL prediction, especially in extended horizons (30 and 60 days) based on daily data without depending on monthly or weekly moving average. This makes the TFT model potentially a new era in time series forecasting.

In this study, we investigated the predictive capabilities of numerous models across multiple forecasting horizons to find the best balance of prediction duration and accuracy. The Temporal Fusion Transformer (TFT) model performed exceptionally well up to a 60-day projection. Extending the prediction horizon beyond this period resulted in a slow but substantial loss in forecast accuracy, as seen by higher RMSE and lower  $R^2$  values. Similarly, the Long Short-Term Memory (LSTM) model had severe performance constraints, which decreased for projections longer than 60 days. These findings indicate that the inherent uncertainty and unpredictability associated with longer-term hydrological forecasts provide significant problems, which are exacerbated by the limitations of the current data utilised to train our models. As a result, we chose to limit our forecasts to 60 days, striking a balance between the requirement for actionable, credible predictions and the technological limits identified. This conclusion is consistent with the practical requirements of water resource management, where a two-month prediction horizon adds considerable operational value.

#### 4. Summary and conclusion

The summary of the main contributions of this study are:

- This study used advanced machine learning techniques to address overfitting and increase forecasting accuracy in Thames Basin. We employed rolling window, holdout sets and walk forward validation to improve the model generalisability and dependability.
- The small difference in RMSE and  $R^2$  between LSTM with Bayesian Optimisation and traditional LSTM suggests that improved accuracy is not always correlated with comprehensive parameter search. This emphasises the necessity of carefully selecting starting parameters, considering computing efficiency, and considering the unique peculiarities of the aquifer system.
- The poor rainfall and GWL correlation in the Chalk and Lower greensand aquifer suggest that other factors mostly influence GWL. Additionally, the weak correlation in the limestone aquifer may point to the influence of adjacent river systems on the aquifer's reaction to precipitation.
- The study showed moderate connections between river levels and rainfall in the Chalk and Lower greensand aquifer. This is especially true in the limestone aquifer, where the presence of river systems influences the aquifer recharge dynamics.

In conclusion, this study is a pioneering effort that not only introduces extended forecasting horizons of 30 and 60 days that does not depend on monthly or weekly moving average, but also delves deeply into the geological complexities of the Thames Basin. Few studies have gone to such lengths to untangle the complexity of the Thames Basin and understand the interplay between its river aquifers and the complicated environment in which they inhabit. Furthermore, this study thoroughly investigates three major aquifer types, providing light on how each aquifer's complexity and geological features may have a substantial impact on the prediction capacities of advanced modelling tools.

#### CRedit authorship contribution statement

Ali J. Ali: Writing – review & editing, Writing – original draft,

Conceptualization. Ashraf A. Ahmed: Supervision. Maysam F. Abbod: Supervision.

#### Declaration of competing interest

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

#### Acknowledgement

We would like to thank the Environment Agency for their valuable support. Thanks also go to the anonymous reviewers, whose comments greatly improved this manuscript. The study was partially funded by UKRI project 10063665.

#### Data availability

Data will be made available on request.

#### References

- Abesser, C., Shand, P., Ingram, J., 2005. Baseline Report Series. 22, the Carboniferous Limestone of Northern England.
- Afzaal, H., Farooque, A.A., Abbas, F., Acharya, B., Esau, T., 2019. Groundwater estimation from major physical hydrology components using artificial neural networks and deep learning. *Water* 12 (1), 5.
- Ahmed, A.A., Sayed, S., Abdoulhalik, A., Moutari, S., Oyedele, L., 2024. Applications of machine learning to water resources management: a review of present status and future opportunities. *J. Clean. Prod.*, 140715.
- Ali, A.J., Ahmed, A.A., 2024. Long-term AI prediction of ammonium levels in rivers using transformer and ensemble models. *Cleaner Water* 2, 100051.
- Alizadeh, B., Bafti, A.G., Kamangir, H., Zhang, Y., Wright, D.B., Franz, K.J., 2021. A novel attention-based LSTM cell post-processor coupled with bayesian optimisation for streamflow prediction. *J. Hydrol.* 601, 126526.
- Amor, L.B., Lahyani, I., Jmaiel, M., 2016. Recursive and Rolling Windows for Medical Time Series Forecasting: a Comparative Study. *IEEE*, pp. 106–113.
- Appels, W.M., Graham, C.B., Freer, J.E., McDonnell, J.J., 2015. Factors affecting the spatial pattern of bedrock groundwater recharge at the hillslope scale. *Hydrol. Process.* 29 (21), 4594–4610.
- Bagheri, M., Bazvand, A., Ehteshami, M., 2017. Application of artificial intelligence for the management of landfill leachate penetration into groundwater, and assessment of its environmental impacts. *J. Clean. Prod.* 149, 784–796.
- Barnett, B., Townley, L., Post, V., Evans, R., Hunt, R., Peeters, L., Richardson, S., Weatherill, D., Werner, A., Knapton, A., 2013. Developing Groundwater Modelling Guidelines for Australia.
- Bearcock, J.M., Smedley, P.L., 2010. Baseline Groundwater Chemistry: the Palaeogene of the Thames Basin.
- Beven, K.J., 2011. Rainfall-runoff Modelling: the Primer. John Wiley & Sons.
- Bricker, S.H., Barron, A.J.M., Hughes, A.G., Jackson, C., Peach, D., 2014. In: Sharp, J.M. (Ed.), From Geological Complexity to Hydrogeological Understanding Using an Integrated 3D Conceptual Modelling Approach—Insights from the Cotswolds, UK. CRC Press, London, UK, pp. 99–114.
- Brouyère, S., 2006. Modelling the migration of contaminants through variably saturated dual-porosity, dual-permeability chalk. *J. Contam. Hydrol.* 82 (3–4), 195–219.
- Buczko, U., Kuchenbuch, R.O., Lennartz, B., 2010. Assessment of the predictive quality of simple indicator approaches for nitrate leaching from agricultural fields. *J. Environ. Manag.* 91 (6), 1305–1315.
- Cerqueira, V., Torgo, L., Mozetič, I., 2020. Evaluating time series forecasting models: an empirical study on performance estimation methods. *Mach. Learn.* 109, 1997–2028.
- Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* (GMD) 7 (3), 1247–1250.
- Chen, C., Zhou, H., Zhang, H., Chen, L., Yan, Z., Liang, H., 2020. A Novel Deep Learning Algorithm for Groundwater Level Prediction Based on Spatiotemporal Attention Mechanism.
- Chen, C., Zhu, X., Kang, X., Zhou, H., 2021. A Deep Learning Algorithm for Groundwater Level Prediction Based on Spatial-Temporal Attention Mechanism. *IEEE*, pp. 716–723.
- Cheng, M., Fang, F., Kinouchi, T., Navon, I.M., Pain, C.C., 2020. Long lead-time daily and monthly streamflow forecasting using machine learning methods. *J. Hydrol.* 590, 125376.
- Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* 7, e623.
- Clevert, D.A., Unterthiner, T., Hochreiter, S., 2015. Fast and accurate deep network learning by exponential linear units (elus). In: arXiv preprint arXiv:1511.07289.
- Condon, L.E., Kollet, S., Bierkens, M.F., Fogg, G.E., Maxwell, R.M., Hill, M.C., Franssen, H. J.H., Verhoef, A., Van Loon, A.F., Sulis, M., Abesser, C., 2021. Global groundwater

- modeling and monitoring: opportunities and challenges. *Water Resour. Res.* 57 (12), e2020WR029500.
- Dauphin, Y.N., Fan, A., Auli, M., Grangier, D., 2017. Language modeling with gated convolutional networks. In: *International Conference on Machine Learning*. PMLR, pp. 933–941.
- Ding, Y., Zhu, Y., Feng, J., Zhang, P., Cheng, Z., 2020. Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing* 403, 348–359.
- Ding, Y., Zhu, Y., Wu, Y., Jun, F., Cheng, Z., 2019. Spatio-temporal Attention LSTM Model for Flood Forecasting. *IEEE*, pp. 458–465.
- Du, L., Gao, R., Suganthan, P.N., Wang, D.Z., 2022. Bayesian optimisation based dynamic ensemble for time series forecasting. *Inf. Sci.* 591, 155–175.
- Duan, Q., Soroshian, S., Gupta, V., 1992. Effective and efficient global optimisation for conceptual rainfall-runoff models. *Water Resour. Res.* 28 (4), 1015–1031.
- Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., Roth, A., 2015. Generalisation in adaptive data analysis and holdout reuse. *Adv. Neural Inf. Process. Syst.* 28.
- Ehteram, M., 2023. An Advanced Deep Learning Model for Predicting Groundwater Level.
- El-Harrouki, K., Ouazar, D., Walters, G.A., Cheng, A.D., 1996. Groundwater optimisation and parameter estimation by genetic algorithm and dual reciprocity boundary element method. *Eng. Anal. Bound. Elem.* 18 (4), 287–296.
- Environment Agency, 2022. *River Basin Management Plans, Updated 2022: Challenges for the Water Environment*. (2022, December 22). GOV, UK. River basin management plans, updated 2022: challenges for the water environment. [www.gov.uk](http://www.gov.uk).
- Fayer, G., Lima, L., Miranda, F., Santos, J., Campos, R., Bignoto, V., Andrade, M., Moraes, M., Ribeiro, C., Capriles, P., Goliatt, L., 2023. A temporal fusion transformer deep learning model for long-term streamflow forecasting: a case study in the funil reservoir, southeast Brazil. *Know. Based Eng. Sci.* 4 (2), 73–88.
- Fetter Jr., C.W., 2001. *Applied Hydrogeology* Fetter Fourth Edition.
- Frederick, K.D., Major, D.C., 1997. Climate change and water resources. *Climatic Change* 37 (1), 7–23.
- Ghaffarian, S., Valente, J., Van Der Voort, M., Tekinerdogan, B., 2021. Effect of attention mechanism in deep learning-based remote sensing image processing: a systematic literature review. *Rem. Sens.* 13 (15), 2965.
- Gulli, A., Pal, S., 2017. *Deep Learning with Keras*. Packt Publishing Ltd.
- Gundu, V., Simon, S.P., 2021. PSO-LSTM for short term forecast of heterogeneous time series electricity price signals. *J. Ambient Intell. Hum. Comput.* 12, 2375–2385.
- Guo, X., Gao, Y., Li, Y., Zheng, D., Shan, D., 2021. Short-term household load forecasting based on Long-and Short-term Time-series network. *Energy Rep.* 7, 58–64.
- Gupta, P.K., Yadav, B., Yadav, B.K., 2019. Assessment of LNAPL in subsurface under fluctuating groundwater table using 2D sand tank experiments. *J. Environ. Eng.* 145 (9), 04019048.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Huang, G., 2021. Missing data filling method based on linear interpolation and lightgbm. In: *Journal of Physics: Conference Series*, vol. 1754. IOP Publishing, 012187. No. 1.
- Hussein, E.A., Thron, C., Ghaziasgar, M., Bagula, A., Vaccari, M., 2020. Groundwater prediction using machine-learning tools. *Algorithms* 13 (11), 300.
- Hydrology Data Explorer - Explore: <https://environment.data.gov.uk/hydrology/>.
- Jebli, I., Belouadha, F.Z., Kabbaj, M.I., Tilioua, A., 2021. Prediction of solar energy guided by Pearson correlation using machine learning. *Energy* 224, 120109.
- Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., Wu, Y., 2016. Exploring the limits of language modeling. In: *arXiv preprint arXiv:1602.02410*.
- Junankar, T., Sondhi, J.K., Nair, A.M., 2023. Wheat yield prediction using temporal fusion transformers. In: *2023 2nd International Conference for Innovation in Technology (INOCON)*. IEEE, pp. 1–6.
- Kaasra, I., Boyd, M., 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10 (3), 215–236.
- Kang, M., Tian, J., 2018. Machine learning: data pre-processing. *Progn. Health Mana. Electro.: Funda. Mach. Learning Inter. Things* 111–130.
- Kennedy, M.P., Milne, J.M., Murphy, K.J., 2003. Experimental growth responses to groundwater level variation and competition in five British wetland plant species. *Wetl. Ecol. Manag.* 11, 383–396.
- Khaki, M., Yusoff, I., Islami, N., 2015. Application of the artificial neural network and neuro-fuzzy system for assessment of groundwater quality. *Clean: Soil, Air, Water* 43 (4), 551–560.
- Khozani, Z.S., Banadkooki, F.B., Ehteram, M., Ahmed, A.N., El-Shafie, A., 2022. Combining autoregressive integrated moving average with Long Short-Term Memory neural network and optimisation algorithms for predicting ground water level. *J. Clean. Prod.* 348, 131224.
- Kombo, O.H., Kumaran, S., Sheikh, Y.H., Bovim, A., Jayavel, K., 2020. Long-term groundwater level prediction model based on hybrid KNN-RF technique. *Hydrology* 7 (3), 59.
- Kong, T., Fang, W., Love, P.E., Luo, H., Xu, S., Li, H., 2021. Computer vision and long short-term memory: learning to predict unsafe behaviour in construction. *Adv. Eng. Inf.* 50, 101400.
- Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.X., Yan, X., 2019. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Adv. Neural Inf. Process. Syst.* 32.
- Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H., Yin, S., 2023. Deep learning attention mechanism in medical image analysis: basics and beyonds. *Intern. J. Network Dyna. Intell.* 93–116.
- Lieskovská, E., Jakubec, M., Jarina, R., Chmulk, M., 2021. A review on speech emotion recognition using deep learning and attention mechanism. *Electronics* 10 (10), 1163.
- Lim, B., Arik, S.Ö., Loeff, N., Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* 37 (4), 1748–1764.
- Liu, H., Sun, W., Zhang, Y., Zhang, W., Han, F., Su, W., 2023. Experimental analysis on the interaction between underground structures and sand layer under groundwater level change. *Undergr. Space* 10, 15–36.
- Liu, P., Wang, J., Guo, Z., 2020. Multiple and complete stability of recurrent neural networks with sinusoidal activation function. *IEEE Transact. Neural Networks Learn. Syst.* 32 (1), 229–240.
- Marcellino, M., Stock, J.H., Watson, M.W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *J. Econom.* 135 (1–2), 499–526.
- Mathers, S.J., Burke, H.F., Terrington, R.L., Thorpe, S., Dearden, R.A., Williamson, J.P., Ford, J.R., 2014. A geological model of London and the Thames Valley, southeast England. *Proc. Geologists' Assoc.* 125 (4), 373–382.
- Maurice, L., Barron, A.J.M., Lewis, M.A., Robins, N.S., 2008. The geology and hydrogeology of the Jurassic limestones in the Stroud-Cirencester area with particular reference to the position of the groundwater divide. *British Geol. Survey. Commissioned Report CR/08/146*.
- May-Lagunes, G., Chau, V., Ellestad, E., Greengard, L., D'Odorico, P., Vahabi, P., Todeschini, A., Girotto, M., 2023. Forecasting groundwater levels using machine learning methods: the case of California's central valley. *J. Hydrol.* X, 100161.
- Mehr, A.D., Kahya, E., Olyaie, E., 2013. Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *J. Hydrol.* 505, 240–249.
- Miller, J.D., Hutchins, M., 2017. The impacts of urbanisation and climate change on urban flooding and urban water quality: a review of the evidence concerning the United Kingdom. *J. Hydrol.: Reg. Stud.* 12, 345–362.
- Mohtashami, A., Akbarpour, A., Mollazadeh, M., 2017. Development of two-dimensional groundwater flow simulation model using meshless method based on MLS approximation function in unconfined aquifer in transient state. *J. Hydroinf.* 19 (5), 640–652.
- Muñoz-Carpena, R., Carmona-Cabrero, A., Yu, Z., Fox, G., Batelaan, O., 2023. Convergence of mechanistic modeling and artificial intelligence in hydrologic science and engineering. *PLOS Water* 2 (8), e0000059.
- Neal, C., Neal, M., Hill, L., Wickham, H., 2006. The water quality of the river thame in the Thames Basin of south/south-eastern England. *Sci. Total Environ.* 360 (1–3), 254–271.
- Neumann, I., Brown, S., Smedley, P.L., Besien, T., 2003. Baseline report series: 7. The great and the inferior oolite of the cotswold district. In: *British Geological Survey Commissioned Report CR/03/202N*.
- Oubagaradin, J.U.K., Sathyamurthy, N., Murthy, Z.V.P., 2007. Evaluation of Fuller's earth for the adsorption of mercury from aqueous solutions: a comparative study with activated carbon. *J. Hazard Mater.* 142 (1–2), 165–174.
- Pandya, H., Jaiswal, K., Shah, M., 2024. A comprehensive review of machine learning algorithms and its application in groundwater quality prediction. *Arch. Comput. Methods Eng.* 1–22.
- Pannu, A., 2015. Artificial intelligence and its application in different areas. *Artif. Intell.* 4 (10), 79–84.
- Pathania, T., Bottacin-Busolin, A., Rastogi, A.K., Eldho, T.I., 2019. Simulation of groundwater flow in an unconfined sloping aquifer using the element-free Galerkin method. *Water Resour. Manag.* 33, 2827–2845.
- Pei, W., Baltrusaitis, T., Tax, D.M., Morency, L.P., 2017. Temporal attention-gated model for robust sequence classification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6730–6739.
- Polomčić, D., Hajdin, B., Stevanović, Z., Bajić, D., Hajdin, K., 2013. Groundwater management by riverbank filtration and an infiltration channel: the case of Obrenovac, Serbia. *Hydrogeol. J.* 21 (7), 1519–1530 pp.147–163.
- Randall, A.D., Albany, N.Y., 1978. Infiltration from tributary streams in the susquehanna river basin, New York. *J. Res. U. S. Geol. Surv.* 6 (3), 285–297.
- Ransom, K.M., Nolan, B.T., Traum, J.A., Faunt, C.C., Bell, A.M., Gronberg, J.A.M., Wheeler, D.C., Rosecrans, C.Z., Jurgens, B., Schwarz, G.E., Belitz, K., 2017. A hybrid machine learning model to predict and visualise nitrate concentration throughout the Central Valley aquifer, California, USA. *Sci. Total Environ.* 601, 1160–1172.
- Roelofs, R., Shankar, V., Recht, B., Fridovich-Keil, S., Hardt, M., Miller, J., Schmidt, L., 2019. A meta-analysis of overfitting in machine learning. *Adv. Neural Inf. Process. Syst.* 32.
- Royle, K.R., Rutter, H.K., Entwisle, D.C., 2009. Property attribution of 3D geological models in the Thames Gateway, London: new ways of visualising geoscientific information. *Bull. Eng. Geol. Environ.* 68, 1–16.
- Rubinato, M., Nichols, A., Peng, Y., Zhang, J.M., Lashford, C., Cai, Y.P., Lin, P.Z., Tait, S., 2019. Urban and river flooding: comparison of flood risk management approaches in the UK and China and an assessment of future knowledge needs. *Water Sci. Eng.* 12 (4), 274–283.
- Saberian, M., Jahandari, S., Li, J., Zivari, F., 2017. Effect of curing, capillary action, and groundwater level increment on geotechnical properties of lime concrete: experimental and prediction studies. *J. Rock Mech. Geotech. Eng.* 9 (4), 638–647.
- Sadeghi-Tabas, S., Samadi, S., Zahabiyou, B., 2017. Application of Bayesian algorithm in continuous streamflow modeling of a mountain watershed. *European Water* 57, 101–108.
- Sakizadeh, M., 2016. Artificial intelligence for the prediction of water quality index in groundwater systems. *Model. Earth Syst. Environ.* 2, 1–9.
- Sameen, M.I., Pradhan, B., Lee, S., 2020. Application of convolutional neural networks featuring Bayesian optimisation for landslide susceptibility assessment. *Catena* 186, 104249.

- Selvadurai, A.P.S., 2019. The Biot coefficient for a low permeability heterogeneous limestone. *Continuum Mech. Therm.* 31, 939–953.
- Shand, P., Cobbing, J.E., Tyler-Whittle, R., Tooth, A., Lancaster, A., 2003b. Baseline report series: 9. The lower greensand of southern England. In: *British Geological Survey Commissioned Report CR/03/273C*.
- Shand, P., Tyler-Whittle, R., Besien, T., Lawrence, A.R., Lewis, O.H., 2003a. Baseline report series: 6. The chalk of the colne and lee river catchments. In: *British Geological Survey Commissioned Report CR/03/069N*.
- Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. Nonlinear Phenom.* 404, 132306.
- Smedley, P.L., Griffiths, K.J., Tyler-Whittle, R., 2003. Baseline Report Series: 5. The Chalk of North Downs, Kent and East Surrey.
- Somasundaram, R.S., Nedunchezian, R., 2011. Evaluation of three simple imputation methods for enhancing preprocessing of data with missing values. *Int. J. Comput. Appl.* 21 (10), 14–19.
- Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Heddami, S., Kim, S., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., Allawi, M.F., 2022. Groundwater level prediction using machine learning models: a comprehensive review. *Neurocomputing* 489, 271–308.
- Taylor, S.D., He, Y., Hiscock, K.M., 2016. Modelling the impacts of agricultural management practices on river water quality in Eastern England. *J. Environ. Manag.* 180.
- United Kingdom Water Industry Research, 2004. Implications of Changing Groundwater Quality for Water Resources and the UK Water Industry, Phase III: Financial and Water Resources Impact. UKWIR, London, p. 70.
- United Nations, 2015. Transforming Our World: the 2030 Agenda for Sustainable Development. United Nations. Available at: [Transforming our world: the 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs \(un.org\)](https://www.un.org/sustainabledevelopment/).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.
- Wada, Y., Van Beek, L.P., Van Kempen, C.M., Reckman, J.W., Vasak, S., Bierkens, M.F., 2010. Global depletion of groundwater resources. *Geophys. Res. Lett.* 37 (20).
- Wan, R., Mei, S., Wang, J., Liu, M., Yang, F., 2019. Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting. *Electronics* 8 (8), 876.
- Wei, C., Kakade, S., Ma, T., 2020. The implicit and explicit regularization effects of dropout. In: *International Conference on Machine Learning*. PMLR, pp. 10181–10192.
- Wu, B., Wang, L., Zeng, Y.R., 2022. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. *Energy* 252, 123990.
- Zektser, S., Loáiciga, H.A., Wolf, J.T., 2005. Environmental impacts of groundwater overdraft: selected case studies in the southwestern United States. *Environ. Geol.* 47, 396–404.
- Zhao, W., Wang, D., Gao, K., Wu, J., Cheng, X., 2023. Large-Scale long-term prediction of ship AIS tracks via linear networks with a look-back window decomposition scheme of time features. *J. Mar. Sci. Eng.* 11 (11), 2132.
- Zivot, E., Wang, J., Zivot, E., Wang, J., 2003. Rolling Analysis of Time Series. *Modeling Financial Time Series with S-Plus®*, pp. 299–346.