https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

Training deep learning models with a multi-station approach and static aquifer attributes for groundwater level simulation: what's the best way to leverage regionalised information?

Sivarama Krishna Reddy Chidepudi[a b], Nicolas Massei[a], Abderrahim Jardani [a], Bastien Dieppois [c], Abel Henriot [b], Matthieu Fournier [a]

[a] Univ Rouen Normandie, UNICAEN, CNRS, M2C UMR 6143, F-76000 Rouen, France

[b] BRGM, 3 av. C. Guillemin, 45060 Orleans Cedex 02, France

[c] Centre for Agroecology, Water and Resilience, Coventry University, Coventry, UK

Correspondence to : Sivarama Krishna Reddy Chidepudi (sivaramakrishnareddy.chidepudi@univrouen.fr)

Abstract. In this study, we used deep learning models with recurrent structure neural networks to simulate large-scale groundwater level (GWL) fluctuations in northern France. We developed a multi-station collective training for GWL simulations, using both "dynamic" variables (i.e. climatic) and static basin characteristics. This large-scale approach offers the possibility of incorporating dynamic and static features to cover more reservoir heterogeneities in the study area. Further, we investigated the performance of relevant feature extraction techniques such as clustering and wavelet transform decomposition in the aim of simplifying network learning using regionalised information. Several modelling performance tests were conducted. Models specifically trained on different types of GWL, clustered on the basis of the spectral properties of the data, performed significantly better than models trained on the whole dataset. In fact, clustering-based modelling reduces complexity in the training data and targets relevant information more efficiently. Applying multi-station models without prior clustering can lead the models to learn the dominant station behavior preferentially, ignoring unique local variations. In this respect, wavelet pre-processing was found to partially compensate clustering, bringing outcommon temporal and spectral characteristics shared by all available time series even when these characteristics are "hidden" because of too small amplitude. When employed along with prior clustering, thanks to its capability of capturing essential features across all time-scales (high and low), wavelet decomposition used as a pre-processing technique provided significant improvement in model performance, particularly for GWLs dominated by low-frequency variations. This study advances our understanding of GWL simulation using deep learning, highlighting the

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

importance of different model training approaches, the potential of wavelet preprocessing, and the value of incorporating static attributes.

1. Introduction

Understanding the large-scale hydrological functioning of a hydrosystem is the best approach for grasping a more global view of water reserves and implementing appropriate long-term management strategies (Kingston et al., 2020; Massei et al., 2020). However, this approach requires the construction of a large-scale hydrological model capable of capturing interactions over large areas, while respecting hydraulic continuity across the hydrosystem. The model must be able to analyze and test, for example, the effects of different modes of exploitation or any other human interventions, as well as the effects of climate change over the long term. Building the large scale model implies collecting and processing a massive database to accurately capture all the forcings - geological, oceanic, climatic and anthropogenic - that drive groundwater flow.

However, the numerical, physics-based representation of all the physical processes occurring during the hydrological cycle in the subsurface remains an extremely complex task to achieve rigorously, in paticular in the framework of large-scale modelling(Paniconi & Putti, 2015). Although progress has been made in this field, applications of physics-based models are still mainly focused on aquifers in relatively small watersheds.

Under these conditions, data-driven tools have emerged as an interesting alternative (or complement) for capturing the complex interactions that occur on different time and space scales, including large ones. They rely on efficiently processing a large database without having to rely on numerical physical representations of the non-linear physical processes that link climatic and hydraulic signals(Hauswirth et al., 2021). These processes are efficiently approximated on the basis of small and simple weight matrices defined to reproduce the observed hydraulic signals, either at aquifer or river(Vu et al., 2023). The application of AI algorithms, and DL in particular, is growing in the geosciences and especially in the hydrosciences(Nourani et al., 2014, 2023; Rajaee et al., 2019), thanks to the increase in computational resources, but also to the growing availability of global datasets for different hydrological variables(Addor et al., 2017; Kratzert et al., 2023), which are making it possible to better address issues related to the understanding and management of hydrological systems (Muñoz-Carpena et al., 2023).This has been confirmed in several recent studies that have highlighted the potential of deep learning tools for hydrological simulations(Fang et al., 2022; Klotz et al., 2022; Kratzert et al., 2019, 2021; Nourani et al., 2021) and forecasting tasks (Jahangir et al., 2023; Momeneh

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

& Nourani, 2022; Sina Jahangir & Quilty, 2023; Vu et al., 2023). Most often, these approaches are applied to rainfall-runoff modelling due to the availability of long-term runoff data. This is not always the case for aquifers, due to the high cost of installing piezometers. Furthermore, the highly heterogeneous nature of underground reservoirs leads to complex hydrodynamic behaviors on a regional or continental scale, which cannot be captured by a limited number of piezometers. Consequently, the few applications of DL to groundwater flow, whether in simulation(Chidepudi et al., 2023a) or forecasting (Bai & Tahmasebi, 2023a; Collados-lara et al., 2023; Rahman et al., 2020; Vu et al., 2023; Wunsch et al., 2021), are on a local scale and involve only single station models on a small number of piezometers in the construction of neural networks.

DL models have proved effective on a local scale,and are also on a larger scale by collectively training a significant number of piezometers(Chidepudi et al., 2023b; Heudorfer et al., 2024). This collective approach involves the use and processing of all available piezometric stations to learn about relationships or events likely to occur at the target station, even if they have not yet been observed at that station. This approach also requires the use and extraction of the relevant global climate signal and the tracking of its effects. This can have a delayed effect on piezometric fluctuations, making DL models more effective for long-term forecasting.

The DL models have proved effective at local scale and are also proving more effective on a larger scale. Working with groundwater data also presents additional challenges compared to runoff data, such as 1) complex and heterogeneous geological factors influencing GWLs, 2) difficulty in linking the available data to the appropriate well (for surface water this is easily done through catchment delineation, but this isn't the case for aquifer delineation), 3) slow response time (longer time series needed, i.e. data availability issue as mentioned above), 4) sensitivity to human activities (e.g. pumping).

In some hydrological studies, the term 'global models' is used to describe models trained from multiple wells or stations. However, this term can be misleading in the groundwater context as it suggests a broader scope than intended. Therefore, in this study, we use the term "multi-station approach", for models trained on data from different wells with external input variables, which more accurately reflects their scope and methodology.

Efforts to use data from multiple GWL stations in model training have been limited and have often focused on forecasting or reconstruction using data from nearby GWL wells as input. For example, Vu et al. (2021) used data from nearby stations to reconstruct the GWLs at a single station, albeit using GWLs from nearby stations only while training individual models for each station. Another recent study (Patra et al., 2023) developed so-called 'global models' for GWL forecasting and not simulations,

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

i.e. these models only use past GWL data to forecast future GWLs. (Bai & Tahmasebi, 2023) used graph neural networks for groundwater level forecasting to capture the spatial dependencies of nearby wells and compared their performance with the single station gated recurrent unit (GRU) and long short-term memory (LSTM). A recent study by (Gholizadeh et al., (2023) used LSTM alongside static attributes and demonstrated its applicability for simulating both streamflow discharge and groundwater levels. However, the scope of the study for GWL simulations was limited to only two dynamic variables: precipitation and temperature. This approach was used to simulate 21 GWL wells across Alabama, for the period 1990 to 2021. Notably, the study focused on annually varying GWLs, which may not represent the most difficult GWL variations to model. (Cai et al., (2021) in their study conducted in the central eastern continental United States, showed that GRU performed better when it was informed by hydrogeological characteristics expected to affect groundwater response along with dynamic input variables (in this case, precipitation and streamflow).

Several studies on groundwater modeling also demonstrated the potential of clustering methods (Nourani et al., 2022) in hybrid models along with AI approaches such self-organising map (Nourani et al., 2016; Wunsch et al., 2022b), K-means (Ahmadi et al., 2022; Kardan Moghaddam et al., 2021; Kayhomayoon et al., 2021, 2022; Nourani et al., 2023), Fuzzy C-means ((Jafari et al., 2021; Nourani & Komasi, 2013; Rajaee et al., 2019; Zare & Koch, 2018). However, most of these studies mainly focused on autoregressive approaches which rely on using previous GWL or nearby wells GWL data as input for forecasting or reconstruction.

The regionalisation of GWLs, a process that could involve clustering and training of DL models using non-autoregressive approach of learning from external input variables on comprehensive datasets, remains underexplored. The potential of multi-station approaches, particularly those that integrate static attributes and dynamic data or use clustering/pre-clustering, remains largely unevaluated in the context of GWL simulations. While these methods have proven effective in runoff modelling(Fang et al., 2022; Hashemi et al., 2022; Klotz et al., 2022), their application to GWL simulation is still questionable. A comprehensive evaluation of their strengths and weaknesses is essential to unlock their full potential in the simulation of GWLs. This includes a detailed investigation of the performance of these models in various GWL simulation scenarios. In addition, techniques such as wavelet preprocessing, such as BC-MODWT (Chidepudi et al., 2023a), have shown promise in single-station models, but have not been extensively tested on a regional-scale simulations. Given this background, the current study aims to address several research questions:

a) How do the generalised (multi-station) models compare with the specialised (single-station) models in simulating GWLs?

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

b) Can wavelet pre-processing techniques improve the performance of models for different types of GWLs when trained with data from all available stations?

c) To what extent do static attributes or one-hot encoding techniques help models to generalize across different GWL behaviours? Is using a combination of these methods more effective than using them individually? Furthermore, how do these models compare to those trained on GWL stations grouped by similar spectral temporal statistical characteristics?

d) What are the key variables that influence the learning of these models, particularly in terms of capturing low-frequency variability while it is buried into high-frequency-dominated explanatory signals?

By addressing these questions, this study aims to provide a comprehensive evaluation of regional modelling approaches for GWL simulations and to compare their performance with the local approaches. We refer to (Beven and Young, 2013), for differences in the use of the terms simulation and forecasting. To achieve this, we test different approaches for multi-station models, while including static attributes and compare the results with those obtained using local models. Furthermore, we evaluate the impact and usefulness of integrating wavelet pre-processing with multi-station deep learning models. All our experiments are conducted only under the gauged scenario similar to (Li et al., 2022)

The rest of the paper is structured as follows: Section 2 details the datasets used; Section 3 presents the methodology and experimental design for the different approaches. Section 4 discusses the ability of the models and robustness to capture different variations in groundwater levels and input scenarios. Section 5 deals with the discussion on the interpretability of the obtained results. Section 6 presents our main conclusions and perspectives.

2. Data

We used the forcing data from ERA5 (Hersbach et al., 2020) with a spatial resolution of 0.25 degrees to obtain the dynamic climate variables. In particular, we extract seven atmospheric variables: 10m zonal (W-E) U-wind component (u10), 10m meridional (S-N) V-wind component (v10), 2m air temperature (t2m), evaporation (e), mean sea level pressure (msl), surface net solar radiation (ssr), total precipitation (tp). These variables are among the most commonly used inputs for hydrological and land surface models as they are representative of atmospheric conditions and circulation, moisture fluxes and radiative forcing. In this work, we also included static attributes (Table1) to assess whether such informative data would help to better represent small differences between GWL time series owing to different contexts (e.g., type of porosity, overall geological context, lithology, location (lon, lat)). Such data were retrieved from the French national database BDLISA (ht

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

tps://bdlisa.eaufrance.fr/) ; they would be related to filtering capabilities of the aquifers with respect to the input signals (e.g. precipitation). Although they seem somehow redundant, they are expected to provide complimentary information about the hydrogeological nature of the hydrosystems. In addition, the available groundwater levels of climate-sensitive wells (i.e. not strongly affected by human activities and sensitive to climate variability (Baulon, Allier, et al., 2022)) with high data quality until the end of 2022 were obtained from the ADES database (Winckel et al., 2022).

Table 1: Description of the static attributes used in the current study.

Variable	Description	Possible values and details
type of porosity	Type of environment for a hydrogeological entity characterised based on the level of porosity: porous, karstic, fracture....	https://id.eaufrance.fr/nsa/353
geological context at large scale	Hydrogeological entity theme based on the different geological formations: alluvial, sedimentary, volcanic...	https://id.eaufrance.fr/nsa/348
lithology	Dominant rock types associated with the well location: limestone, clay...	https://id.eaufrance.fr/nsa/165
co-ordinates	latitude and longitude of the well location	

In addition, the GWL data were clustered into three different clusters following the methodology outlined by Baulon et al. (2022), which is based on spectral properties (i.e. characteristic time scales of variability inherent to each cluster). These clusters are identified as annual, mixed, and inertial, as depicted in Figure 1. Specifically, the first cluster showcased in Figure 1 exhibits a pattern that is predominantly influenced by the annual cycle, indicating an annual behaviour. The second cluster, the mixed, shows characteristics of both annual and interannual variability. The third cluster, the inertial, is mainly characterised by its low-frequency variability.

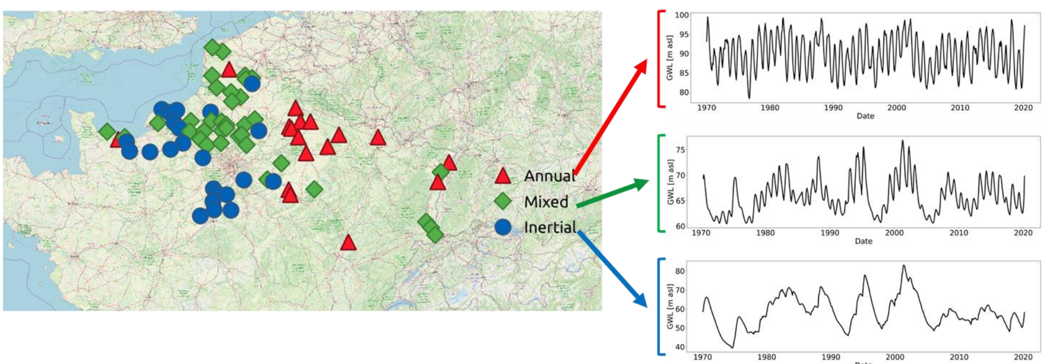

Figure 1: Clustering of GWL timeseries data (Background layer: © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.) based on the spectral statistical properties (Baulon et al., 2022)

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

3. Methodology: from single station to multi-station training

3.1 Theoretical modelling background

In the current study, we explored the use of recurrent-based deep learning models to simulate groundwater levels (GWLs) across multiple stations using different approaches as described in the section 3.2. We apply three types of recurrent neural networks: Long Short-Term Memory (LSTM, Hochreiter & Schmidhuber, 1997), Gated Recurrent Unit (GRU, Cho et al., 2014), and Bidirectional LSTM (BiLSTM), alongside a wavelet pre-processing strategy (BC-MODWT). Each of these methods is designed to process data that changes over time, capturing patterns and dependencies that occur over extended periods of time. In brief, LSTM has a single memory cell and three gates (forget, input, and output) to manage the flow of information. GRU simplifies this design, with only two gates (reset and update) to increase computational efficiency by reducing the number of parameters compared to LSTM. BiLSTM further optimises data analysis by simultaneously processing sequences in both forward and backward directions. These models are particularly good at identifying various patterns in data sequences, making them ideal for simulating GWLs, that change over time. We also explored the potential of wavelet decomposition (BC-MODWT) to decompose the data into components of varying frequencies (Figure 2), from high to low, to provide more detailed input to the DL models to better simulate the GWLs. As explained in Chidepudi et al. (2023a), decomposition depth (i.e. the choice of the number of components) was constrained by the trade-off between 1- achieving a sufficient high level of decomposition to ensure the low-frequency variability is properly reached, and 2- keeping the number of coefficients affected by boundary conditions as low as possible since these have to be ultimately removed from the input time series. Figure 2 illustrates that a 4-level decomposition of one precipitation time series (tp) efficiently extracted the first 4 so-called wavelet details (tp_1 to tp_4) while the last fifth (so-called "smooth”) tp_5 component remains of sufficiently low frequency. It is clearly visible that tp_5, almost invisible in the original tp precipitation time series, corresponds well to the variability of the most interial GWL types (Figure.2, in red, with a few month time lag with respect to tp).

To maintain consistent comparison criteria across all methods evaluated in the study, Bayesian optimisation was used for hyperparameter tuning. Furthermore, the range of hyperparameters used for optimisation was standardised across all methods, following the best practices outlined for both standalone and wavelet-assisted models, as detailed in Chidepudi et al. (2023a) and Quilty and Adamowski (2018).

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

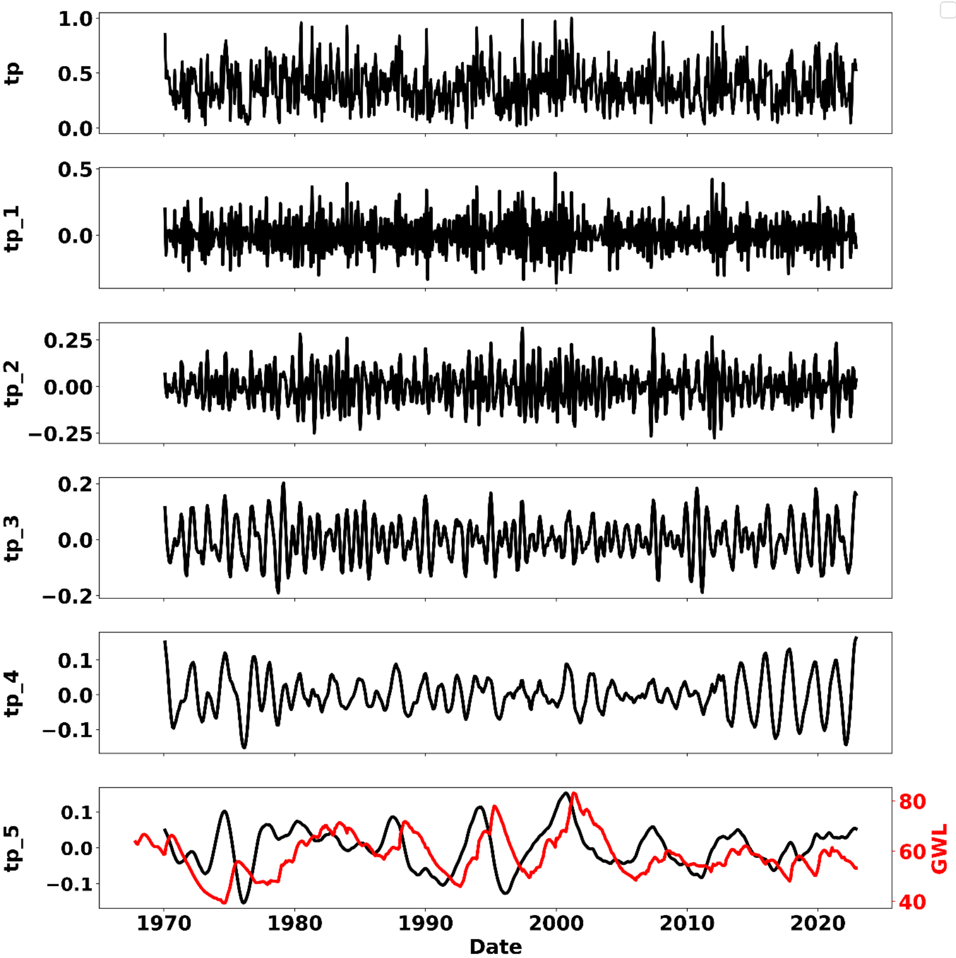

Figure 2: Total precipitation(tp) and its wavelet components: High(tp_1) to low frequency(tp_5) and GWL (in red).

However, we made an important update to the model architecture by setting the number of layers to one for all models, rather than optimising it. This decision was based on findings (Figure 3) that optimising the number of layers did not significantly improve performance and was in line with recent studies in related fields like rainfall-runoff modelling (Kratzert et al., 2019, 2021). Other adjustments included reducing the number of initialisations to 10 and setting the number of trials in the Bayesian optimisation to 30. These changes were aimed at reducing the computational requirements of our approach, making it more efficient without significantly affecting the quality of our results and are consistent with recent studies(Wunsch et al., 2022a). The intricacies and specific technical details of the architectures these models are well documented in the existing body of deep learning research

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

applied to hydrological simulations, as detailed in several studies (Chidepudi et al., 2023a; Chidepudi et al., 2024; Fang et al., 2022; Kratzert et al., 2021; Li et al., 2022; Vu et al., 2023).

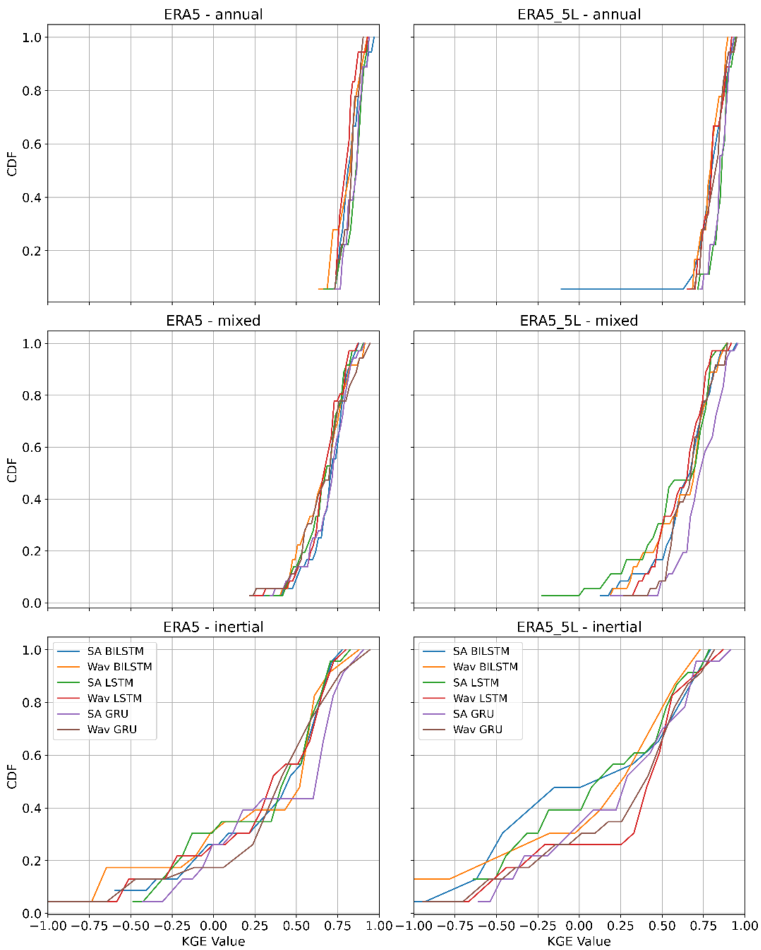

Figure 3: Comparison of performance of single layer DL models (left column) and multiple-layer DL models (right column) with respect to single station model as a reference. SA represents Standalone models while Wav represents Wavelet-assisted models.

To further interpret and decrypt the results for better understanding, we used the SHAP approach(Lundberg & Lee, 2017), which is an increasingly popular game-centric approach for explaining the outcomes of deep learning models. SHAP, or Shapley Additive Explanations, explains how each input feature influences the model's simulations. It does this by highlighting two key aspects: the importance of each variable, where a higher mean absolute SHAP value indicates a greater impact, and the nature of that impact, whether positive or negative.

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

3.2 Experimental design

This section details the experimental design used to assess the effectiveness of training models using data from all available stations. Our study uses different strategies to incorporate both numerical and categorical data into the models. The aim is to improve the accuracy of groundwater level simulations by exploring ways of incorporating regional variability into the models. The experimental setup is structured to test different modelling strategies, as described below and visualised in Figure 4:

1. **Single station** or local models (models trained and tested individually per station): These models are trained and evaluated on data from individual stations. As a baseline, their performance provides a benchmark for evaluating the effectiveness of more generalised models. This approach is dominant in the development of data-driven models for groundwater simulations and is discussed in detail in Chidepudi et al. (2023a; 2024).
2. **Multi-station** (models trained and tested together on many stations): These models are trained using data aggregated from multiple stations and tested with different input configurations:
    a. **NO (dynamic inputs only):** Models are trained on all stations using dynamic variables only, excluding static attributes and one-hot encoding.
    b. **OHE (One-Hot Encoding):** This method involves one-hot encoding to represent individual station ID information as binary vectors, to ensure that the specific information is obtained from collective training, similar to the one-hot vector strategy developed in rainfall-runoff modelling (Li et al., 2022). This study showed that one-hot vector (one hot encoding using basin ID) can produce similar results to using catchment attributes in gauged basin scenarios. One-hot encoding serves as an alternative to incorporating static attributes directly into the model (Table 2).

Table 2: Example of one hot encoding based on different wells

WELL	Dynamic variables	Well_ID_1	Well_ID_2	Well_ID_3
1	...	**1**	**0**	**0**
2	...	**0**	**1**	**0**
3	...	**0**	**0**	**1**

    c. **STAT (Static attributes and dynamic Variables):** Models include both static attributes (e.g., latitude, longitude) and dynamic variables as inputs, with categorical variables encoded similarly to one-hot encoding but represented in separate columns for each unique value or class (Table 3).

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

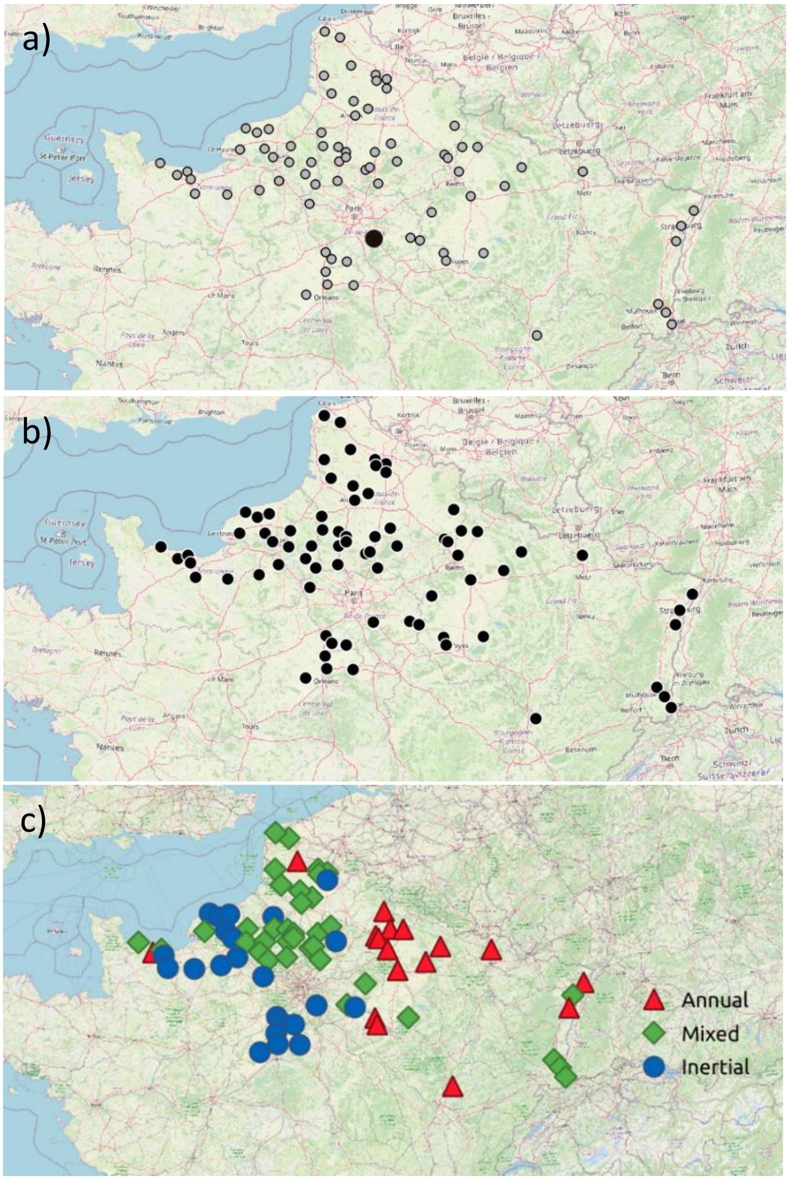

Figure 4: Comparison of different approaches adopted in the current study: a) single station (Top) b) multistation without clustering (Middle) c) multistation with clustering based on spectral properties(bottom). (Background layer: © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.)

3. **STAT_OHE (Static attributes, one-hot encoding, and dynamic variables):** This configuration combines static attributes, one-hot encoding for well IDs, and dynamic variables to provide a comprehensive dataset for model training. In other words, a combination of the two input strategies above.

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

Table 3: Example with static attributes of numeric and categorical types

WELL	Dynamic variables	Static_1 (Lattitude)	Static_2 (Longitude)	Category_ 1 (Alluvial)	Category 2 (sedimentary)	Category 3 (Mountainous)
1	...	5.1	9.5	1	0	0
2	...	2.8	10.8	0	1	0
3	....	5.4	9.2	0	0	1

In addition to these configurations, we investigated the performance of multi-station models trained on groundwater levels (GWLs) with similar spectral statistical properties. This approach assesses the effectiveness of models tailored to specific GWL behaviours compared to more generalised models using the aforementioned strategies. In this study, KGE is preferred over NSE and other metrics because it offers a more comprehensive evaluation by integrating three aspects of model error: correlation, bias, and the ratio of standard deviations.

Our methodology for comparing single station and multi-station approaches, both with and without prior clustering based on spectral properties, is consistent with the research conducted in rainfall-runoff modelling by Hashemi et al. (2022), where the catchments were divided into five subsets according to hydrological regimes. This comprehensive experimental design aims to identify the most effective strategies for using multi-station data in the simulation of groundwater level variations.

4. Capabilities, performances and interpretability of multi-station approaches

4.1 Different strategies for multi-station approach

Among the models tested in the case of this study, the GRU stood out for its superior performance in predicting GWLs, leading us to focus our analysis on this model. Details of this comparison for LSTM and BiLSTM can be found in the appendix A.

Figure 5 shows the results of different GRU model configurations for simulating groundwater levels (GWLs). The first row shows the performance of the standalone GRU model for different GWL categories, while the second row shows the wavelet-assisted GRU results.

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

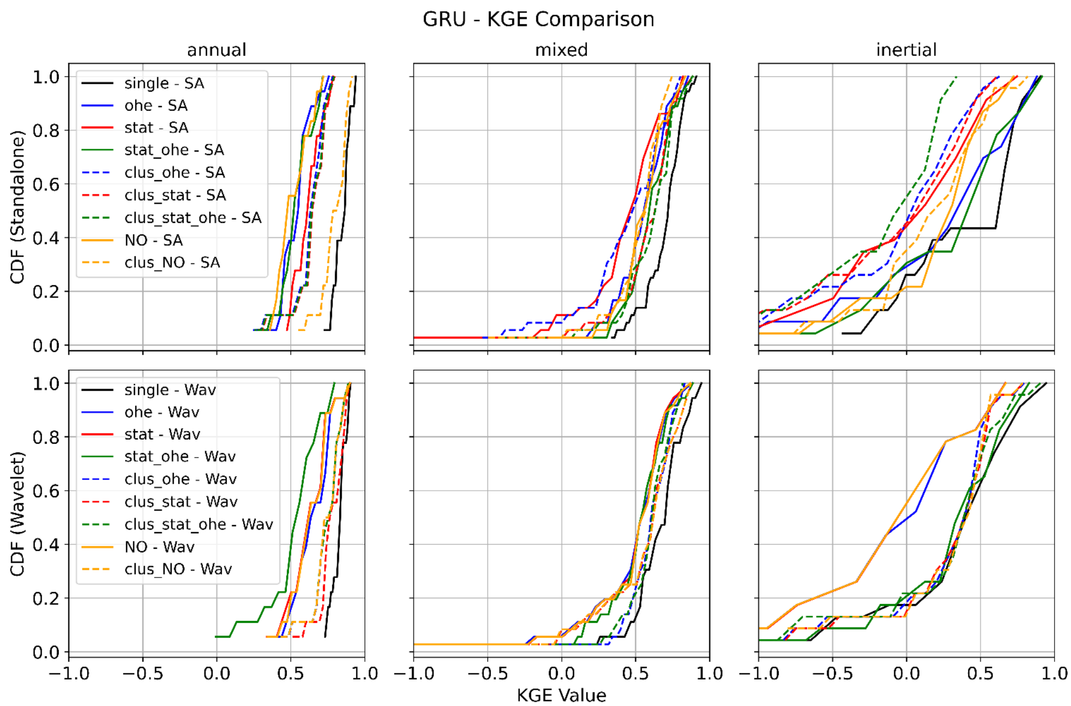

Figure 5: CDF Comparison of KGE values of the GRU With different approaches and GWL types.

Several observations can be made from Figure 5. Wavelet pre-processing generally improves model performance, especially in the inertial GWL category, where cumulative distribution functions (CDFs) are steeper and shifted to the right, indicating a higher proportion of simulations with high performance. This is in line with previous findings as already reported in our previous works (Chidepudi et al., 2023a & 2024). This demonstrates the wavelet decomposition ability to extract "hidden" inertial dynamics features which facilitates their assimilation by the model in the learning process. In other words, the improvement attributed to wavelet pre-processing becomes more pronounced as we move from annual to mixed, and then further to inertial behaviour. This is because in the case of annual-type GWL, the dominant variability (annual cycle) is already well expressed in several input variables (e.g. t2m, msl, ssr). In the case of mixed- and inertial GWL types, the dominant low-frequency variability, while also present, is barely expressed, almost "hidden", in the input data, and becomes prominent in GWL due to the low-pass filtering action of aquifers(Baulon et al., 2022; Schuite et al., 2019). Wavelet decomposition allows unraveling such hidden information, helping the neural networks to reach it for enhanced learning. This is illustrated in figure 2 with low-frequency component of precipitation (tp5) matching the variations of one intertial-type GWL (in red, with a few month-lag time), whereas it is masked by other higher-frequency components in the original

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

precipitation time series (tp).The combination of static attributes and OHE gives competitive results, particularly in the inertial category, demonstrating the effectiveness of this method without the need for prior clustering of GWL behaviour. Multi-station models, when trained separately for each GWL cluster, generally outperform those trained on aggregated data. This is reflected in higher KGE values for cluster-specific models, suggesting a better representation of the unique characteristics of each GWL type. However, this advantage diminishes for mixed GWLs, which are the majority in the study area. Although single station models perform best for all GWL types, some multi-station models approach or match their performance, highlighting their potential for regional-scale groundwater simulations. For the annual GWL category, models trained on mixed GWL data without wavelet pre-processing and relying solely on static attributes do not show significant performance improvements, suggesting that static features alone may not adequately represent the dynamic nature of groundwater behaviour.

Figures 6-8 show the best GWL simulations obtained of different types (annual, mixed and inertial) for single and multi-station models. While single station models perform best, multi-station models are valuable where single station modelling is impractical either due to data limitations or computational requirements.

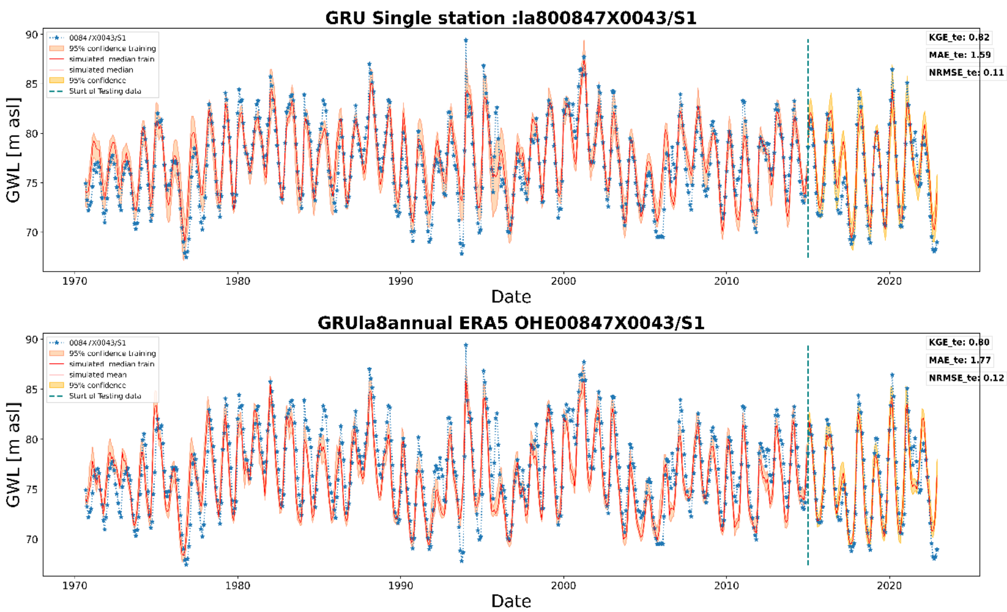

Figure 6: Results with wavelet assisted GRU in annual type of GWLs through a) Single station (top) and b) Multi station model trained on annual type of GWLs with static and ohe (bottom)

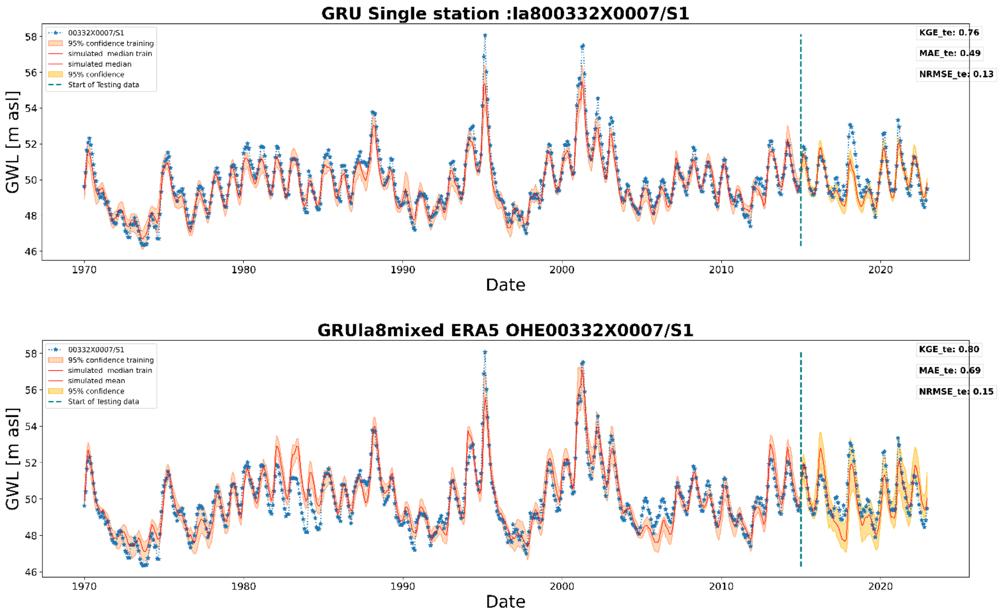

Figure 7: Results with wavelet assisted GRU in mixed type of GWLs through a) Single station (top) and b) Multi-station model trained on mixed type of GWLs with static and ohe (bottom)

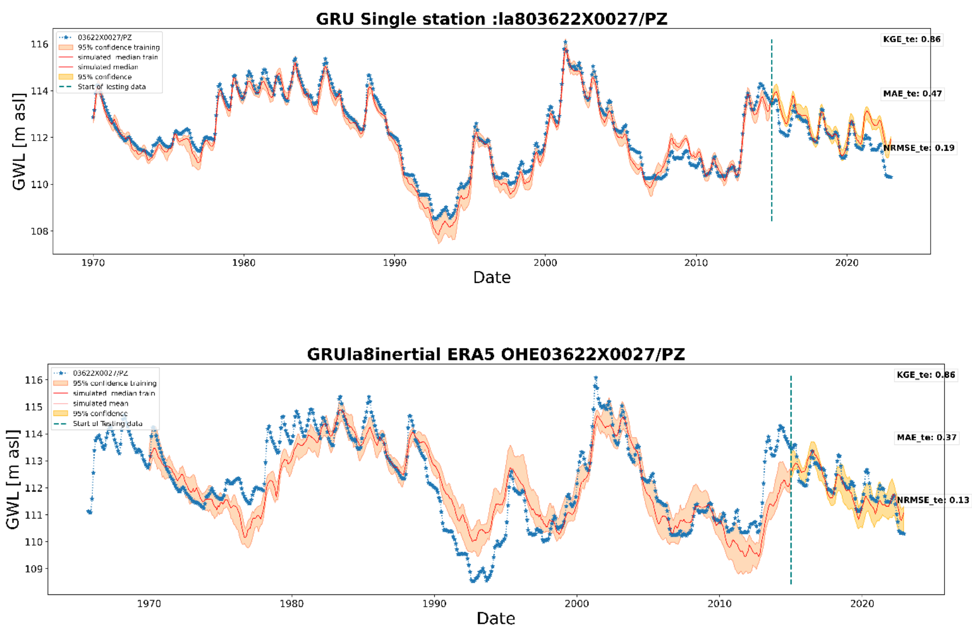

Figure 8: Results with wavelet assisted GRU in inertial type of GWLs through a) Single station (top) and b) Multi-station model trained inertial type of GWLs with static and ohe (bottom)

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

In summary, wavelet-assisted GRU models are particularly effective, especially for low-frequency dominated GWL behaviour, and multi-station models designed for specific GWL types (i.e. training over specific pre-clustered datasets) generally outperform generalised models. The multi-station approach is sensitive to the dominant GWL type in the training dataset, with the best results seen in models trained for the predominant mixed GWL type in the study region, consisting of 36 mixed, 23 inertial and 18 annual stations. To address the issue of model learning dominant behaviour in collective training of multistation approaches, possible future investigation may involve generating synthetic time series with randomised amplitude changes of constituting frequencies to increase the dataset while balancing all the important behaviours. This could also help in understanding the influence of the size of dataset on using multistation approaches.

4.2 Understanding GWL Simulations Through SHAP Interpretability

This section deals with the deeper understanding of the simulations from the insights obtained from the SHAP analysis on model's interpretability. In this study, we investigated the key contributing factors for GWL simulations in different approaches that were previously evaluated above in terms of accuracy.

Figure 9a shows the SHAP summary plot for the standalone models using a single station approach. These plots highlight the influence of different variables/attributes on the final simulation. In particular, the distribution of data points in the SHAP summary plot (figure 9), with more points to the right (coloured red) indicating positive influences, and the opposite indicating negative relationships.

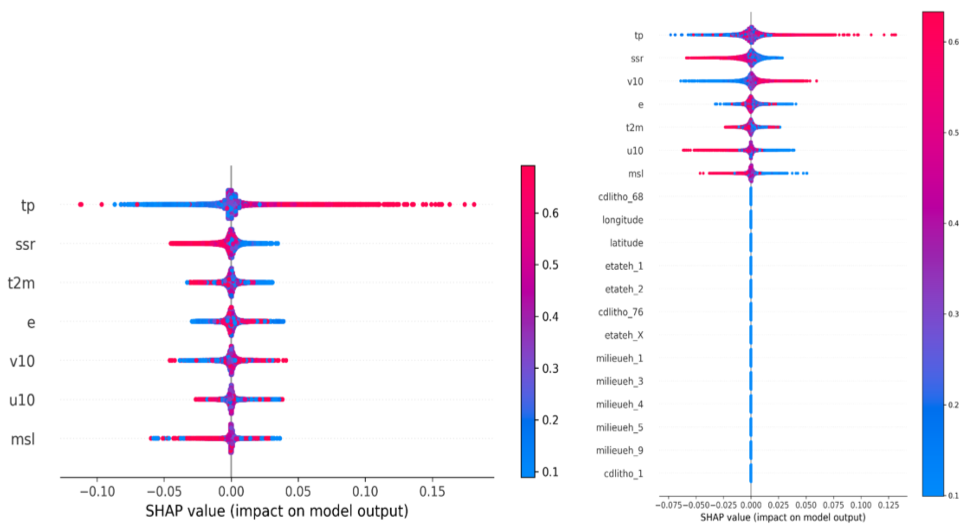

Figure 9: SHAP summary plot examples for single station model and multi station model with static attributes

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

The insights from the SHAP results help us to understand the following aspects: First, when do the models effectively simulate the “diverse” clusters of GWLs? What does SHAP interpretability reveals on addressing this issue of diversity (with focus on low-frequency variability)?. Secondly, what can we learn from training the DL models using the clustering based on spectral properties? How does it relate it behave along with the wavelet pre-processing in dealing with low-frequency variability?

From the analysis of Figure 10 and Figure 11, several notable patterns emerge regarding the contribution of different variables to groundwater level (GWL) simulations using standalone models and those with wavelet pre-processing, and the impact of clustering as well as pre-clustering based on spectral statistical properties.

In single station standalone models, SHAP analysis shows that certain variables consistently influence GWL simulations, although their order of importance can change. Total Precipitation (TP) emerges as the key factor, with Surface Net Solar Radiation (SSR) occasionally overtaking in mixed GWL cluster, especially in models trained on clusters, along with static features, or one-hot encoding (OHE). Nonetheless, TP and SSR are the primary drivers in these simulations.

In multi-station standalone models without clustering, TP and SSR lead in importance, followed by wind speed at 10 meters (v10), evaporation (e), and air temperature close to the ground (2-meter temperature, t2m), which vary in their influence. Notably, v10 plays a bigger role in models in multi-station approaches. When models are trained on clusters, evaporation becomes more significant, yet the impact of clustering on variable importance is generally minor.

The spectral statistical characteristics (amplitude of high and low frequencies) were used for the pre-clustering of GWLs. These characteristics are related the filtering of the input signal by the physical properties of the hydrological system. This highlights the importance of pre-clustering in capturing the physical characteristics of basins and suggests that it may be preferable to cluster based on these properties rather than relying on static attributes, especially when the relevance of static attributes is uncertain.

SHAP analyses show that standalone models maintain similar variable importance rankings even after clustering with static attributes and OHE. However, wavelet pre-processing shifts the importance towards dynamic components, reducing the contributions of static features or OHE. When clustering is combined with wavelet preprocessing, low-frequency precipitation components emerge as key contributors, improving model performance.

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

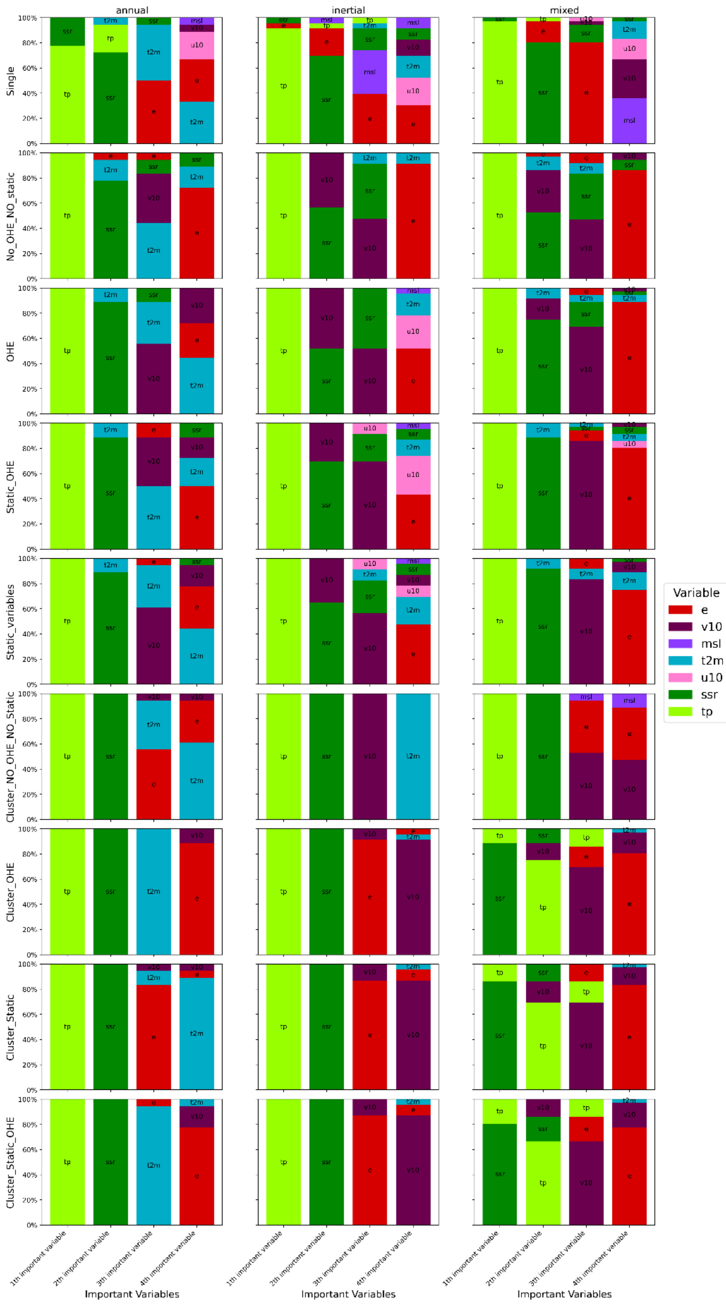

Figure 10: Top four important variables by cluster for standalone GRU models with different approaches. On Y-axis, Percentage of stations for each variable within in the cluster.

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

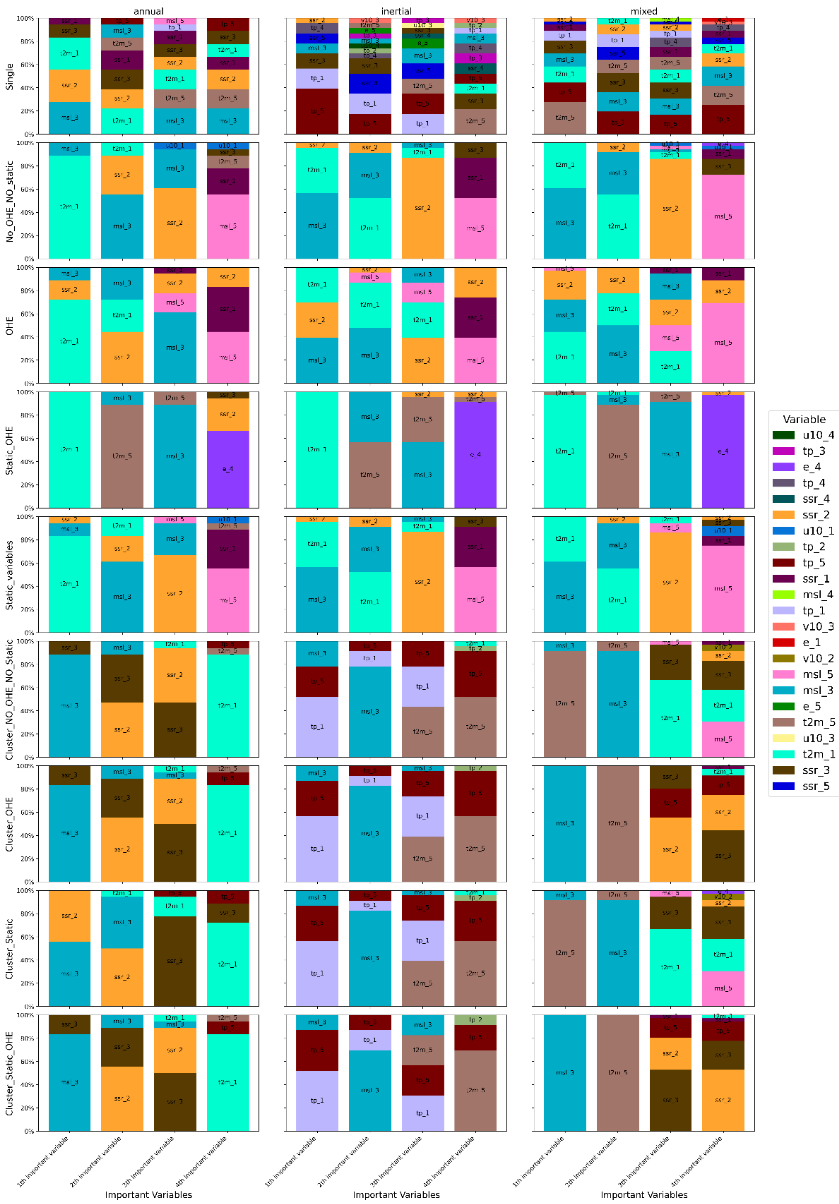

Figure 11: Top four important variables in regional GRU wavelet assisted model trained with different approaches for different classes

410 When models are trained after clustering, low-frequency components (e.g. tp_5, t2m_5) are prioritised in mixed and inertial clusters: components not seen without clustering. Annual types

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

prioritise relevant frequencies (1 to 3), consistent with single station model patterns. The addition of static attributes to the OHE does not significantly alter the contributions, suggesting a dominance of dynamic variables after decomposition. Also, differences among multi-station approaches after clustering are minimal for both standalone and wavelet models.

Wavelet pre-processing performs a similar function to pre-clustering based on spectral properties by revealing information across all frequencies, including low amplitude frequencies that may be obscured. The order of best approaches based on the results: wavelet plus pre-clustering, followed by pre-clustering only, then wavelet only, and finally standalone highlighting the effectiveness of this approach.

There is a clear pattern when clustering is applied; without clustering the high frequency component of the 2-meter temperature (T2m_1) is dominant. Multi-station models show less diversity in variable contributions than single station models. The exception is the Stat_OHE without clustering approach, which uniquely captures low-frequency information from T2m_5 and e_4. Otherwise, the static and No_ohe_no_stat approaches gave similar results.

The influence of static attributes or OHE appears to be minimal, possibly due to the high dimensionality introduced by numerous dynamic and static attributes. This observation suggests that future research could investigate alternative methods, such as target encoding, to address this dimensionality issue.

Although the added value of static variables was found to be marginal in the current study, they may prove useful in settings where no measurement is available. Further research is required to determine their utility in simulating such ungauged hydro systems. The approaches presented (except OHE) may be applicable to ungauged aquifers but require validation in a pseudo-ungauged environment. The use of data from multiple stations can enrich the dataset, improving the representation of groundwater systems and the robustness of the models. This multi-station approach also allows the model to be applied to areas without groundwater level (GWL) monitoring, thereby capturing regional dynamics. However, single station modelling remains important for understanding local interactions. The choice of method should therefore be guided by research objectives, data availability and the hydrogeological context. Where clustering results in too many groups, future studies should consider fine-tuning the general model for each cluster, following the approach of Mohammed & Corzo (2024).

5. Concluding remarks

This study has demonstrated the different multi-station approaches to groundwater level simulations with emphasis on the use of static attributes, one-hot encoding and the combination of both while

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

training on all available data or by training on each GWL type based on the clustering. The study also highlights the potential of these approaches compared to the traditional single station approach with and without the use of BC-MODWT. Key findings from this research highlight the advantages of clustering based on spectral properties, which has been shown to significantly improve the results of multi-station models, surpassing those of general models. Clustering is preferred over the use of static attributes, as the use of static attributes alone may not be sufficient to effectively distinguish different behaviours. Wavelet pre-processing is very effective at extracting relevant information at all levels, from high to low frequency, allowing low frequency dominated GWLs to be handled with increased accuracy. The combination of clustering and wavelet pre-processing produced the most accurate simulations, indicating that wavelet pre-processing is likely to capture key information needed for accurate modelling.

The study also showed that a multi-station approach, without clustering, should be used with caution, as models tend to adopt dominant behaviour, which may not always be desirable. In scenarios where wavelet pre-processing is not applied, the combination of static attributes and OHE demonstrated promising results, particularly for GWLs dominated by low frequencies. However, the minimal effect of static attributes or OHE observed in wavelet-assisted models may be due to the high-dimensional nature of these variables, suggesting a potential avenue for future research to explore alternative encoding strategies, such as target encoding. SHAP (SHapley Additive exPlanations) analyses consistently identified key contributors across models, with clustered models highlighting the pivotal role of low-frequency components, especially precipitation and temperature, in achieving superior simulations for inertial and mixed types of GWL.

In summary, although the study has led to a better understanding of GWL simulation approaches with limited static attributes, further research is needed in the following areas, also exploring other physical basin parameters such as: thickness of overlying formations, altitude, distance from the sea, etc. It should also be pointed out that clustering can be a source of information on the physical properties of the basin. Indeed, the three groups determined in this study on the basis of spectral properties indirectly carry information on the modalities of water transfer in the shallow formations and aquifer, which are controlled by the hydraulic properties of the basin. The study of the importance of using static data in groundwater flow modelling using deep learning tools needs to be extended to cover level prediction at sites with no piezometers. The insights gained here pave the way for future efforts to simulate GWLs in unmonitored or new locations, taking advantage of the robustness offered by multi-station models, while recognising the value of single-station models for capturing local-scale interactions. Finally, it is noticeable through our study that the overall approach is compatible with a frugal AI approach (keeping in mind that our datasets are very small compared to classsical big

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

datasets from other fields like natural language processing etc): compact networks were tested and preferred (one layer), Bayesian optimisation was used instead of grid search for hyperparameter tuning. In addition, multi-station approaches pave the way for transfer learning, which will reduce the need for specialised models and retraining new models. The way forward is clear: to improve the groundwater simulations in an efficient manner, we may need to adopt a nuanced mix of clustering (although it should not be needed if large enough datasets capturing diverse behaviours are available), efficient input signal pre-processing and potentially new encoding strategies to incorporate all possible knowledge of the system.

Competing interests. The contact author has declared that none of the authors has any competing interests

Data availability and Acknowledgement

We acknowledge the computational resources provided by CRIANN to carry out the experiments carried out as part of our ongoing project.

CRediT authorship contribution statement

Sivarama Krishna Reddy Chidepudi: Data curation; Formal analysis; Writing and conceptualization of original draft; model development and model runs; Investigation;

Nicolas Massei: Funding acquisition; Supervision; Writing and co-conceptualisation of the original draft- review & editing; Project administration. Abderrahim

Abderrahim Jardani: Supervision; writing, review & editing; Project administration.

Bastien Dieppois: review & editing;

Abel Henriot : Supervision; review & editing; Project administration.

Matthieu Fournier: review & editing

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

Appendix A:

Results from LSTM and BILSTM

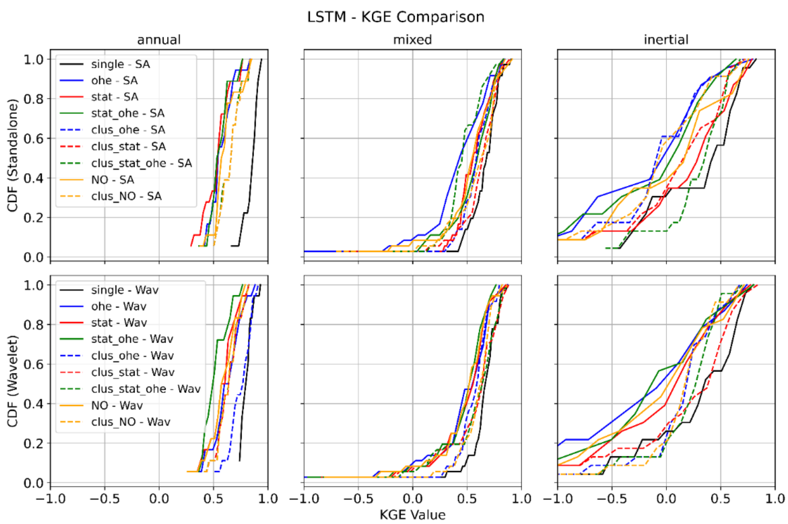

Figure A1: CDF Comparison of KGE values of the LSTM With different approaches and GWL types.

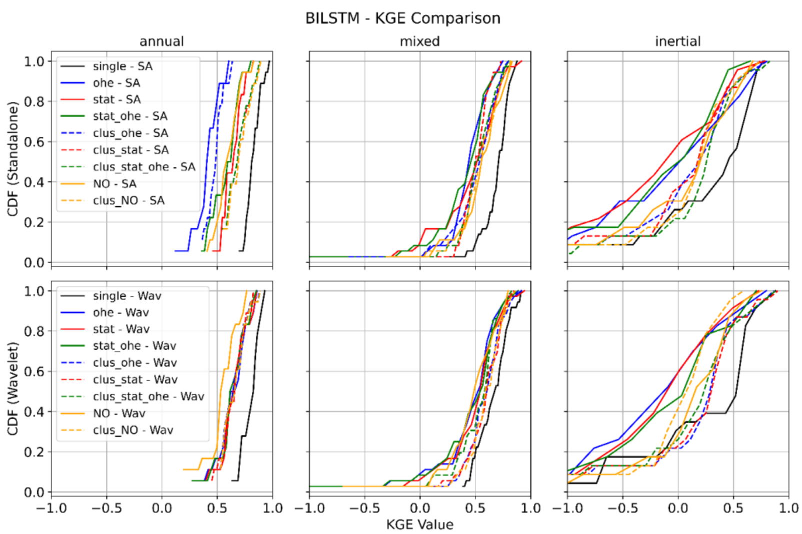

Figure A2: CDF Comparison of KGE values of the BiLSTM With different approaches and GWL types.

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

6. References

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/HESS-21-5293-2017

Ahmadi, A., Olyaei, M., Heydari, Z., Emami, M., Zeynolabedin, A., Ghomlaghi, A., Daccache, A., Fogg, G. E., & Sadegh, M. (2022). Groundwater Level Modeling with Machine Learning: A Systematic Review and Meta-Analysis. *Water (Switzerland)*, *14*(6), 949. https://doi.org/10.3390/W14060949/S1

Bai, T., & Tahmasebi, P. (2023a). Graph neural network for groundwater level forecasting. *Journal of Hydrology*, *616*(August 2022), 128792. https://doi.org/10.1016/j.jhydrol.2022.128792

Bai, T., & Tahmasebi, P. (2023b). Graph neural network for groundwater level forecasting. *Journal of Hydrology*, *616*, 128792. https://doi.org/10.1016/J.JHYDROL.2022.128792

Baulon, L., Allier, D., Massei, N., Bessiere, H., Fournier, M., & Bault, V. (2022). Influence of low-frequency variability on groundwater level trends. *Journal of Hydrology*, *606*, 127436. https://doi.org/10.1016/j.jhydrol.2022.127436

Baulon, L., Massei, N., Allier, D., Fournier, M., & Bessiere, H. (2022). Influence of low-frequency variability on high and low groundwater levels: example of aquifers in the Paris Basin. *Hydrology and Earth System Sciences*, *26*(11), 2829–2854. https://doi.org/10.5194/hess-26-2829-2022

Cai, H., Shi, H., Liu, S., & Babovic, V. (2021). Impacts of regional characteristics on improving the accuracy of groundwater level prediction using machine learning: The case of central eastern continental United States. *Journal of Hydrology: Regional Studies*, *37*(September), 100930. https://doi.org/10.1016/j.ejrh.2021.100930

Chidepudi, S. K. R., Massei, N., Henriot, A., Jardani, A., & Allier, D. (2023, February 22). Local vs regionalised deep learning models for groundwater level simulations in the Seine basin. *EGU General Assembly 2023b, EGU2023-3535*. https://doi.org/10.5194/EGUSPHERE-EGU23-3535

Chidepudi, S. K. R., Massei, N., Jardani, A., & Henriot, A. (2024). Groundwater level reconstruction using long-term climate reanalysis data and deep neural networks. *Journal of Hydrology: Regional Studies*, *51*, 101632. https://doi.org/10.1016/J.EJRH.2023.101632

Chidepudi, S. K. R., Massei, N., Jardani, A., Henriot, A., Allier, D., & Baulon, L. (2023a). A wavelet-assisted deep learning approach for simulating groundwater levels affected by low-frequency variability. *Science of the Total Environment*, *865*, 161035. https://doi.org/10.1016/j.scitotenv.2022.161035

Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111. https://doi.org/10.3115/v1/w14-4012

Collados-lara, A. J., Pulido-velazquez, D., Ruiz, L. G. B., Pegalajar, M. C., Pardo-igúzquiza, E., & Baena-ruiz, L. (2023). A parsimonious methodological framework for short-term forecasting of

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

groundwater levels. *Science of the Total Environment*, *881*(April), 163328. https://doi.org/10.1016/j.scitotenv.2023.163328

Fang, K., Kifer, D., Lawson, K., Feng, D., & Shen, C. (2022). The Data Synergy Effects of Time-Series Deep Learning Models in Hydrology. *Water Resources Research*, *58*(4), e2021WR029583. https://doi.org/10.1029/2021WR029583

Gholizadeh, H., Zhang, Y., Frame, J., Gu, X., & Green, C. T. (2023). Long short-term memory models to quantify long-term evolution of streamflow discharge and groundwater depth in Alabama. *Science of the Total Environment*, *901*. https://doi.org/10.1016/j.scitotenv.2023.165884

Hashemi, R., Brigode, P., Garambois, P.-A., & Javelle, P. (2022). How can we benefit from regime information to make more effective use of long short-term memory (LSTM) runoff models? *Hydrol. Earth Syst. Sci*, *26*, 5793–5816. https://doi.org/10.5194/hess-26-5793-2022

Hauswirth, S. M., Bierkens, M. F. P., Beijk, V., & Wanders, N. (2021). The potential of data driven approaches for quantifying hydrological extremes. *Advances in Water Resources*, *155*, 104017. https://doi.org/10.1016/J.ADVWATRES.2021.104017

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Heudorfer, B., Liesch, T., & Broda, S. (2024). On the challenges of global entity-aware deep learning models for groundwater level prediction. *Hydrol. Earth Syst. Sci*, *28*, 525–543. https://doi.org/10.5194/hess-28-525-2024

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jafari, M. M., Ojaghlou, H., Zare, M., & Schumann, G. J. P. (2021). Application of a novel hybrid wavelet-anfis/fuzzy c-means clustering model to predict groundwater fluctuations. *Atmosphere*, *12*(1), 1–15. https://doi.org/10.3390/atmos12010009

Jahangir, M. S., You, J., & Quilty, J. (2023). A quantile-based encoder-decoder framework for multi-step ahead runoff forecasting. *Journal of Hydrology*, *619*, 129269. https://doi.org/10.1016/J.JHYDROL.2023.129269

Kardan Moghaddam, H., Ghordoyee Milan, S., Kayhomayoon, Z., Rahimzadeh kivi, Z., & Arya Azar, N. (2021). The prediction of aquifer groundwater level based on spatial clustering approach using machine learning. *Environmental Monitoring and Assessment*, *193*(4), 1–20. https://doi.org/10.1007/s10661-021-08961-y

Kayhomayoon, Z., Ghordoyee Milan, S., Arya Azar, N., & Kardan Moghaddam, H. (2021). A New Approach for Regional Groundwater Level Simulation: Clustering, Simulation, and Optimization. *Natural Resources Research*, *30*(6), 4165–4185. https://doi.org/10.1007/s11053-021-09913-6

Kayhomayoon, Z., Ghordoyee-Milan, S., Jaafari, A., Arya-Azar, N., Melesse, A. M., & Kardan Moghaddam, H. (2022). How does a combination of numerical modeling, clustering, artificial intelligence, and evolutionary algorithms perform to predict regional groundwater levels?

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

Computers and Electronics in Agriculture, *203*, 107482. https://doi.org/10.1016/J.COMPAG.2022.107482

Kingston, D. G., Massei, N., Dieppois, B., Hannah, D. M., Hartmann, A., Lavers, D. A., & Vidal, J. P. (2020). Moving beyond the catchment scale: Value and opportunities in large-scale hydrology to understand our changing world. *Hydrological Processes*, *34*(10), 2292–2298. https://doi.org/10.1002/hyp.13729

Klotz, D., Kratzert, F., Gauch, M., Sampson, A. K., Brandstetter, J., Klambauer, G., Hochreiter, S., & Nearing, G. (2022). Uncertainty estimation with deep learning for rainfall-runoff modeling. *Hydrol. Earth Syst. Sci*, *26*, 1673–1693. https://doi.org/10.5194/hess-26-1673-2022

Kratzert, F., Klotz, D., Hochreiter, S., & Nearing, G. S. (2021). A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall-runoff modeling. *Hydrology and Earth System Sciences*, *25*(5), 2685–2703. https://doi.org/10.5194/hess-25-2685-2021

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., & Matias, Y. (2023). Caravan - A global community dataset for large-sample hydrology. *Scientific Data*, *10*(1), 1–11. https://doi.org/10.1038/s41597-023-01975-w

Li, X., Khandelwal, A., Jia, X., Cutler, K., Ghosh, R., Renganathan, A., Xu, S., Tayal, K., Nieber, J., Duffy, C., Steinbach, M., & Kumar, V. (2022). Regionalization in a Global Hydrologic Deep Learning Model: From Physical Descriptors to Random Vectors. *Water Resources Research*, *58*(8). https://doi.org/10.1029/2021WR031794

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *2017-Decem*, 4766–4775. https://github.com/slundberg/shap

Massei, N., Kingston, D. G., Hannah, D. M., Vidal, J. P., DIeppois, B., Fossa, M., Hartmann, A., Lavers, D. A., & Laignel, B. (2020). Understanding and predicting large-scale hydrological variability in a changing environment. *Proceedings of the International Association of Hydrological Sciences*, *383*, 141–149. https://doi.org/10.5194/piahs-383-141-2020

Momeneh, S., & Nourani, V. (2022). Forecasting of groundwater level fluctuations using a hybrid of multi-discrete wavelet transforms with artificial intelligence models. *Hydrology Research*, *53*(6), 914–944. https://doi.org/10.2166/NH.2022.035

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. https://doi.org/10.1371/journal.pwat.0000059

Nourani, V., Alami, M. T., & Vousoughi, F. D. (2015). Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. *Journal of Hydrology*, *524*, 255–269. https://doi.org/10.1016/J.JHYDROL.2015.02.048

Nourani, V., Ghaneei, P., & Kantoush, S. A. (2022). Robust clustering for assessing the spatiotemporal variability of groundwater quantity and quality. *Journal of Hydrology*, *604*, 127272. https://doi.org/10.1016/J.JHYDROL.2021.127272

Nourani, V., Ghareh Tapeh, A. H., Khodkar, K., & Huang, J. J. (2023). Assessing long-term climate change impact on spatiotemporal changes of groundwater level using autoregressive-based and ensemble machine learning models. *Journal of Environmental Management*, *336*, 117653. https://doi.org/10.1016/J.JENVMAN.2023.117653

Nourani, V., Gökçekuş, H., & Gichamo, T. (2021). Ensemble data-driven rainfall-runoff modeling using multi-source satellite and gauge rainfall data input fusion. *Earth Science Informatics*, *14*(4), 1787–1808. https://doi.org/10.1007/s12145-021-00615-4

Nourani, V., Hosseini Baghanam, A., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. *Journal of Hydrology*, *514*, 358–377. https://doi.org/10.1016/j.jhydrol.2014.03.057

Nourani, V., & Komasi, M. (2013). A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. *Journal of Hydrology*, *490*, 41–55. https://doi.org/10.1016/J.JHYDROL.2013.03.024

Nourani, V., Mohammad, ;, Alami, T., & Daneshvar Vousoughi, F. (2016). *Hybrid of SOM-Clustering Method and Wavelet-ANFIS Approach to Model and Infill Missing Groundwater Level Data*. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001398

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, *51*(9), 7090–7129. https://doi.org/10.1002/2015WR017780

Patra, S. R., Chu, H. J., & Tatas. (2023). Regional groundwater sequential forecasting using global and local LSTM models. *Journal of Hydrology: Regional Studies*, *47*, 101442. https://doi.org/10.1016/j.ejrh.2023.101442

Rahman, A. T. M. S., Hosono, T., Quilty, J. M., Das, J., & Basak, A. (2020). Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms. *Advances in Water Resources*, *141*(April). https://doi.org/10.1016/j.advwatres.2020.103595

Rajaee, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, *572*(December 2018), 336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037

Schuite, J., Flipo, N., Massei, N., Rivière, A., & Baratelli, F. (2019). Improving the Spectral Analysis of Hydrological Signals to Efficiently Constrain Watershed Properties. *Water Resources Research*, *55*(5), 4043–4065. https://doi.org/10.1029/2018WR024579

Sina Jahangir, M., & Quilty, J. (2023). Generative deep learning for probabilistic streamflow forecasting: conditional variational auto-encoder. *Journal of Hydrology*, 130498. https://doi.org/10.1016/J.JHYDROL.2023.130498

Vu, M. T., Jardani, A., Massei, N., Deloffre, J., Fournier, M., & Laignel, B. (2023). Long-run forecasting surface and groundwater dynamics from intermittent observation data: An evaluation for 50 years. *Science of the Total Environment*, *880*(April). https://doi.org/10.1016/j.scitotenv.2023.163338

https://doi.org/10.5194/egusphere-2024-794
Preprint. Discussion started: 13 May 2024

Vu, M. T., Jardani, A., Massei, N., & Fournier, M. (2021). Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. *Journal of Hydrology*, *597*(November 2020). https://doi.org/10.1016/j.jhydrol.2020.125776

Winckel, A., Ollagnier, S., & Gabillard, S. (2022). Managing groundwater resources using a national reference database: the French ADES concept. *SN Applied Sciences*, *4*(8), 1–12. https://doi.org/10.1007/s42452-022-05082-0

Wunsch, A., Liesch, T., & Broda, S. (2021). Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrology and Earth System Sciences*, *25*(3), 1671–1687. https://doi.org/10.5194/hess-25-1671-2021

Wunsch, A., Liesch, T., & Broda, S. (2022a). Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. *Nature Communications*, *13*(1), 1–13. https://doi.org/10.1038/s41467-022-28770-2

Wunsch, A., Liesch, T., & Broda, S. (2022b). Feature-based Groundwater Hydrograph Clustering Using Unsupervised Self-Organizing Map-Ensembles. *Water Resources Management*, *36*(1), 39–54. https://doi.org/10.1007/s11269-021-03006-y

Zare, M., & Koch, M. (2018). Groundwater level fluctuations simulation and prediction by ANFIS- and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: Application to the Miandarband plain. *Journal of Hydro-Environment Research*, *18*, 63–76. https://doi.org/10.1016/J.JHER.2017.11.004