



HAL
open science

Groundwater level reconstruction using long-term climate reanalysis data and deep neural networks

Sivarama Krishna Reddy Krishna Reddy Chidepudi, Nicolas Massei, Abderrahim Jardani, Abel Henriot

► **To cite this version:**

Sivarama Krishna Reddy Krishna Reddy Chidepudi, Nicolas Massei, Abderrahim Jardani, Abel Henriot. Groundwater level reconstruction using long-term climate reanalysis data and deep neural networks. *Journal of Hydrology: Regional Studies*, 2024, 51, pp.101632. 10.1016/j.ejrh.2023.101632 . hal-04364434

HAL Id: hal-04364434

<https://hal.science/hal-04364434>

Submitted on 27 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

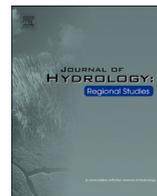
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Groundwater level reconstruction using long-term climate reanalysis data and deep neural networks

Sivarama Krishna Reddy Chidepudi^{a,b,*}, Nicolas Massei^a, Abderrahim Jardani^a, Abel Henriot^b

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

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

ARTICLE INFO

Keywords:

Reconstruction
BiLSTM
Groundwater
Reanalysis
Deep learning

ABSTRACT

Study region: Northern Metropolitan France.

Study focus: Assessing long-term changes in groundwater is crucial for understanding the impacts of climate change on aquifers and for managing water resources. However, long-term groundwater level (GWL) records are often scarce, limiting the understanding of historical trends and variability. In this paper, we present a deep learning approach to reconstruct GWLs up to several decades back in time using recurrent-based neural networks with wavelet pre-processing and climate reanalysis data as inputs. GWLs are reconstructed using two different reanalysis datasets with distinct spatial resolutions (ERA5: 0.25° x 0.25° & ERA20C: 1° x 1°) and monthly time resolution, and the performance of the simulations were evaluated.

New insights: Long term GWL timeseries are now available for northern France, corresponding to extended versions of observational timeseries back to early 20th century. All three types of piezometric behaviours could be reconstructed reliably and consistently capture the multi-decadal variability even at coarser resolutions, which is crucial for understanding long-term hydroclimatic trends and cycles. GWLs' multidecadal variability was consistent with the Atlantic multidecadal oscillation. From a synthetic experiment involving a modified long-term observational time series, we highlighted the need for longer training datasets for some low-frequency signals. Nevertheless, our study demonstrated the potential of using DL models together with reanalysis data to extend GWL observations and improve our understanding of groundwater variability and climate interactions.

1. Introduction

Understanding variations in groundwater levels (GWLs) is crucial for water resource management, especially under changing climatic conditions. Recent droughts across Europe have led to new restrictions on water use (Toreti et al., 2023). Hydrological reconstruction can be a way forward to address the long-term statistical properties of climate variability (Massei et al., 2020) and provide a complementary approach to better assess the impacts of climate change on hydrosystems (Bonnet et al., 2017a, 2020; Devers et al., 2021; Dieppois et al., 2016).

Traditionally, reconstruction studies are carried out using simulation approaches that make use of long-term reanalysis products

* Corresponding author at: Univ Rouen Normandie, UNICAEN, CNRS, M2C UMR 6143, F-76000 Rouen, France.

E-mail address: sivaramakrishnareddy.chidepudi@univ-rouen.fr (S.K.R. Chidepudi).

<https://doi.org/10.1016/j.ejrh.2023.101632>

Received 17 August 2023; Received in revised form 14 December 2023; Accepted 16 December 2023

2214-5818/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

such as NOAA 20-CR and ERA20C. [Bonnet et al. \(2017b\)](#) used ERA 20C and NOAA 20CR for hydrological reconstruction in France using a hydrological model. [Bonnet et al. \(2020\)](#) used a land surface model to reconstruct river flows using SAFRAN and NOAA 20CR datasets to understand the influence of multi-decadal variability in high and low flows in the Seine basin. [Jackson et al. \(2016\)](#) used a lumped conceptual model (Aqumod) to reconstruct multi-decadal groundwater level time series, and [Ascott et al. \(2020\)](#) did the same for sub-Saharan Africa and showed that multi-decadal variability plays a role in GWLs. These approaches require downscaling and bias correction when using global datasets, which is a resource-intensive process: Deep learning (DL) can help to avoid this task. For example, [Hagen et al. \(2023\)](#) recently showed that using deep learning models with ERA20C atmospheric variables as input could be a complementary approach to reconstruct streamflow without having to perform downscaling and bias correction as intermediate steps.

However, DL in hydrology is mostly used in the context of forecasting or hindcasting, and less often in simulation. As we have already mentioned in ([Chidepudi et al., 2023](#)) and according to ([Beven and Young, 2013](#)), “simulation is defined as quantitative reproduction of system behaviour without reference to observed output; “forecasting” is defined as reproduction of system behaviour ahead of time with observed outputs up until the onset of forecast included”. In other words, forecasting uses past values of the target variable as input to the model, i.e. prior knowledge of the target variable is required, whereas simulation depends only on the external variables used as input to the model. In this context, ‘reconstruction’ corresponds to simulation. The paper by ([Beven and Young, 2013](#)) provides detailed information as well as a thorough explanation of all these terms, which are theoretically well defined, not always well used and often used interchangeably, which should be avoided.

Our review of the literature showed that only a few studies dealt with hydrological reconstruction (i.e. simulation) highlighting the need for further research in this area. Given the immense potential of using DL to reconstruct hydrological variables, some studies have demonstrated the usefulness of these approaches. For example, [Uz et al. \(2022\)](#) used deep learning to reconstruct GRACE total water storage anomalies (TWSA) using reanalysis data (ERA5), and [Satish Kumar et al. \(2023\)](#) used artificial neural networks (ANN) to do the

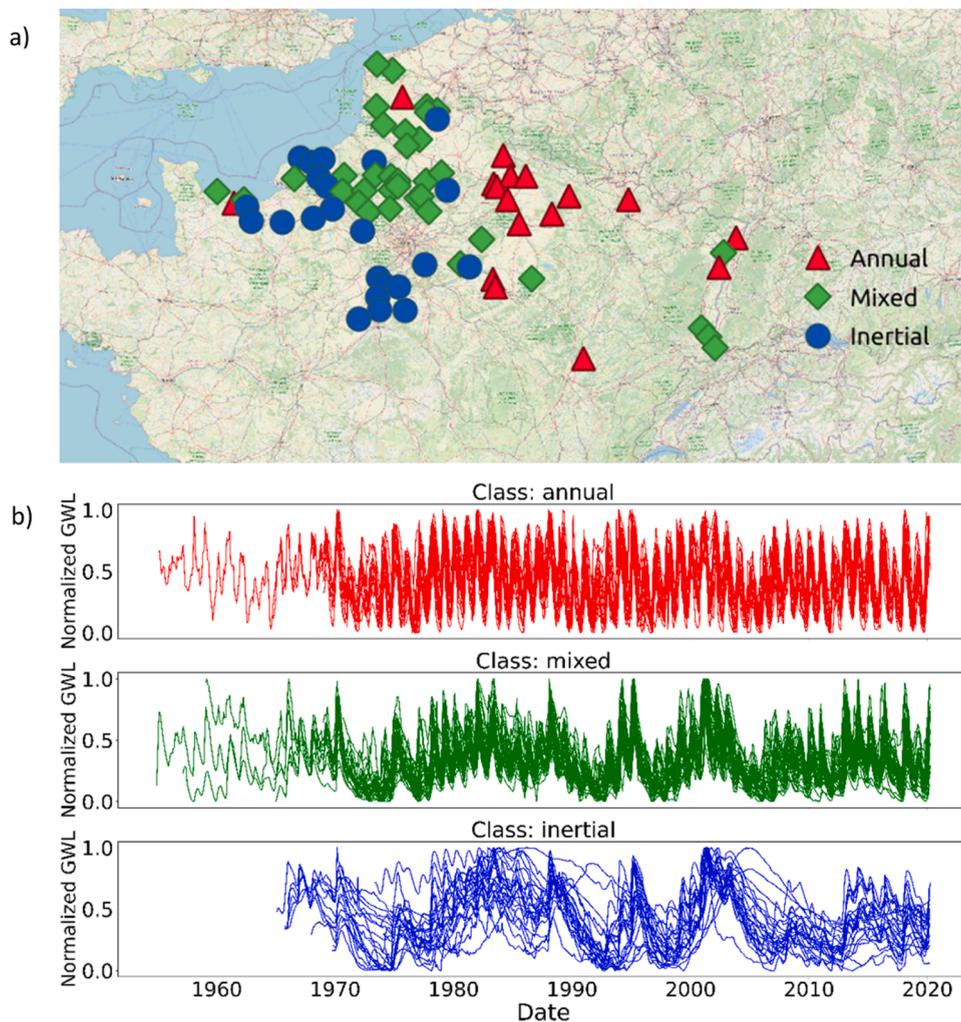


Fig. 1. Classification of long-term (1970–2020) groundwater level (GWL) monitoring stations based on GWL types: annual (red), mixed (green), and inertial (blue). Figure shows GWL station locations (a) and normalised GWLs of all selected stations based on class (b).

same. Kalu et al. (2022a) proposed a DL approach to reconstruct climate-driven terrestrial water storage (TWS) using global indices and hydroclimatological datasets to assess the influence of global climate on hydrological fluxes. Jing et al. (2023) showed that DL models outperformed tree-based models in simulating groundwater levels from GRACE data, even when influenced by human activities. Xiong et al. (2022) used random forest (RF), support vector machine (SVM), and ANN to reconstruct monthly GWLs from GRACE data, but such data are not long enough to be suitable for long-term reconstruction. Wunsch et al. (2022) used convolutional neural networks which allowed them to develop projections ('what-if' simulation according to (Beven and Young, 2013)) and investigate the influence of climate change on groundwater resources in the 21st century using only direct meteorological inputs. Vu et al. (2021) used groundwater levels from nearby stations to fill in missing values at piezometric stations, but this type of approach requires long-term GWL observations to be available in the vicinity of the target station. It is then necessary to explore the use of external input variables to enable long term reconstruction in the context of deep learning models. In recent years, DL models have also shown some potential in reconstructing missing GWL data (Vu et al., 2021) and forecasting surface and groundwater levels (Jahangir et al., 2023; Kalu et al., 2022b; Vu et al., 2023), but there is still a lack of studies that have evaluated the capabilities of DL in reconstructing GWLs over longer periods, up to a century.

Finally, the use of DL for simulation in groundwater studies is still relatively new, but has shown great potential in simulating groundwater levels, and recently Chidepudi et al. (2023) used recurrent-based deep learning models (long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional LSTM (BiLSTM)) to simulate groundwater levels affected by low-frequency variability using meteorological input data from high-resolution SAFRAN reanalysis. This study also showed that the use of some wavelet pre-processing (the so-called BC-MODWT: Boundary corrected-maximum overlap discrete wavelet transform) would improve the simulations compared to standalone (i.e. without any pre-processing) deep learning models. As a follow-up to this study, in this work we have analysed the capabilities of recurrent-based neural networks to simulate and reconstruct different types of GWL dynamics using large-scale climate reanalysis data as input. To the best of our knowledge, this is the first study to exploit the potential of deep learning for long-term groundwater level reconstruction, especially when GWLs contain both low-frequency (up to decadal) and annual cycle timescales.

Several research questions and scientific locks need to be addressed. First, we examine the impact of the spatial resolution of the input reanalysis data, as we know that there are different types of reanalysis products that serve different purposes. For example, the coarse resolution ERA20C reanalysis is classically used to study hydroclimate variability over the entire 20th century, while the finer resolution ERA5 only goes back to 1940. We also assess the ability of such models to produce consistent reconstructions of the GWL that still contain meaningful or even crucial information about hydrological variability over more than several decades using much shorter lengths of observed data (less than three decades). Finally, we discuss the influence of the length of the training data on the performance of the models.

2. Data and methodology

2.1. Study area

The study area considered is shown in Fig. 1a and covers mainly the north of France and was selected on the basis of data availability. Indeed, the spatial distribution of GWL time series duration across France is characterised by a pronounced north-south divide, with the longest time series located in the Paris Basin and northeastern France, while time series are much shorter from the extreme west to the south of France. This striking north-south divide can be attributed to the delayed establishment of the piezometric network in central and southern France, which occurred in the 1990s. In northern France, on the other hand, the piezometric network was already in place in the 1970s.

In the area of northern France considered in the present study, most of the GWL time series consist of piezometers located in the Seno-Turonian Chalk of the Paris Basin, which is the most represented aquifer among the selected piezometers, and in the Jurassic and Eocene limestones at the edge of the Paris Basin, respectively at the eastern and southwestern edges of the basin. As described in the next section, a major advantage of this area is that it contains 3 clearly contrasting GWL time series behaviours, despite its more limited spatial coverage.

2.2. Datasets

In this study, both the existing GWL data available over contiguous France and external climate variables from ERA reanalysis products were used. The GWL data were classified into 3 different classes according to Baulon et al. (2022a), according to the characteristic time scales that define their respective variability. These 3 types correspond to annual, mixed and inertial, as shown in Fig. 1b. The first GWL time series in Fig. 1b has an annual behaviour, i.e. the signal is strongly influenced by the annual cycle; the second is the mixed GWL time series with both annual and interannual components, and finally the inertial GWL time series, which was dominated by the low-frequency component. Understanding the dynamics of these three main behaviours is crucial for water resources management, as they are also related to the physical properties of aquifers, such as recharge rate, permeability and storage, etc. Recent studies have highlighted the importance of low-frequency variability and how its amplitude plays a crucial role in terms of estimated trends or extremes (Baulon et al., 2022a, 2022b).

Of these three types, mixed behaviour is the most interesting, as it combines the dynamics of both annual and low frequency dominated behaviour. The only long observational record (over 1 century) of groundwater levels available in the database that we could use to assess the ability of our models to develop consistent backward extensions (i.e. reconstructions) of models that are of such

a mixed type.

The input variables were taken from ECMWF ERA 20C (Poli et al., 2016) and ERA 5 (Hersbach et al., 2020), as summarised in Table 1. ERA 20C and ERA 5, being the global datasets with comparatively better spatio-temporal resolution, become valuable inputs to the models. ERA 20C covers a significant period of historical records dating back to 1900.

The selected input variables 10 m zonal (W-E) U-wind component (u10), 10 m méridional (S-N) V-wind component (v10), 2 m temperature (t2m), evaporation (e), mean sea level pressure (msl), surface net solar radiation (ssr), total precipitation (tp), with spatial resolution of 0.25 degrees (ERA5) and 1° (ERA20C) and monthly time resolution. The input variables used in this study are commonly used forcing data for hydrological and land surface models as they are representative of atmospheric conditions and circulation, moisture fluxes and radiative forcing (Kratzert et al., 2023; Mishra et al., 2018). This would also allow future studies to make fair comparisons between reconstructions from data-driven models and other models (conceptual, hydrological, land surface models, etc.). Furthermore, the datasets used in this study are globally available and widely used for various hydroclimatological applications.

The use and potential of these global datasets in the context of groundwater simulations remains unexplored. The GWLs used in this study are from a database developed with wells that have not been strongly affected by human activities (water abstraction) and remain sensitive to climate variability (Baulon et al., 2022a). Using these datasets, we reconstructed groundwater levels using the period 1970–1996 for model training and 1997–2010 for testing. The reconstructed period was 1940–1970 for ERA5 and 1900–1970 for ERA 20C.

The wells used in this study were selected on the basis of data quality, with a minimum of continuous gaps or abnormal changes in the data, and with data available from 1976 onwards. The models were trained and tested on the same data source (ERA20C or ERA5), using standalone models and MODWT pre-processing.

2.3. Methodology

2.3.1. Recurrent based deep neural networks and wavelet pre-processing approach for the simulation of the GWL

In the current study, we trained recurrent-based deep learning models (long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional LSTM (BiLSTM)) designed to handle sequential data and capture long-term dependencies. It is not the purpose here to explain in detail how these neural network architectures work, and the reader is referred to the abundant literature on this topic (Kratzert et al., 2019; Vu et al., 2021; Wunsch et al., 2022) or to our previous work using these three types of deep neural network models (Chidepudi et al., 2023). In short, LSTM has one memory cell and three gates to regulate the flow of information: forget, input, and output. GRU, on the other hand, is a simplified version of LSTM with only two gates (reset and update), designed to improve the computational efficiency of LSTM (reduced number of parameters compared to LSTM). BiLSTM, on the other hand, trains two LSTM models that learn forward and backward states, allowing even better exploitation of the available data. It should also be noted that the BiGRU models, not used in this work, can also be used for the same objective (2-way exploration of available information and reduced number of parameters). All these models are known for their ability to capture the non-linear relationship between input and output variables, such as atmospheric or climatological inputs and, in our case, groundwater levels.

Bayesian optimisation was used to tune the hyperparameters for all models, using the same range of values as described in Table 1 of (Chidepudi et al., 2023). As a critical parameter for LSTM or GRU recurrent-based neural networks, the sequence length was set to 48 months, which was determined based on previous works on hydrological variability over northern France (Baulon et al., 2022a, 2022b; Massei et al., 2010, 2017). These works showed that low-frequency multiannual variability in precipitation, originating from large-scale climate variability, very often controls either streamflow or GWL variations. Although the amplitude of such low-frequency temporal patterns remains very small in precipitation and other climate signals, their strong amplitude in GWL variability underscores the need to provide recurrent networks with the ability to use information going back several years. While many climate models, such as ENSO or NAO would show significant variability on 2 to 8 year time scales, a sequence length of 48 months (i.e. the number of continuous time steps used in the form of a sequence to compute each point output) proved to be quite consistent to achieve the best results (Chidepudi et al., 2023).

In addition, maximal overlap discrete wavelet transform (MODWT) pre-processing was used to decompose the original input signal into multiple wavelet and scaling coefficients, representing the high-frequency and low-frequency components. The appropriate use of this pre-processing technique aims to improve the groundwater level simulations by capturing the low-frequency variability, as shown in (Chidepudi et al., 2023), following the best practices recommended by (Quilty and Adamowski, 2018). To maintain the uniformity of the results and to avoid losing a large amount of data due to boundary conditions, it was decided to always use the least asymmetric filter with a width of 8 (La8) for the wavelet models.

The results obtained from the different models with and without wavelet pre-processing were compared using different metrics in the form of cumulative distribution function (CDF) curves to understand the performance across all the stations considered in the

Table 1
Summary of data sources used in this study.

Type of data	Data source	Years	Temporal resolution	Spatial resolution
Reanalysis	ERA 20C (Poli et al., 2016)	1900-2010	Monthly	1.0*1.0
	ERA 5 (Hersbach et al., 2020)	1940-2022		0.25*0.25
Observed GWL data	ADES (Winckel et al., 2022)	1970-2020	Monthly	Station data
	Long GWL timeseries (Tincques)	1903-2020		

study. The study also used a multimodel approach for uncertainty quantification, with different models, each with different initialisations. In addition, the study considered the median to calculate metrics instead of the mean to be robust against outliers. The metrics used to evaluate the performance of the models were NRMSE (normalised root mean square error), MAE (mean absolute error) and KGE (Kling Gupta efficiency) (Gupta et al., 2009).

3. Results

We investigated how recurrent-based neural networks perform in reconstructing long-term (i.e. several decades back in time) time series, where only a few decades are available for both training and testing the models, typically ~ 25 and 15 years respectively. ERA5 and ERA20C, which differ in length and spatial resolution, were tested.

All models with and without MODWT were tested for all GWL stations. The results obtained are then plotted separately for each type of GWL. For a synthetic presentation, these results are presented as empirical cumulative distribution functions of Kling-Gupta efficiency (KGE) values for all the different model combinations for both ERA5 (in the first column) and ERA 20C (in the second column) (Fig. 2). In the current study, KGE is considered over NSE or other metrics as it is more comprehensive than NSE and combines

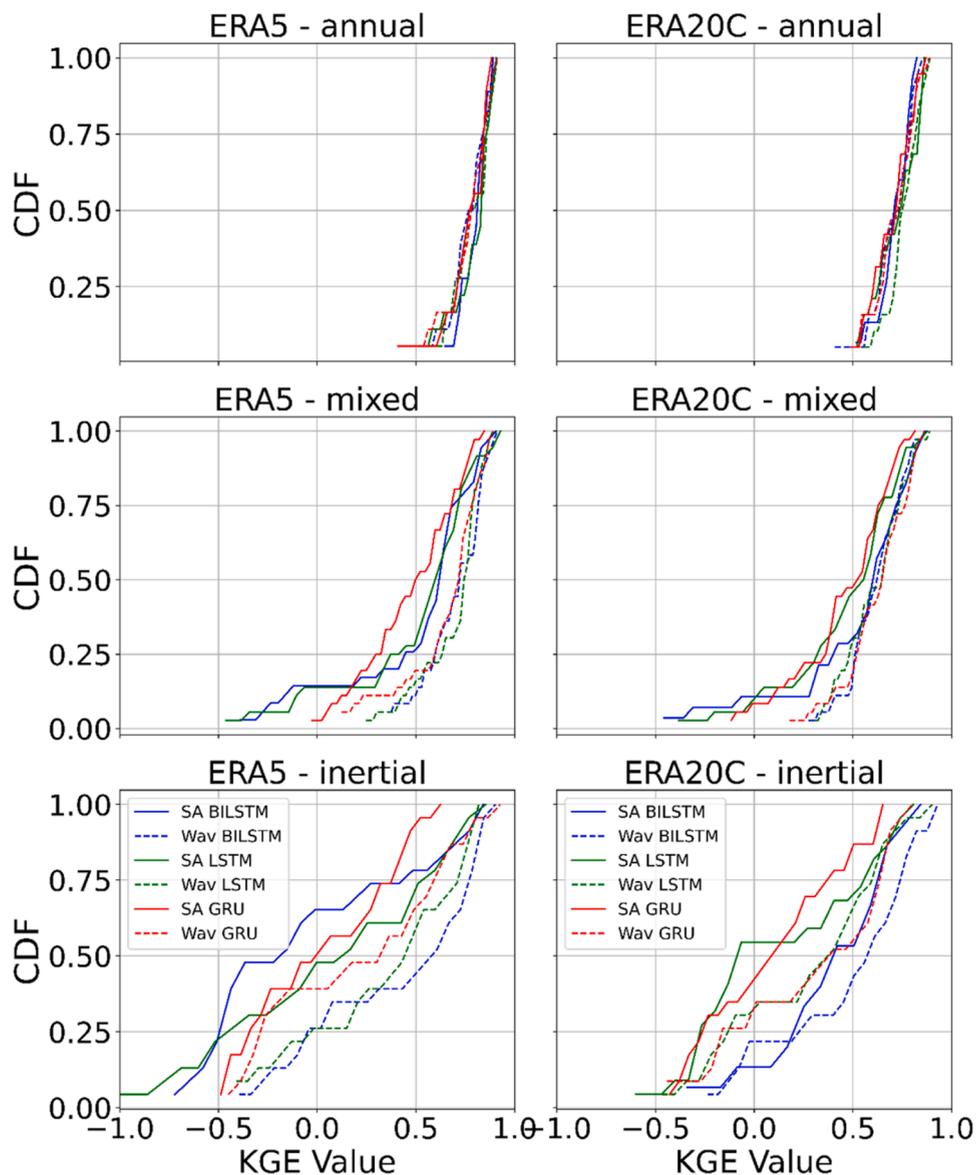


Fig. 2. Cumulative density functions of the KGE for standalone (SA) and wavelet-assisted (Wav) models for both ERA5 (first column) and ERA 20C (second column). Each model is represented by a different colour: BILSTM (blue), LSTM (green), and GRU (red) for standalone models, and corresponding dotted lines for wavelet models.

three components of model error (correlation, bias, & ratio of standard deviations).

The standalone models are marked ‘SA’, and the wavelet-assisted models are marked ‘Wav’. Each model is represented by a different colour: for example, the standalone models BiLSTM, LSTM and GRU are represented by sky blue, light green and violet respectively, while the wavelet models BiLSTM, LSTM and GRU are represented by orange, red and brown respectively.

It is clear from Fig. 2 that all types of architectures (LSTM, BiLSTM, GRU) often perform similarly overall, with only minor differences for annual GWLs with KGE greater than 0.5, while the differences are more pronounced for mixed and inertial GWL types. Second, the wavelet-assisted models (henceforth referred to as Wav models in the text) systematically outperformed the standalone models in all cases, confirming the findings of (Chidepudi et al., 2023). Indeed, it is easy to see from Fig. 2 that Wav LSTM/BiLSTM/GRU always had higher KGE values. Fig. 2 also shows that BiLSTM with wavelet pre-processing would slightly outperform the other models in many cases (orange lines in Fig. 2), suggesting that this approach is effective in capturing long-term temporal patterns. We then preferred to use this model for further analysis and simulations. There did not appear to be any substantial difference in KGE values when comparing ERA5 and ERA20C, although it looks like the results are slightly better for ERA5, especially for annual and mixed GWLs; this is related to the effect of spatial resolution which is discussed further below. Nevertheless, the KGE values seem to be more inconsistent for the inertial type of GWLs. This will be investigated below to assess the reasons for this inconsistency.

While the main objective of this study is to develop suitable models for reconstructing the GWL as far back as possible, another aim of the study was to assess the impact of the spatial resolution of the reanalysis data (i.e. ERA5 vs. ERA20C) on the GWL simulations. To better understand the difference between the ERA5 and ERA20C reconstructions, the performance of the Wav-BiLSTM model is compared in both cases, as shown in Fig. 3. ERA5 gave slightly better test results than ERA20C in the case of annual and mixed GWL types, while very similar results were observed in the case of inertial GWL types. Although there are no studies comparing ERA5 and ERA20C in groundwater simulations, a recent work dedicated to streamflow reconstruction (Hagen et al., 2023) showed that the scale issues are a softer constraint for machine learning than in traditional hydrological modelling. In addition, climate impact studies using global climate outputs may require the use of coarse resolution reanalysis datasets and longer periods of climate records. This makes it imperative to use ERA20C until sufficient long-term fine-resolution data become available.

Furthermore, validating the model performance of long available GWL observations using the long-term climate reanalysis data can provide a basis for identifying the robust models for reconstructing GWL. While there are quite a few studies on runoff reconstruction from reanalysis datasets (Ghiggi et al., 2019; Hagen et al., 2023; Nasreen et al., 2022), there are only a few studies on groundwater levels.

Using the only long, consistent and reliable ‘Tincques’ time series available over France (Tincques station), we then aimed to validate the approach for long-term reconstruction. Fig. 4a–4c shows the CDF plots of simulations versus observations over the period (1996–2010) for all three wavelet-based models. From Fig. 4 the GWLs are slightly better captured in the case of BiLSTM, although low levels are slightly underestimated, and Fig. 4d shows the reconstruction of the Tincques long-term observations in the form of a time series plot using wav-BiLSTM. In addition, the reconstruction results in this case are further validated on the historical period before 1970 when the long-term observations are available, further confirming the performance of the model in producing reliable reconstructions into the past. The results on the test set (KGE:0.68, MAE:1.77 & NRMSE:0.15) are in this case comparable to those on the reconstruction set (KGE:0.69, MAE:1.77 & NRMSE:0.16).

As demonstrated above, our reconstruction approach based on training with ~20–25 years of observed GWL revealed the capability of achieving a quite good reconstruction over 60 years back in time (i.e. between 1910 and 1970). The only available observed time series (Tincques station) dating back to 1903 was used to assess the quality of the extension in the past. Fortunately, the Tincques time series is of the so-called mixed type, which means that it was a perfect candidate for evaluating the reliability of the developed models in dealing with both annual and low-frequency variability. On the other hand, for time series of the inertial type (i.e. dominated only by multiannual to decadal variability), in some cases no satisfactory results could be obtained, which need to be analysed and discussed. It should also be noted that systemic changes or changes due to human influence are not part of the current study.

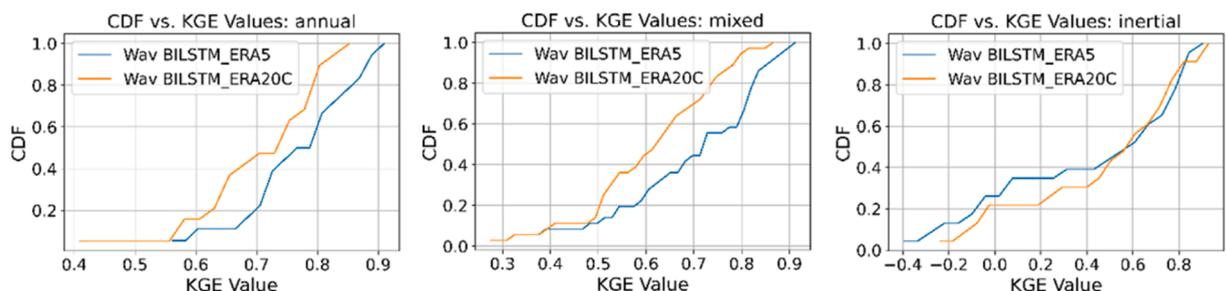


Fig. 3. CDF comparison of KGE values of the BiLSTM with wavelet (La8) for different GWL types.

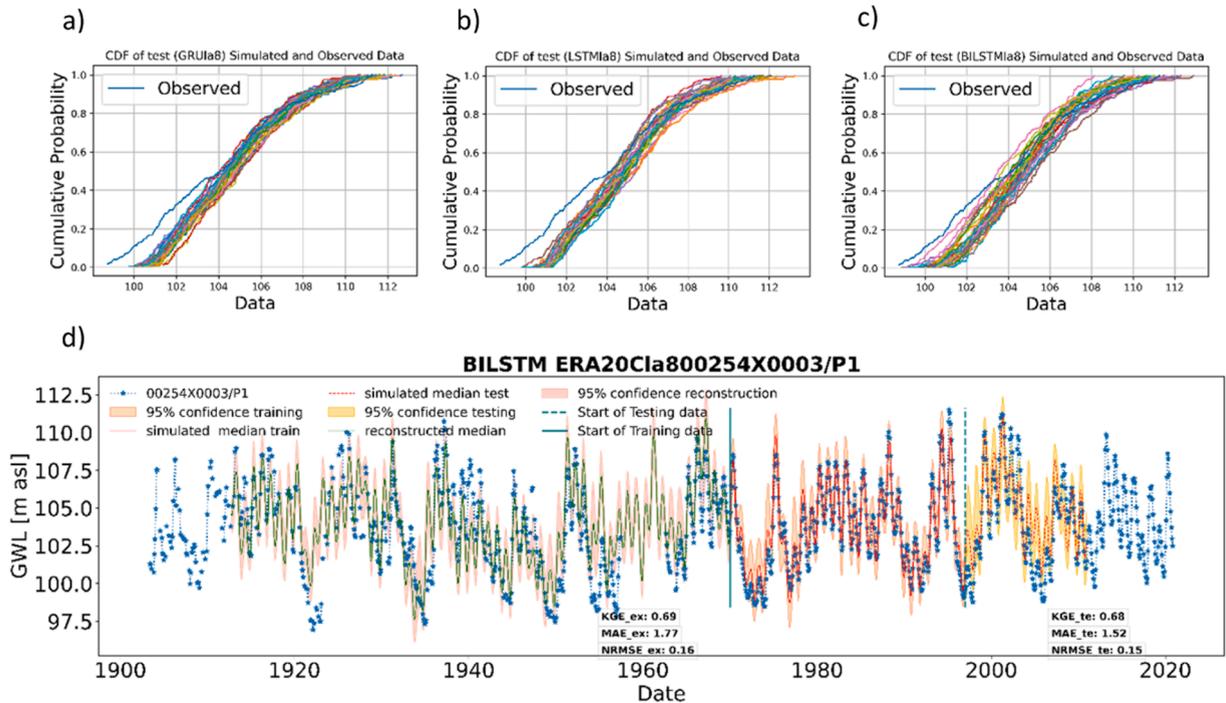


Fig. 4. Cumulative density functions (CDF) of simulated vs observed (skyblue) Tincques on the test set(a-c). Reconstruction of the longest GWL level time series available for double validation (d).

4. Discussion and conclusion

4.1. Identification of multidecadal variability in GWL

We then generated long-term reconstructions for all the short-term time series in the database, which are now available for any further investigation. Three examples of such reconstructions, one for each type of piezometric behaviour, are shown in Fig. 5. Despite recent studies in defence of metrics, it is challenging to rely entirely on metrics, as different metrics can lead to different conclusions, as shown below in Fig. 5a - 5c, which compare the comparison of the best reconstructed groundwater levels in each of the three main GWL types, and in the Appendix, Fig. 8, which compares reconstructions with low KGE values.

Different objectives can be pursued that require long time series. From a hydroclimatological point of view, assessing trends or changes in extremes over the long term is of great interest. However, as emphasised in the introduction of this article, a particular challenge remains to identify the contribution of multidecadal variability in hydroclimatic time series. As highlighted in several previous studies (Boé and Habets, 2014; Bonnet et al., 2017a), multidecadal variability due to natural climate variability may either mask, reduce or amplify the effects of anthropogenic climate change. This can affect interpretations or conclusions drawn in retrospective studies and increase uncertainty in projections or long-term forecasts. Using the same long-term Tincques station as in our study, Baulon et al. (2022b) showed that the Atlantic Multidecadal Variability (AMV), also known as the Atlantic Multidecadal Oscillation (AMO), would play a role in modulating the occurrence of groundwater extremes over the past century. We then investigated how well our reconstructions would capture such multidecadal variability for all three types of GWL time series.

To extract such a low-frequency oscillation, all reconstructed and Tincques observed time series were smoothed with a LOESS smoothing filter. Fig. 6 shows the comparison of such smoothed time series for all reconstructions (all types combined, grey) and the long-term observed data (Tincques, black line). For comparison and presentation purposes, all time series have been normalised (which obviously gives an exaggerated perception of the true amplitude of this oscillation in the time series). All reconstructed and observed time series show the same long-term multidecadal pattern. It is interesting to note that when only shorter time series are available (e.g. from 1970 onwards), such multidecadal variability can easily be mistaken for a trend. This multidecadal pattern was compared with the AMO index, confirming a potential influence of the AMO on groundwater levels as suggested by Baulon et al. (2022a, 2022b) or Bonnet et al. (2020).

The main limitation of the proposed approach identified concerned GWLs time series that were completely dominated by very low frequency variability. In some of these cases, the deep learning models did not achieve satisfactory results, even when assisted by wavelet pre-processing.

Although our results seemed quite consistent overall, we had identified some difficulties for the models to achieve good simulations for the most low-frequency GWL time series. For instance, a substantial number of low KGE values were obtained for inertial-type

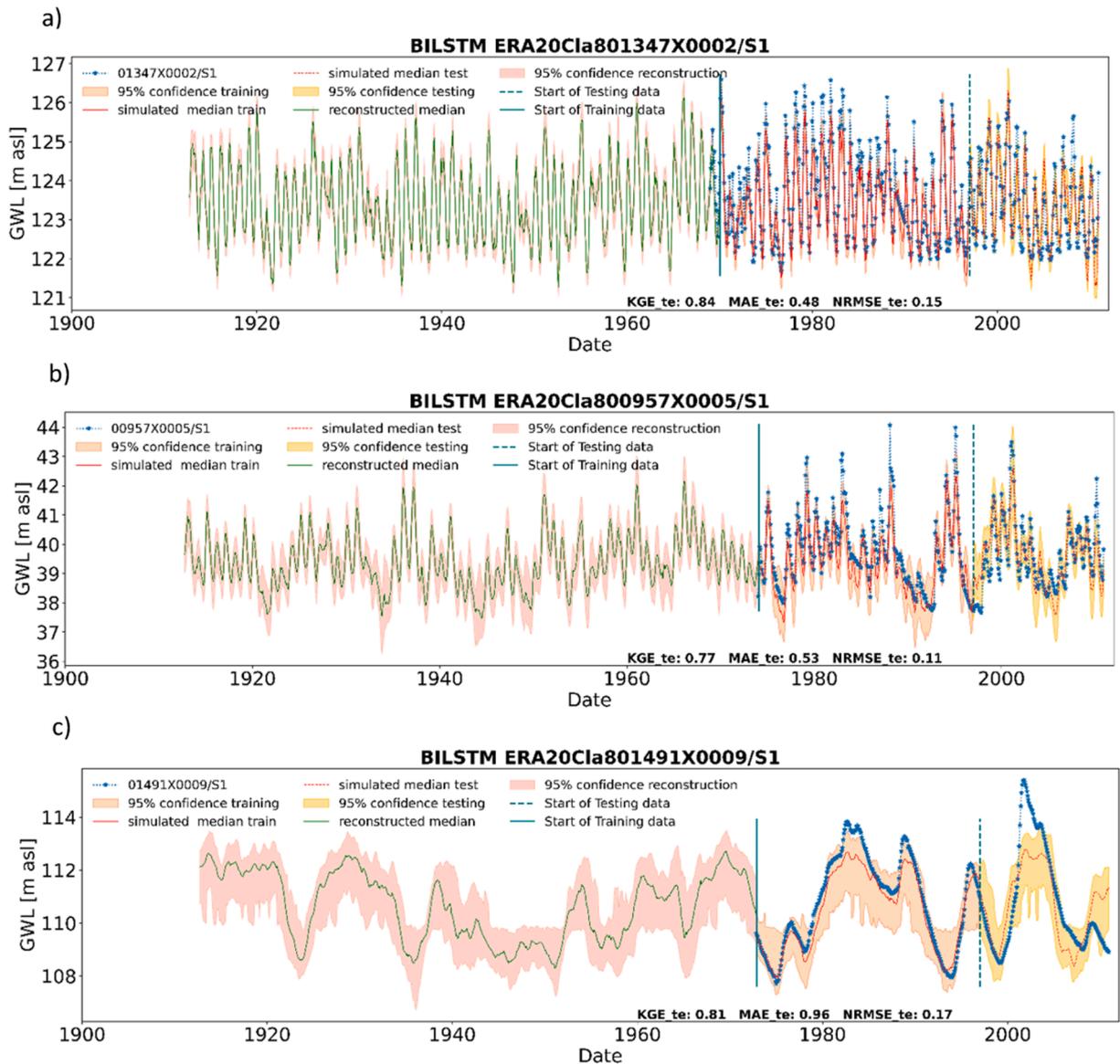


Fig. 5. Comparison of reconstructed annual (a), mixed (b) and inertial (c) groundwater level (GWL) variations with best Kling-Gupta efficiency (KGE) values.

GWLs (Figs. 2 and 3). The inconsistency of the reconstruction for the highly inertial type of GWLs, as shown in Fig. 7a, could be due to not learning enough relevant data patterns in the data. Given the lack of data available in this case, it is not surprising that they did not work: the information we get for this case would not be enough to achieve acceptable reconstructions. To illustrate this hypothesis, we carried out further experiments with synthetic time series that would replicate the most inertial time series by applying LOESS smoothing to the available Tincques long time series. The observational Tincques time series were smoothed until a highly inertial behaviour was observed, that would be comparable to the most inertial observed (blue line in Fig. 7b and c). Fig. 7b shows the reconstruction obtained with the same training data as the high inertial time series which, as expected, did not give good results. However, when the training data was increased by a further 20 years, the results improved significantly: for example, the KGE values would increase from -0.30 to 0.61 (Fig. 7c). This confirms the need for longer training data to achieve a better reconstruction in this case. While it is not always practical or possible to increase the amount of training data over time, another alternative could be to train on data from multiple stations, leading to regional models: in this way, much more information would be available for the model to learn such low-frequency variability. Our preliminary experiments with regional models showed improved simulations for this type of GWL. However, this approach would require further analysis for generalisation and is therefore beyond the scope of the current study.

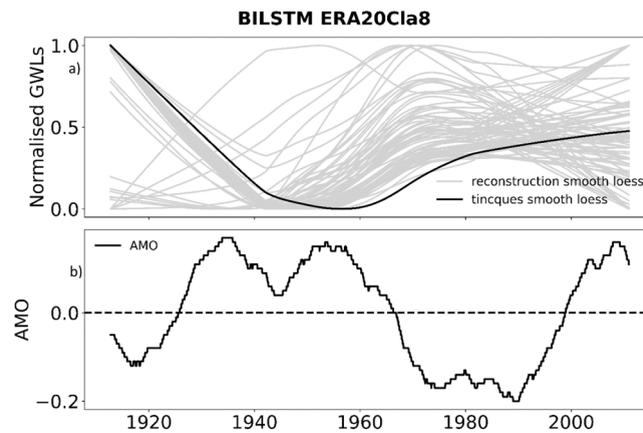


Fig. 6. Comparison of loess of smoothing of reconstructed median (light grey) from all stations with smoothing of long-term series available(a) and the AMO (b).

4.2. Concluding remarks

This study demonstrated the potential of deep learning models to reconstruct groundwater levels (GWLs) from reanalysis data with and without wavelet pre-processing. The DL models used input variables from two different reanalysis datasets (ERA20C and ERA5) separately. The performance of these models was compared for three different types of groundwater level behaviour (annual, mixed, and inertial). To further validate the approach for long-term reconstructions, the only long, consistent, and reliable time series available in France was used. This validation on the historical period before 1970, for which observations are available, resulted in results comparable to those of the original test set after 1996, confirming the reliability of the models in providing reliable reconstructions into the past. The ability of the models to capture multi-decadal variability was also evaluated for all reconstructions of groundwater levels in northern France. Overall, the models presented in this study proved capable of significantly extending the temporal information to produce reliable reconstructions. For example, these reconstructions were able to capture a meaningful trend that corresponded to multidecadal variability originating from the large-scale climate.

Both ERA5 and ERA20C would be useful in different contexts because while ERA5 has a high resolution and leads to a slightly better reconstruction, these data are currently only available up to 1940, while ERA20C is available up to 1900, but ends in 2010. Therefore, ERA5 could be a good alternative for more recent periods in the future. While wavelet-assisted models outperformed standalone models in all cases, wav-BiLSTM consistently outperformed other wavelet models. Multi-decadal variability was found in all wav-BiLSTM reconstructions. Inertial GWLs have a high variance in uncertainty compared to annual and mixed. An experiment with an artificial long-term observational time series suggested that this is logically due to the amount of data available for training that would be required to capture enough of the information characterising GWL variability. We expect that highly inertial types of GWL would require longer training data for the models to achieve acceptable results, and that training the models at multiple stations simultaneously may help to overcome this issue. Our preliminary results with the multi-station approach showed that there is an improvement in the case of the inertial type of groundwater levels, even when very low frequency variability would dominate the GWL signals.

Although the results presented here are only reconstructions, their effectiveness for this purpose would lead us to expect that this type of model would be useful for generating projections into the next few decades, as the similar types of variables are also available in GCM outputs. This would open the possibility of assessing future changes in groundwater resources in relation to climate change. In this context, the ability of deep learning models to cope with biases in GCM outputs without requiring bias-correction techniques as a prerequisite would also be an interesting research topic to address. Future research can also focus on developing even more frugal AI to reduce model complexity, testing the proposed models for minimal layers and other simple loss functions, and using large spatial grids to capture climate regimes or identifying relevant grids using correlation or more sophisticated approaches (Massei et al., 2017).

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-conceptualization of the original draft- review & editing; Project administration. **Abderrahim Jardani:** Supervision; review & editing; Project administration. **Abel Henriot:** Supervision; review & editing; Project administration.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sivarama Krishna Reddy Chidepudi reports financial support was provided by Normandy Region and BRGM. We acknowledge the HPC resources provided by CRIANN to carry out the experiments carried out as part of our ongoing project.

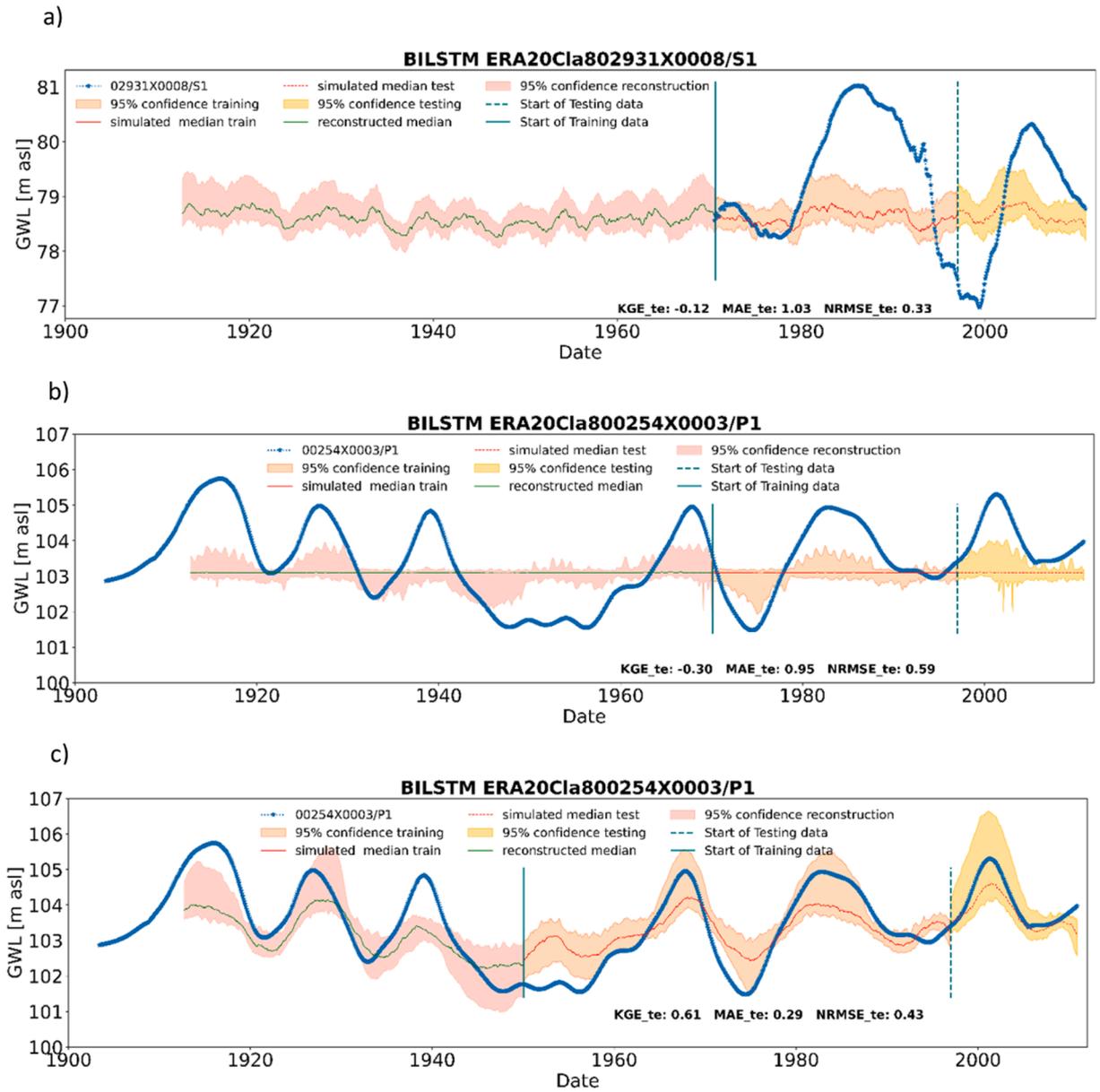


Fig. 7. Reconstruction of high inertial type GWLs time series (a) and reconstruction of Tincques based highly inertial time series (obtained by LOESS smoothing of observed Tincques GWL) with training date starting in 1970 and 26 years of training (b) and 1950 (c).

Data availability

The authors do not have permission to share data.

Appendix

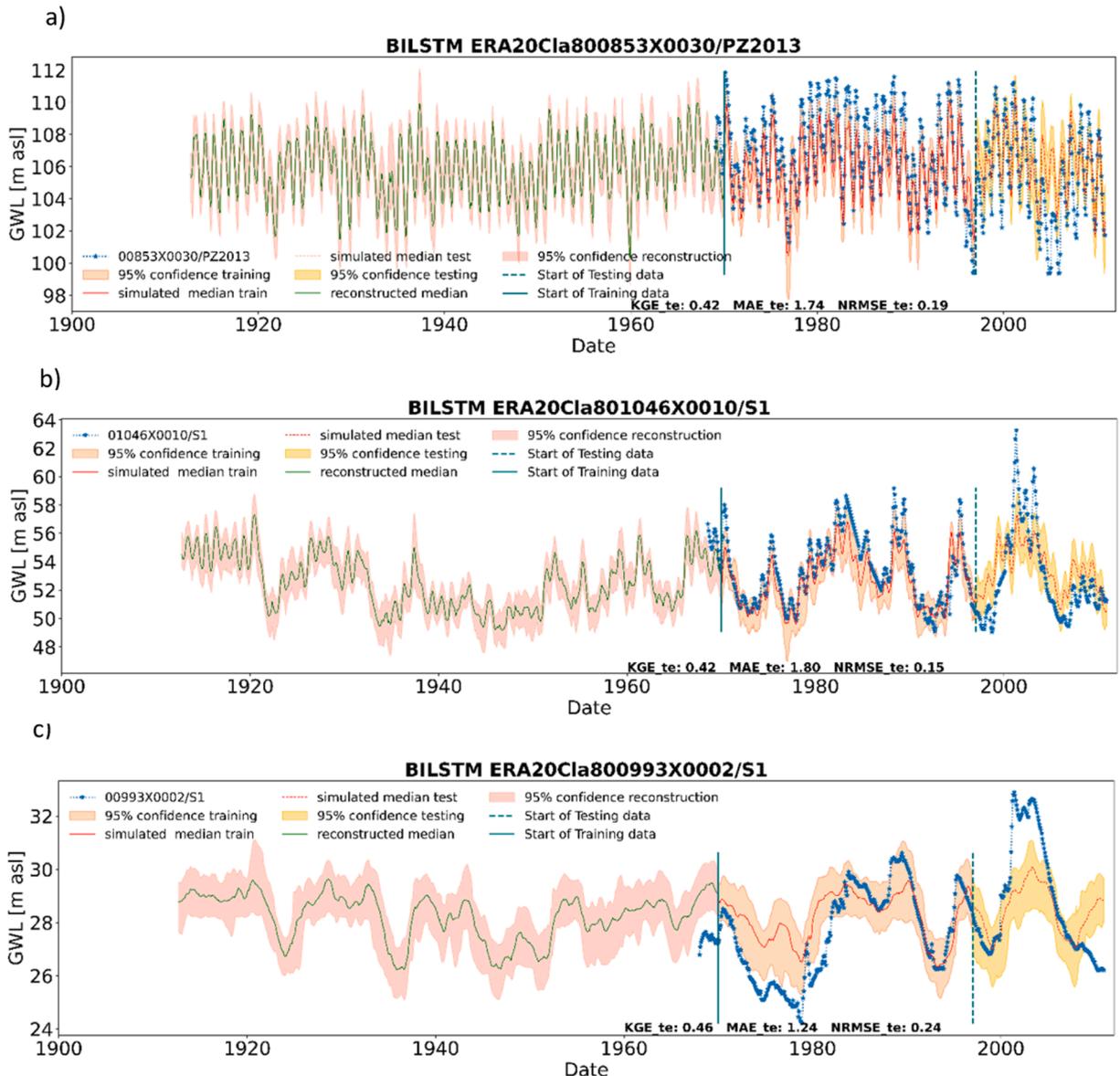


Fig. 8. Comparison of reconstruction when KGE values are in the range of 0.4–0.5 for all three types (a-c).

References

- Ascott, M.J., Macdonald, D.M.J., Black, E., Verhoef, A., Nakouh, P., Tirogo, J., Sandwidi, W.J.P., Bliefernicht, J., Sorensen, J.P.R., Bossa, A.Y., 2020. In situ observations and lumped parameter model reconstructions reveal intra-annual to multidecadal variability in groundwater levels in Sub-Saharan Africa. *Water Resour. Res.* 56 (12), e2020WR028056. <https://doi.org/10.1029/2020WR028056>.
- Baulon, L., Allier, D., Massei, N., Bessiere, H., Fournier, M., Bault, V., 2022. Influence of low-frequency variability on groundwater level trends. *J. Hydrol.* 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. *Hydrol. Earth Syst. Sci.* 26 (11), 2829–2854. <https://doi.org/10.5194/hess-26-2829-2022>.
- Beven, K., Young, P., 2013. A guide to good practice in modeling semantics for authors and referees. *Water Resour. Res.* 49 (8), 5092–5098. <https://doi.org/10.1002/wrcr.20393>.
- Boé, J., Habets, F., 2014. Multi-decadal river flow variations in France. *Hydrol. Earth Syst. Sci.* 18 (2), 691–708. <https://doi.org/10.5194/hess-18-691-2014>.
- Bonnet, R., Boé, J., Dayon, G., Martin, E., 2017a. Twentieth-century hydrometeorological reconstructions to study the multidecadal variations of the water cycle over France. *Water Resour. Res.* 53 (10), 8366–8382. <https://doi.org/10.1002/2017WR020596>.

- Bonnet, R., Boé, J., Dayon, G., Martin, E., 2017b. Twentieth-century hydrometeorological reconstructions to study the multidecadal variations of the water cycle over France. *Water Resour. Res.* 53 (10), 8366–8382. <https://doi.org/10.1002/2017WR020596>.
- Bonnet, R., Boé, J., Habets, F., 2020. Influence of multidecadal variability on high and low flows: the case of the Seine basin. *Hydrol. Earth Syst. Sci.* 24 (4), 1611–1631. <https://doi.org/10.5194/hess-24-1611-2020>.
- Chidepudi, S.K.R., Massei, N., Jardani, A., Henriot, A., Allier, D., Baulon, L., 2023. A wavelet-assisted deep learning approach for simulating groundwater levels affected by low-frequency variability. *Sci. Total Environ.* 865 (November 2022), 161035 <https://doi.org/10.1016/j.scitotenv.2022.161035>.
- Devers, A., Vidal, J.P., Lauvernet, C., Vannier, O., 2021. FYRE Climate: a high-resolution reanalysis of daily precipitation and temperature in France from 1871 to 2012. *Climate* 17 (5), 1857–1879. <https://doi.org/10.5194/cp-17-1857-2021>.
- Dieppois, B., Lawler, D.M., Slonosky, V., Massei, N., Bigot, S., Fournier, M., Durand, A., 2016. Multidecadal climate variability over northern France during the past 500 years and its relation to large-scale atmospheric circulation. *Int. J. Climatol.* 36 (15), 4679–4696. <https://doi.org/10.1002/JOC.4660>.
- Ghiggi, G., Humphrey, V., Seneviratne, S.I., Gudmundsson, L., 2019. GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth Syst. Sci. Data* 11 (4), 1655–1674. <https://doi.org/10.5194/essd-11-1655-2019>.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* 377 (1–2), 80–91. <https://doi.org/10.1016/J.JHYDROL.2009.08.003>.
- Hagen, J.S., Hasibi, R., Leblos, E., Lawrence, D., Sorteberg, A., 2023. Reconstructing daily streamflow and floods from large-scale atmospheric variables with feed-forward and recurrent neural networks in high latitude climates. *Hydrol. Sci. J.* <https://doi.org/10.1080/02626667.2023.2165927>.
- 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. *Q. J. R. Meteorol. Soc.* 146 (730), 1999–2049. <https://doi.org/10.1002/qj.3803>.
- Jackson, C.R., Wang, L., Pachocka, M., Mackay, J.D., Bloomfield, J.P., 2016. Reconstruction of multi-decadal groundwater level time-series using a lumped conceptual model. *Hydrol. Process.* 30 (18), 3107–3125. <https://doi.org/10.1002/HYP.10850>.
- Jahangir, M.S., You, J., Quilty, J., 2023. A quantile-based encoder-decoder framework for multi-step ahead runoff forecasting. *J. Hydrol.* 619, 129269 <https://doi.org/10.1016/J.JHYDROL.2023.129269>.
- Jing, H., He, X., Tian, Y., Lancia, M., Cao, G., Crivellari, A., Guo, Z., Zheng, C., 2023. Comparison and interpretation of data-driven models for simulating site-specific human-impacted groundwater dynamics in the North China Plain. *J. Hydrol.* 616, 128751 <https://doi.org/10.1016/j.jhydrol.2022.128751>.
- Kalu, I., Ndehedehe, C.E., Okwuashi, O., Eyoh, A.E., Ferreira, V.G., 2022a. An assimilated deep learning approach to identify the influence of global climate on hydrological fluxes. *J. Hydrol.* 614, 128498 <https://doi.org/10.1016/J.JHYDROL.2022.128498>.
- Kalu, I., Ndehedehe, C.E., Okwuashi, O., Eyoh, A.E., Ferreira, V.G., 2022b. A new modelling framework to assess changes in groundwater level. *J. Hydrol.: Reg. Stud.* 43, 101185 <https://doi.org/10.1016/J.EJRH.2022.101185>.
- 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. *Hydrol. Earth Syst. Sci.* 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. *Sci. Data* 10 (1), 1–11. <https://doi.org/10.1038/s41597-023-01975-w>.
- Massei, N., Laignel, B., Deloffre, J., Mesquita, J., Motelay, A., Lafite, R., Durand, A., 2010. Long-term hydrological changes of the Seine River flow (France) and their relation to the North Atlantic Oscillation over the period 1950–2008. *Int. J. Climatol.* 30 (14), 2146–2154. <https://doi.org/10.1002/joc.2022>.
- Massei, N., Dieppois, B., Hannah, D.M., Lavers, D.A., Fossa, M., Laignel, B., Debret, M., 2017. Multi-time-scale hydroclimate dynamics of a regional watershed and links to large-scale atmospheric circulation: application to the Seine river catchment, France. *J. Hydrol.* 546, 262–275. <https://doi.org/10.1016/j.jhydrol.2017.01.008>.
- 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. *Proc. Int. Assoc. Hydrol. Sci.* 383, 141–149. <https://doi.org/10.5194/piahs-383-141-2020>.
- Mishra, V., Shah, R., Azhar, S., Modi, P., Kumar, R., 2018. Reconstruction of droughts in India using multiple land-surface models (1951–2015). *Hydrol. Earth Syst. Sci.* 22, 2269–2284. <https://doi.org/10.5194/hess-22-2269-2018>.
- Nasreen, S., Součková, M., Vargas Godoy, M.R., Singh, U., Markonis, Y., Kumar, R., Rakovec, O., Hanel, M., 2022. A 500-year annual runoff reconstruction for 14 selected European catchments. *Earth Syst. Sci. Data* 14 (9), 4035–4056. <https://doi.org/10.5194/essd-14-4035-2022>.
- Poli, P., Hersbach, H., Dee, D.P., Berrisford, P., Simmons, A.J., Vitart, F., Laloyaux, P., Tan, D.G.H., Peubey, C., Thépaut, J.N., Trémolet, Y., Hólm, E.V., Bonavita, M., Isaksen, I., Fisher, M., 2016. ERA-20C: an atmospheric reanalysis of the twentieth century. *J. Clim.* 29 (11), 4083–4097. <https://doi.org/10.1175/JCLI-D-15-0556.1>.
- Quilty, J., Adamowski, J., 2018. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *J. Hydrol.* 563 (February), 336–353. <https://doi.org/10.1016/j.jhydrol.2018.05.003>.
- Satish Kumar, K., AnandRaj, P., Sreelatha, K., Sridhar, V., 2023. Reconstruction of GRACE terrestrial water storage anomalies using Multi-Layer Perceptrons for South Indian River basins. *Sci. Total Environ.* 857 (February 2022), 159289 <https://doi.org/10.1016/j.scitotenv.2022.159289>.
- Toreti, A., Bavera, D., Acosta Navarro, J., Arias-Muñoz, C., Avanzi, F., Marinho Ferreira Barbosa, P., De Jager, A., Di Ciollo, C., Ferraris, L., Fioravanti, G., Gabellani, S., Grimaldi, S., Hraest Essenfelder, A., Isabellon, M., Jonas, T., Maetens, W., Magni, D., Masante, D., Mazzeschi, M., McCormick, N., Meroni, M., Rossi, L., Salamon, P. and Spinoni, J., Drought in Europe March 2023, EUR 31448 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978–92-68–01068-6, doi: 10.2760/998985, JRC133025.
- Uz, M., Atman, K.G., Akyilmaz, O., Shum, C.K., Keleş, M., Ay, T., Tandogdu, B., Zhang, Y., Mercan, H., 2022. Bridging the gap between GRACE and GRACE-FO missions with deep learning aided water storage simulations. *Sci. Total Environ.* 830 <https://doi.org/10.1016/j.scitotenv.2022.154701>.
- 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. *J. Hydrol.* 597 (November 2020) <https://doi.org/10.1016/j.jhydrol.2020.125776>.
- 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. *Sci. Total Environ.* 880 (April) <https://doi.org/10.1016/j.scitotenv.2023.163338>.
- Winckel, A., Ollagnier, S., Gabillard, S., 2022. Managing groundwater resources using a national reference database: the French ADES concept. *SN Appl. Sci.* 4, 217. <https://doi.org/10.1007/s42452-022-05082-0>.
- Wunsch, A., Liesch, T., Broda, S., 2022. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. *Nat. Commun.* 13 (1), 1–13. <https://doi.org/10.1038/s41467-022-28770-2>.
- Xiong, J., Abhishek, Guo, S., Kinouchi, T., 2022. Leveraging machine learning methods to quantify 50 years of dwindling groundwater in India. *Sci. Total Environ.* 835 (March), 155474 <https://doi.org/10.1016/j.scitotenv.2022.155474>.