Deep learning for the simulation, reconstruction & projection of groundwater level variations

Sivaramakrishnareddy Chidepudi

Normandie Université

UNIVERSITÉ DE ROUEN NORMANDIE

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité **SCIENCES DE L'UNIVERS**

Préparée au sein de l'**Université de Rouen Normandie**

Deep learning for the simulation, reconstruction & projection of groundwater level variations

Présentée et soutenue par

SIVARAMAKRISHNAREDDY CHIDEPUDI

Thèse soutenue le 14/11/2024

devant le jury composé de :

M. NICOLAS MASSEI	Professeur des Universités - Université de Rouen Normandie	Directeur de thèse
M. ABDERRAHIM JARDANI	Professeur des Universités - Université de Rouen Normandie	Co-directeur de thèse
MME NATHALIE DORFLIGER	Chargé de Recherche - Danone	Membre du jury
M. ABEL HENRIOT	Chargé de Recherche - Bureau de Recherches Geologiques et Minieres	Membre du jury
M. JULIAN KOCH	Directeur de Recherche - GEUS	Membre du jury
MME TANJA LIESCH	Professeur Associé - Université de Karlsruhe (ALLEMAGNE)	Membre du jury
M. JEAN-RAYNALD DE DREUZY	Directeur de Recherche - UNIVERSITE RENNES 1	Rapporteur du jury
MME REHANA SHAIK	Professeur Associé - IIIT Hyderabad, India	Rapporteur du jury

Thèse dirigée par **NICOLAS MASSEI** (MORPHODYNAMIQUE CONTINENTALE ET COTIERE) et **ABDERRAHIM JARDANI** (Université de Rouen Normandie)

ED 497 NBISE
Normande de Biologie Intégrative, Santé, Environnement

M2C

Abstract

Accurate groundwater level (GWL) simulations facilitate reconstructions and projections for analysing historical and future groundwater trends and variability at the decadal scale. In this thesis, we investigate the use of deep learning (DL) approaches for GWL simulations, reconstructions, and projections, with a focus on capturing low-frequency variability and leveraging climate reanalysis and GCM model outputs. A wavelet-assisted DL framework was developed, using the Maximal Overlap Discrete Wavelet Transform (MODWT) as a pre-processing step to decompose input signals. We specifically evaluated advanced DL models, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM), for single-station and multi-station approaches. The single station approach results indicated that MODWT-assisted GRU models allowed for extracting low-frequency information and significantly outperformed standalone models in simulating GWLs, particularly for inertial-type GWL. The Shapley Additive Explanations (SHAP) technique was used to interpret model outputs and highlight important input features. For long-term GWL reconstructions, DL models were trained on ERA5 and ERA20C climate reanalysis datasets, enabling reconstructions up to 1900 and 1940, respectively. These DL-based models were able to capture multi-decadal variability in all reconstructed GWLs. Several multi-station training approaches and clustering were used for large-scale GWL simulations, incorporating dynamic climatic variables and static aquifer characteristics. Models specifically trained on different GWL types, clustered by spectral properties, performed significantly better than those trained on the whole dataset. Finally, A multi-station GRU model trained for each GWL type with boundary-corrected MODWT (BC-MODWT) pre-processing was used to generate projections until 2100. Future changes show decreasing trends in groundwater levels and variability, intensifying from SSP2-4.5 to SSP5-8.5, despite projected groundwater levels being higher on average compared to the historical period in all scenarios. We explain this seemingly counter-intuitive result by the fact that projected levels are systematically much higher at the beginning of the future period (up to ~2050) compared to the historical period. Finally, our results indicate that the variability of annual-type aquifers has increased for all emission scenarios.

Résumé

Des simulations précises du niveau des eaux souterraines (GWL) sont indispensables pour générer les reconstructions et les projections servant à analyser les tendances et la variabilité historiques et futures des eaux souterraines à long terme. Dans cette thèse, nous étudions l'utilisation d'approches d'apprentissage profond (DL) pour les simulations, reconstructions et projections du niveau des eaux souterraines, en mettant l'accent sur les questions liées à la représentation de la variabilité à basse fréquence interannuelle à décennale, et en utilisant divers produits de réanalyses climatiques et sorties de GCM. Une approche de pré-traitement par ondelettes assistant les modèles DL a été développée, en particulier à partir de transformée en ondelettes discrète à chevauchement maximal (MODWT) en une étape de décomposition les signaux d'entrée. Les modèles récurrents à mémoire long- et court- terme (LSTM) et leurs développements plus récents (unité récurrente à porte GRU et LSTM bidirectionnels BiLSTM) ont été plus spécifiquement utilisés et évalués, pour développer des approches d'apprentissage à station unique et à stations multiples. Les résultats de l'approche à station unique ont indiqué que les modèles GRU assistés par MODWT permettaient d'extraire des informations à basse fréquence et surpassaient considérablement les modèles « simples » (i.e. sans pré-traitement) dans la simulation des GWL, en particulier pour les GWL de type inertiel. La méthode SHAP a été utilisée pour appréhender l'interprétabilité des résultats des modèles et le fonctionnement des modèles eux-mêmes, mettant ainsi notamment en évidence les caractéristiques d'entrée les plus importantes. Pour les reconstructions GWL à long terme, les modèles DL ont été construits en utilisant les ensembles de données de réanalyse climatique ERA5 et ERA20C du centre européen de prévisions météorologiques à moyen terme (ECMWF), permettant des reconstructions jusqu'en 1940 et 1900, respectivement. Ces modèles basés ont pu capturer avec succès la variabilité multidécennale dans tous les niveaux de nappe reconstruits, un enjeu important en contexte de changement climatique dans la mesure où la variabilité multidécennale peut fortement interférer avec les effets du changement climatique. Plusieurs approches d'apprentissage multi-stations et de clustering ont été utilisées pour les simulations GWL à grande échelle, intégrant des variables climatiques dynamiques et des caractéristiques statiques des aquifères. Les modèles spécifiquement entraînés sur différents types de GWL, regroupés sur la base de leurs propriétés spectrales, ont obtenu des résultats significativement meilleurs que ceux entraînés sur l'ensemble des données. Enfin, un modèle GRU multi-stations entraîné pour chaque type de GWL avec un prétraitement MODWT avec correction des effets de bord (BC-MODWT) a été utilisé pour générer des projections jusqu'en 2100. Les changements

futurs indiquent des tendances à la baisse des niveaux et de la variabilité des eaux souterraines, s'intensifiant de SSP2-4.5 à SSP5-8.5, malgré des niveaux des eaux souterraines projetés plus élevés en moyenne par rapport à la période historique dans tous les scénarios. Nous expliquons ce résultat apparemment contre-intuitif par le fait que les niveaux projetés sont systématiquement bien plus élevés en début de période future (jusqu'à ~2050) par rapport à la période historique. Nos résultats indiquent enfin que la variabilité des aquifères de type annuel a augmenté pour tous les scénarios d'émission.

Abbreviations and Acronyms

AI - Artificial Intelligence

ANN - Artificial Neural Network.

GWL - Groundwater Level

DL - Deep Learning

ML – Machine Learning

DDMs - Data-driven models

RNN -Recurrent Neural networks

CNN -Convolutional Neural networks

GNN -Graph Neural Networks

NARX - Nonlinear Autoregressive network with eXogenous inputs

LSTM - Long Short-Term Memory

GRU - Gated Recurrent Unit

BiLSTM - Bidirectional Long Short-Term Memory

XGBOOST - Extreme Gradient Boosting

RF - Random Forest

SVM - Support Vector Machine

MLR - Multiple linear regression

CMIP6 - Coupled Model Intercomparison Project Phase 6

SSP - Shared Socioeconomic Pathway

ECMWF - European Centre for Medium-Range Weather Forecasts

ERA - ECMWF Re-Analysis

SAFRAN - Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige

BC-MODWT - Boundary Corrected Maximal Overlap Discrete Wavelet Transform

TWSA - Total Water Storage Anomalies

TWS - Terrestrial Water Storage

SHAP - SHapley Additive exPlanations

NLDAS - North American Land Data Assimilation System

KGE - Kling-Gupta Efficiency

MAE - Mean Absolute Error

RMSE - Root Mean Square Error

NRMSE – Normalised Root Mean Square Error

NSE - Nash-Sutcliffe efficiency

AMO - Atlantic Multidecadal Oscillation

NOAA - National Oceanic and Atmospheric Administration

ENSO - El Niño-Southern Oscillation

NAO - North Atlantic Oscillation

GRACE - Gravity Recovery and Climate Experiment

Table of Contents

List of Figures

List of Tables

Introduction

Groundwater resources are crucial in global water security, supporting ecosystems, agriculture, and human populations worldwide. In the context of climate change, understanding, simulating, and developing scenarios for the temporal evolution of variations in groundwater levels (GWLs) over large spatial and temporal scales is critical, particularly due to the complex nature of aquifer systems and their response to climatic variability. This necessitates the availability of appropriate modelling tools. While understanding and simulating long-term GWL variations is important, current modelling approaches face several limitations. Distributed modelling methods are difficult to implement on large spatial scales, and the scarcity of long-term piezometric data (exceeding fifty years) hinders comprehensive analysis (Barthel & Banzhaf, 2016). Several approaches are often used to model the hydrological responses of aquifers on these scales: conceptual modelling (Jackson et al., 2016; Ascott et al., 2020), physically-based modelling (Halloran et al., 2023; Vergnes et al., 2023) and statistical modelling (Coulibaly et al., 2001; Wunsch et al., 2021). At large scales, conceptual models are generally unsuitable, and the implementation of physically-based models, while still possible, becomes very complex due to the need for a large amount of geological information (Barthel & Banzhaf, 2016; Condon et al., 2021). On the scale of mainland France, the AquiFR project (Habets et al., 2015), a national hydrogeological modelling platform project that has been underway for several years, combines different physics-based and conceptual modelling approaches with this objective and clearly demonstrates the difficulties inherent in developing a relevant large-scale modelling tool for decision support. Data-driven models (DDMs) and, in particular, deep learning (DL) methods have been receiving increasing attention in the field of hydrology for several years (Fang et al., 2022; Klotz et al., 2022; Kratzert et al., 2018, 2019; Vu et al., 2021; Wunsch et al., 2021). They offer the possibility of tackling complex problems for highly non-linear cases in hydrology and provide potential solutions for hydrological time series classification or regression problems for simulation. However, most of these works remained focused on surface water, and only a few studies addressed the issue of long-term simulation of GWL, including reconstruction or projection. It then appears particularly interesting to analyse the capability of DL models to address this issue.

Aquifers act as low-pass filters, with long-term variabilities sometimes strongly controlling piezometric variations, as well as long-term changes in streamflow, in particular in large watersheds. Low-frequency multi-year to multi-decadal variability is inherited from large-scale climatic patterns controlled by atmospheric and ocean circulation (Baulon et al.,

2022a & b; Massei et al., 2010 and 2017; Rust et al., 2018; Neves et al., 2019; Liesch & Wunsch, 2019) and may significantly influence the estimation of piezometric trends and extremes as shown in Baulon et al. (2022a & b) or Rust et al. (2019). Such low-frequency variability is also known as a potentially significant contributor to uncertainties in future time hydrological projections, as it may, in certain circumstances, mask or amplify the effect of climate change, as underlined in Boé and Habets (2014). Although groundwater systems involve complex interactions between many variables like climate, geology, etc., DL models should be able to capture the subsequent nonlinear relationships and be typically good at extracting patterns at multiple spatio-temporal scales. Hence, such models need to simulate groundwater levels that can be dominated by any of the annual, infra-annual, multiannual, decadal variability or any mix of those.

Several scientific and technical gaps exist. Most previous studies exploring DDM and DL models predominantly focused on forecasting, with previous GWLs being used as input rather than simulation, relying heavily on autocorrelation in GWL values, which should be avoided. Further, there is a clear lack of studies looking at either forecasting or simulating GWL time series affected by low-frequency interannual to decadal variability with a significant amplitude. In this framework, the major challenge then consists of producing consistent long-term GWL simulations using only external input variables (i.e. not using previous data of the target variable) and limited observed target variable data (e.g. GWL time series that would be too short to display enough low-frequency "oscillations" for training the DL model). Different approaches must be developed to ensure the training datasets contain all the necessary information for the model to learn effectively. The most adequate architectures and approaches should then be capable of properly describing and simulating low-frequency variability and long-term trends in groundwater systems. Finally, such models should then permit to address one ultimate scientific question about the possible evolution of groundwater levels characterised not only by seasonal but also by low-frequency climate-induced variations following different climate change scenarios from different climate models.

This thesis explores how deep learning approaches can provide alternatives and complement traditional physical-based or conceptual modelling techniques to address these challenges. We then aimed to develop robust local to regional DL approaches to establish reconstructions and projections of different types of piezometric time series at the scale of northern France, leveraging large-scale hydrological and climatic information to address issues of trends and variability over the last century in the past and up to 2100 into the future. To summarise, we aimed to address the following main research questions:

1. How can we develop deep learning models that account for various types of groundwater level variations beyond annual cyclicities and are not designed only for forecasting for a limited number of time steps (and not based on previous values of target variables)?
2. How does spatial resolution of explanatory input climate variables from reanalysis datasets impact simulations or reconstruction? How accurate are these in capturing different variabilities?
3. Can such DL models consistently capture trends, long-term oscillations, and variability? What relevant strategies are needed to ensure this?
4. What does DL simulated GWL time series tell about how different types of aquifers (e.g. annual dominated or low-frequency dominated) may respond to climate change scenarios?

The work carried out is part of a wider framework aimed at understanding groundwater level variations over metropolitan France, their characteristics and origins and developing data-driven models suitable for simulating (reconstructing and projecting) their evolution on long time scales, past or future. The work presented here builds on previous works carried out in the thesis of Baulon (2023), which consisted of understanding the climate-forcing factors of groundwater level variability. For this purpose, Baulon (2023) curated a database of piezometric stations with minimal human influence (i.e., mainly pumping) and the longest available records. It eventually resulted in 76 stations which were mainly located in northern France.

The above research questions are addressed in the following five chapters as described below and in Figure 0.1:

In Chapter 1, we review the state-of-the-art DL techniques applied to groundwater studies, highlighting current challenges such as limited global datasets, complex geological influences, and sensitivity to human activities. It then also explores recent advances in DL, including multi-station training and transfer learning, to bolster model generalisation and address data scarcity.

In Chapter 2, we present a wavelet-assisted deep learning approach to model groundwater levels that allows for accurately taking into account the low-frequency variability affecting some time series. We evaluate and compare the respective performance of wavelet-assisted models and standalone DL models. This chapter has been published in the Science of The Total Environment Journal (https://doi.org/10.1016/j.scitotenv.2022.161035).

In Chapter 3, we explore the capability of the DL models developed for reconstructing groundwater levels (GWLs) in northern France during the past century, using long-term climate reanalysis data. We used DL models with and without wavelet pre-processing to simulate GWLs from different reanalysis products, namely ERA5 and ERA20C datasets. This chapter was published in the Journal of Hydrology: Regional Studies (https://doi.org/10.1016/j.ejrh.2023.1016).

Chapter 4 is dedicated to exploring the so-called "multi-station" approach, which consists of training DL models on multiple stations simultaneously to try to include as much information as possible in the training process compared to training using only one station. The chapter introduces different multi-station approaches incorporating dynamic climatic variables and static aquifer characteristics to account for groundwater variations across the study area (northern France). Both clustering and wavelet transform decomposition were tested to leverage regionalised information along with static attributes and a one-hot encoding approach. The content of this chapter has been submitted and is currently under review after acceptance with major revisions to Hydrology and Earth System Sciences (https://doi.org/10.5194/egusphere-2024-794).

Finally, in Chapter 5, we use a most suited deep learning approach to generate groundwater level projections for northern France under various climate change scenarios until 2100, comparing different state-of-the-art climate models. Specifically, we investigate the possible changes in GWL average and variability for different GWL types (i.e. more or less dominated by low frequency). This chapter corresponds to an article submitted to Earth's Future (https://doi.org/10.22541/essoar.172526712.23981307/v1) and is currently under review.

This thesis was conducted in the framework of the HYDROREC project (Deep learning for long-term and large-scale groundwater reconstruction) in collaboration with M2C CNRS and BRGM, with the financial support of the Normandie Region and BRGM under the agreement No. 21E01959.

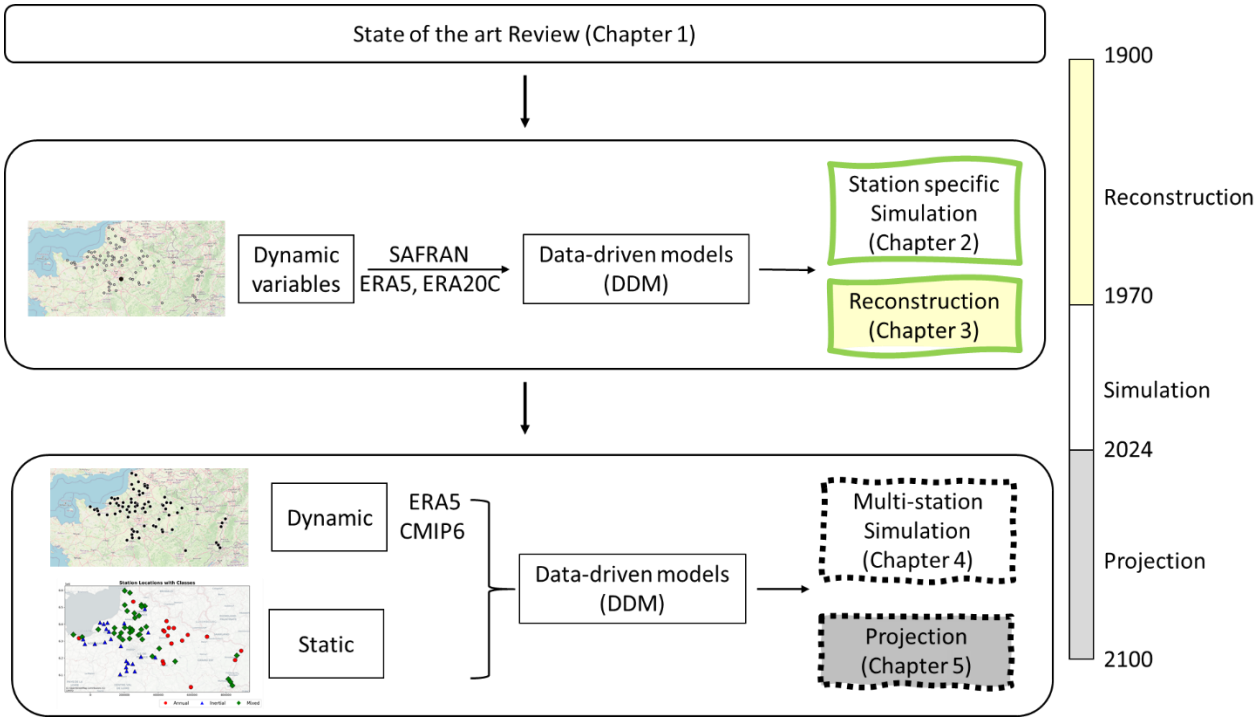

Figure 0.1: Simplified overview of thesis structure

Chapter 1. Deep learning for modelling groundwater level variations: Challenges and opportunities

Deep learning has gained wide interest in the hydrological community in the last few years but mostly focussed on surface hydrology. It then appeared to us that there is a need for a comprehensive review of groundwater applications of DL, in particular to highlight the current scientific and technical barriers in this field. This chapter is based on a manuscript in preparation for a review paper focused on understanding the current state of the art in DL for groundwater level variations and exploring future possibilities.

1.1. Introduction

Trends and variations in Groundwater level (GWL) are a critical aspect of water resource management, particularly in the context of climate changes (Baulon et al., 2022; Jasechko et al., 2024; Massei et al., 2020). GWL simulations are vital in creating reconstructions to investigate past fluctuations in groundwater resources or developing seasonal to multidecadal climate change impact projections (Chidepudi et al., 2024). Evaluating long-term shifts in GWL is vital for assessing the effects of climate change on aquifers and for effective water resource management. For instance, in regions experiencing prolonged droughts or increased water demand, accurate simulations of GWL can inform sustainable extraction strategies and prevent aquifer depletion (Scanlon et al., 2016).

1.1.1. Importance of GWL Simulations

The significance of GWL and the necessity for precise simulations cannot be overstated. Accurate simulations provide valuable insights, enabling us to manage this critical resource better (Chidepudi et al., 2023). However, the lack of comprehensive GWL observations significantly limits our understanding of historical GWL trends, variations, and the intricate relationship between GWL and climate. This lack of data poses a significant challenge to our ability to analyse patterns, observe spatial and temporal fluctuations, and establish robust correlations between groundwater dynamics and climatic factors (Chidepudi et al., 2024a; Dorigo et al., 2021; Taylor et al., 2013).

Deep learning (DL) has revolutionised various aspects of hydrology (e.g., flood forecasting: Nearing et al., 2024) and climatology (e.g., weather forecasting and climate projections: Bauer et al., 2023). The recent advancements have significantly contributed to the progress in developing a high-resolution digital twin for the entire Earth (Bauer et al., 2024) and its terrestrial water cycle (Brocca et al., 2024). While significant advancements have been made in surface hydrology, there remains a need to explore groundwater studies within the context of DL applications. Global GWL modelling experiments have historically faced challenges due to the complexity of subsurface processes, data scarcity, and computational limitations (Condon et al., 2021; Heudorfer et al., 2024). While providing valuable insights, traditional physics-based models often struggle to capture the full range of heterogeneity and non-linearity in groundwater systems, especially at large scales (Chidepudi et al., 2024b). DL offers promising solutions to these challenges through its ability to handle complex, non-linear relationships and integrate diverse data sources (Kratzert et al., 2021).

1.1.2. Objectives of the Review

In groundwater, research leveraging DL is still emerging (Heudorfer et al., 2024), highlighting the critical need to draw comprehensive insights from the progress made in DL applications across other relevant fields of hydrology and climatology. Even though surface water entities (rivers and lakes) are typically treated separately from groundwater, the two are intrinsically linked, interacting at points such as streambeds, floodplains, wetlands, and springs, where water can move between surface and underground sources (Scanlon et al., 2023). As the surface hydrology community continues to make significant strides, it becomes increasingly evident that a parallel advancement in subsurface hydrology is imperative. A comprehensive understanding of the current state-of-the-art technologies and methodologies becomes indispensable in advancing groundwater studies within the broader framework of large-scale hydrology.

In this review, we aim to provide an overview of the current state-of-the-art research using deep learning approaches for simulating, reconstructing, and projecting groundwater levels, highlighting the associated challenges and opportunities. In Section 1.2, we discuss the evolution of data-driven models used in groundwater studies. In Section 1.3, we highlight the current state of the art in DL for groundwater with all different possible applications. In Sections 1.4-1.6, we discuss the recent advances in using DL from surface hydrology applications and their relevance to groundwater. We highlight the opportunities and future directions in Section 1.7. Finally, we provide a summary and conclusion in Section 1.8

1.2. Evolution of Data-Driven Models in Hydrological Simulations and Forecasting

GWL simulations have often been developed using physically based and conceptual models such as MODFLOW (Mcdonald et al., 1988), Aquimod (Jackson et al., 2016), ParFlow (Maxwell et al., 2015) and LISFLOOD (Trichakis et al., 2017). While originally designed for surface water, LISFLOOD is now also being adapted for groundwater. These models offer significant benefits, including their ability to represent complex groundwater flow physics through mathematical equations, incorporate diverse hydrogeological parameters, and provide insights into underlying hydrological processes. They also allow for better generalisation to new scenarios by following physical laws and hydrological principles. However, these approaches are often difficult to implement at the global or regional scale, which is essential for detecting large-scale hydrological changes (Kingston et al., 2020). Due to the substantial computational and specific data requirements associated with these models, data-driven methods have become increasingly popular complements or sometimes alternatives in recent years (Chidepudi et al., 2023; Hauswirth et al., 2021; Wunsch et al., 2022a).

While previous review papers (Ahmadi et al., 2022; Boo et al., 2024; Nourani et al., 2024; Rajaee et al., 2019; Tao et al., 2022; Uc-Castillo et al., 2023) have predominantly focused on the application of artificial intelligence (AI), including machine learning (ML) and DL, in groundwater studies, their emphasis has always been on forecasting as shown in Table 1.1. It is important to note that while ML encompasses a wide range of algorithms, such as regression trees, random forests, XGBoost, and linear regression, DL specifically refers to models based on neural networks. In contrast, our review paper distinguishes itself by comprehensively examining forecasting and simulations, reconstructions, and projections within the context of DL applied to GWL time series. Furthermore, while traditional forecasting methods, as defined by Beven & Young (2013), have relied primarily on past target values in GWL, a very recent study has underscored the potential benefits of integrating external variables (Precipitation, temperature...) for ungauged regions, even within forecasting frameworks (e.g., Nearing et al., 2024).

Nevertheless, existing review papers have yet to undertake a comparative analysis of these models under varying conditions, such as autoregressive (mainly used for forecasting) and non-autoregressive approaches (suitable for simulation, reconstruction and projections).

Our paper seeks to address this gap by delving into a detailed discussion on the distinctiveness of various DL approaches, elucidating their respective limitations and opportunities. Through this comprehensive exploration, our review aims to overview the application of DL in groundwater studies and discuss potential avenues for future research and practical implementation within large-scale hydrology.

Table 1.1: Recent review papers in groundwater AI applications

Study	Focus	Models	Articles reviewed
(Rajaee et al., 2019)	Modelling (Forecasting):Review	AI methods (2001-2018)	67
(Ahmadi et al., 2022)	Forecasting: Review	Machine learning (2010-2020)	197
(Tao et al., 2022)	Forecasting: Review	Machine learning (2008-2020)	138
(Uc-Castillo et al., 2023)	Forecasting-Review	Machine learning (2000-2023)	168
(Nourani et al., 2024)	Forecasting (GRACE):Review	ML & DL (2002-2023)	90
(Boo et al., 2024)	Forecasting: Review	Machine learning (2017-2023)	142

1.3. DL Models for GWL Studies

DL techniques have gained significant traction in GWL modelling due to their ability to capture complex, non-linear dynamics by leveraging diverse data sources and features. These techniques encompass a wide range of neural network architectures, including Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), Bidirectional LSTMs

(BiLSTMs), CNN (Wunsch et al., 2021), GNN (Bai & Tahmasebi, 2023). LSTMs and GRUs are variants of RNNs that are particularly effective in capturing long-term dependencies and temporal patterns, making them well-suited for time series forecasting tasks. CNNs excel at extracting spatial features and patterns, which can be beneficial for incorporating spatial information in groundwater models. GNNs, on the other hand, are adept at handling graph-structured data and capturing spatial relationships, making them suitable for large-scale groundwater modelling applications.

Table 1.2: Examples of recent studies using DL for GWL forecasting

Reference study	Lead time	Model used
(Collados-Lara et al., 2023)	1-6 months	ANN (NAR, NARX,ELMAN)
(Bai & Tahmasebi, 2023)	One week	Graph neural networks, GRU, LSTM
(Patra et al., 2023)	Three days	LSTM – TL (Transfer learning)
(Vu et al., 2023)	30 days	BiLSTM

1.3.1. Simulations, Forecasting, Projections, and Reconstructions

While simulation, forecasting, projections, and reconstructions are often used interchangeably in groundwater studies, there is a compelling need to distinguish them to prevent ambiguity. According to Beven & Young (2013), forecasting involves quantitatively reproducing a system's behaviour in advance, utilising observations of inputs and relevant state variables up to the present time, known as the forecasting origin. Conversely, simulation entails replicating the behaviour of a system based on predetermined inputs without considering observed outputs. Projections, often referred to as 'what-if' simulations Beven & Young (2013), involve simulating the future behaviour of a system based on prior assumptions about future input data.

1.3.1.1. Forecasting

GWL forecasting using DL is still an active field of research with growing interest exponentially in recent years. However, the focus of this study is not on forecasting. A few examples of recent GWL forecasting studies, lead time, and model evaluation are shown in Table 1.2. It is important to notice that while using previous target values has been quite usual in forecasting studies, these studies have limited applicability. For instance, they are not transferable to unmonitored locations or used in the periods that GWL data is not available. Hence it has been recommended to explore external influencing variables as often done in rainfall-runoff modelling.

1.3.1.2. Reconstruction

Although Beven & Young (2013) do not explicitly define reconstruction, it can be interpreted as applying simulation techniques to historical data, utilising external variables. In some cases, nearby groundwater well data can be employed for near-historical reconstruction when available, as demonstrated by (Vu et al., 2021). Nevertheless, for long-term reconstruction, leveraging external variables from historical reanalysis datasets would offer greater advantages (Chidepudi et al., 2024; Hagen et al., 2023; Massei et al., 2020). Studies on simulations and reconstruction using data-driven models are prevalent on streamflow, as shown in Table 1.3, yet a very limited studies have been done on groundwater. In a recent effort, Chidepudi et al. (2023) employed recurrent-based DL models, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM), to simulate groundwater levels affected by interannual to decadal variability. This study utilised meteorological input data from the high-resolution SAFRAN reanalysis dataset. Additionally, the study found that applying a wavelet preprocessing technique, specifically the boundary-corrected maximum overlap discrete wavelet transforms (BC-MODWT), improved the simulations compared to using DL models alone, without any pre-processing. In a subsequent study, Chidepudi et al. (2024) used recurrent neural network architectures (LSTM, BiLSTM, GRU) with and without wavelet preprocessing on groundwater level (GWL) data from the ERA5 and ERA20C climate reanalysis datasets. The wavelet-assisted models, especially the BiLSTM, consistently outperformed the standalone models in reconstructing GWLs, capturing annual, interannual and decadal variability. The reconstructions using the higher-resolution ERA5 dataset slightly outperformed those using the coarse-resolution ERA20C dataset. The study validated the long-term GWL reconstructions against the longest observed time series available (Tincques station), confirming the models' ability to capture multidecadal variability linked to the Atlantic Multidecadal Variability (Klavans et al., 2022; O'Reilly et

al., 2023). More such studies would be needed in diverse groundwater level variations and regions for better generalisation.

Table 1.3: Recent papers in reconstruction

(Vu et al., 2021)	Reconstruction of missing GWL	LSTM	Only with nearby GWLs
(Tursun et al., 2024)	Reconstruction of missing Stream flow	Data integration with LSTM	In human-regulated catchments along with global reanalysis streamflow data (GloFAS)
(Chuphal & Mishra, 2023)	Reconstruction of streamflow (1951-2021)	Land surface hydrological and river routing models	Observed meteorological variables (P, tmax and tmin)
(Mishra et al., 2018)	Reconstruction of droughts (1951-2015)	Land surface models	Meteorological
(Jackson et al., 2016)	Reconstruction of multidecadal GWL	Lumped conceptual model	Obs Rainfall and PET
(Chidepudi et al., 2024)	Reconstruction of long-term GWL	DL (LSTM, GRU, BiLSTM) along with BC-MODWT	Reanalysis (ERA20C & ERA5)
(Hagen et al., 2023)	Reconstructing daily stream flow and floods	DL	Large-scale atmospheric variables

1.3.1.3. Projecting Future Trends Under Climate Change Scenarios

DL models can be used to assess the impact of climate change on groundwater systems by incorporating climate projections as input data. These assessments can provide insights into potential changes in groundwater levels under different climate scenarios. Recent studies have explored using ML/DL methods for groundwater projections, mostly CMIP 5 data, as shown in Table 1.4.

Table 1.4: Recent studies on groundwater projections

(Secci et al., 2023)	Projections	NARX, CNN, LSTM	CMIP5
(Wunsch et al., 2022a)	Projections	CNN	CMIP5
(Chakraborty et al., 2021)	Projections	XAI (XGBoost, LR)	CMIP5
(Xiong et al., 2022)	Projections	SVM, RF, ANN	CMIP5
(Nourani et al., 2023)	Projections	Autoregressive and ensemble ML	CMIP6

Leveraging potential global climate datasets (e.g., historical Reanalysis and climate models like CMIP6) is relatively underexplored in groundwater studies. DL techniques have been extensively applied in surface hydrology to reconstruct streamflow records and project future streamflow under various climate change scenarios (Hagen et al., 2023; Wi & Steinschneider, 2022). These studies have demonstrated the potential of DL models to capture complex spatiotemporal patterns and non-linear relationships, outperforming traditional statistical methods in many cases. However, groundwater-focused reconstruction and projection studies have been relatively scarce, partly due to the limited availability of long-term groundwater level data and the inherent complexity of groundwater systems.

1.3.2. Data limitations

While DL shows great promise, it is important to note that these techniques require large, high-quality datasets for training and validation. Groundwater studies face significant challenges due to limited global datasets, complex geological factors, and sensitivity to human activities (Condon et al., 2021; Heudorfer et al., 2024). Spatially and temporally,

data scarcity poses difficulties in training accurate DL models and capturing intricate groundwater dynamics. The influence of intricate geological factors and human activities introduces additional complexity and sensitivity in simulations. Furthermore, the need for historical data, transferability and generalisation in data-scarce regions, and addressing interannual to decadal variability in groundwater levels poses additional challenges. Barthel et al. (2021) highlighted that groundwater system properties are difficult to obtain and require additional efforts to make them suitable for structured analysis. Further, the catchment delineation from surface water systems does not apply to groundwater. Another challenge in groundwater modelling is the difficulty of operating physics-based (PB) or conceptual models on large spatial scales. As the scale of the study area increases, PB models become increasingly complex and computationally demanding, often requiring detailed parameterisation that may not be available across vast regions. In contrast, DL modelling approaches may be more suitable for large-scale applications, as they can implicitly account for various processes without explicitly modelling each one. Addressing data scarcity is a significant challenge in groundwater modelling and simulations, particularly for large-scale or regional applications. Furthermore, uncertainty analysis is also crucial in groundwater simulations and projections when models deal with limited data, as it helps quantify and account for the various sources of uncertainty inherent in these complex systems. For model uncertainty, the ensemble DL models trained with different initialisations can be used to estimate uncertainties and quantify model robustness.

1.3.3. Model Interpretability and Explainability

The interpretability and explainability of DL models remain challenging and essential for building trust and understanding the underlying mechanisms driving groundwater dynamics (Chidepudi et al., 2023; Jung et al., 2024; Niu et al., 2023). Explainability and interpretability are crucial when applying DL models for groundwater simulations due to the inherent complexity of groundwater systems involving intricate physical processes and interactions. Several techniques have emerged to enhance the interpretability of DL models in groundwater simulations. Such as 1) Shapley Additive exPlanations (SHAP), a game-centric approach whose values can be aggregated across datasets to provide a global understanding of the model's behaviour, revealing the importance of different features in groundwater simulation tasks (Chidepudi et al., 2023). SHAP values can be visualised using techniques like summary plots, dependence plots, and force plots, aiding in interpreting the model's behaviour and understanding complex relationships between input features (e.g., precipitation, temperature) and groundwater predictions. 2) Local Interpretable Model-agnostic Explanations (LIME), used to get local explanations by approximating the complex model with a simpler, interpretable model around a specific simulation. This

approach can be used to identify the most influential features for individual hydrological simulations, providing localised insights into model behaviour (Althoff et al., 2021; Cheng et al., 2022).

Another possibility to enhance interpretability is by building hybrid models. The term "hybrid modelling" has been used in various ways in hydrological studies. In general, any model that combines two or more different approaches is considered a hybrid model. Some examples of such hybrid models in relevant subfields of hydrology include physical-based models combined with data-driven models (DDMs) (Xu et al., 2024), conceptual models combined with DDMs (Espinoza et al., 2024), global hydrological models combined with DDMs (Kraft et al., 2022), preprocessing techniques (e.g., BC-MODWT) combined with DDMs (Momeneh & Nourani, 2022; Nourani et al., 2014; Rehana & Rajesh, 2023), combination of multiple DDMs (e.g., CNN-LSTM, KNN-LSTM) (Barzegar et al., 2021; Rehana & Rajesh, 2023), and hybrid Physics-AI approaches (Jiang et al., 2020).

In recent years, there has been a growing interest in combining different groundwater modelling techniques to create hybrid approaches. So far, these hybrid methods have primarily focused on combining data-driven methods such as DL architectures like LSTMs and CNNs with pre-processing techniques like wavelet transform. However, there is still much to explore in the development of hybrid approaches that combine data-driven models with physics-based groundwater models, benefiting from the complementary advantages of both methodologies (Kraft et al., 2022). The data-driven components, powered by machine learning or DL algorithms, can fill gaps or provide boundary conditions for the physics-based models rooted in the fundamental principles and equations governing groundwater flow. This coupling allows the data-driven models to capture intricate patterns and relationships from relevant influencing variables, while the physics-based models contribute physical realism and interpretability to the overall modelling approach (Tahmasebi et al., 2020; Tripathy & Mishra, 2024).

Furthermore, multi-model approaches involve ensemble techniques, where multiple models (either physics-based, data-driven, or a combination) are employed, and their outputs are combined through various strategies, such as model averaging, Bayesian model combination, or machine learning-based fusion (Thébault et al., 2024). This approach can leverage the strengths of individual models and potentially improve overall predictive performance and robustness. Hybrid and multi-model approaches have shown promising results in hydrological applications, particularly when data availability is limited,

or the physical processes are complex and challenging to accurately model using a single approach.

1.4. Approaches for Regionalisation and Large-Scale Modelling

Large-scale modelling in hydrology has been a long-standing challenge, particularly in data-scarce regions or ungauged catchments. Recent advancements in deep learning techniques, such as transfer learning and multi-catchment training, have opened up new possibilities for addressing these challenges. Review of the significant advancements in the application of data-driven models, particularly in areas such as multi-station training and transfer learning.

1.4.1. Transfer Learning

Pre-trained DL models from data-rich regions can be fine-tuned or adapted to the data-scarce areas, reducing the need for extensive local data (Ghobadi et al., 2024; Xu et al., 2024; Yao et al., 2023). This approach has shown promising results in rainfall-runoff modelling, where models trained on multiple catchments can be transferred and tested for their capabilities in simulating pseudo-ungauged catchments. Recently Sun et al. (2021) showed that Graph Neural Networks (GNNs) are robust in learning spatiotemporal dependencies, resulting in similar or better performance than models trained using LSTMs. GNNs' ability to capture spatial relationships and leverage graph-structured data makes them well-suited for large-scale hydrological modelling, where spatial dependencies play a crucial role.

1.4.2. Multi-Catchment Training

With the growing availability of large-scale datasets spanning multiple catchments, recent studies have explored multi-station training for data-driven models in hydrology, where models are trained simultaneously on data from diverse catchments and stations, leveraging information from varied hydro-meteorological conditions. This approach leads to more robust and transferable models with improved generalisation performance, especially in data-scarce regions, enabling the capture of broader hydrological patterns and dynamics. It enhances transferability to ungauged or data-sparse basins, addressing the limitations of traditional single-catchment training approaches. DL models trained on multiple catchments have shown increased interest and outperformed their single-catchment counterparts (Kratzert et al., 2019; Klotz et al., 2022; Nearing et al., 2024).

1.5. Comparison and Benchmarking Studies

Benchmarking and comparison across models are necessary to understand the advancements and benefits of different approaches. Recent studies have compared the performance of LSTM models against conceptual and physically-based hydrological models in various contexts, including ungauged basins, historical simulations, climate impact studies, and extreme flow estimations (Hauswirth et al., 2021). These comparisons aim to assess the strengths and limitations of different modelling approaches and identify areas for improvement. Lees et al. (2021) and Kratzert et al. (2019) have benchmarked LSTM models against lumped conceptual models (e.g., TOPMODEL, VIC) in Great Britain and across the continental United States, respectively. LSTMs have demonstrated better out-of-sample performance than conceptual models' in-sample performance, particularly in capturing complex hydrological patterns and dynamics. However, conceptual models maintain water mass balance constraints, which LSTMs do not inherently enforce. Comparisons between LSTMs and physically-based models are less common, but some studies have explored hybrid approaches that combine the two (Kraft et al., 2022; Liu et al., 2024). Recent studies also have highlighted the potential benefits of using differentiable models that combine process-based models with machine learning techniques like neural networks (Feng et al., 2023; Shen et al., 2023).

LSTMs have demonstrated their ability to generalise and perform well in ungauged basins or for historical simulations, where traditional models may struggle due to limited data availability. The ability to simulate hydrological processes in ungauged catchments is a critical requirement for large-scale modelling. Several studies have explored the potential of DL models in this context. Arsenault et al. (2023) showed that LSTM outperformed the traditional hydrological models for streamflow prediction at ungauged catchments. (Nearing et al., 2024) showed that AI-based forecasting is reliable even for extreme events in ungauged basins with up to five days lead time and performed better than nowcasts from GLOFAS. The study did not use previous streamflow data as inputs to forecast in ungauged situations with no streamflow data. (Barthel et al., 2021) proposed to build upon the learning from the progress in ungauged basins (Streamflow) with similarity-based approaches, which are still not widely adopted in groundwater studies for unmonitored sites. While these studies are not yet completely evaluated for groundwater, recently, Haaf et al. (2023) presented a method to estimate daily groundwater level time series at unmonitored sites by linking groundwater dynamics to local hydrogeological system

controls. The approach uses regression analysis (XGBoost and MLR) to estimate groundwater head duration curves at unmonitored locations using site descriptors, then constructs hydrographs using time series from neighbouring monitored locations. More such benchmark studies with collaborative efforts would be needed globally to exploit the potential of the latest advancements in DL fully.

1.6. Bridging surface water and groundwater research

Surface and groundwater systems are interconnected, and their interactions play a crucial role in various hydrological processes, including floods, droughts, and water resources management. Acknowledging the interconnected nature of surface water and groundwater systems is vital for effective water resource management (Scanlon et al., 2023). Recent studies have highlighted the importance of accounting for these interactions and the interconnected nature of streamflow and groundwater. For instance, Berghuijs et al. (2022) conducted a global synthesis of field-estimated groundwater recharge and found that recharge rates are strongly linked to climatic aridity. Their study suggests higher estimates of global groundwater recharge, implying that groundwater contributes more significantly to evapotranspiration and streamflow than previously represented by global hydrological models. This finding highlights the need to improve the representation of groundwater processes in large-scale models to capture water fluxes and availability accurately. (Barclay et al., 2023) demonstrated that incorporating groundwater discharge processes into stream temperature models using process-guided DL can significantly enhance predictive accuracy, reduce seasonal bias, and better inform habitat management strategies. This study underscores the importance of considering groundwater-surface water interactions in ecological and environmental applications, such as managing aquatic habitats and assessing the impacts of climate change. De Graaf et al. (2024) revealed that approximately 20% of globally pumped groundwater stems from diminished streamflow, highlighting the significant impact of groundwater extraction on surface water systems. This finding emphasises the need to account for groundwater-surface water interactions in large-scale water resources assessments and management strategies to ensure sustainable water use and maintain ecosystem health. (Slater et al., 2024) developed an approach based on quantile regression forests to assess the spatial variability of flood drivers across Great Britain. They found that increasing precipitation and urbanisation amplify flood magnitude significantly more in catchments with high baseflow contribution and low runoff ratio, which are often groundwater-dominated. This finding emphasises the need to account for groundwater dynamics in flood risk assessments and management strategies, particularly in areas with significant groundwater influence. Sharma &

Mujumdar (2024) highlighted the significant baseflow contributions to river floods in India and noted that groundwater is often overlooked in flood studies. Their research underscores the importance of considering groundwater-surface water interactions in flood modelling and risk assessment, as neglecting these interactions can lead to inaccurate predictions and inadequate flood management strategies. Hence, future research could explore the simulation of groundwater levels, considering interactions with surface hydrology (e.g., streamflow).

1.7. Opportunities & Future Directions

Extending the advancements in surface hydrology to groundwater studies can open new avenues for addressing challenges in data-scarce regions, incorporating physical principles, and improving the representation of groundwater-surface water interactions in models. The advancements in AI research have progressed at different paces within the surface water and groundwater communities, with several potential reasons contributing to this disparity. The widespread use of conceptual models has facilitated a smoother transition toward DL techniques in the surface water community. This transition is evident from the rise in benchmarking studies that compare DL models with conceptual and physics-based models for rainfall-runoff modelling. However, such comprehensive benchmarking studies are scarce in groundwater, partly due to the lack of globally available datasets akin to CAMELS or CARAVAN for streamflow data. Moreover, modelling groundwater systems is often not simply a matter of input/output with a single, highly localised outlet response. Instead, groundwater modelling requires a fully spatialised representation of hydraulic head and flow based on physical laws such as Darcy's law and derived diffusivity equation. This inherent complexity in groundwater systems poses unique challenges for applying DL techniques, further contributing to the slower adoption of AI in this field.

The ability of DL models to capture complex, non-linear dynamics by leveraging diverse data sources and features has made them attractive for groundwater modelling applications. However, it is essential to pre-process the input data carefully, employ appropriate architectures tailored to the specific task, and integrate domain knowledge and physical constraints to ensure reliable and interpretable results. Several open research areas have been highlighted to summarise recent advancements yet to be explored for groundwater in DL applications but have been extensively used for rainfall-runoff modelling. These include reconstruction and projections (Hagen et al., 2023; Nasreen et al., 2022; Zhao et al., 2023), constrained DL for projections (Feng et al., 2022; Tahmasebi et al., 2020; Tripathy & Mishra, 2024), extreme flow estimations (Hauswirth et al., 2023;

Nearing et al., 2024), and differential modelling (Feng et al., 2023; Shen et al., 2023; Tsai et al., 2021; Wang et al., 2024).

Efforts are underway to address data availability issues; however, the groundwater community has traditionally been more interested in modelling groundwater flow processes and recharge estimation rather than directly predicting observed groundwater levels at wells. The focus on recharge estimation and the inherent complexity of groundwater systems have likely contributed to the slower adoption of DL techniques in this field. Moreover, the effect of pumping activities is often more pronounced in groundwater levels, and the lack of comprehensive data on pumping rates and locations hinders the usage of DL models across wide geographic locations, presenting a challenge in developing robust and generalisable models for groundwater level simulations.

Hybrid modelling approaches that can combine DL with classical groundwater models hold immense potential for leveraging the strengths of both methodologies. Integrating these approaches can lead to more accurate and interpretable simulations, reconstructions, and projections of groundwater dynamics. Enhancing the interpretability of DL models, handling uncertainty in groundwater data and models, leveraging domain-specific knowledge, and developing scalable and computationally efficient DL models are crucial aspects for actionable insights from groundwater studies.

1.8. Conclusion

Deep learning (DL) techniques hold immense potential for the simulation, reconstruction, and projection of groundwater behaviour. These methods enable accurate reconstructions of historical groundwater levels and future projections under various climate change scenarios. However, spatial and temporal data limitations pose significant hurdles for training accurate DL models and capturing the complex dynamics of groundwater systems. Integration of physical hydrogeological knowledge and the incorporation of physical constraints into data-driven frameworks are crucial for ensuring realistic simulations. It is essential to enhance interpretability through techniques like SHAP and hybrid approaches, as this is vital for obtaining insights. Bridging the gap between surface water and groundwater research is imperative, given the interconnected nature of these systems and their significant implications for water resource management. Leveraging recent advancements in surface hydrology, such as multi-station training, transfer learning, and physics-guided neural networks, can open up new avenues for groundwater studies and enhance the representation of groundwater-surface water interactions in models.

Addressing uncertainty in GWL data and models through ensemble techniques and uncertainty quantification methods is crucial for reliable simulations. Collaborative efforts involving global benchmarking, interdisciplinary collaborations, and data-sharing platforms are essential for advancing groundwater research, facilitating knowledge exchange, and establishing data standards. In conclusion, the application of DL techniques in groundwater level (GWL) simulations, reconstructions, and projections presents significant opportunities to enhance our understanding of groundwater dynamics. By addressing data limitations, enhancing interpretability, bridging the gap with surface water research, developing hybrid modelling approaches, and quantifying uncertainties, the groundwater community can harness the power of DL to develop more accurate, and actionable models.

Chapter 2. A wavelet-assisted deep learning approach for simulating groundwater levels affected by low-frequency variability

The current state of the art and potential opportunities of deep learning techniques in modelling groundwater level variations are outlined in Chapter 1. While most of the previous groundwater studies focussed on short-term forecasting using previous GWL values as input and relied heavily on autocorrelation, as a first step, we aim to assess the capabilities of DL for simulations using only external variables in this chapter. More specifically, we aimed to develop suitable DL models to simulate groundwater level variations affected not only by typical annual variability but also by longer-term multi-annual to decadal variations inherited from large-scale climate variability.

Chapter 2 is a paper that was published in the Science of the Total Environment.

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. *Science of The Total Environment*, *865*, 161035. https://doi.org/10.1016/j.scitotenv.2022.161035

Abstract

Groundwater level (GWL) simulations allow the generation of reconstructions for exploring the past temporal variability of groundwater resources or provide the means for generating projections under climate change on decadal scales. In this context, analysing GWLs affected by low-frequency variations is crucial. In this study, we assess the capabilities of three deep learning (DL) models (long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional LSTM (BiLSTM)) in simulating three types of GWLs affected by varying low-frequency behaviour: inertial (dominated by low-frequency), annual (dominated by annual cyclicity) and mixed (in which both annual and low-frequency variations have high amplitude). We also tested if maximal overlap discrete wavelet transform pre-processing (MODWT) of input variables helps to better identify the frequency content most relevant for the models (MODWT-DL models). Only external variables (i.e., precipitation, air temperature as raw data, and effective precipitation (EP)) were used as input. Results indicate that for inertial-type GWLs, MODWT-DL models with raw data were notably more accurate than standalone models. However, DL models performed well for annual-type GWLs, while using EP as input, with MODWT-DL models exhibiting only minor improvements. Using raw data as input improved MODWT-DL models compared to standalone models; nevertheless, all models using EP performed better for annual-type GWLs. For mixed-type GWLs, while using EP as input, MODWT-DL models performed well, with substantial improvements over standalone models. Using raw data as input, improvement of MODWT-DL models is marginal compared to that of standalone models; nevertheless, they perform better than standalone models with EP. The Shapley Additive exPlanations (SHAP) approach used to interpret models highlighted that they preferentially learned from low-frequency in precipitation data to achieve the best simulations for inertial and mixed GWLs. This study showed that MODWT-based input pre-processing is highly suitable to better simulate low-frequency varying GWLs.

2.1. Introduction

Accurate and reliable groundwater level (GWL) simulations and forecasting play vital roles in water resource management (Liu et al., 2022; Rahman et al., 2020). While the terms simulations and forecasting are inherently different, they are often used synonymously in an inappropriate way in groundwater studies (e.g., (Ghazi et al., 2021)), even after (Beven & Young, 2013) provided clear definitions of these terms. According to (Beven & Young, 2013), "simulation" is defined as the quantitative reproduction of system behaviour without

reference to observed output; "forecasting" is defined as the reproduction of system behaviour ahead of time with observed outputs up until the onset of forecast included.

Traditionally, GWL simulations have been performed using physically based models, such as MODFLOW (Mcdonald et al., 1988) and ParFlow (Maxwell et al., 2015), in which a set of physical properties of the aquifer is implicitly used in the numerical solution of the groundwater equation. However, this type of calculation is often cumbersome and complicates the calibration of the model for approximating the hydrodynamic behaviour of the aquifer expressed in observational data. Owing to the high computational and data requirements (Maxwell et al., 2015) of these models, data-driven approaches have gained traction in recent years as viable alternatives (Rahman et al., 2020; Rajaee et al., 2019; Tao et al., 2022). Furthermore, earlier studies on these approaches mainly focused on the classical multi-layer perceptron (MLP), i.e., a basic form of artificial neural network (ANN), which has outperformed MODFLOW in groundwater flow simulations(Mohanty et al., 2013). This simple network, which is considered to be the cornerstone of deep learning (DL), has demonstrated through several studies its effectiveness to approximate in black box form the highly nonlinear relationships that can link piezometric fluctuations to climate signals (Coulibaly et al., 2001; Wunsch et al., 2021). However, the popularity of MLP networks in processing time series is decreasing in favour of a new generation of neural networks (such as LSTM and GRU). These neural networks include memory effect modules that handle temporal dependencies not accounted for in traditional neural networks, such as the MLP. This popularity is reflected in the numerous articles published in recent years that use this type of approach and highlight the effectiveness of standalone LSTMs in GWL simulations (Zhang et al., 2018), forecasting (Bowes et al., 2019), and reconstructing missing values (Vu et al., 2021).

To further improve the GWL forecasting performance of artificial intelligence (AI) algorithms, a new generation of algorithms called hybrids has been developed. These algorithms combine time-series pre-processing tools, such as principal component analysis (Cai et al., 2021), singular spectrum analysis (Yadav et al., 2020), DWT (Wu et al., 2021), and maximal overlap discrete wavelet pre-processing (MODWT) (Rahman et al., 2020) with AI tools. Among these, wavelet transform based pre-processing coupled with data-driven models has shown increased efficiency regarding GWL forecasting because of its ability to extract time-varying behaviour (Rahman et al., 2020; Rajaee et al., 2019). However, a recent study by Quilty & Adamowski (2018) highlighted the frequent incorrect usage of wavelet transform in hydrological forecasting and recommended using boundary-corrected MODWT. This approach is now gaining traction in hydrological time series forecasting; for

example, (Mouatadid et al., 2019) combined LSTM with BC-MODWT for irrigation flow forecasting and found it to perform better than standalone LSTM. Rahman et al. (2020) used BC-MODWT along with random forest and XGBoost and found the approach promising for forecasting GWLs. Finally, a more recent study by Barzegar et al. (2021) coupled BC-MODWT with CNN-LSTM and found it promising for multiscale lake water level forecasting. However, it further highlighted that efforts are required to find a suitable wavelet family, filter length, and decomposition level. Even after (Quilty & Adamowski, 2018) suggested using BC-MODWT or AT, a few studies (Liang et al., 2021; C. Wu et al., 2021) continued to use DWT coupled with LSTM for groundwater forecasting. Other pre-processing techniques have also been used for GWL forecasting; for example, (M. Wu et al., 2021) combined variational mode decomposition with an extreme learning machine (ELM), and (W. Liu et al., 2021) coupled CEEMDAN with the deep belief network (DBN) model.

All of these pre-processing techniques coupled with DL are mainly focused on forecasting, i.e., they use previous target values in the input. A recent study by (Bahmani & Ouarda, 2021) coupled EEMD and CEEMD with GEP and M5 with the aim of simulating GWL; however, they used GWLs as input, which contradicts the definition of simulation by (Beven & Young, 2013a). Hence, the primary purpose of this study is to appropriately assess these approaches in simulations.

As models grow in complexity, there is increasing interest in their explanation and interpretation through approaches like SHapley Additive exPlanations (SHAP) which is mainly helpful in understanding the influence of input variables on the model simulations. According to recent literature, this is particularly true for ML and DL methods; for example, (Q. Liu et al., 2022) used the SHAP approach to interpret machine-learning models for near-term GWL simulations. This approach allowed the authors to identify that the flow volume and distance to the river and reservoir played significant roles in groundwater changes. (Anderson & Radić, 2022) also showed that DL models learn the contribution of glacial runoff from meteorological variables in streamflow modelling. Even though research in this direction is only beginning, there is significant potential for interpreting DL models, thereby increasing their credibility for wider adoption across different decision-making levels. However, SHAP interpretation is still not very common in hydrology; in particular, it has not been used to explore the impacts of pre-processing.

We identified the following lack of scientific and technical information in the existing literature: Most of the previous studies exploring wavelet usage were limited to forecasting, with previous GWLs being used as input. Consequently, these models relied heavily on high

autocorrelation in previous GWL values, resulting in high accuracy. In contrast, our study focuses on simulation (not forecasting), in which only external factors influencing GWLs are used as input. Furthermore, we noticed a clear lack of studies looking at either forecasting or simulating GWL time series with the low-frequency variability (i.e., interannual to decadal) explaining most of the GWL variations. Indeed, being able to reproduce and establish simulations (e.g., projections) is critical for accounting for the influence of low-frequency climate variability, which can either mask or aggravate the effects of climate change on hydrology (Boé & Habets, 2014; Bonnet et al., 2022; Kingston et al., 2020a).

We hypothesise that wavelet expansion, by extracting the most relevant high- to low-frequency variability information from external input variables (i.e., precipitation, air temperature, and effective precipitation), will help in achieving better simulations of GWL variations, even when these are dominated by low-frequency variability that is barely visible or even invisible in input variables. In this study, we evaluate the effectiveness of coupling recurrent-based DL algorithms (i.e., LSTM, bidirectional LSTM (BiLSTM), and GRU) with BC-MODWT signal decomposition for simulating GWL variations in the case of three piezometric time series representative of the most contrasting temporal behaviours in northern France (Baulon et al., 2022a).

To test this hypothesis, we 1) evaluate the effectiveness of using raw data (i.e., precipitation and air temperature) as input against using processed data, such as effective precipitation (EP), as input, 2) compare DL methods with and without wavelet pre-processing while quantifying the performance of each method, and 3) investigate the internal functioning and plausibility testing of the DL models through the SHAP interpretative approach. The remainder of this chapter is structured as follows: Section 2.2 presents the data and normalisation. Section 2.3 presents the theoretical background of the methods and pre-processing techniques employed in this study. Section 2.4 presents the analysis and interpretation of the results. Section 2.5 presents a detailed investigation on what models learn using SHAP. Section 2.6 presents our conclusions.

2.2. Data

We used a database consisting of a relatively long GWL time series initially taken from the ADES (Accès aux Données sur les Eaux Souterraines) database (https://ades.eaufrance.fr/; (Winckel et al., 2022)). These GWL data were especially selected because the anthropogenic impact on them was relatively low (Baulon et al.,

2022a). From this database, we selected three GWL time series that fluctuated over a 50-year period (i.e., 1970–2020) in contrasting manners: the first GWL time series has an inertial nature where the low-frequency component dominates the signal; the second is the so-called mixed GWL time series because the fluctuations are reflected by the annual and interannual components, with the latter being largely influenced by the annual cycle. Thus, the numerical simulation tools presented here are confronted with the prediction of piezometric fluctuations in aquifers with well-contrasted hydrodynamic regimes.

Figure 2.1: Study area (Red dots indicate the stations GWL1 (Inertial), GWL2 (Annual) and GWL3 (Mixed)).

The locations of the three stations are shown in Figure 2.1. The GWL time series of the three stations are shown in Figure 2.2.

The current work uses precipitation, air temperature, or EP as input variables (Figure 2.3: Example for one station). The numerical simulation consists of training different types of DL networks with meteorological data as input. Specifically, EP is used as one input variable on the one hand, and precipitation and air temperature combined are used as input on the

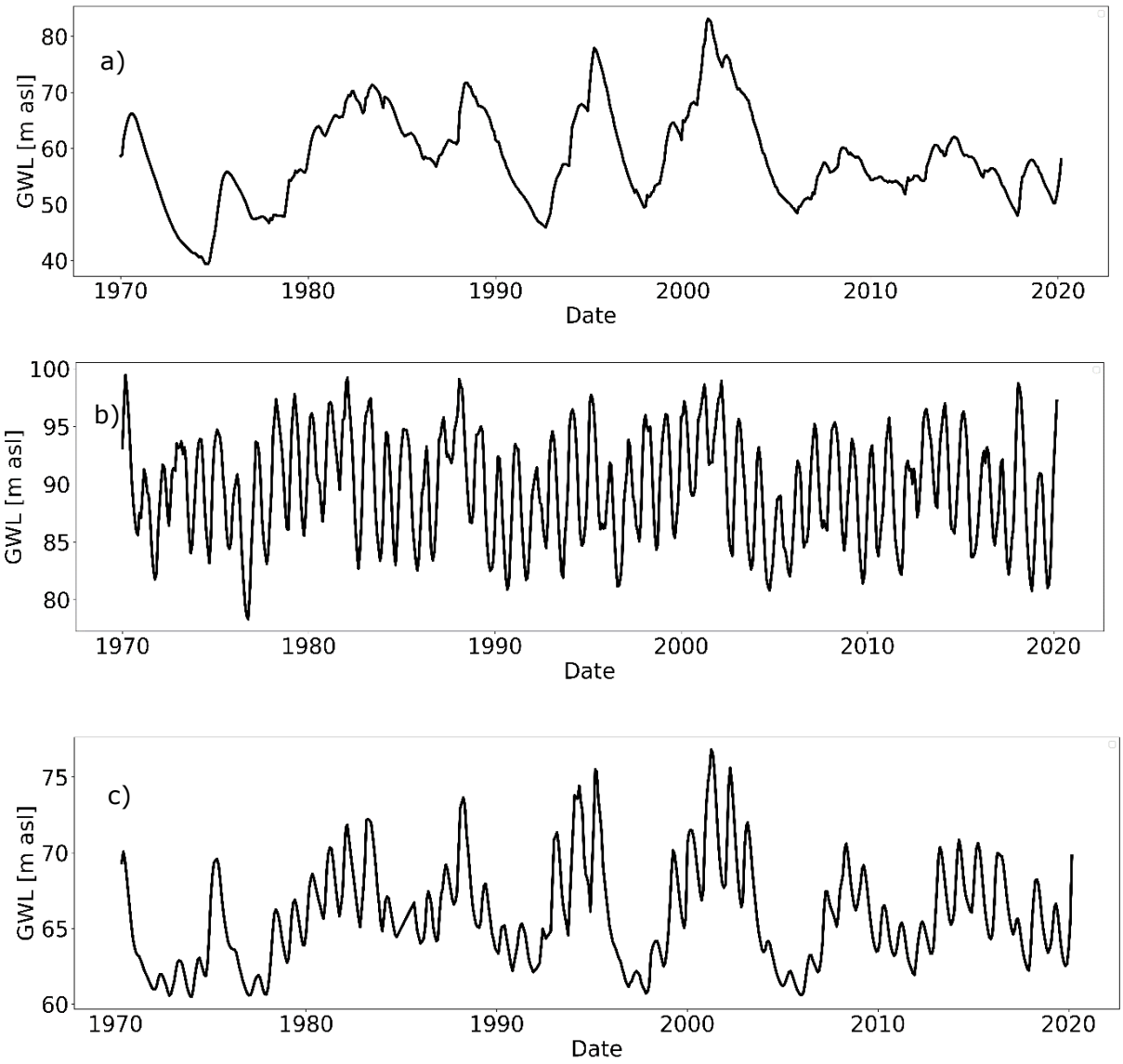

Figure 2.2: a) Inertial, b) annual, and c) mixed types of GWLs.

other hand; the output variable is piezometric responses.

Precipitation and mean air temperature data were retrieved from the SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) Reanalysis (Vidal et al., 2010), which is available at a daily time step with a spatial resolution of 8 × 8 km^2. Although the input time series is available at the daily time step, only the monthly averages

between 1970–2020 were used to match the monthly time step of the piezometer data. EP was derived from a calculation using the water budget method of (Edijatno & Michel, 1989).

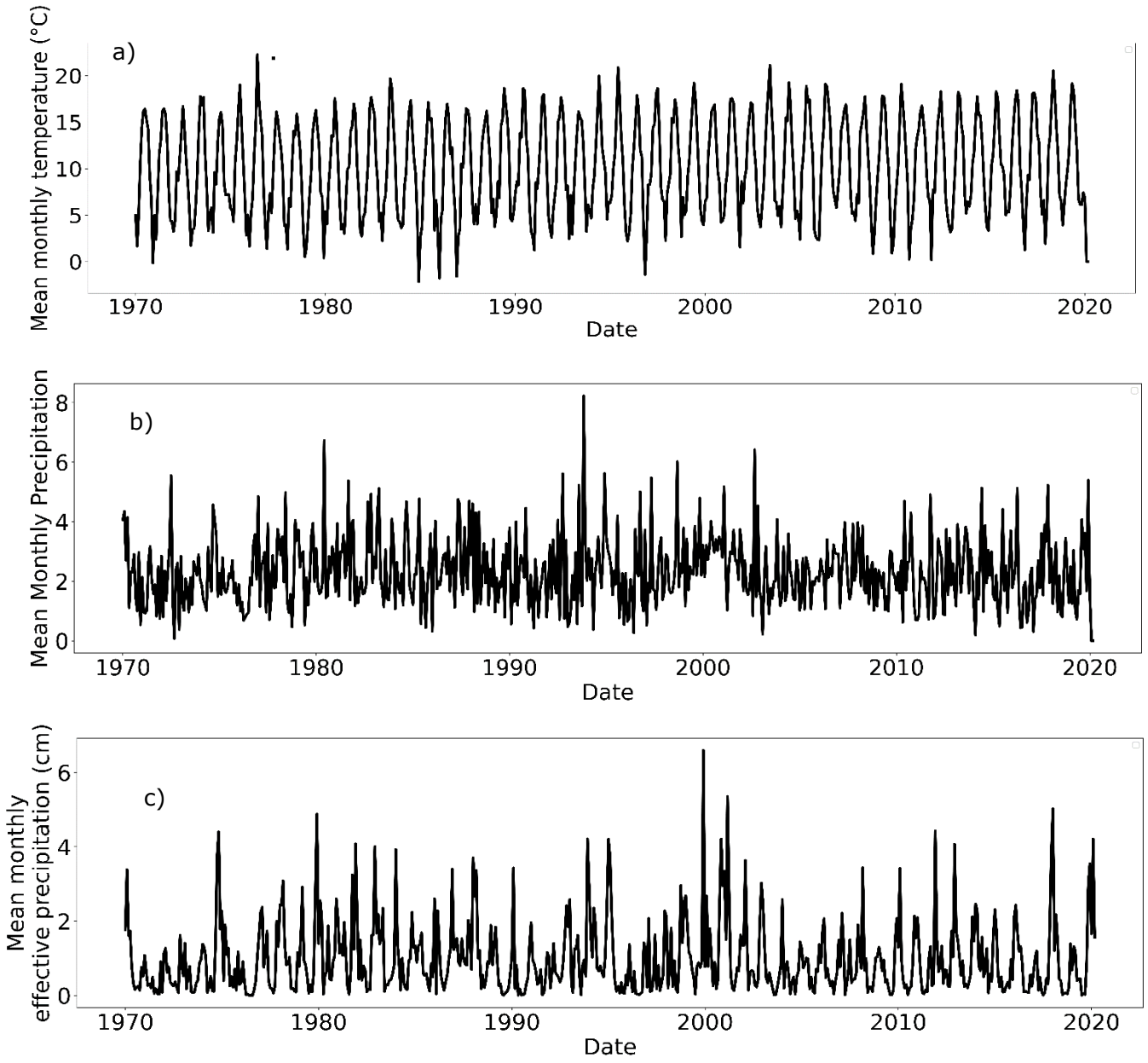

Figure 2.3: Time series of a) air temperature, b) precipitation, and c) effective precipitation.

Conversely, the processed data came from areas in France with limited anthropogenic impacts; hence, the piezometric variations were essentially due to the climatic signal.All data used as input and output are normalised into the network to facilitate stable convergence in the learning phase. The following equation is used for this normalisation so that all data can vary in (0, 1) intervals:

$$x_{scale} = (x - x_{min})/(x_{max} - x_{min}), \quad (1)$$

Where x_{min} and x_{max} represent the minimum and maximum values of the data, respectively, whereas x and x_{scale} denote the original and scaled data, respectively.

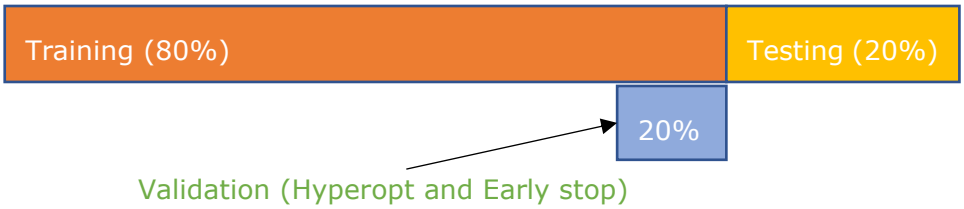

Figure 2.4: Data partitioning and the corresponding proportions.

We note that there is a common pitfall in this step, i.e., where the previous studies scaled all available data—including test data—together; this step results in data leakage by allowing the model to be influenced by the test data, which should not have been made available at this stage. Consequently, this can lead to overestimated results and undesirable expectations regarding the model's performance. In this study, this was taken into consideration, and we scaled out each input variable individually to be in the (0,1) range. After normalisation, the database is divided into two main subsets: the training and testing sets (constituting 80% and 20% of the database, respectively). Furthermore, the last 20% of the training set, which is equivalent to 16% of the total dataset available, was used for validation, as shown in Figure 2.4. The purpose of each split is as follows: the training data are used for fitting the model on them; the validation data are used for hyperparameter optimisation and early stopping; and the testing data are used for obtaining the performance of the models on the unseen data. Using only meteorological variables as input in this study utilises the long time series of gridded data available at a finer spatiotemporal resolution across metropolitan France, with no missing values. For instance, SAFRAN reanalysis data are available on a daily scale for 70 years (i.e., 1951–2019), which is not the case for the GWL time series data, which are heterogeneous with missing values.

2.3. Theoretical background

The overall modelling methodology developed here involves the use and comparison of three types of recurrent-based deep neural networks, as well as a MODWT pre-processing approach to improve the recovery of the most useful information from the input data.

2.3.1. Wavelet decomposition pre-processing

In this study, we intended to test the usefulness of wavelet decomposition for detecting scale-dependent information to be used as input signals in high- to low-frequency wavelet components, thereby helping AI models to better simulate GWL variations. For instance, pre-processing approaches using wavelet transform have received particular attention in recent years; however, one significant contribution to the field (i.e., Quilty & Adamowski, 2018) highlighted the incorrect use of wavelet transform as a pre-processing tool for hydrological forecasting. In particular, the authors revealed associated future data issues when using DWT-MRA and MODWT-MRA; they suggested some best practices while testing them on wavelet-based ML methods for urban water demand forecasting.

In this methodological framework, we aim to develop the so-called direct approach (Quilty & Adamowski, 2018) and address the crucial constraints emphasised in this paper. In this approach, wavelet decomposition of the training dataset using MODWT is applied to the input signals only, thereby resulting in several high- to low-frequency wavelet components, each being used as input to the selected models. When using MODWT-assisted DL modelling, all wavelet components of precipitation (Figure 2.5) and air temperature or EP time series (not shown) were used as input.

The MODWT algorithm was used to decompose the input variables into various scale levels and extract the variability of the decomposed signal at each time scale. This method is well-suited to real-world signals, as it enables the decomposition of a given signal (here, precipitation and air temperature or EP) into several components across different time scales (i.e., from high to low-frequencies), while keeping the amplitudes of the transform aligned with the amplitude in the original signal (Percival & Walden, 2000). This means that the discrete wavelet component still bears some "physical" meaning compared to the original decimated and non-redundant discrete wavelet transform. Decomposition can be achieved up to the maximum decomposition level depending on the length of the time-series; however, the decomposition depth (or filter length) must be constrained to avoid computing values being affected by boundary effects, which is even more critical when decomposition is included in a modelling framework (Quilty & Adamowski, 2018). In this

study, the boundary-affected coefficients (L_J) were removed from the beginning of the input and target variables using the equation recommended by Quilty and Adamowski (2018): $L_J = (2^J-1)(L-1)+1$.

Here, J is the number of decomposition levels, which has been set to 4, and L is the length of the filter used, i.e., eight wavelet/scaling values for La8, 12 for La12, and so on. Regarding the choice of wavelet filter, we chose the least asymmetric filters (i.e., La8 to La16), with periodic boundary conditions selected for the wavelet transform, meaning that the original signal is repeated after its last value is reached. The combined use of the MODWT and La filters, along with circular shifting of the computed wavelet and scaling

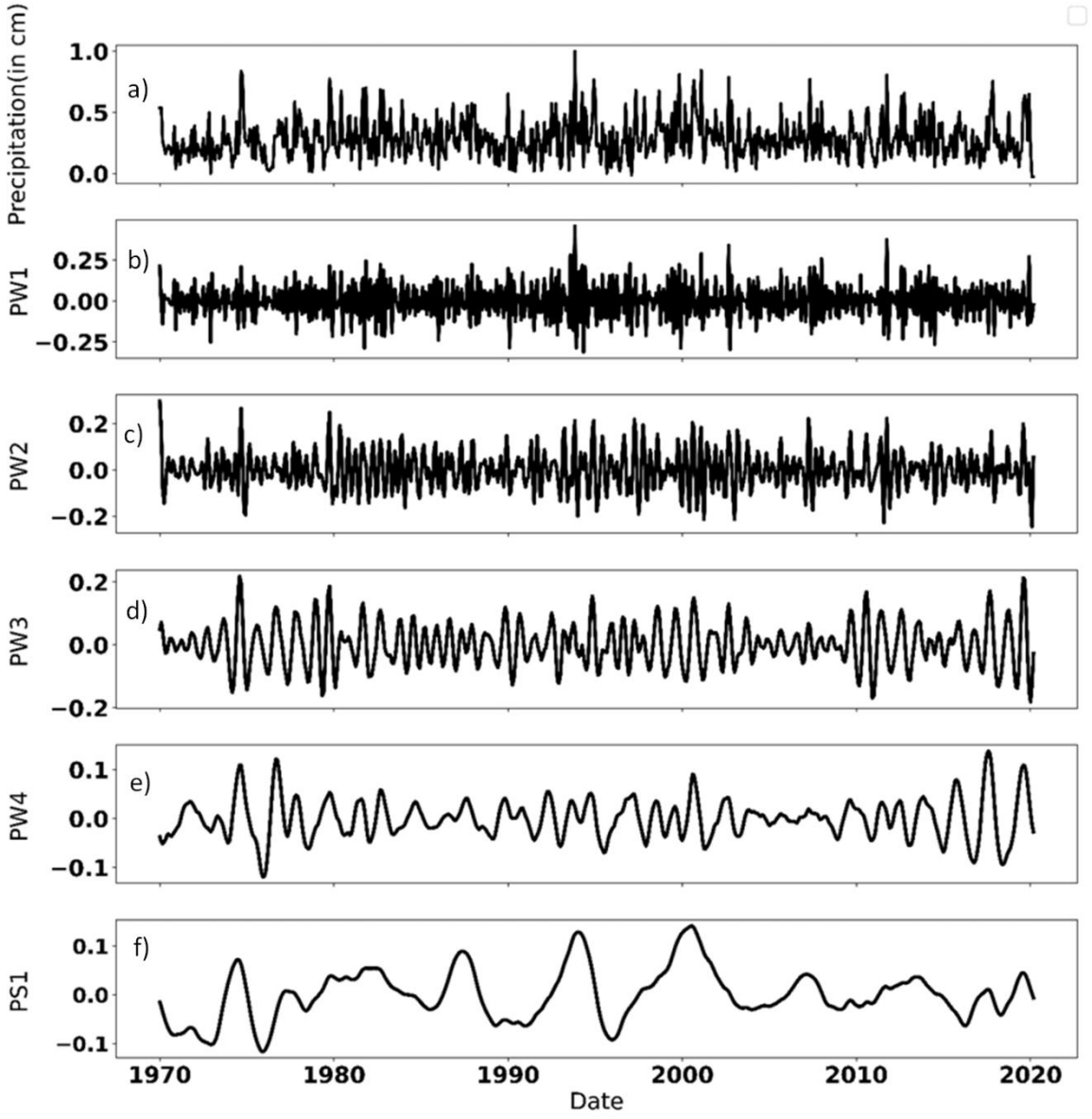

Figure 2.5: Decomposed components of precipitation with la10 wavelet. (a) Original signal of precipitation b-e) Wavelet coefficients (PW1 to PW4) f) Scaling coefficients

coefficients, ensured the preservation of the phase alignment between the data and the calculated wavelet and scaling coefficients at each scale as much as possible. Scaling coefficients are not retained at all levels, but only at the final level; in other words, all MODWT components except the last one correspond to wavelet coefficients, while the last component corresponds to scaling coefficients. Figure 2.5 shows the decomposed components of the precipitation at the conservative level. Further explanation of the MODWT and its main interest in hydrological applications can be found in Baulon et al. (2022a) and Massei et al. (2017); for full mathematical details, we refer to Cornish et al. (2006) and Percival and Walden (2000).

2.3.2.Recurrent-based deep neural networks models

The long-term memory deep neural network (LSTM,(Hochreiter & Schmidhuber, 1997)) is a recurrent network designed to overcome leakage gradient problems and preserve long-term dependence through the inclusion of a hidden state, c_t, that retains historical information. In addition, LSTM has internal mechanisms in the form of control gates—namely forget, input, and output gates—for regulating the flow of information.

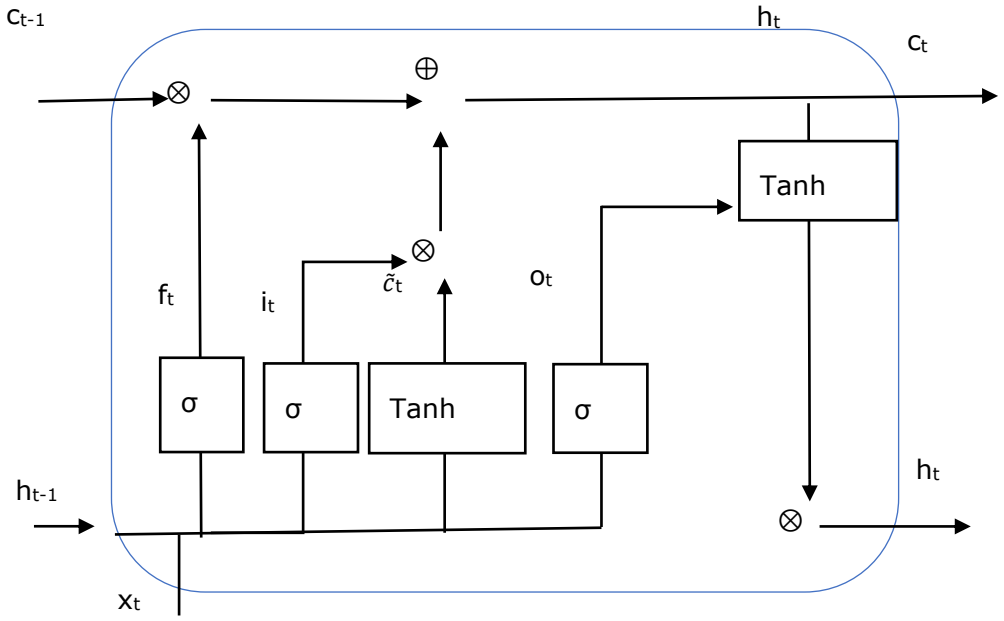

Figure 2.6: Schematic representation of the simple LSTM cell.

The flow of processed information in simple LSTM is described in Figure 2.6 and formulated mathematically below in three steps: In step 1, the forget gate manages the information from the prior cell state, c_{t-1}, while subsequently adding it to the present state with the

help of the element-wise multiplication operator ($\otimes$) in the form of $f_t \otimes c_{t-1}$. This gate gives the binary output [0,1], with 0 and 1 indicating the deletion and retention of all previous information, respectively.

$$f_t = \sigma\big(W_f x_t + U_f h_{t-1} + b_f \big). \quad (2)$$

In the second step, the present cell state is computed in three phases. The first phase involves converting the values of x_t and h_{t-1} into the range of [-1,1] to obtain a new cell state, $\tilde{c}_t$, using an activation function (tanh).

$$\tilde{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c). \quad (3)$$

In the second phase, values resulting from the input gate (i_t) are used to reorganise the present cell state, c_t, as $i_t \otimes \tilde{c}_t$. The input gate regulates both the sequence of input data at present (x_t) and the hidden state information at t–1 (h_{t-1}), which is incorporated into the cell state as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i). \quad (4)$$

In the final phase, the new cell state, c_t, is obtained by adding a revised cell state in step 1 ($f_t \otimes c_{t-1}$) with the updated cell state in previous phases (2. b) ($i_t \otimes \tilde{c}_t$).

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t. \quad (5)$$

In the third and last step, the information in the new cell state that must pass as an output of the present LSTM and the new hidden state to the upcoming cell is managed by the output gate (Alizadeh et al., 2021).

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \quad (6)$$

$$h_t = o_t \otimes tanh(c_t), \quad (7)$$

Where W and U are the network weight matrices; f_t, i_t, and o_t constitute the output of the forget, input, and output gates, respectively; c_t, c_{t-1} are the cell states at t and t−1; σ is the sigmoid activation function; h_t and h_{t-1} are the current and previous hidden states, respectively; $\tilde{c}_t$ is the cell candidate value; and b is the bias vector.

Bidirectional LSTM trains two LSTM models, as shown in Figure 2.7. The first model learns the input sequence, i.e., through the forward state. In contrast, the second model learns from the opposite direction of the input sequence, i.e., through the backward state (Saeed et al., 2020), as depicted in Figure 2.7.

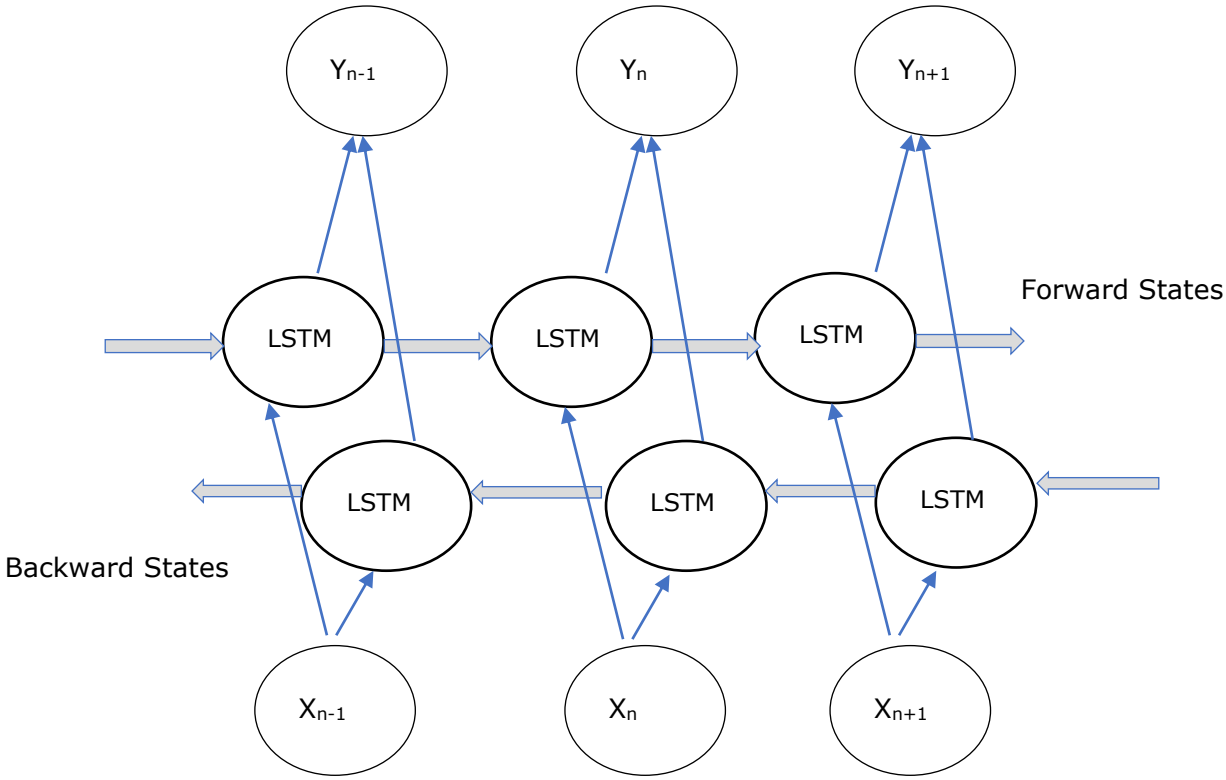

Figure 2.7: Schematic representation of the bidirectional LSTM (adapted and modified from Saeed et al. (2020)).

Both models are merged using the concatenation mechanism by default. In other words, BiLSTMs include an additional layer of training data compared with simple LSTMs. Siami-Namini et al. (2019) showed that BiLSTMs outperformed regular LSTMs and ARIMA in time series forecasting because of the additional training layer in BiLSTM, which improves the

learning of long-term dependencies. The internal processes of the cells used in BiLSTMs are explained above for the LSTM.

The gated recurrent unit (GRU) was developed by (Cho et al., 2014) to address the complexity of LSTM and improve computational efficiency. While GRU is relatively similar to LSTM as shown in Figure 2.8, unlike LSTM, it has on gates (i.e., reset (r_t) and update (u_t)), as it does not have separate memory cells. The cell structure of GRU is shown in Figure 2.8. GRU and, LSTM are modified versions of RNN, where RNN is a type of ANN used for sequential data. Moreover, the GRU is computationally less expensive, with a faster learning curve owing to the lower number of learnable parameters.

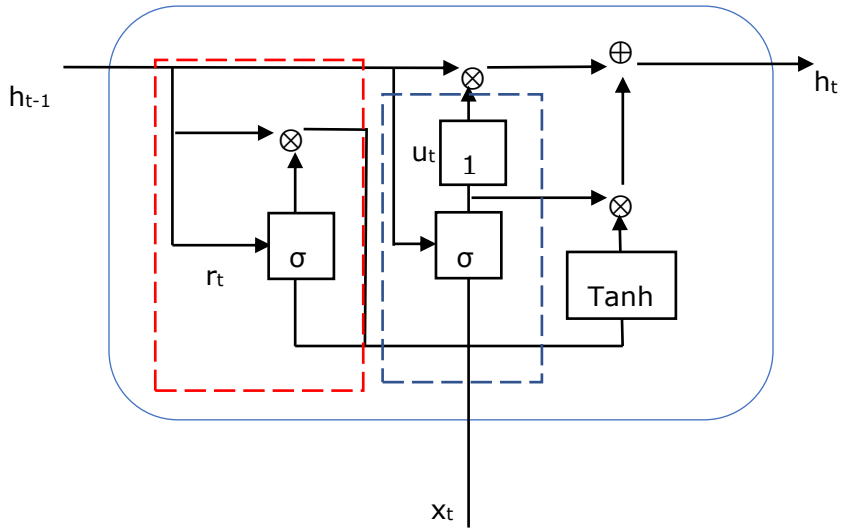

Figure 2.8: Schematic representation of the simple GRU cell.

The GRU phases are summarised as follows:

$$u_t = \sigma(W_u x_t + U_u h_{t-1} + b_u), \tag{8}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), \tag{9}$$

$$\tilde{c}_t = tanh(W_c x_t + U_c(r_t \otimes h_{t-1}) + b_c), \tag{10}$$

$$c_t = (1 - u_t) \otimes c_{t-1} + u_t \otimes \tilde{c}_t, \tag{11}$$

Where u_t and r_t are the update and reset gates, respectively; and b is the bias vector.

2.3.3.Hyperparameter tuning

Data-driven models involve using hyperparameters that must be defined during model initialisation. Hyperparameters play a crucial role in controlling the overall training behaviour of the model and substantially impact its performance; therefore, it is necessary to determine their optimal values.

Table 2.1: Hyperparameter values.

Hyperparameter	Value considered
Sequence length	48
Dropout	0.2
Optimizer	ADAM
Early stopping	50
Number of layers	(1,2,3,4,5,6)
Hidden neurons	(10, 20, …,100) by 10
Learning rate	(0.001,0.01) (log values)
Batch size	(16, 32, …,256) by powers of 2
Epoch	(50, 100, …,500)

There are multiple approaches used in data-driven models for optimising hyperparameters, the most common of which are the trial-and-error approach (Zhang et al., 2019), grid search (Afan et al., 2021), and random search. However, in recent years, an informed approach known as Bayesian optimisation has gained traction in hydrological forecasting (Barzegar et al., 2021; Quilty et al., 2022; Rahman et al., 2020; Wunsch et al., 2022a). The main advantage of this informed approach is that it takes less time to get comparable

results as in random search, as it learns from previous iterations while allowing automatic hyperparameter selection. Here, Bayesian optimisation was performed to minimise the mean squared error. The hyperparameters of all the models were tuned using Bayesian optimisation, with the range of values shown in Table 2.1.

Based on hyperparameter space exploration, as well as on previous works revealing substantial hydrological variabilities at scales between 2–4 and 5–9 years over metropolitan France (Baulon et al., 2022a; Fossa et al., 2021), northern France (Massei et al., 2010, 2017), and even Great Britain (Rust et al., 2019), the sequence length was eventually set to 48 months (i.e., four years), as shown in Table 2.1. Such variability is immediately evident, particularly in Figure 2.2 *(a, c),* and was found to originate from precipitation and large-scale climate variability in the studies referenced earlier. For instance, such variability was also found in the North Atlantic Oscillation index (Massei et al., 2010; Massei & Fournier, 2012) or more directly related to particular sea-level pressure patterns over the Euro-Atlantic sector(Massei et al., 2017). The dropout approach was adapted to avoid overfitting the models during training. We subsequently tested values from 0.1 to 0.5 within the Bayesian optimisation but noticed minimal changes, so we finally opted for the commonly used dropout value of 0.2, i.e., dropping out 20% of the neurons.

After identifying the best hyperparameters from 100 trials of Bayesian optimisation using the validation set, we trained an ensemble of 30 pseudo-randomly initialised models and fitted them to the training set. This was performed to deal with epistemic uncertainty, which was mainly due to the initial model weights obtained from the random number generator seed; in other words, it addressed the uncertainties generated by the model structure. These ensemble models were tested on an unseen test set to evaluate their performances. Simulations obtained with different initialisations were used to compute confidence intervals. The confidence interval limits were computed by adding and subtracting the 1.96 times standard deviation to the mean of the resulting distribution at each timestep.

2.3.4. Evaluation and interpretation

The current framework involved two models: 1. MODWT-assisted, i.e., including MODWT pre-processing, and 2. standalone models, i.e., without pre-processing. Three common performance evaluation statistics were chosen to evaluate the models on both the training and test sets, namely the mean absolute error (MAE), root-mean-squared error (RMSE), and squared Pearson's correlation coefficient (R^2). Furthermore, the RMSE percentage

change for MODWT-assisted models was computed with respect to standalone models to compare the performance improvement.

$$MAE = \frac{1}{N}\sum_{i=1}^{n} |GWL_{sim} - GWL_{obs}|, \qquad 0 \leq MAE < +\infty, \tag{12}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(GWL_{sim} - GWL_{obs})^2}{n}}, \qquad 0 \leq RMSE < +\infty, \tag{13}$$

$$R^2 = 1 - \frac{\sum(GWL_{sim} - GWL_{obs})^2}{\sum(GWL_{sim} - GWL_{obs_{mean}})^2}, \qquad -\infty < R^2 < 1 \tag{14}$$

$$\%\ \text{Improvement in RMSE} = \left(\frac{\text{RMSE}_{\text{SA}} - \text{RMSE}_{\text{WT}}}{RMSE_{SA}}\right) * 100. \qquad 0 < \%\text{Improv in RMSE} < 100 \tag{15}$$

Finally, the SHAP approach (Lundberg & Lee, 2017) was implemented to interpret the results. Shapley additive explanation (also known as SHAP) is a game-centric approach gaining traction in interpreting DL models. The SHAP summary plot helps to explain the contribution of each input feature to the final simulated/predicted value through two major aspects: i) the relative importance of each variable through the magnitude of the effect, wherein a higher mean of absolute SHAP value means a higher influence and ii) the direction of the influence, whether it is a positive or negative relationship, where a higher number of points on the right side (indicated in red) shows positive relationships, and vice versa. The basic framework of the methodology is shown in Figure 2.9.

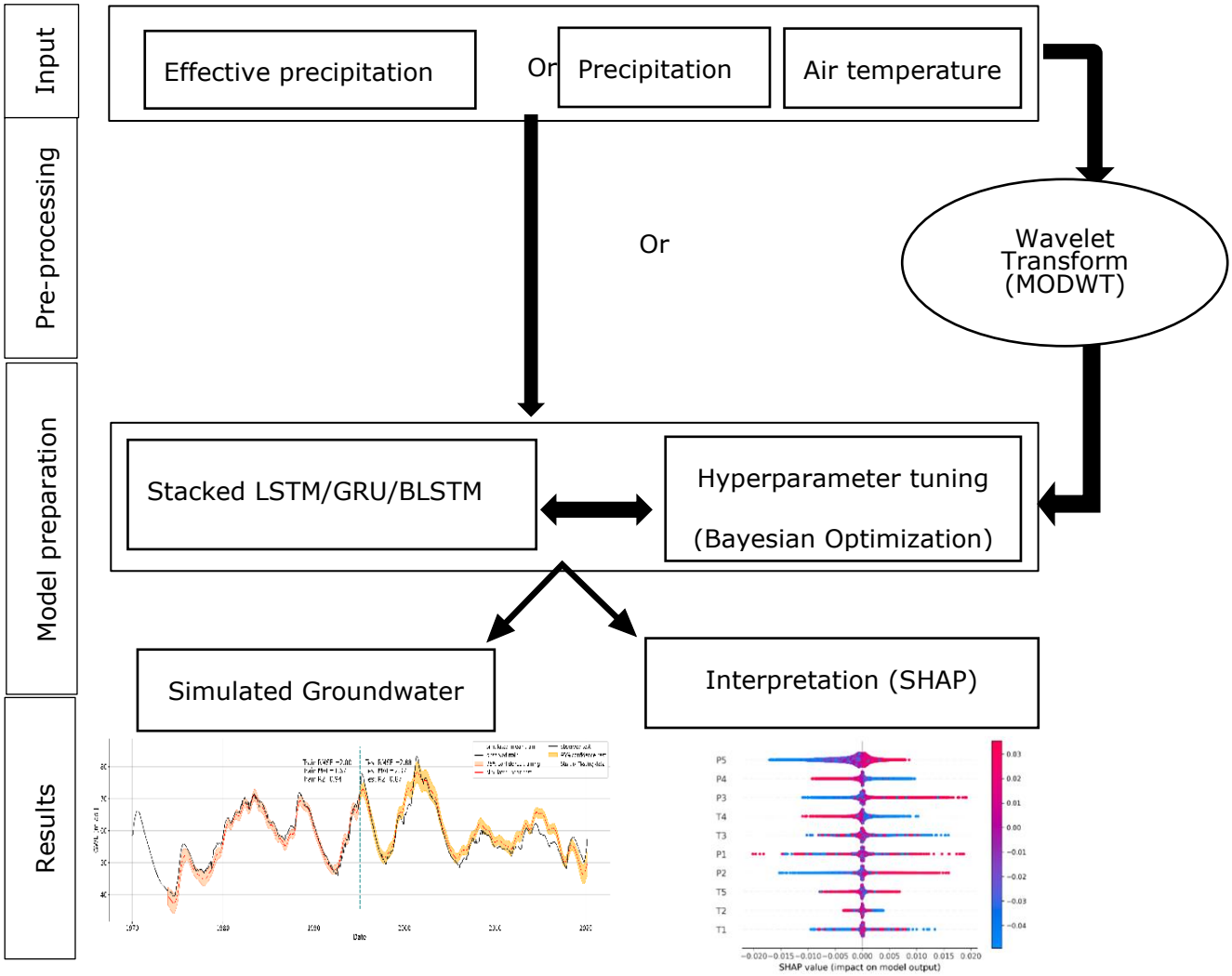

Figure 2.9: Basic framework of single station methodology

2.3.5. Software used

Deep-learning models were built using TensorFlow (Abadi et al., 2016) and Keras (Chollet, 2015). All figures were prepared using Matplotlib (Hunter, 2007), pandas (McKinney, 2010), and NumPy (Van Der Walt et al., 2011). Bayesian optimisation was performed using the Optuna software(Akiba et al., 2019). All this work was conducted in Python version 3.8.13, using a Dell workstation with an NVIDIA Quadro RTX 5000 GPU and 128GB RAM.

2.4. Performance and interpretability of the developed models

2.4.1. On the performances of standalone and MODWT-assisted models for GWL simulations

All MODWT-assisted and standalone models using different inputs were tested to simulate the GWLs of three types of variability. Figure 2.10, Figure 2.11, and Figure 2.12 present the comparison between observed and simulated GWL for each variability type (i.e., inertial, mixed, and annual) for different models and input types; the yellow-shaded region represents the 95% confidence interval obtained from simulations with different initialisations (as explained in section 2.3.3). The red line shows the mean of these different simulations, whereas the black line represents the observed GWL. In addition, Table 2.2 & Table 2.3 show a comparison summary of the performance evaluation metrics (i.e., MAE, RMSE, and R2) on the test set for each of the three DL methods tested (i.e., GRU, BiLSTM, and LSTM) for each variability type. Optimal hyperparameters obtained for each of the tests are shown in Table S10-S12.

From the overall results, the following key aspects are evident. For the inertial type, in simulations using precipitation and air temperature as input, the performance of the MODWT-assisted models ($1.52 < RMSE < 1.80$; Figure 2.10d-2.10f) is substantially improved compared to that of the standalone models ($2.46 < RMSE < 2.93$; Figure 2.10a-2.10c). In addition, the MODWT-GRU model (RMSE = 1.52; Figure 2.10d) outperformed the standalone GRU (RMSE = 2.93; Figure 2.10a). The standalone GRU performed similarly in both cases, i.e., with EP (Figure 2.10g) and precipitation and air temperature as input, with RMSEs of 2.89 and 2.93, respectively. In contrast, improvement in MODWT-assisted models was only apparent with precipitation and air temperature as input. So, in the case of the inertial type of GWLs, models with raw data (i.e., precipitation and air temperature) as input (Figures 2.10a-2.10f) outperformed models with processed data (i.e., effective precipitation) (Figures 2.10g-2.10l)

Table 2.2: Performance metric comparison for models with effective precipitation as input.

GWL Type	Metrics	EP			EP-WT		
		GRU	LSTM	BILSTM	GRU-WT	LSTM-WT	BILSTM-WT
Inertial	R2	0.15	-0.57	-0.13	-0.09	-0.01	-0.04
	MAE	2.59	3.61	3.03	2.93	2.73	2.96
	RMSE	2.89	3.93	3.33	3.70	3.14	3.20
	RMSE improv. (%)**	-	-	-	-13.30	19.97	3.91
Annual	R2	0.82	0.84	0.86	0.83	0.85	0.86
	MAE	1.54	1.51	1.41	1.55	1.43	1.33
	RMSE	1.92	1.82	1.72	1.86	1.74	1.69
	RMSE improv. (%)	-	-	-	3.15	4.55	2.10
Mixed	R2	0.43	0.66	0.50	0.74	0.71	0.59
	MAE	1.36	1.01	1.19	0.87	0.98	1.12
	RMSE	1.76	1.37	1.65	1.19	1.26	1.51
	RMSE improv. (%)	-	-	-	32.18	8.29	8.68

** RMSE improv. (%) is an improvement in RMSE in the MODWT-assisted model against the corresponding standalone model.

Table 2.3:Performance metrics comparison for models with precipitation and air temperature as input.

GWL Type	Metrics	PT			PT-WT*		
		GRU	LSTM	BILSTM	GRU-WT	LSTM-WT	BILSTM-WT
Inertial	R2	0.12	0.25	0.38	0.76	0.71	0.67
	MAE	2.30	1.73	1.72	1.17	1.28	1.44
	RMSE	2.93	2.71	2.46	1.52	1.67	1.80
	RMSE improv. (%)	-	-	-	48.09	38.25	26.83
Annual	R2	0.74	0.10	0.62	0.79	0.75	0.71
	MAE	1.89	3.62	2.28	1.63	1.71	1.93
	RMSE	2.30	4.28	2.79	2.06	2.25	2.44
	RMSE improv. (%)	-	-	-	10.58	47.39	12.54
Mixed	R2	0.57	0.48	0.59	0.58	0.54	0.64
	MAE	1.15	1.27	1.17	1.24	1.31	1.06
	RMSE	1.53	1.68	1.50	1.52	1.59	1.41
	RMSE improv. (%)	-	-	-	0.74	5.17	6.00

* Here, PT-WT indicates precipitation and air temperature as input along with the wavelet transform.

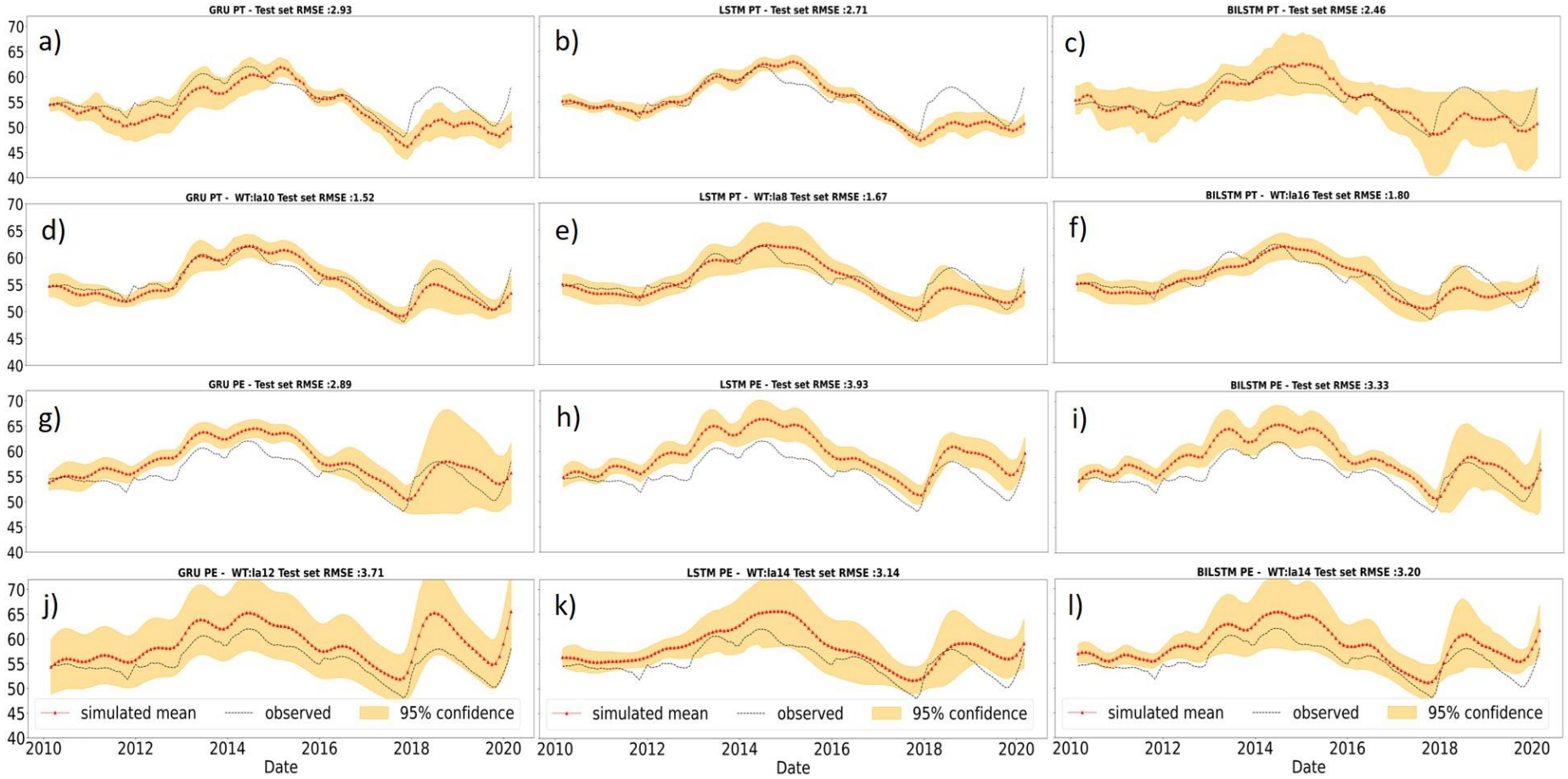

Figure 2.10: Results obtained for different inputs for inertial type with GRU, LSTM, and BILSTM: precipitation and air temperature (PT)(a-c), PT with wavelet transform (PTWT)(d-f), effective precipitation (PE)(g-i), and PE with wavelet transform (PEWT)(j-l). Red lines indicate the mean of the simulations and black lines indicate the observed GWL. Yellow shading represents the 95% confidence interval.

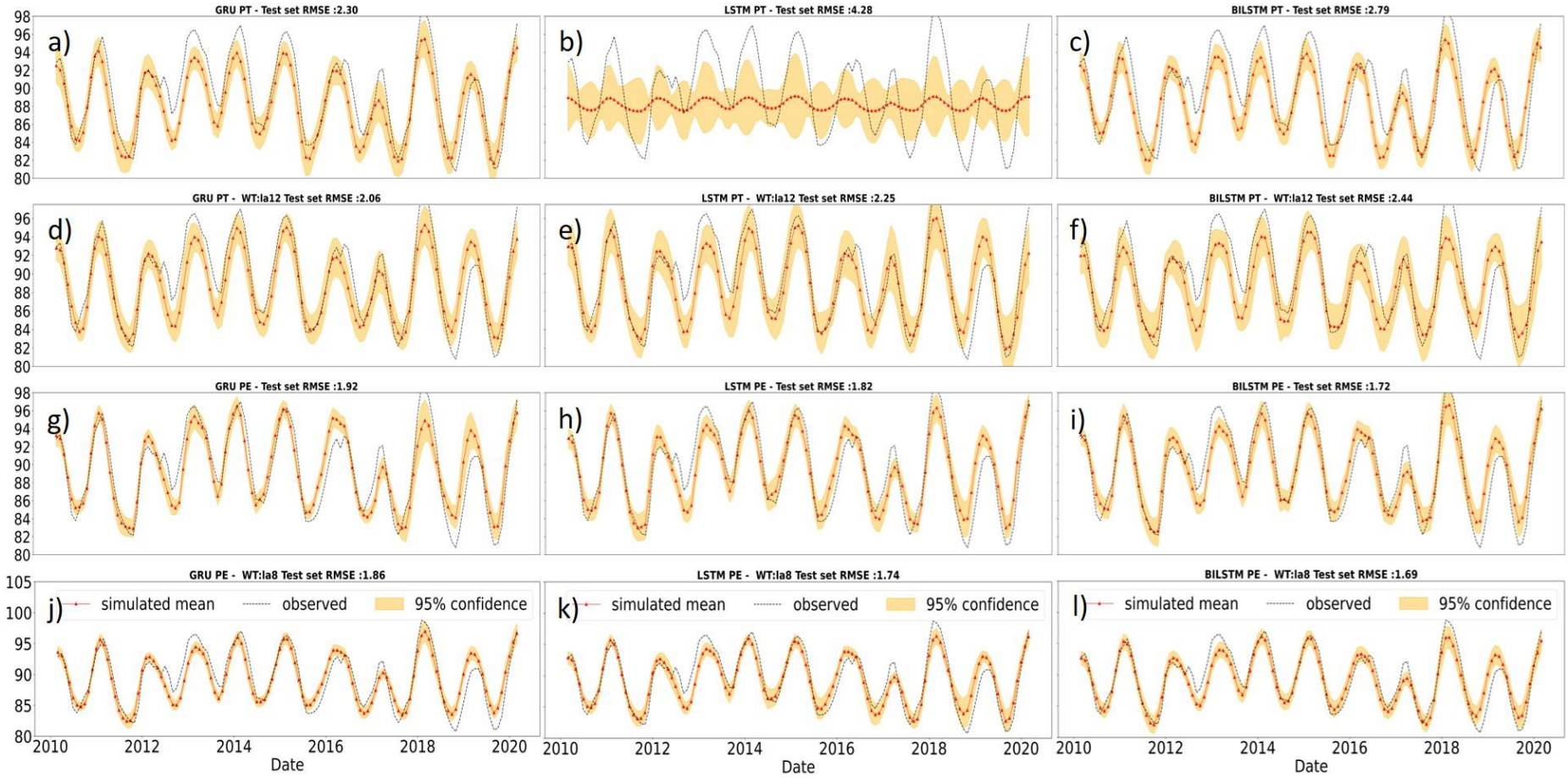

Figure 2.11: Results obtained for the annual type with GRU, LSTM, and BILSTM: precipitation and air temperature (PT)(a-c), PT with wavelet transform (PTWT)(d-f), effective precipitation (PE)(g-i), and PE with wavelet transform (PEWT)(j-l). Red lines indicate the mean of the simulations and black lines indicate the observed GWL. Yellow shading represents the 95% confidence interval.

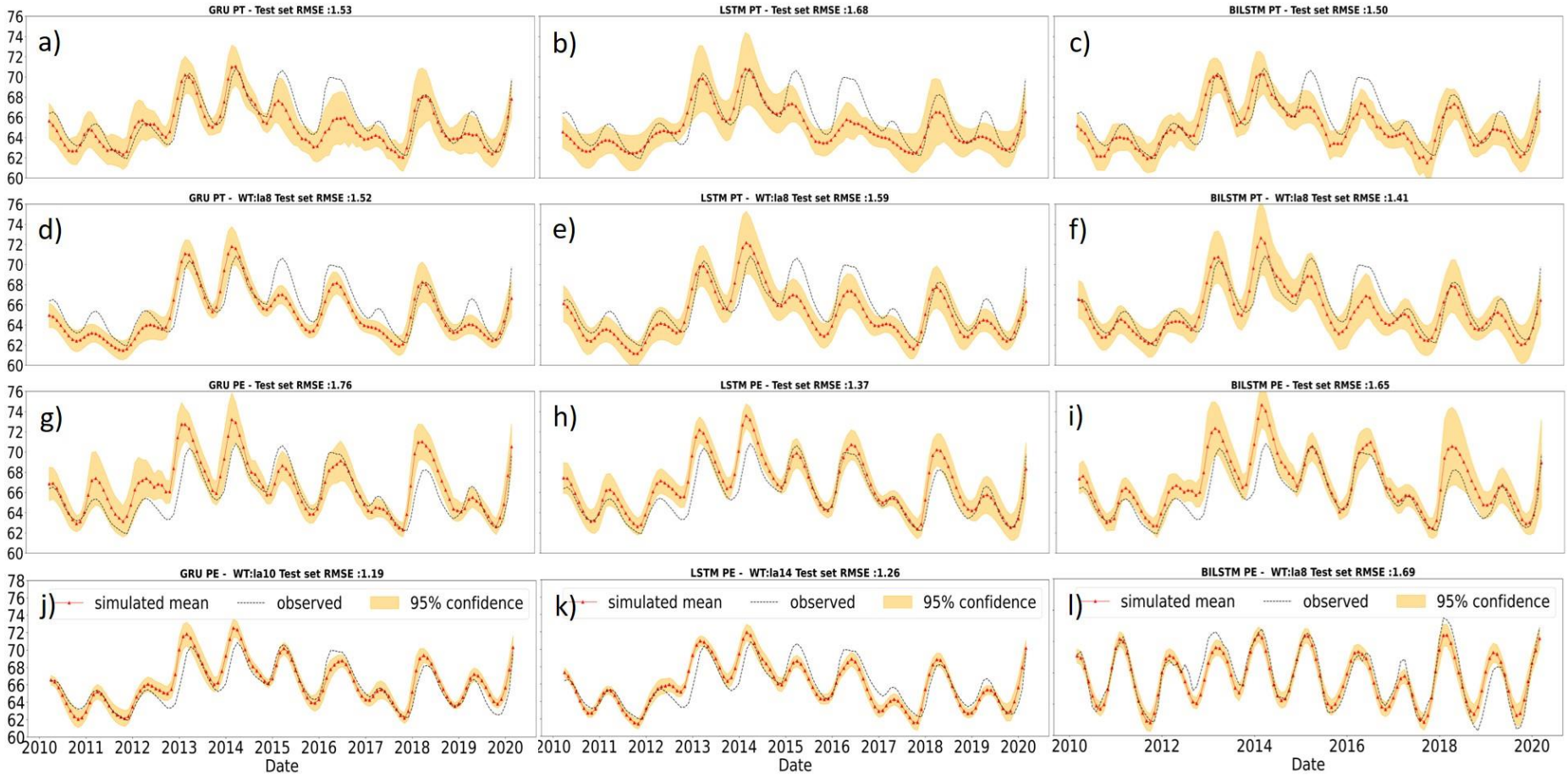

Figure 2.12: Results obtained for the mixed type with GRU, LSTM, and BILSTM: precipitation and air temperature (PT)(a-c), PT with wavelet transform (PTWT)(d-f), effective precipitation (PE)(g-i), and PE with wavelet transform (PEWT)(j-l). Red lines indicate the mean of the simulations and black lines indicate the observed GWL. Yellow shading represents the 95% confidence interval.

For the annual type, both the standalone (1.72<RMSE<1.92; Figures 2.11g-2.11i) and MODWT-assisted models (1.69<RMSE<1.86; Figures 2.11j-2.11l) with EP as input always performed better than those with precipitation and air temperature as input (Figures 2.11a-2.11f). However, all these models resulted in better simulations, with the differences among these models with EP being very small. On the other hand, standalone models (2.30<RMSE<4.28; Figures 2.11a-2.11c) with precipitation and air temperature as input yielded less accurate results; the MODWT-assisted models (2.06<RMSE<2.44; Figures 2.11d-2.11f) improved the results but were not better than those obtained with EP as input.

For the mixed type, such as the annual, models with EP as input performed well; however, there were improvements in MODWT-assisted models (1.19<RMSE<1.51; Figures 2.12j-2.12l) with EP as input as against standalone models (1.37<RMSE<1.76; Figures 2.12g-2.12i) with EP as input. However, MODWT-assisted models (1.41<RMSE<1.59; Figures 2.12d-2.12f) with precipitation and air temperature gave comparable results with standalone models (1.50<RMSE<1.68; Figures 2.12a-2.12c) with little to no improvement.

To summarise, the improvement in the inertial type of GWL simulations using precipitation and air temperature in MODWT-assisted models was consistently high against standalone models, with an improvement in mean RMSE ranging between 26% and 48%, as shown in Table 2.3. While improvement varied from 10% to 47% for the annual type, it was much lower for the mixed type, varying from 0.7% to 6%. On the other hand, Using EP, MODWT-assisted models improved consistently only with annual (8%-32%) and is very minor in mixed type (2%-4%). Assuming that we can process data (i.e., EP), the best combination, along with metrics, is shown in Table 2.4. When only precipitation and air temperature are available, the best combination, along with metrics, is shown in Table 2.5. Table 2.6 shows the best model combination for each GWL variability. This table shows that GRU combined with MODWT decomposed inputs (precipitation and air temperature) leads to the best results for the inertial type. Conversely, BiLSTM combined with MODWT with effective precipitation was superior for the annual type. Finally, GRU combined with MODWT and effective precipitation gave better results for mixed type. But overall, GRU seemed to perform better than other models.

Table 2.4: Best model combinations for each type of GWL, provided that effective precipitation is available.

GWL variability	Pre-Processing	Model	Metrics
Annual	EP-WT	BILSTM	R^2=0.86, MAE=1.33, RMSE=1.69
Inertial	EP	GRU	R^2=0.15, MAE=2.59, RMSE=2.89
Mixed	EP-WT	GRU	R^2=0.74, MAE=0.87, RMSE=1.19

Table 2.5: Best model combinations for each type of GWL when only raw data (i.e., precipitation and air temperature) are available.

GWL variability	Pre-processing	Model	Metrics
Annual	PT-WT	GRU	R^2=0.79, MAE=1.63, RMSE=2.06
Inertial	PT-WT	GRU	R^2=0.76, MAE=1.17, RMSE=1.52
Mixed	PT-WT	BILSTM	R^2=0.64, MAE=1.06, RMSE=1.41

Table 2.6: Overall best models for each type of GWL variability

GWL variability	Best model	Metrics
Inertial	GRU-PT-WT	R^2=0.76, MAE=1.17, RMSE=1.52
Annual	BILSTM-EP-WT	R^2=0.86, MAE=1.33, RMSE=1.69
Mixed	GRU-EP-WT	R^2=0.74, MAE=0.87, RMSE=1.19

It is noticeable that even though effective precipitation considers both precipitation and air temperature in its computation, it did not appear capable of capturing the low-frequency variability even after MODWT decomposition. Instead, precipitation and air temperature inputs combined with MODWT-based input pre-processing seemed necessary to achieve the best simulations. Although the results indicate that MODWT-assisted models perform well, we did not find any consistent way to select the most appropriate filter length among those used in our study.

2.5. Towards a deeper investigation of what and how the models learn

Figure 2.13a shows the simple summary plot of standalone models with precipitation and air temperature. Figure 2.13b shows the SHAP summary plot for the inertial GWL type, indicating the order of importance of each decomposed component. Here, P1 to P5 represent the MODWT components of precipitation, i.e., P1 to P4 are wavelet coefficients, with P1 and P4 being the highest and lowest-frequency coefficients, respectively, while P5 represents the retained scaling coefficient. The same applies to the air temperature components T1–T5. Using the terminology defined earlier, SHAP summary plots were generated for all three GWL variability types, and three DL models were tested, as shown in Figure 2.14. The SHAP summary plots show that the order of importance of different

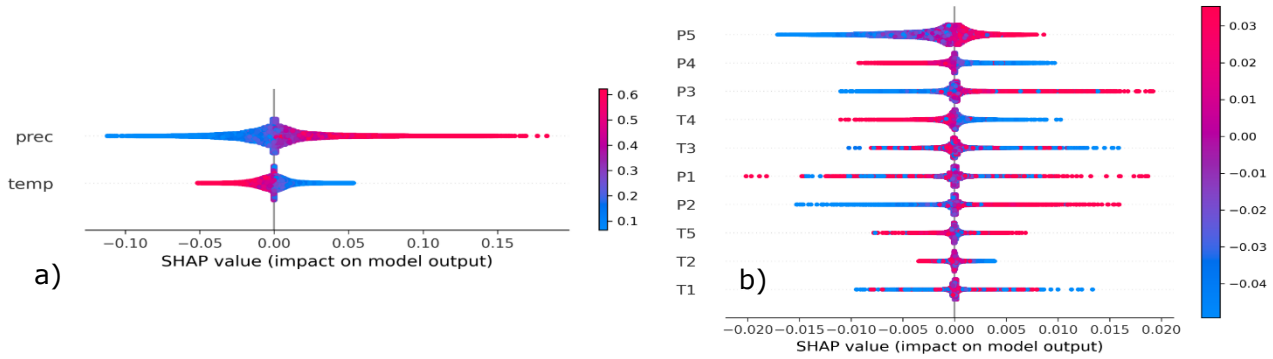

Figure 2.13: SHAP summary results with feature importance of each variable in the GRU: a) with precipitation and air temperature (PT) as input and b) with PT with la10 wavelet as input for the inertial type. Here, P1 to P4 and T1 to T4 represent wavelet coefficients for precipitation and air temperature, respectively. P5 and T5 represent scaling coefficients of precipitation and air temperature.

features was the same for LSTM and GRU but was slightly different for the bidirectional LSTM. This might be attributed to the basic difference in the architectures of LSTM, GRU, and BiLSTM, i.e., in LSTM and GRU, information flows in only one direction

In contrast, in bidirectional LSTM, information flows in both directions. For example, In mixed type, SHAP results showed that only BiLSTM could capture the low-frequency variability in precipitation (P5) as a major contributor to the simulations. Simulation results showed that, in this case, BiLSTM was the most efficient model (RMSE=1.41; Figure 2.12f).

The SHAP plots show that the inertial type always had low-frequency precipitation components as major contributors. We note that the difference in performance among the MODWT-assisted DL models is relatively small, indicating that retrieving useful information plays a more important role than the model itself. Our results (Figure 2.14) show that the models mainly use the air temperature input to explain/represent the annual periodic variability within GWL; the air temperature is mainly dominated by annual cyclicity. This is confirmed by the SHAP analysis, which clearly shows that the MODWT component (T3) corresponding to the annual time scale has the highest impact on simulating the annual GWL type. Barzegar et al. 2021 highlighted that LSTM is efficient in modelling seasonality, which explains why standalone models result in high accuracy in annual GWL time series but failed to perform accurate simulations in the case of low-frequency dominated (inertial) GWLs. However, as emphasised in Table 2.6, the best results for the so-called inertial GWL type could be achieved using precipitation and air temperature time series as input and a MODWT-assisted model (best results obtained with GRU). For low-frequency dominated GWL, the DL models must access low-frequency information in the input variables to learn from it. Our experiments highlighted that, in such cases, pre-processing using MODWT helped access the information required. Figure 2.15a shows the normalised precipitation used as input and the lowest-frequency component it contains (red line). Although the variability explained by the low-frequency P5 component remains relatively low compared to that of the original precipitation signal (standard deviation of resp. 0.05 and 0.16 in Figure 2.15a), it can be easily seen that the P5 component (displayed in its full range in Figure 2.15b) matches very well the targeted GWL time series (Figure 2.15c). One should notice here that in this case, although la10-based filters were used to ensure limited boundary effects, pre-processing still revealed capable of reaching the desired low-frequency information required to feed the model.

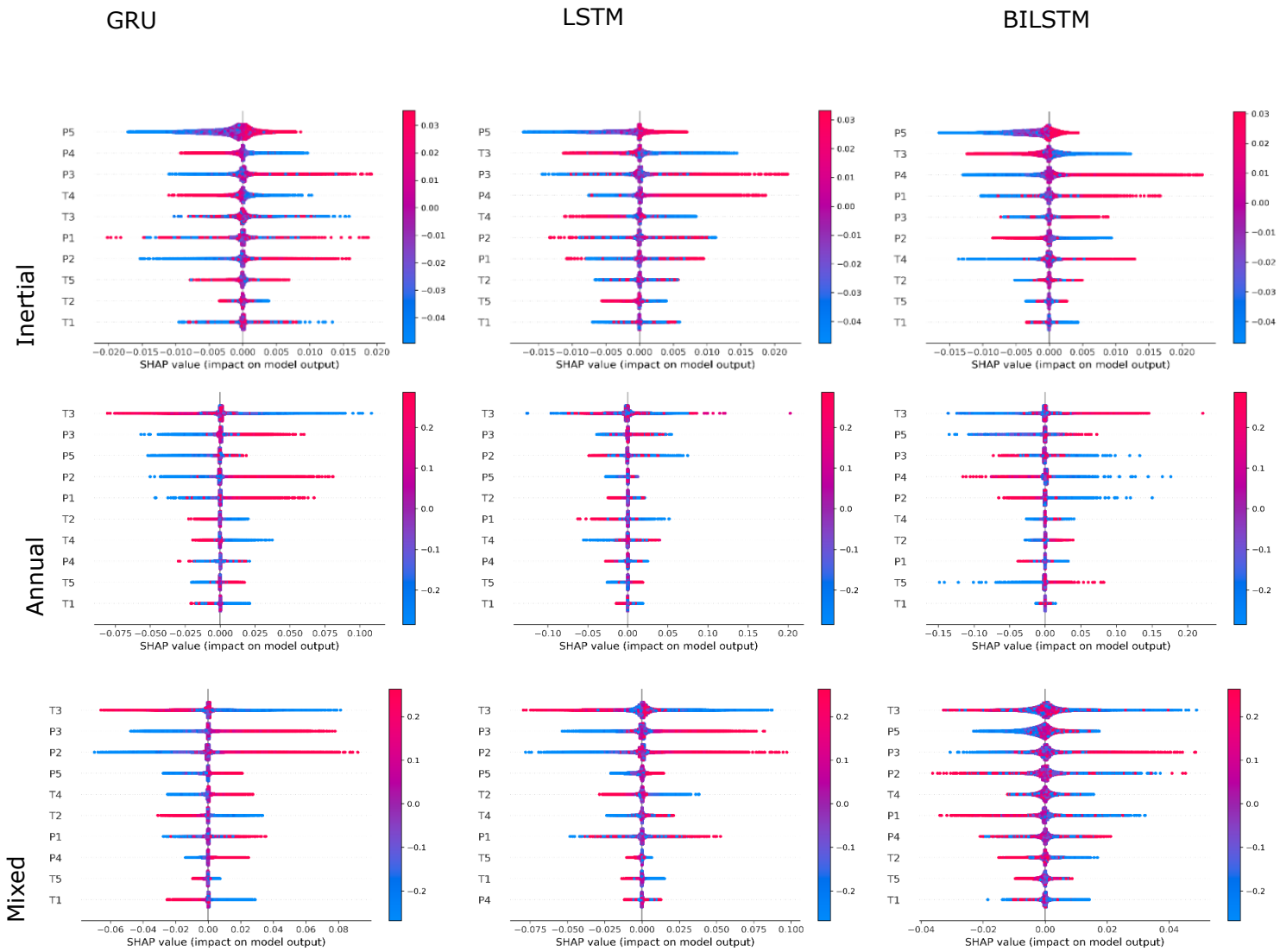

Figure 2.14: Comparison of SHAP summary plots for different types of GWL time series and different deep learning models. Here P1 to P4 and T1 to T4 represents wavelet coefficients for precipitation and air temperature respectively. P5 and T5 represents scaling coefficients of precipitation and air temperature.

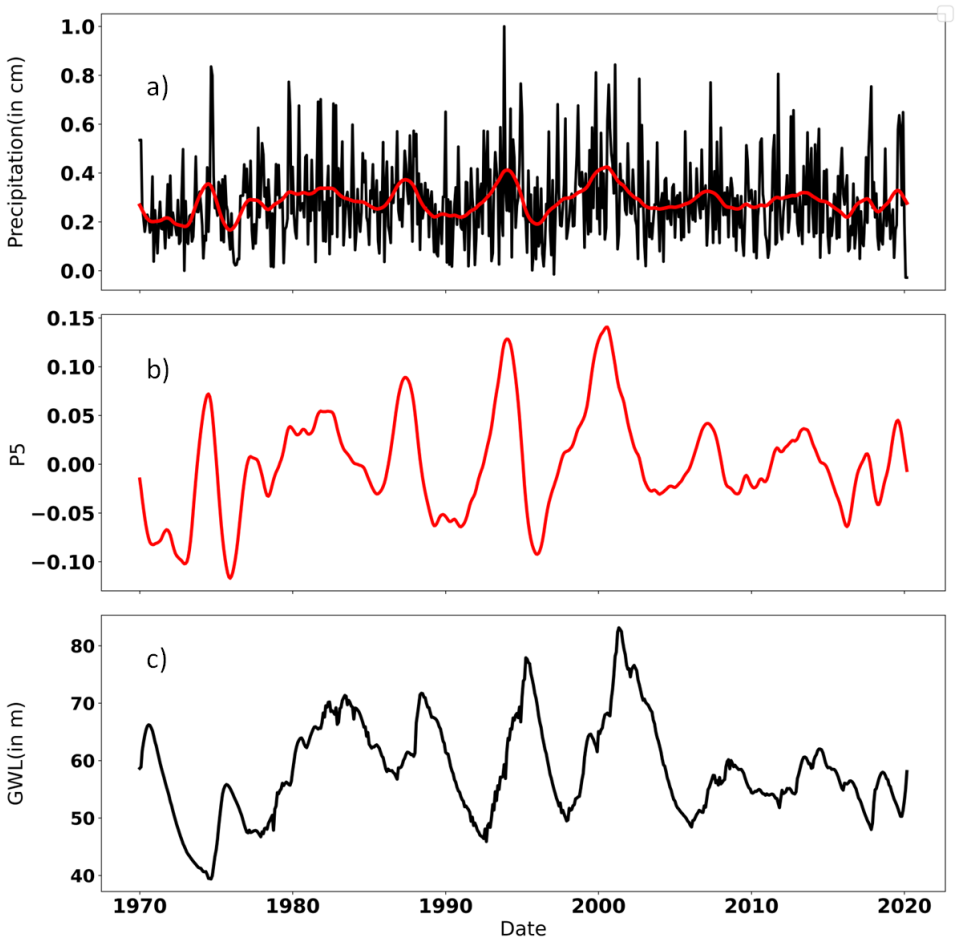

Figure 2.15: Comparison of a) precipitation, b) the last MODWT component (approximation) of precipitation (P5), and c) the original GWL timeseries.

The GWL data used in this study are relatively free from anthropogenic influences; hence, pumping was not considered. Nevertheless, for further generalisation, wherever possible, pumping data should also be considered. The GWL data can also be used for long-term historical reconstructions or direct downscaling using large-scale climate/atmospheric variables (e.g., NAO) and climate fields from reanalyses, which could also help study the hydrological evolution of GWL time series over the last century. In a benchmark study, Hagen et al. (2021) used machine learning models to identify drivers of atmospheric variables for the direct downscaling of streamflow and highlighted the need for such benchmark studies for DL. Similar studies focusing on interpretable DL for GWL downscaling are required to exploit the full potential of these models.

2.6. Concluding remarks

This study assessed the effectiveness of DL models for simulating different types of GWLs, including inertial, annual, and mixed. The DL models considered different input types, 1) precipitation and air temperature or 2) effective precipitation, and also considered the case where the inputs were pre-processed by the MODWT. The MODWT-assisted GRU seemed to perform well with all three types of GWLs, mainly when only precipitation and air temperature data were used as input. However, MODWT-assisted BiLSTM performed slightly better for mixed type; interestingly, although the improvement was very small (approximately 7% improvement in RMSE), SHAP results indicated the BiLSTM prioritised low-frequency better than GRU in this case. When effective precipitation was used as input, both standalone and MODWT-assisted models consistently gave better results for annual and mixed types but underperformed for inertial type (i.e., in such a case, using effective precipitation as input did not improve the simulation results). However effective precipitation relies on evapotranspiration, a complex variable by itself which may also be affected by high uncertainty owing to its assessment and computation. This is why capturing relevant information directly from source variables may be a better option. Nevertheless, the MODWT was still helpful in improving the results of all three models in almost all cases. Improvements were more substantial when using precipitation and air temperature as input, while they were much smaller when effective precipitation was used.

The whole framework presented in this study is flexible and reproducible, which means that each of the internal steps can be modified or replaced to check for further advancements, and can be used to adapt different types of GWL that are not currently considered here. For example, pre-processing steps can be replaced with other types of signal analysis and processing techniques, or interpretability with SHAP can be replaced with the local interpretable model-agnostic explanations or integrated gradient approach. In conclusion, using relevant information and pre-processing techniques, such as MODWT, helps DL models generate better GWL simulations.

The models presented in this study used only meteorological variables. However, in the current situation of growing concerns over frequent extreme events, such as heat waves and droughts, other input variables, such as climate indices or large-scale climate projections, should be considered when simulating the historical and future GWLs, thereby improving the decision-making processes. Hence, the models developed in the current study can project near long-term GWL simulations under different climate scenarios using GCM projections as input. In addition, this study addressed only the epistemic uncertainty

owing to the randomness in model weights. Future studies should also consider other sources of uncertainty, particularly in input data (Evapotranspiration model when using EP. Precipitation and temperature uncertainty associated with the reanalysis products used) using the current approach.

In light of this research, several questions can be raised that could serve as a basis for future studies. Does having additional variables (e.g., regional-scale hydrological/hydrometeorological variables, temporally static variables indicative of basins' physical properties, large-scale climate indices, or large-scale climate field variables from gridded reanalysis data) still require pre-processing to extract consistent information? Would other types of signal processing techniques (such as empirical/variational mode decomposed-based methods, successive LOESS or Savitzky-Golay smoothing of input data to extract low-frequency content) improve the simulations even further? Although approaches such as multi-basin training are being studied in rainfall-runoff modelling, they are yet to be explored in the context of GWL simulations. Future work should determine whether one global model is sufficient for learning various GWL variations across multiple sites simultaneously without developing a single model for each site. Concerning DL approaches, attention-based models are receiving increasing interest in recent years. However, further studies would then be required to assess whether such models would successfully account for low-frequency variations in GWLs without any help from signal pre-processing.

Chapter 3. Groundwater level reconstruction using long-term climate reanalysis data and deep neural networks

Building upon the learning from the wavelet-assisted deep learning approach for groundwater level simulations in Chapter 2, Chapter 3 extends this methodology to reconstruct long-term groundwater levels using climate reanalysis data and validate across northern France. By leveraging recurrent neural networks like LSTM, BiLSTM, and GRU in conjunction with wavelet pre-processing, we seek to evaluate the ability of DL models to capture multidecadal variability and reconstruct groundwater levels dating back to the early 20th century. This long-term reconstruction capability addresses the need for historical data and provides a valuable data reconstruction approach for assessing hydroclimatic trends and variability.

The results of this chapter were published in the Journal of Hydrology: Regional Studies.

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.2023.1016

Abstract

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. Long term GWL time series are now available for northern France, corresponding to extended versions of observational time series back to the 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.

3.1. 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 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 (Aquimod) 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 reconstructing 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 & Young, 2013), *"simulation is defined as quantitative reproduction of system behaviour without reference to observed output; "forecasting" is defined as the 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 & Young, 2013b) 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 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. (2022a) used convolutional neural networks, which allowed them to develop projections ('what-if' simulation according to (Beven & Young, 2013b)) 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., 2021b) 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.

3.2. Data and Methodology

3.2.1. Study Area

The study area considered is shown in Figure 3.1a 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.

3.2.2. 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 Figure 3.1b. The first GWL time series in Figure 3.1b 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).

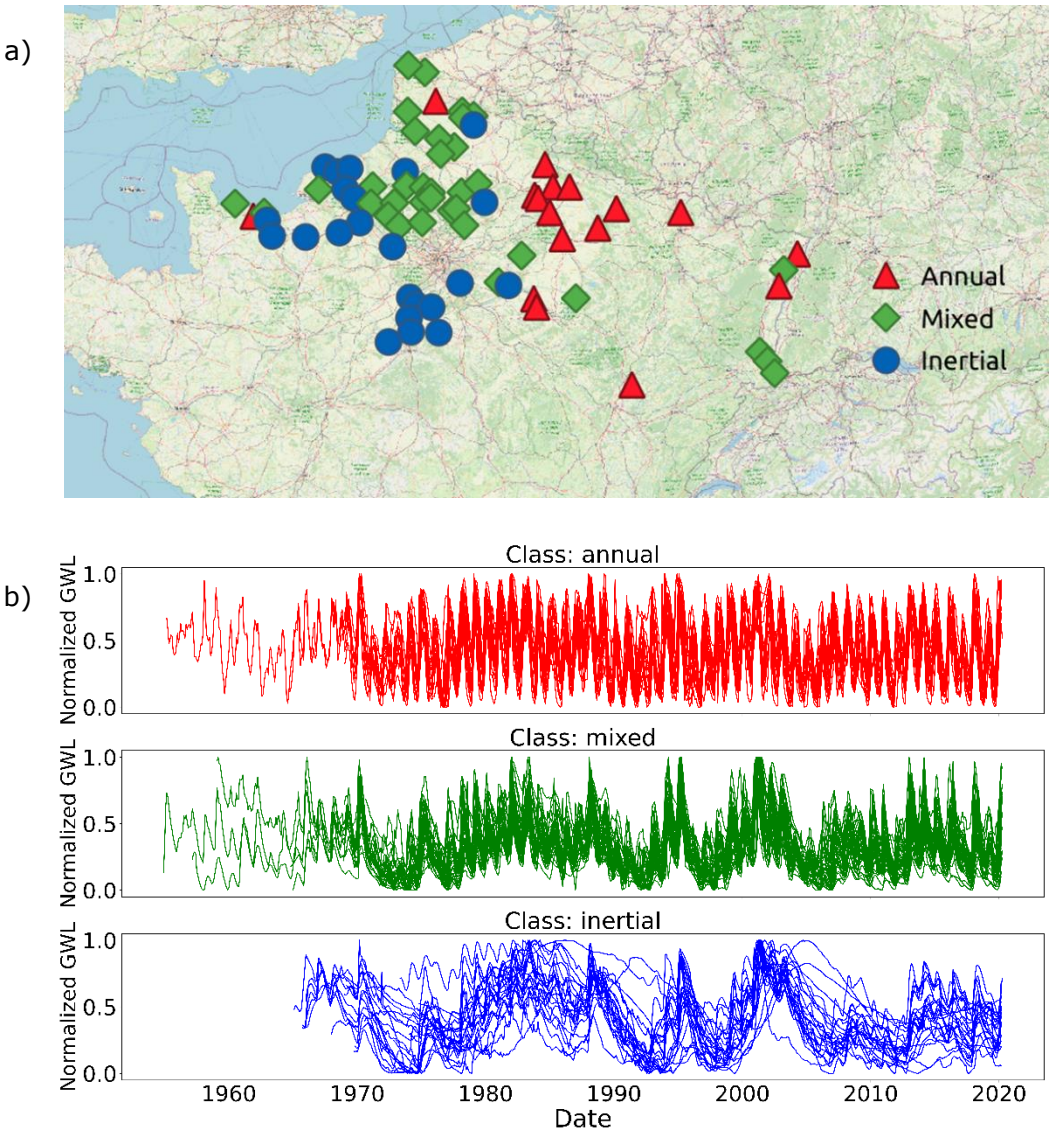

Figure 3.1: Classification of long-term (1970-2020) groundwater level (GWL) monitoring stations based on GWL types: annual (red), mixed (green), and inertial (blue). Figure shows GWL station locations (a) and normalised GWLs of all selected stations based on class (b).

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., 2020a), as summarised in Table 3.1. 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.

Table 3.1: 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		

The selected input variables 10m zonal (W-E) U-wind component (u10), 10m méridional (S-N) V-wind component (v10), 2m 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 degree (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 remain 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.

3.3. Methodology

3.3.1. 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., 2022a) 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 2.1 of (Chidepudi et al., 2023a). 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 and b; Massei et al., 2010 and 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., 2023a).

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., 2023a), following the best practices recommended by (Quilty & 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 study. The study also used a multimodal 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).

3.4. 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) (Figure 3.2). In the current study, KGE is considered over Nash-Sutcliffe efficiency (NSE) or other metrics as it is more comprehensive than NSE and combines three components of model error (correlation, bias, and 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 Figure 3.2 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., 2023a). Indeed, it is easy to see from Figure 3.2 that Wav LSTM/BiLSTM/GRU always had higher KGE values. Figure 3.2 also shows that BiLSTM with wavelet pre-processing would

slightly outperform the other models in many cases (orange lines in Figure 3.2), suggesting that this approach is effective in capturing long-term temporal patterns.

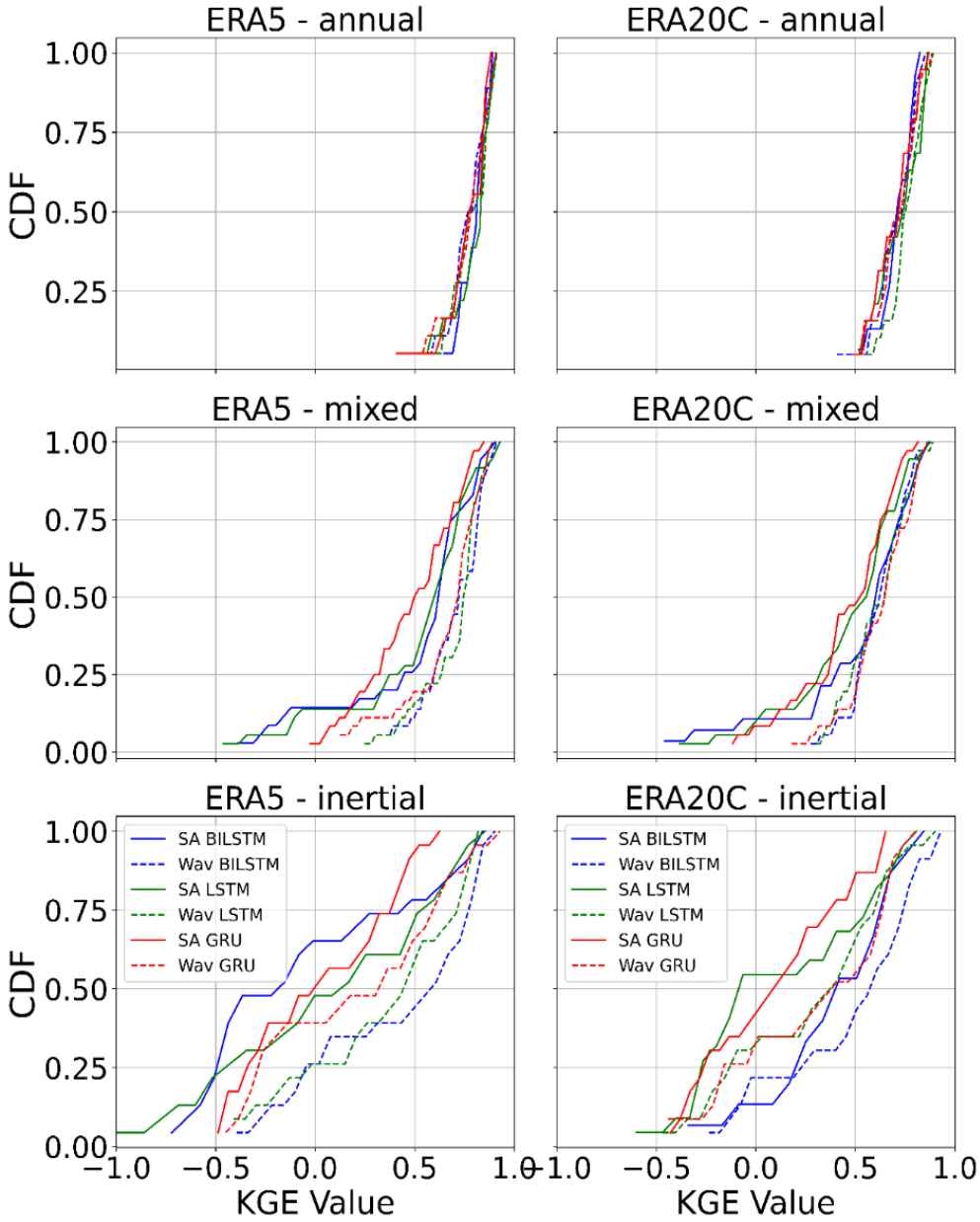

Figure 3.2: Cumulative density functions of the KGE for standalone (SA) and wavelet-assisted (Wav) models for both ERA5 (first column) and Model is represented by a different colour: BILSTM (blue), LSTM (green), and GRU (red) for standalone models and corresponding dotted lines for wavelet models.

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 Figure 3.3. 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.

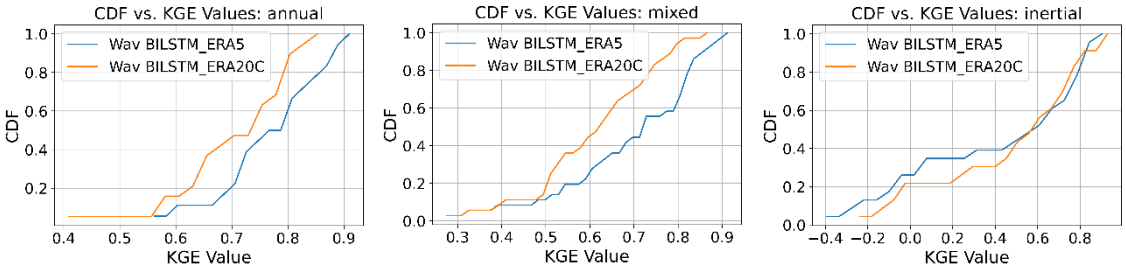

Figure 3.3: CDF comparison of KGE values of the BiLSTM with wavelet (La8) for different GWL types.

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.

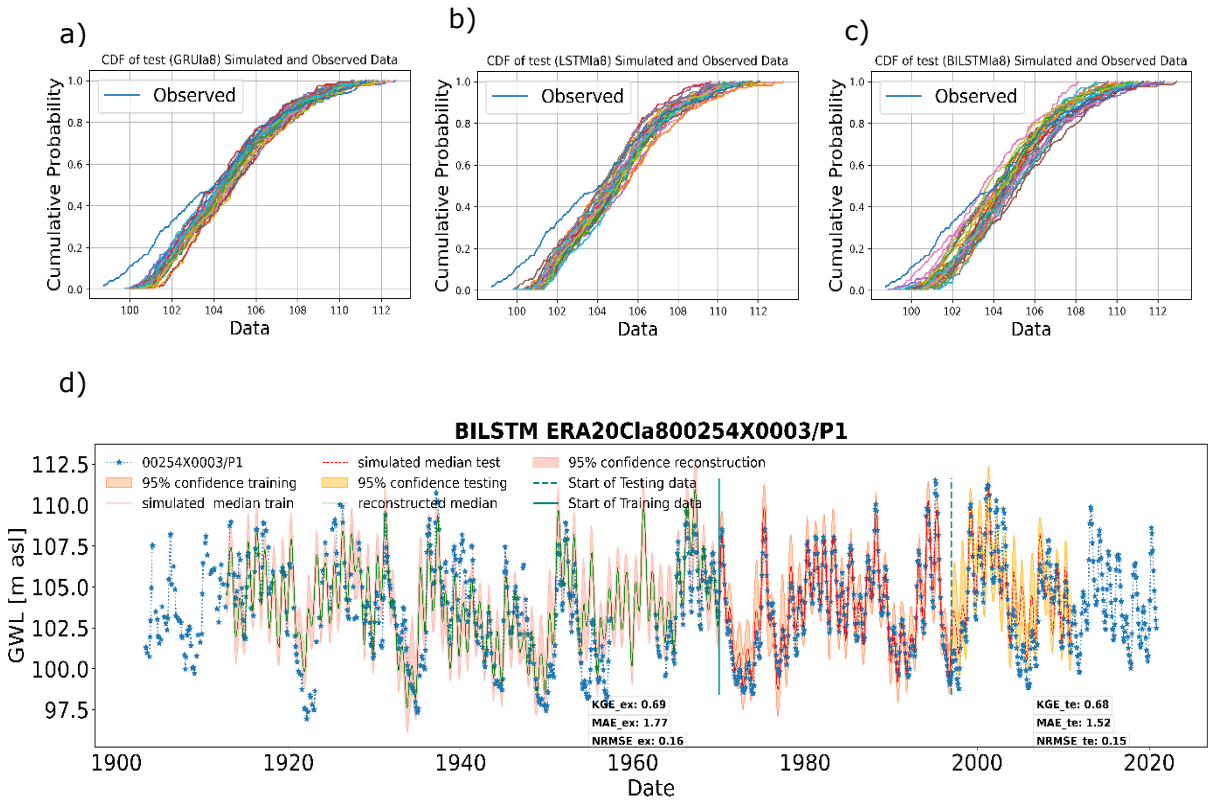

Figure 3.4: Cumulative density functions (CDF) of simulated vs observed (skyblue) Tincques on the test set(a-c). Reconstruction of the longest GWL level time series available for double validation (d).

Figures 3.4a-3.4c show the CDF plots of simulations versus observations over the period (1996-2010) for all three wavelet-based models. From Figure 3.4, the GWLs are slightly better captured in the case of BiLSTM, although low levels are slightly underestimated, and Figure 3.4d 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:068, 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, in some cases, no satisfactory results could be obtained for the time series of the inertial type (i.e. dominated only by multiannual to decadal variability), which needs 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.

3.5. Discussion: Trends and multidecadal variability in reconstructed GWL

We then generated long-term reconstructions for all the short-term time series in the database, which are now available for further investigation. Three examples of such reconstructions, one for each type of piezometric behaviour, are shown in Figure 3.5.

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 Figures 3.5a – 3.5c, which compare the comparison of the best reconstructed groundwater levels in each of the

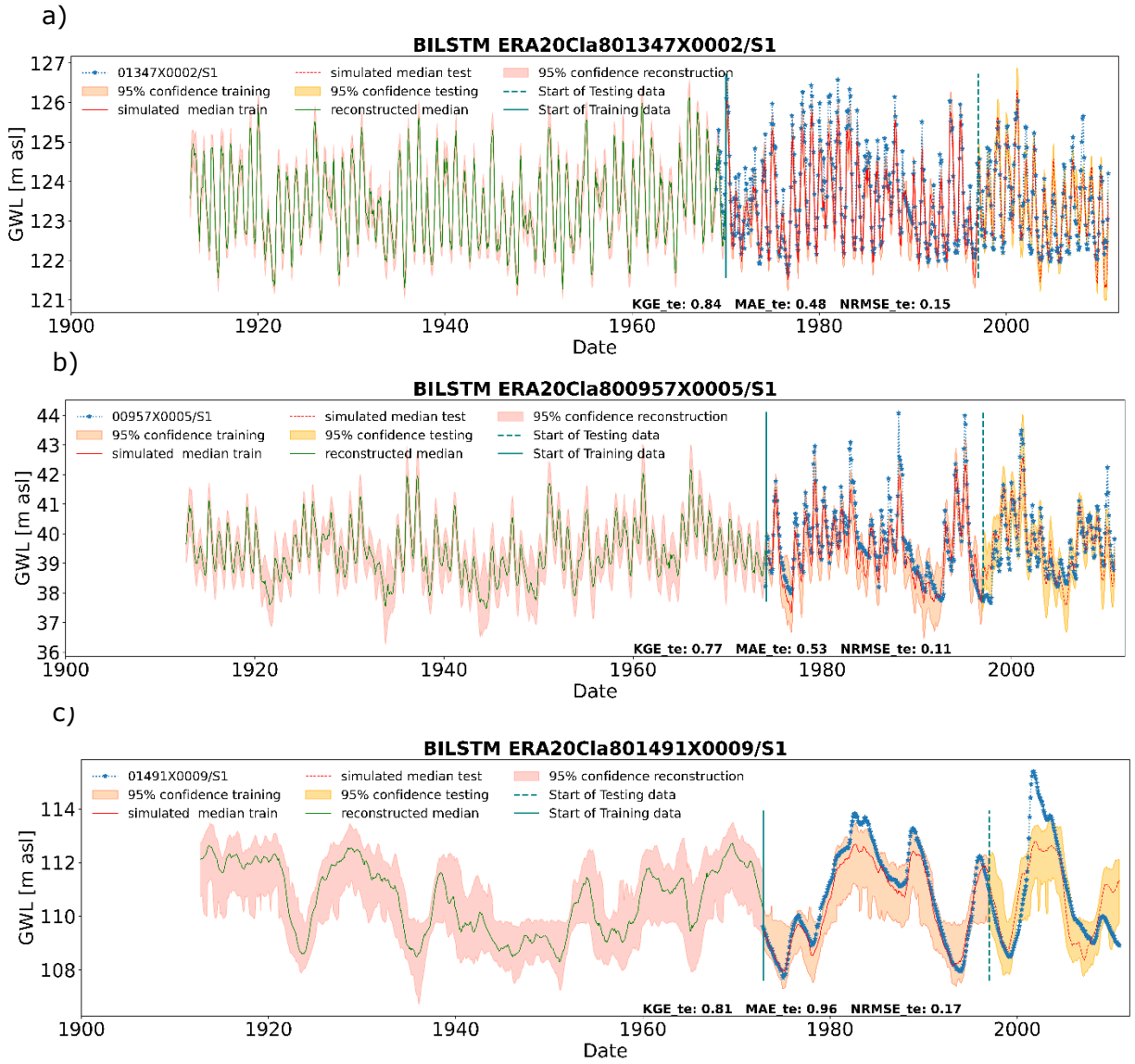

Figure 3.5: Comparison of reconstructed annual (a), mixed (b) and inertial (c) groundwater level (GWL) variations with best Kling-Gupta efficiency (KGE) values

three main GWL types, and in Figure 3.6, which compares reconstructions with low KGE values.

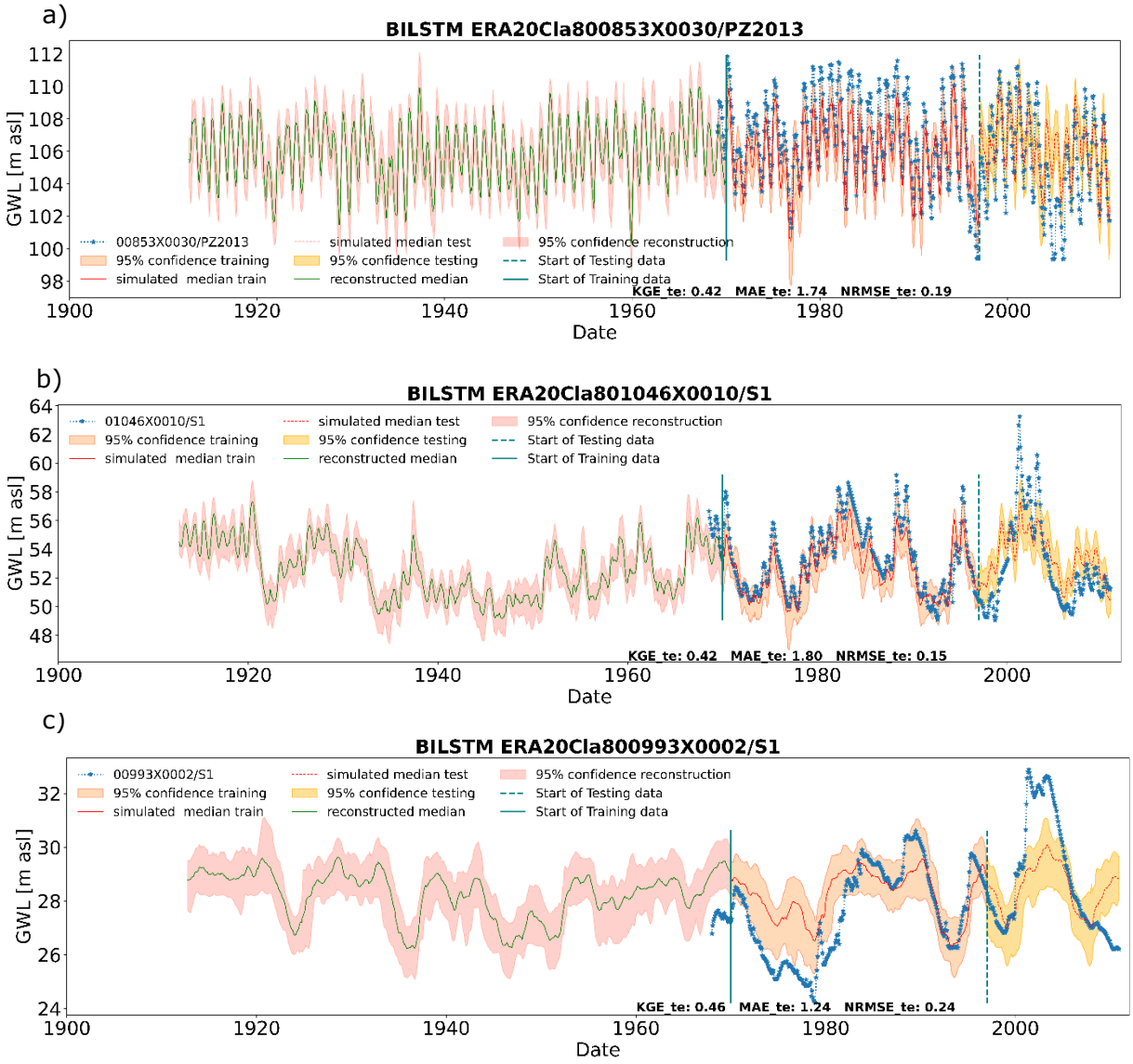

Figure 3.6: Comparison of reconstruction when KGE values are in the range of 0.4-0.5 for all three types (a-c)

Different objectives can be pursued that require long time series. From a hydro climatological point of view, assessing trends or changes in extremes over the long term is of great interest. We then carried out trend analysis using correlated seasonal Mann-Kendall test (Explained in detailed later in Section 5.3.2). Overall from the trend analysis, in the reconstructed groundwater levels from ERA20C and deep learning models, there is no significant trend in

reconstructed GWLs (Figure 3.7). 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 & Habets, 2014; Bonnet et al., 2017), multidecadal variability due to natural climate variability may either mask, reduce or amplify the effects of anthropogenic climate change.

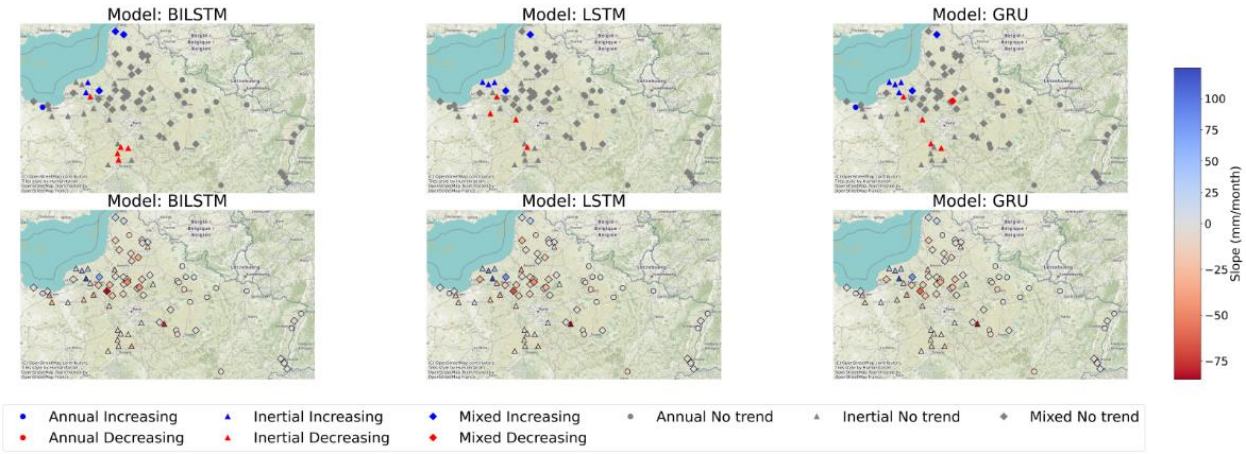

Figure 3.7: Trend direction (top row) and slope (bottom row) magnitude of reconstructed groundwater levels from different models using ERA20C (1900-1970) as input with different models

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. Figure 3.8 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 multi-decadal pattern was compared with the AMO index, confirming a potential influence of the AMO on groundwater levels, as suggested by Baulon et al. (2022a) or Bonnet et al. (2020).

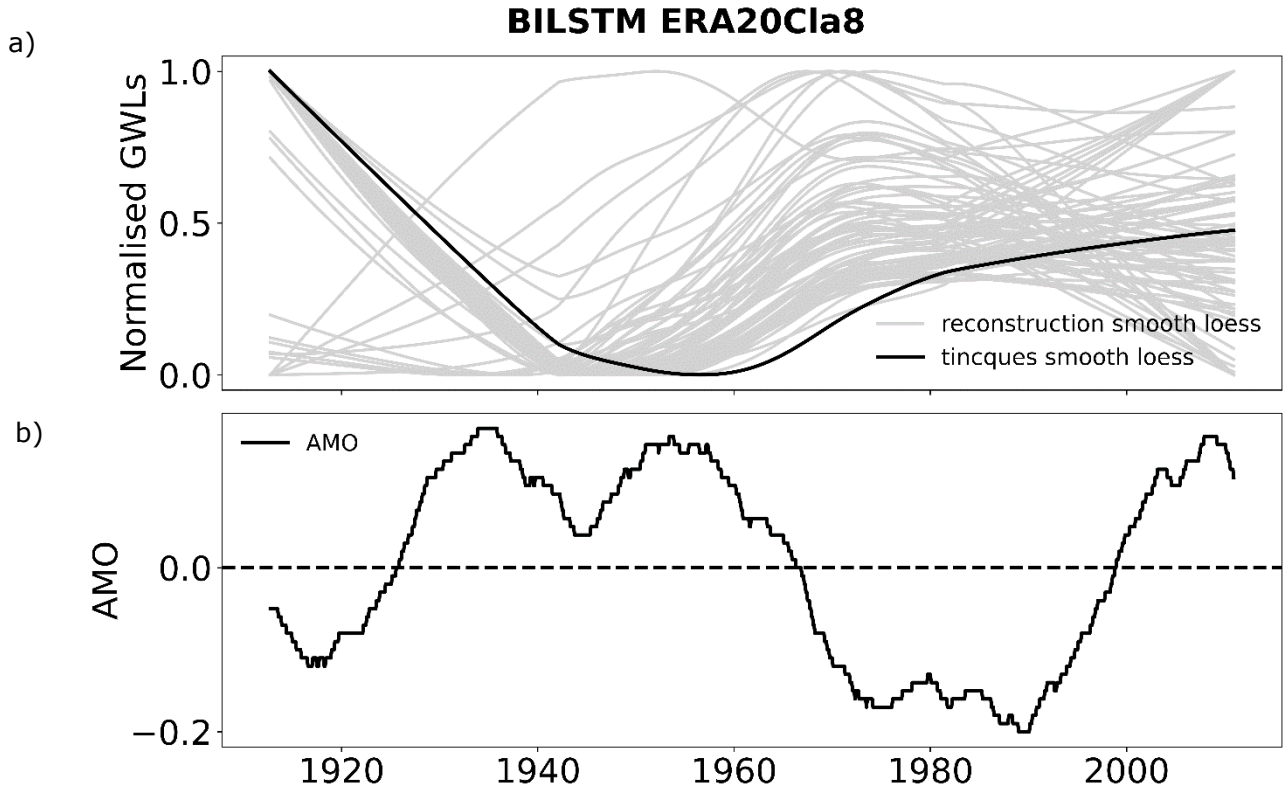

Figure 3.8: Comparison of loess of smoothing of reconstructed median (light grey) from all stations with smoothing of long-term series available(a) with the AMO (b).

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.

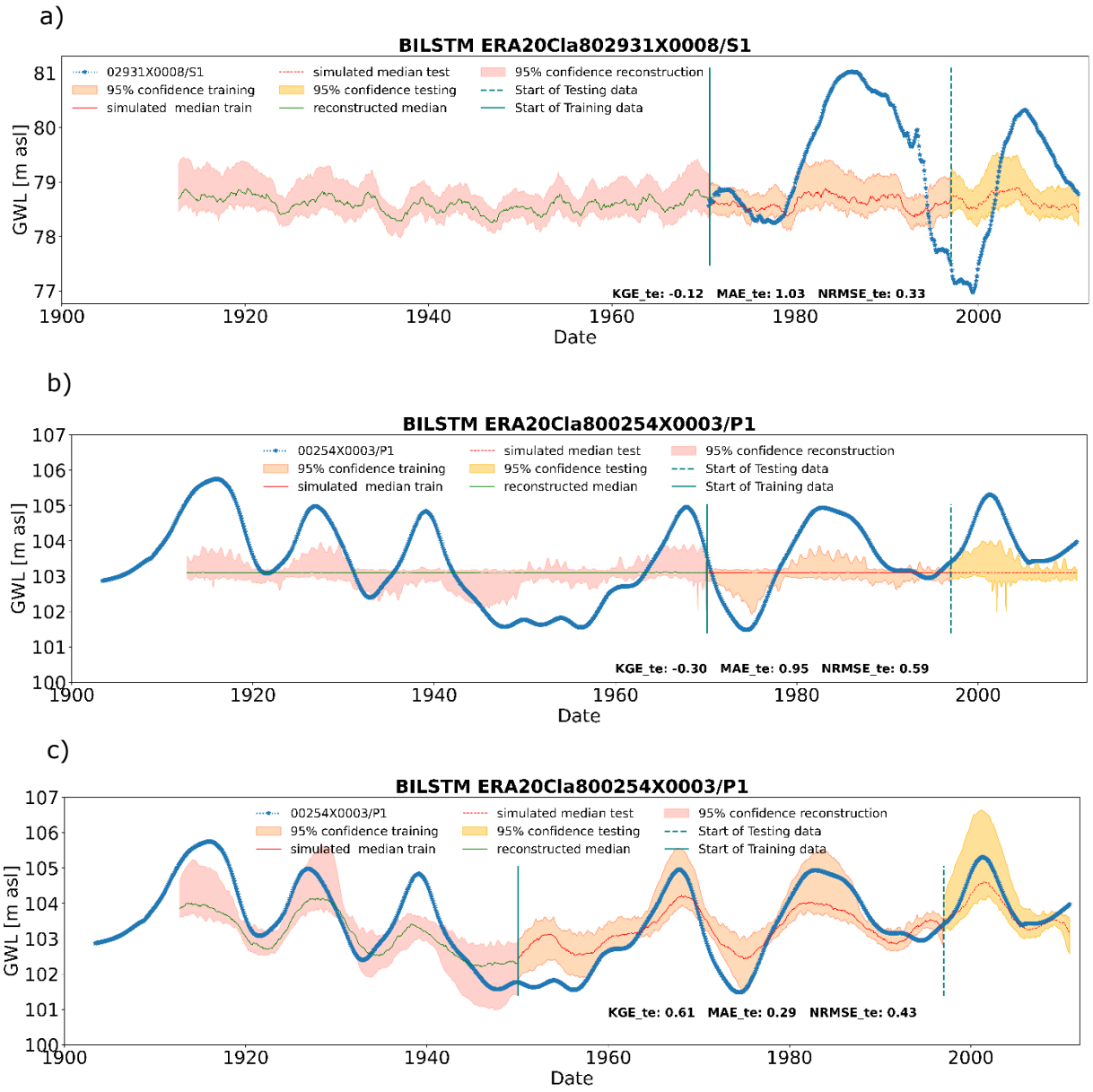

Figure 3.9: Reconstruction of high inertial type GWLs time series (a) and reconstruction of Tincques based highly inertial time series (obtained by LOESS smoothing of observed Tincques GWL) with training date starting in 1970 and 26 years of training(b) and 1950 (c).

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 GWLs (Figures 3.2 and 3.3). The inconsistency of the reconstruction for the highly inertial type of GWLs, as shown in Figure 3.9a, 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 figures 3.9b and 3.9c).

Figure 3.9b 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 (Figure 3.9c). 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.

3.6. 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).

Chapter 4. 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?

While Chapter 3 focused on reconstructing groundwater levels at individual locations, Chapter 4 expanded the scope to large-scale simulations across northern France. By employing a multi-station collective training approach, Chapter 4 incorporates both dynamic climatic variables and static aquifer characteristics, enabling the consideration of groundwater heterogeneities across the study area and then a regional approach for model training. Using clustering along with wavelet transform decomposition techniques, we now aim at testing the ability to leverage regionalised information and capture local variations, building upon the DL wavelet pre-processing methods introduced in Chapters 2 and 3.

Chapter 3 corresponds to an article accepted in Hydrology and Earth System Sciences (HESS) with major revision required, and currently under review in its revised version. It is available as preprint as:

Chidepudi, S. K. R., Massei, N., Jardani, A., Dieppois, B., Henriot, A., and Fournier, M.: 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?, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-794 , 2024.

Abstract

In this study, we used deep learning models with advanced variants of recurrent neural networks, specifically Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM), to simulate large-scale groundwater level (GWL) fluctuations in northern France. We developed a multi-station collective training for GWL simulations, using "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 with the aim of simplifying network learning using regionalised information. Several modelling performance tests were conducted. Models specifically trained on different types of GWL, clustered based on the spectral properties of the data, performed significantly better than models trained on the whole dataset. 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 behaviour preferentially, ignoring unique local variations. In this respect, wavelet pre-processing was found to partially compensate clustering, bringing out common 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 importance of different model training approaches, the potential of wavelet pre-processing, and the value of incorporating static attributes.

4.1. 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 constructing 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 analyse 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 geological, oceanic, climatic, and anthropogenic forcings 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, particularly in 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 an aquifer or river(Vu et al., 2023). The application of artificial intelligence (AI) algorithms, and deep learning (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 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 growing interest 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 & Nourani, 2022; 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, which 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 behaviours on a regional or continental scale, which cannot be captured by a limited number of piezometers. Consequently, the few applications of DL to groundwater, whether in simulation (Chidepudi et al., 2023) or forecasting (Bai & Tahmasebi, 2023; 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 using and processing 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 using and extracting the relevant global climate signal and tracking its effects. This can have a delayed effect on piezometric fluctuations, making DL models more effective for long-term forecasting.Working with groundwater data also presents unique 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) distinct sensitivities to human activities (e.g. pumping), which differ from those affecting runoff data, like river straightening and dam construction.

In some hydrological studies, the term 'global models' is being 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, i.e. these models only use past GWL data to forecast future GWLs. (Bai & Tahmasebi, 2023) used graph neural networks for GWL 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 GWL. 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 from 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 modelling also demonstrated the potential of clustering methods (Nourani et al., 2022) in hybrid models along with AI approaches such as self-organising map (Nourani et al., 2015, 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 that 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 the 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 not fully explored or validated across diverse hydrogeological settings. 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 pre-processing, such as BC-MODWT (Chidepudi et al., 2023a), have shown promise in single-station models but have not been extensively tested on regional-scale simulations. Given this background, the current study aims to address several research questions:

1) How do the generalised (multi-station) models compare with the specialised (single-station) models in simulating GWLs?
2) Can wavelet pre-processing techniques improve the performance of models for different types of GWLs when trained with data from all available stations?

3) To what extent do static attributes or one-hot encoding techniques help models generalise 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 and temporal statistical characteristics?
4) 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 want to highlight that the present study is not dedicated to 'forecasting' as is the case in most applications of DL to groundwater modelling. The reader can be referred to Beven & Young (2013) for distinctions between 'simulation' and 'forecasting'. In brief, according to their framework, 'simulation' means reproducing system behaviour without using observed outputs, while 'forecasting' involves reproducing system behaviour ahead of time based on past observations. This study focuses on simulation to understand GWL dynamics rather than forecasting future levels. This distinction is important for framing our approach and interpreting our results. To achieve this, we test different approaches for multi-station models while including static attributes and comparing 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 4.2 details study area the datasets used, and Section 4.3 presents the methodology and experimental design for the different approaches. Section 4.4 discusses the ability of the models and robustness in capturing different variations in GWLs and input scenarios. Section 4.5 deals with the discussion on the interpretability of the obtained results. Section 4.6 presents our main conclusions and perspectives.

4.2. Study Area and Data

The study area is approximately 80,000 km2 of Northern France, as depicted in Figure 4.1. The available GWLs of climate-sensitive wells (i.e., not strongly affected by human activities and sensitive to climate variability (Baulon et al., 2022a)) with high data quality until the end of 2022 were obtained from the ADES (Accès aux Données sur les Eaux Souterraines) database (https://ades.eaufrance.fr/; Winckel et al., 2022). All the wells considered in the study are in unconfined aquifers. In addition, the GWL data were clustered into three different clusters following the methodology outlined by Baulon et al. (2022b), 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 4.1.

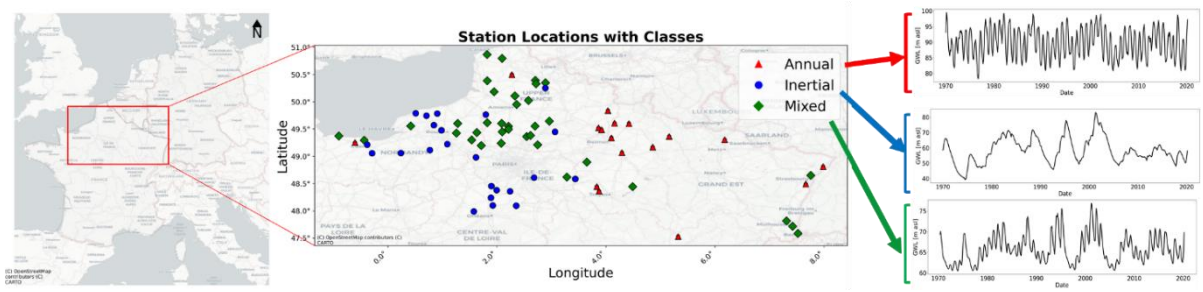

Figure 4.1: Clustering of GWL timeseries data (Background layer: © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.) based on the spectral statistical properties (Baulon et al., 2022b)

Specifically, the first cluster showcased in Figure 4.1 exhibits a pattern 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, as shown in Figure 4.1. The dataset consists of 35 mixed, 23 inertial and 18 annual stations. All the wells considered in the study are in unconfined aquifers. A comprehensive list of all analysed wells, including their identifiers, GWL type and coordinates, is available in the supplement (Table S1).

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, representing atmospheric conditions and circulation, moisture fluxes and radiative forcing. ERA5 is the best available global reanalysis with the data available from 1940 and is generally considered adequate for capturing regional and global hydrometeorological variations. Addressing the uncertainty issue of ERA5 is beyond the scope of this paper and can be considered a complete research work. ERA5 Reanalysis data have uncertainty related to potential regional biases; this and their use for hydrological modelling is still ongoing research, particularly in "large-sample hydrology" (Maria Clerc-Schwarzenbach et al., 2024.). Precipitation is considered to have more bias than temperature. However, recent studies conducted recently concluded that ERA5 temperature and precipitation biases had been consistently reduced compared to ERA-Interim and were found to be quite accurate for hydrological modelling, for instance, in the case of conterminous US (Tarek et al., 2020). Gualtieri (2022) highlighted that ERA5 uncertainties were greater in mountainous and particularly in coastal locations located less than 15 km from the coastline (in the study presented herein, only 1 station out of 76 is located within the 10-15 km range identified in Gualtieri (2022)). Finally, one recent study (Lavers et al., 2022) conducted by ECMWF on evaluating ERA5 precipitation for climate monitoring concluded that using ERA5 precipitation should be recommended for extra-tropical regions. However, for our study area, we have been evaluating different potential alternative reanalysis products, such as the SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) reanalysis developed specifically for France (Vidal et al., 2010). ERA 5 and SAFRAN precipitation appeared to have the same low-frequency components as detected in the GWL time series, as displayed in Figure 4.3 (this chapter) and Figure 2.15 in Chidepudi et al. 2023a. ERA 5, then, is suitable for our purpose.

Static attributes are available for different ranges of aquifer classes with different resolutions; we took the static attribute's value corresponding to each well's location—associated with the Well IDs. Static attributes, coming from the BDLISA (Base de Donnée des Limites des Systèmes Aquifères) (https://bdlisa.eaufrance.fr/) database, are point-scale information, i.e., each well received set of attributes given different possible methods (geographical imputation, rule-based, human expertise). BDLISA is based on a mix of information from geological maps, piezometric maps, and hydrochemistry at a scale of 25km. BDLISA was originally designed at a 25km scale and later upscaled to larger scales. For our study, we kept information coming

from BDLISA at its original scale (25km), which means aquifer static attributes have a resolution of 25km. This information from BDLISA should be understood as a local-to-regional description of aquifers.

Table 4.1: Summary of the static attributes used in the current study. Comprehensive explanation of all descriptions can be found at the URLs provided in the 3rd column.

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 this work, we also included static attributes (Table 4.1 and Figure 4.2) 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); they would be related to the 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. Exact details of static attributes for each GWL station can be found in the supplement (Table S1).

The decision to include the relevant static attributes comes from a trade-off between the transposability of models and the availability of attributes, as we have to make sure that all those variables are widely available at the required resolution. Also, for some attributes like hydraulic conductivity, it might not be straightforward to get the most relevant resolution, which is needed to account for the most appropriate characteristic describing the well. For instance, a 25km resolution might not be relevant when aquifers are highly heterogeneous. Exploring the role of static attributes in more detail would require much further work than what was conducted in this study.

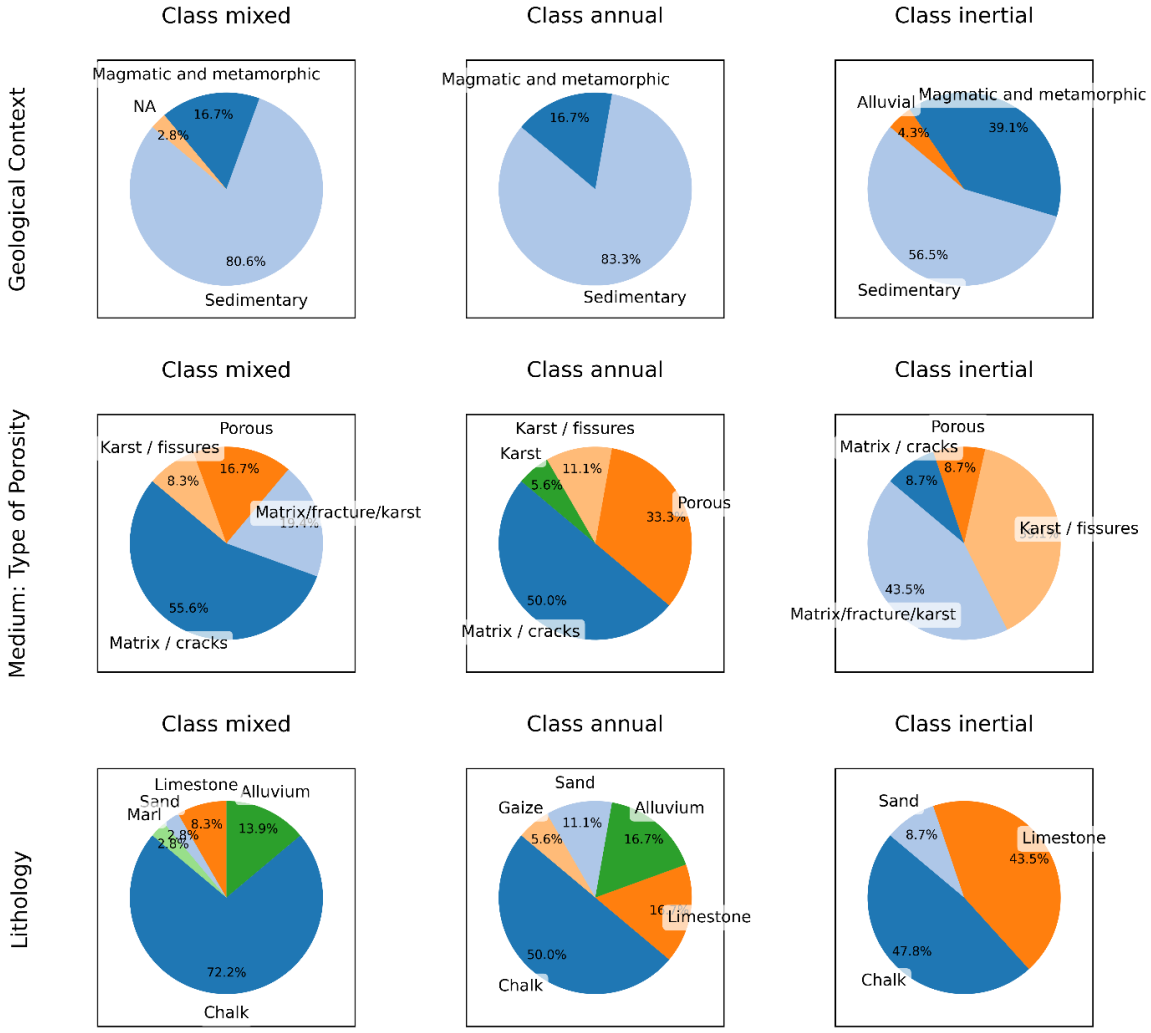

Figure 4.2: Distribution of Geological Features by Class

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

4.3.1. Theoretical modelling background

In the current study, we explored the use of recurrent-based deep learning models to simulate GWLs across multiple stations using different approaches as described in 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, Graves & Schmidhuber, 2005), 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. 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 (Vu et al., 2023).

We also explored the potential of wavelet decomposition (BC-MODWT) to decompose the data into components of varying frequencies (Figure 4.3), 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 sufficiently 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. All input time series were decomposed using BC-MODWT, with a decomposition depth of 4, as in Chidepudi et al. (2023a). Figure 4.3 illustrates the decomposition result for the precipitation time series. A 4-level decomposition 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 inertial GWL types (Figure 4.3, in red, with a few month time lag with respect to tp).

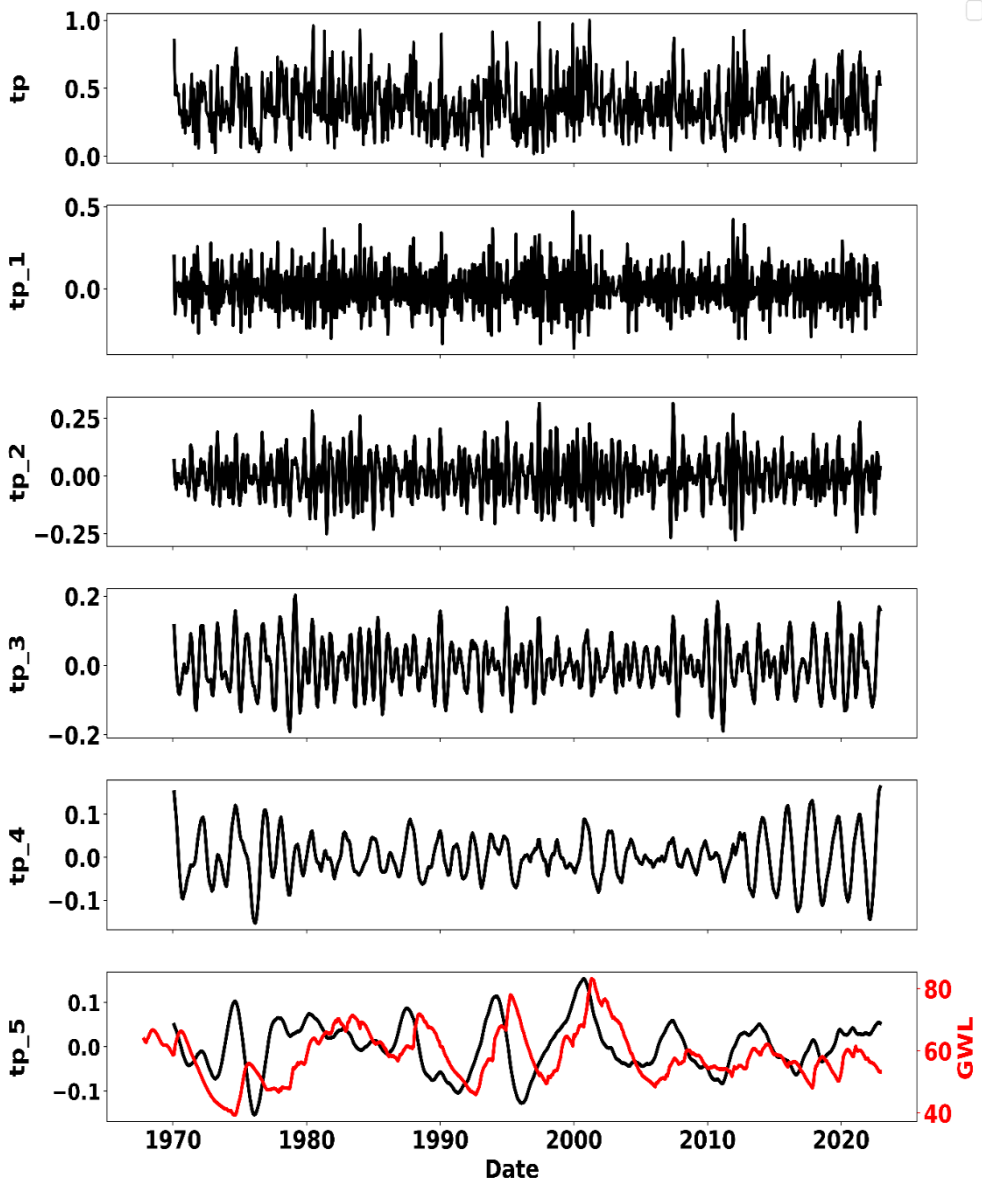

Figure 4.3: Total precipitation(tp) and its wavelet components: High(tp_1) to low frequency(tp_5) and GWL (in red).

4.3.1.1. Model training and evaluation

To maintain consistent comparison criteria across all methods evaluated in the study, Bayesian optimisation was used for hyperparameter tuning. Details of the range of hyperparameters used are shown in Table 4.2

Table 4.2: Hyperparameter details (Modified and adapted from Chidepudi et al. 2023a)

Hyperparameter	Value considered
Sequence length	48
Dropout	0.2
Optimizer	ADAM
Early stopping	50
Number of layers	1
Hidden neurons	(10, 20, …,100) by 10
Learning rate	(0.001,0.01) (log values)
Batch size	(16, 32, …,256) by powers of 2
Epoch	(50, 100, …,500)

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). 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 4.4) 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 applied to hydrological

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

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.

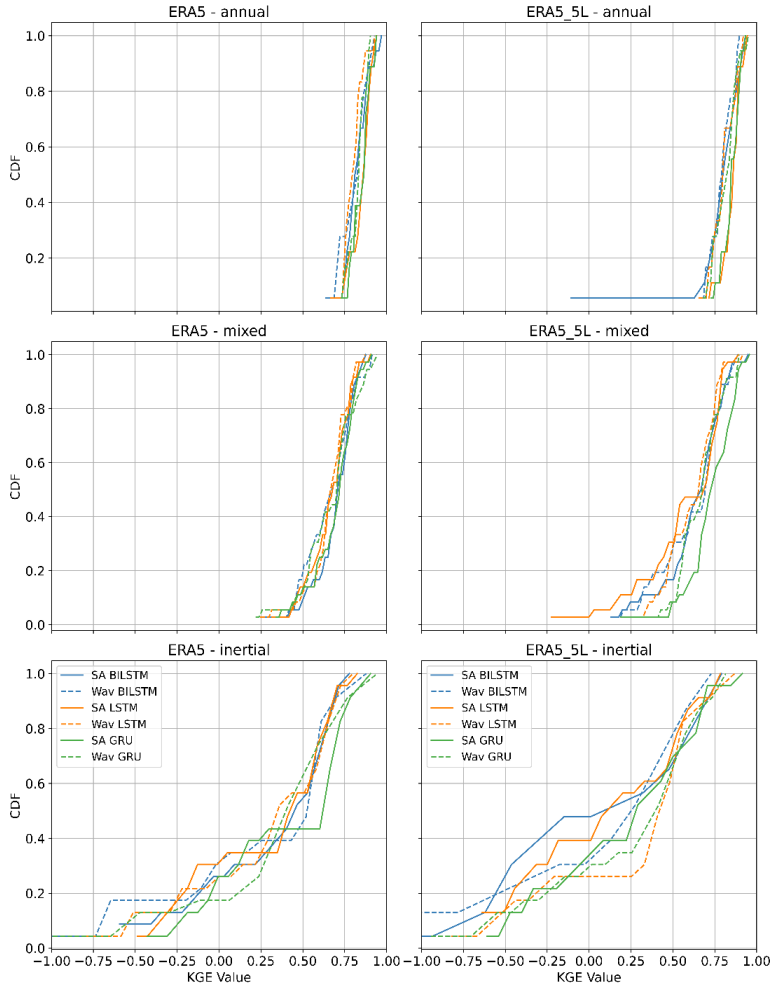

Figure 4.4: Comparison of performance of single layer DL models (left column) and multiple-layer DL models (right column) with respect to single station model as a reference. SA represents Standalone models while Wav represents Wavelet-assisted models.

4.4. 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 numerical and categorical data into the models. The aim is to improve the accuracy of GWL 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 4.5 & Figure 4.6:

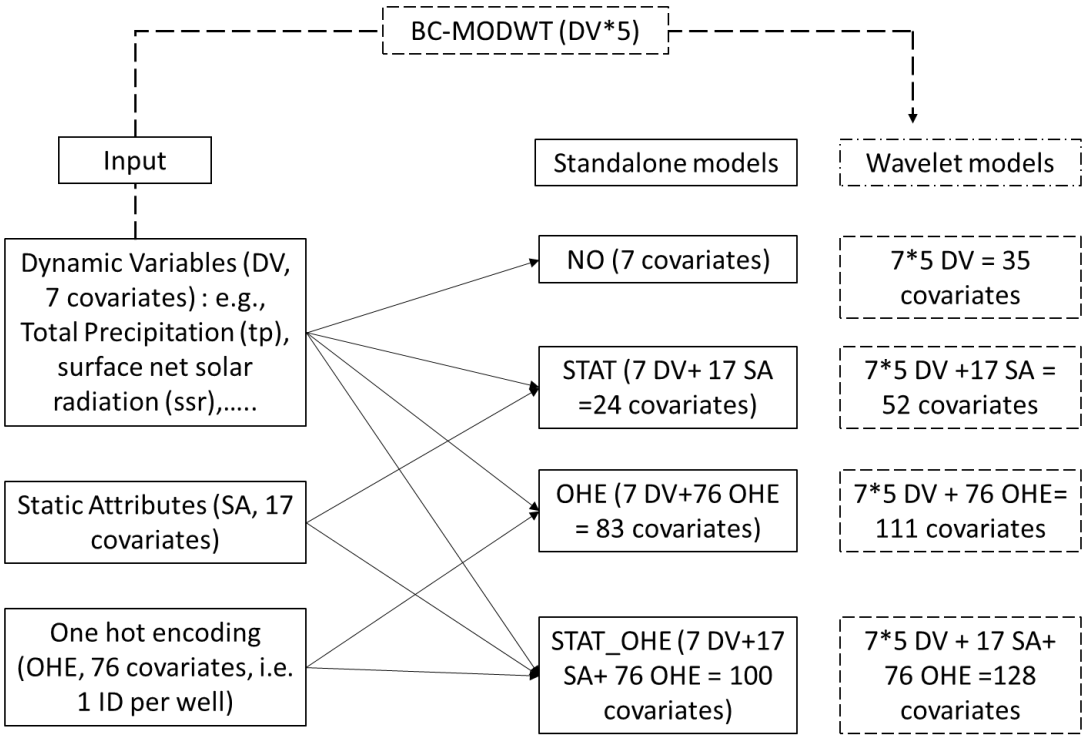

Figure 4.5: Construction of the different multi-station approaches for standalone and wavelet models and associated covariates (input features).

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 developing data-driven models for GWL simulations and is discussed in detail in Chidepudi et al. (2023a; 2024). The

optimal hyperparameters for all standalone and wavelet models in the single-station approach are presented in the supplement (Table S3-S4).

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. The covariates and input shapes for various multi-station approaches are summarised in Figure 24, and the exact shapes of 3D tensors are provided in the supplement (Table S5):

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) could 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 4.3).

Table 4.3: 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 4.4).

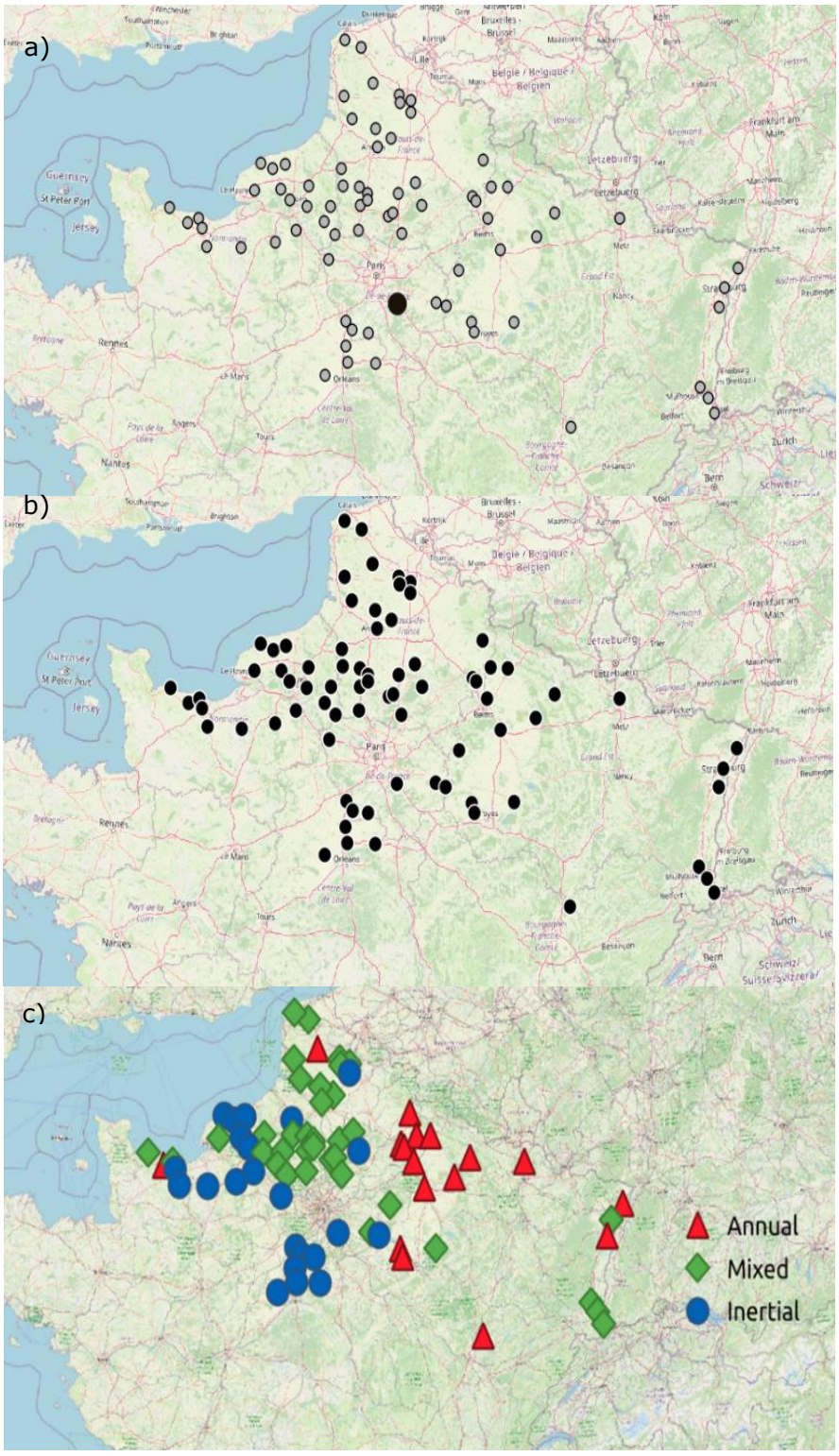

Figure 4.6: Comparison of different approaches adopted in the current study: a) single station (Top) b) multistation without clustering (Middle) c) multistation with clustering based on spectral properties(bottom). (Background layer: © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.)

d) **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, it is a combination of the two input strategies above.

Table 4.4: 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 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, Kling-Gupta efficiency (KGE, Gupta et al. 2009) is preferred over Nash–Sutcliffe efficiency (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.

For the single-station approach, the data was split into training (80%) and testing sets (20%), as described in Chidepudi et al., 2023. Furthermore, to facilitate hyperparameter tuning, the last 20% of the training data was used as a validation set. For the multi-station approach, the train-test split was also performed at each station, following the same procedure as the single-station approach. However, the data from all stations was then collectively combined during the training. The rationale behind the specific train-test split is to ensure that the models capture the multi-annual to decadal variability in GWLs observed in the region. To achieve this, a minimum of 34 years of data (1970-2014) was used for training, while the most recent

8.66 years of data (2015/01-2023/08) were reserved for testing. This split corresponds to approximately 80% of the data for training and 20% for testing. By following this approach, we aimed to ensure that the models were exposed to a sufficiently long period of data during training, enabling them to capture the amplitude and variability of GWL fluctuations over multi-annual to decadal timescales. The testing period was chosen to be the most recent years, allowing for an evaluation of the models' performance on the latest available data. The specific dates and periods used for training and testing at each station are detailed in the supplement (Table S2).

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 to simulate groundwater level variations. Detailed hyperparameters for all the multi-station standalone and wavelet models can be found in the supplement (Tables S6-S9)

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

4.5.1. Different strategies for multi-station approach

All models tested in the case of this study performed more or less equivalently and eventually yielded very satisfactory results. This can be attested by the performance comparison shown in Figure 4.4 (comparison of the 3 model types in single-station mode) and by comparing Figure 4.7 (GRU Multi-station) with Figure 4.8 (LSTM Multi-station) and Figure 4.9 (BiLSTM Multi-station). We finally decided to favour the GRU architecture owing to its recognised computational efficiency over more traditional LSTM-based architectures (Cho et al., 2014; Cai et al., 2021; Chidepudi et al., 2023, 2024). Figure 4.7 shows the results of different GRU model configurations for simulating 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.

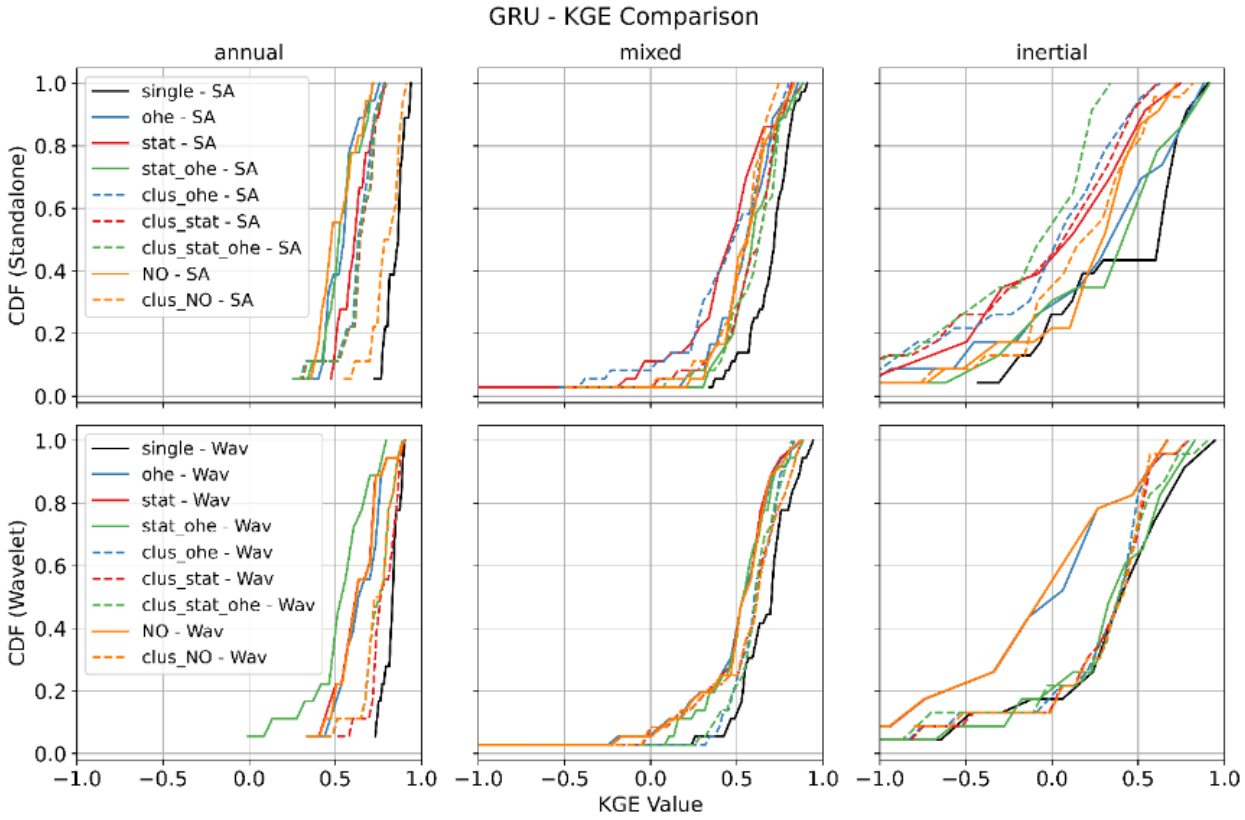

Figure 4.7: CDF Comparison of KGE values of the GRU With different approaches and GWL types.

Several observations can be made from Figure 4.7. 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.

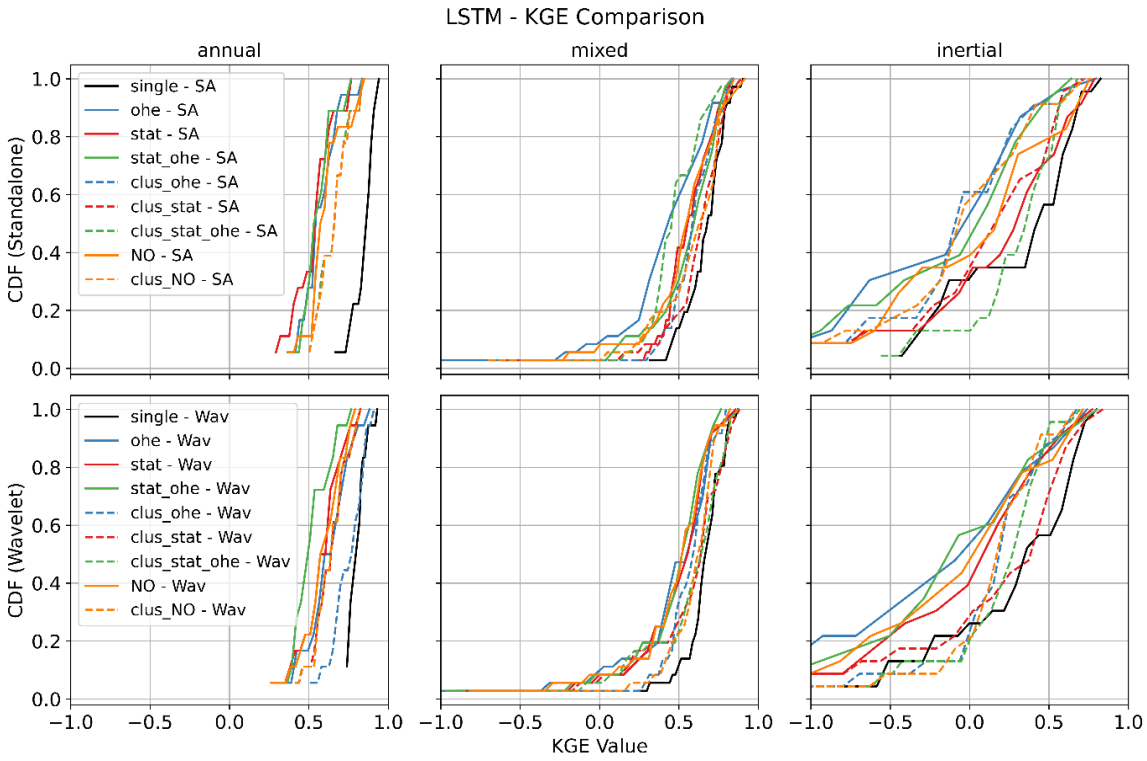

Figure 4.8: CDF Comparison of KGE values of the LSTM With different approaches and GWL types.

This is in line with previous findings as already reported in our earlier works (Chidepudi et al., 2023a & 2024). This demonstrates the wavelet decomposition ability to extract "hidden" inertial dynamics features, which facilitate 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).

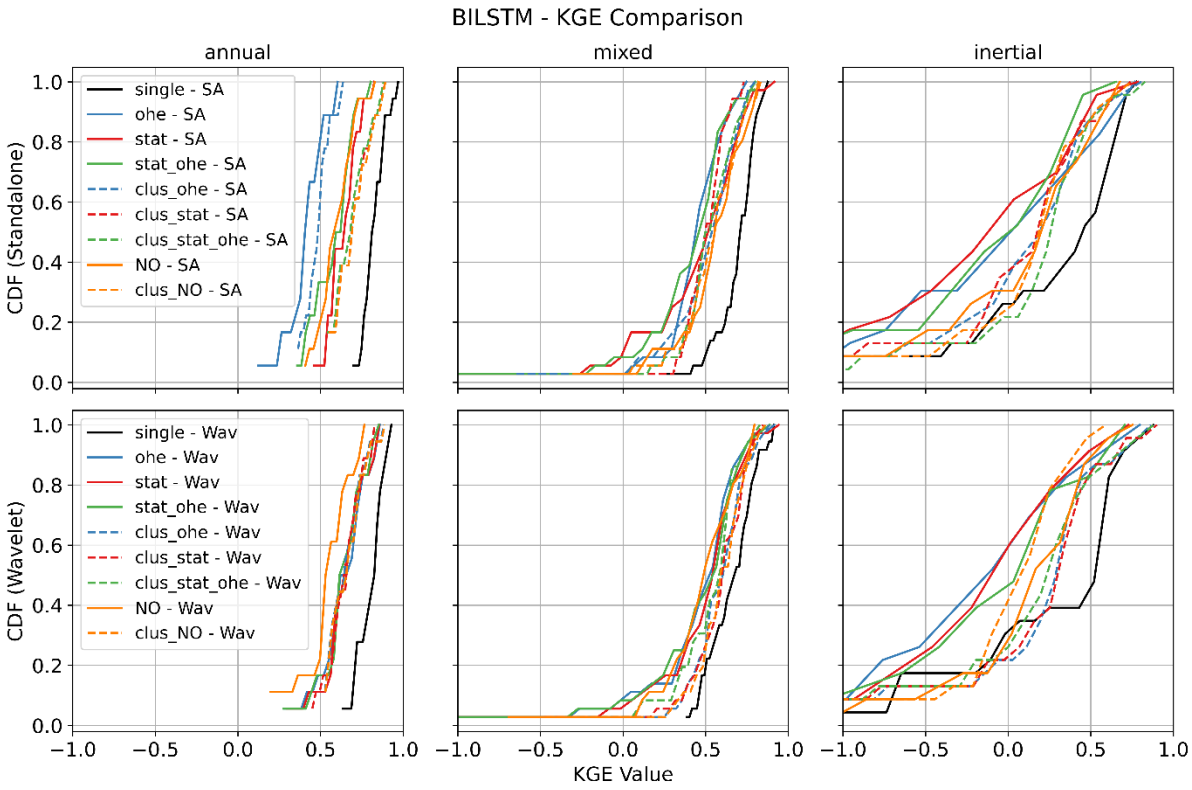

Figure 4.9: CDF Comparison of KGE values of the BiLSTM With different approaches and GWL types.

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., 2022a; Schuite et al., 2019). Wavelet decomposition allows the unravelling of such hidden information, helping the neural networks to reach it for enhanced learning. This is illustrated in Figure 4.3 with the low-frequency component of precipitation (tp5) matching the variations of one inertial-type GWL (in red, with a few month-lag time), whereas it is masked by other higher-frequency components in the original 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 generally outperform those trained on aggregated data when trained separately for each GWL cluster. 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 GWL 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.

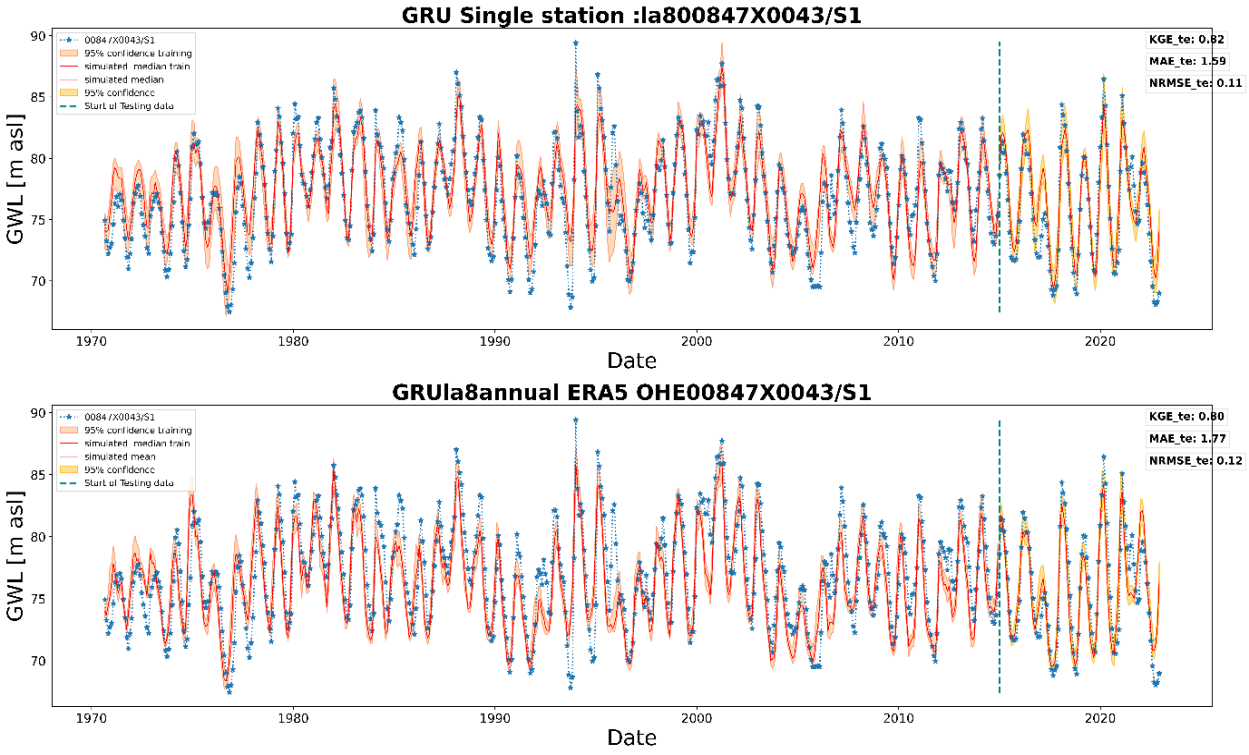

Figure 4.10: Results with wavelet assisted GRU in annual type of GWLs through a) Single station (top) and b) Multi station model trained on annual type of GWLs with static and ohe (bottom)

Figure 4.10-Figure 4.12 show the best GWL simulations obtained of different types (annual, mixed and inertial) for single and multi-station models. While single station models perform

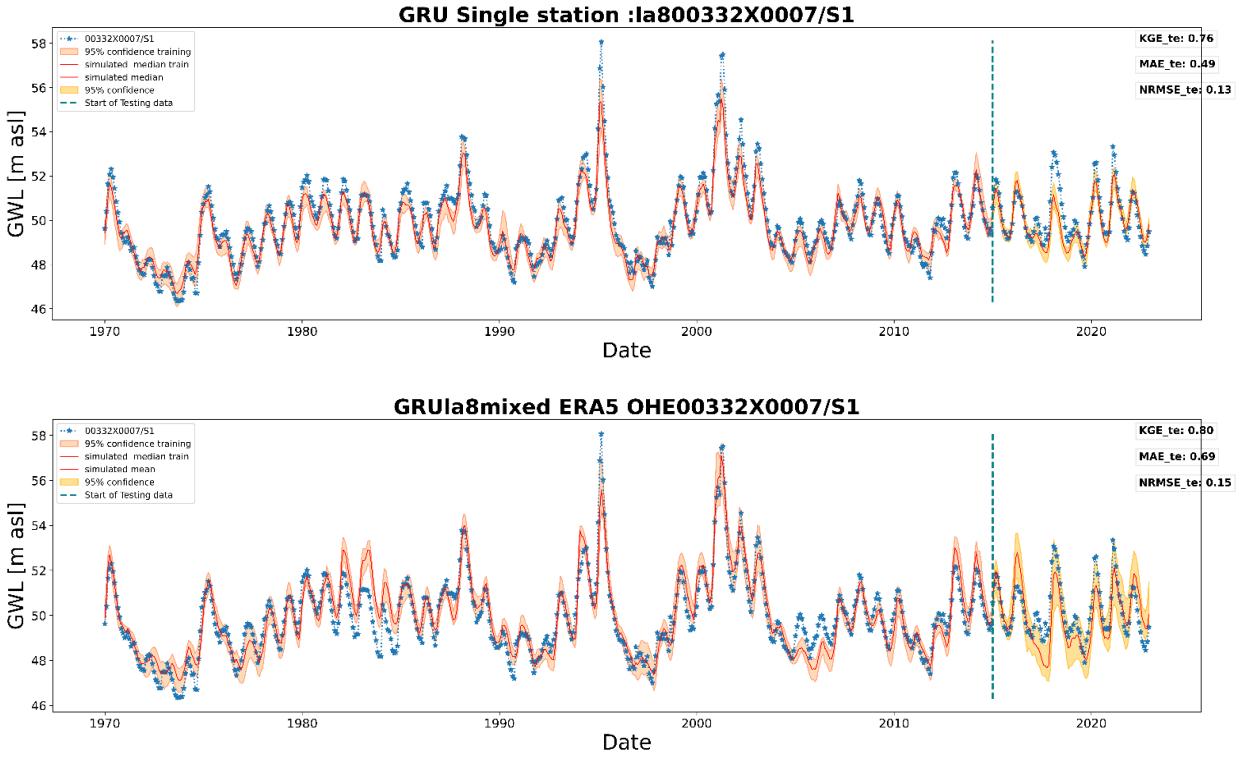

Figure 4.11: Results with wavelet assisted GRU in mixed type of GWLs through a) Single station (top) and b) Multi-station model trained on mixed type of GWLs with static and ohe (bottom)

best, multi-station models are valuable where single station modelling is impractical either due to data limitations or computational requirements.

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. To address the issue of model learning dominant behaviour in collective training of multi-station 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 the dataset on using multi-station approaches.

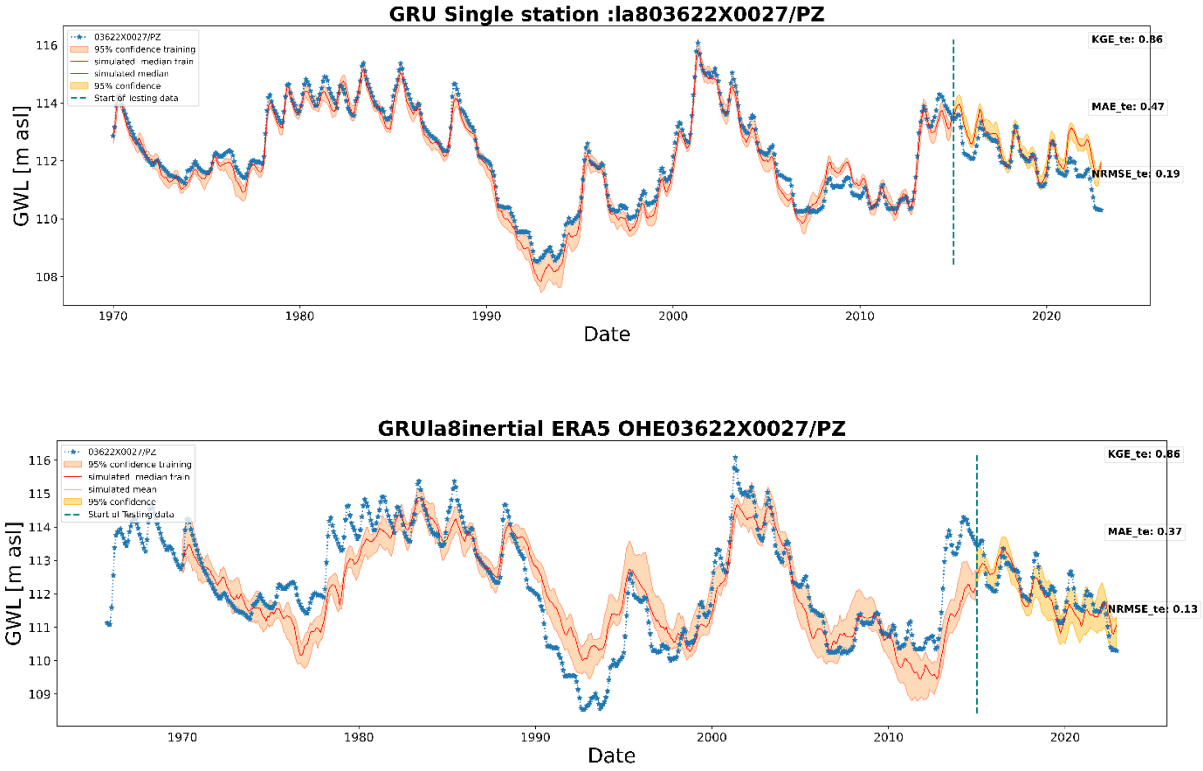

Figure 4.12: Results with wavelet assisted GRU in inertial type of GWLs through a) Single station (top) and b) Multi-station model trained inertial type of GWLs with static and ohe (bottom)

4.5.2. 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.

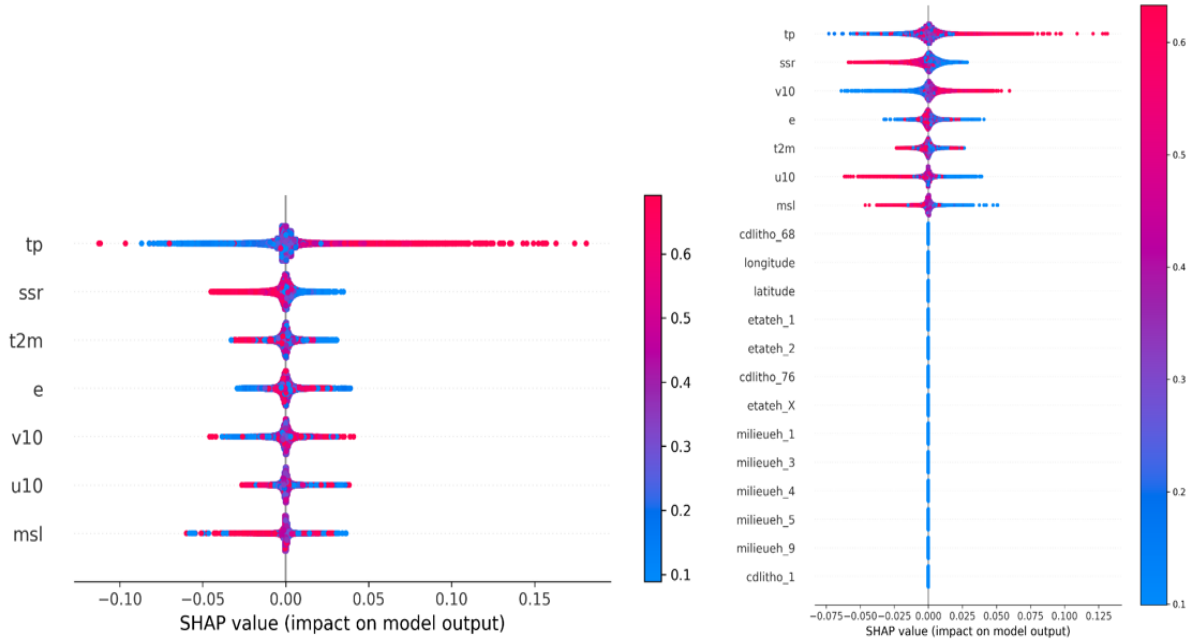

Figure 4.13: SHAP summary plot examples for single station model and multi station model with static attributes

Figure 4.13a shows the SHAP representative 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 plots (Figure 4.13), with more points to the right (coloured red) indicating positive influences and the opposite indicating negative relationships.

From the analysis of Figure 4.14 and Figure 4.15, several notable patterns emerge regarding the contribution of different variables to 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 to 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 pre-processing, low-frequency precipitation components emerge as key contributors, improving model performance. 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 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. Based on the results, the order of best approaches is 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 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.

The purpose of the study presented here was not to determine the forcing factors of GWL variations; in this aim, a more comprehensive evaluation of such links would require specific approaches that have been undertaken and presented in several previous works (Lee et al., 2019; Heudorfer et al., 2019; Liesch & Wunsch, 2019; Haaf et al., 2020; Giese et al., 2020). In some of our previous works (albeit for the Normandy region only), the linkages between GWL variability and potential forcing factors such as the thickness and lithology of surficial formations, aquifer thickness, vadose zone thickness, upstream/downstream location along the flow path, distance to the river, presence of karst, were investigated using dedicated approaches combining multivariate analysis, clustering and time series / spectral analysis and decomposition (Slimani et al., 2009; El Janyani et al., 2012 and 2014), which showed that GWL dynamics could be related to some basin and aquifer properties, although these relationships remained rather complex. In a recent study, Haaf et al. (2023) developed an innovative methodological approach for modelling GWL at unmonitored locations using basin properties and machine learning on a daily time-step basis for alluvial aquifers with probably quite high hydraulic conductivity overall. The models developed performed quite well in representing GWL variations at both intra- and interannual time scales using physiographic, land cover and geological characteristics. However, the amplitude of low-frequency, interannual to decadal variability of the dataset used in their study was much lower than what could be encountered in our monthly time step database. The specific type of aquifer Haaf et al. (2023) investigated likely explains their high sensitivity to many surface processes. In our study, alluvial aquifers only represented approximately 10% of the GWL stations (8 over 76 stations) and were only of annual (3 stations) or mixed (4 stations) types. Almost all other wells were located in chalk or limestones.

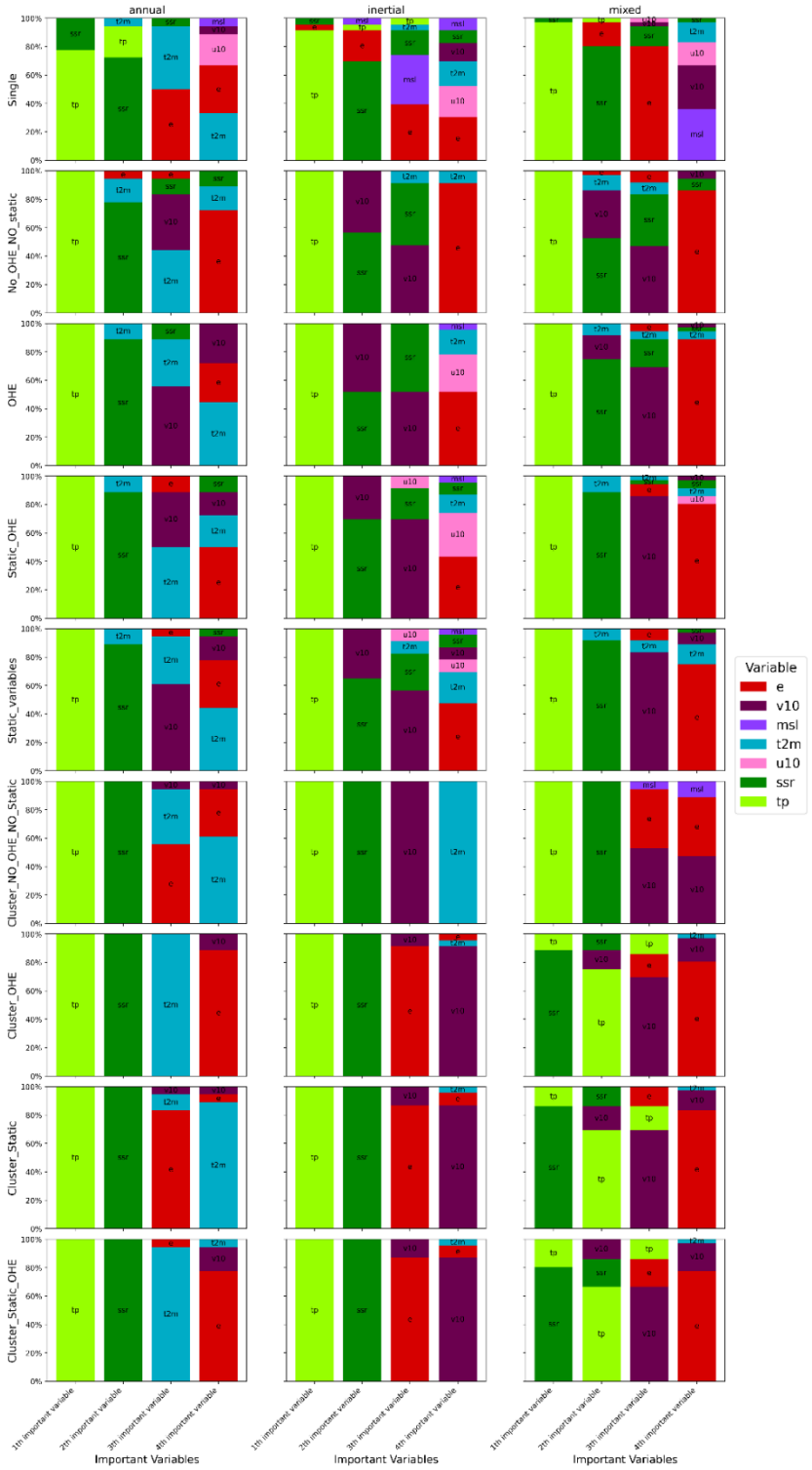

Figure 4.14: Top four important variables by cluster for standalone GRU models with different approaches. On Y-axis, Percentage of stations for each variable within in the cluster.

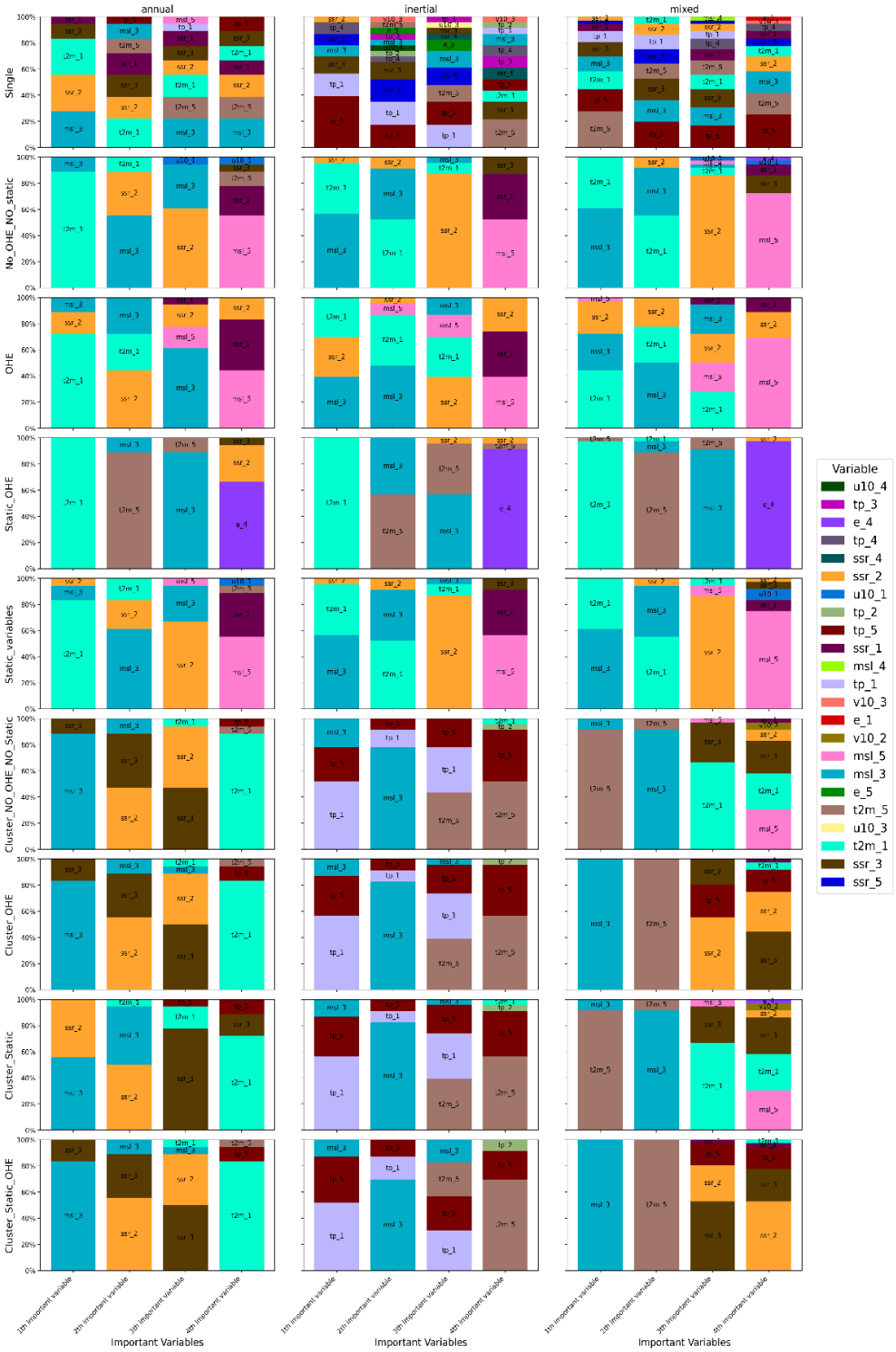

Figure 4.15: Top four important variables in regional GRU wavelet assisted model trained with different approaches for different classes

In the framework of our study, we decided to exclude some relevant characteristics such as vadose or saturated zone thickness: even when averaged over quite long periods (several years), these values actually represent GWL (the target variable). For mixed or inertial types in particular, it would probably make our models irrelevant for simulations over long-term predictions of several years or even decades when used along with climate projections and in another recent and relevant study by Heudorfer et al. (2024) developed entity-aware deep learning models with static attributes such as aquifer type (based on porosity). These authors concluded that the models did not show any entity awareness and eventually utilised static attributes as simple identifiers (almost similar to the OHE approach presented herein), meaning that the models did not make use of the relevant (hydro)geological information.

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 unmonitored 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 GWL monitoring, thereby capturing regional dynamics. However, single-station modelling remains important for understanding local interactions. Therefore, the choice of method should 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).

4.6. Concluding remarks

This study has demonstrated the different multi-station approaches to GWL simulations with emphasis on the use of static attributes, one-hot encoding and the combination of both while 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 have 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 likely captured key information needed for accurate modelling.

The study also showed that a multi-station approach, without clustering, should be used cautiously, 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 (due to wavelet decomposition that increases the number of covariates), suggesting a potential avenue for future research to explore alternative encoding strategies, such as target encoding. SHAP 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 this article, we introduced the following question: "What's the best way to leverage regionalised information?". In light of our results, it then seems like this is highly dependent on the amount and types of static attributes. It is generally expected that a much higher number of static attribute types would allow for a much better improvement of the multi-station simulation approach. However, Heudorfer et al. (2024) found no improvements using around 28 static features (including 18 environmental and ten time series-based). Also, as pointed out by these authors, employing static attributes for model training might be more relevant in applications on larger scales and/or larger datasets. Moreover, one must remember that a trade-off must be found between the amount of static attributes required and data availability, especially for applications at unmonitored sites. However, the use of static attributes and OHE yielded similar results in the gauged scenario (this study) and proved efficient in accounting for local station information, which aligns with the findings of Heudorfer et al. (2024). On the other hand, in the study presented herein, wavelet pre-processing allowed for deciphering the "hidden" dynamic components of GWL variability (i.e. by separating low-frequency variations from annual or intra-annual variability), which eventually

corresponded to taking into account the influence of (hydro)geological, geomorphological and physiographic properties. Ultimately, the latter – which varies across the study region - operates a differential filtering effect of the input signals. Pre-clustering the dataset also yielded significant improvements that were even more noticeable when combined with wavelet pre-processing. However, owing to its capability of leveraging pre-processing the different frequency components in the time series of the whole dataset, wavelet pre-processing somehow acts in the same way as pre-clustering, which consists of grouping inertial (i.e. low-frequency dominated), mixed and annual time series in different clusters.

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 the 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 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 classical big 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, reducing the need for specialised models and retraining new models. The way forward is clear: to improve the GWL simulations efficiently, we may need to adopt a nuanced mix of efficient input signal pre-processing, potentially new encoding strategies to incorporate all possible additional knowledge of the system, and possibly clustering. Yet, we would recommend using advanced pre-processing over clustering, which would allow for leveraging the same type of information while preventing from separating the dataset and reducing its size.

Chapter 5. Groundwater level projections for aquifers affected by annual to decadal hydroclimate variations

The multi-station training strategy and wavelet pre-processing techniques developed in Chapter 4 provided a robust foundation for generating groundwater level projections under different climate change scenarios. By leveraging 16 CMIP6 models and 3 Shared Socioeconomic Pathways (SSPs), Chapter 5 extends the groundwater simulations to future projections, enabling the assessment of potential impacts on groundwater resources.

Chapter 5 has been submitted to Earth's Future and is currently under review and available as preprint as follows:

Sivarama Krishna Reddy Chidepudi, Nicolas Massei, Abderrahim Jardani, et al. Groundwater level projections for aquifers affected by annual to decadal hydroclimate variations. ESS Open Archive. https://doi.org/10.22541/essoar.172526712.23981307/v1

Abstract

In a context where anticipating future trends and long-term variations in water resources is crucial, improving our knowledge about most types of aquifer responses to climate variability and change is necessary. Aquifers with variability dominated by seasonal (marked annual cycle) or low-frequency variations (interannual to decadal variations driven by large-scale climate dynamics) may encounter different sensitivities to climate change. We investigated this hypothesis by generating groundwater level projections using deep learning models for annual, inertial (low-frequency dominated) or mixed annual/low-frequency aquifer types in northern France from 16 CMIP6 climate model inputs in an ensemble approach. Generated projections were then analysed for trends and changes in variability. Generally, groundwater levels tended to decrease for all types and scenarios across the 2030-2100. The variability of projections showed slightly increasing variability for annual types for all scenarios but decreasing variability for mixed and inertial types. As the severity of the scenario increased, more mixed and inertial-type stations appeared to be affected by decreasing variability. Focusing on low-frequency confirmed this observation: while a significant amount of stations showed increasing variability for the less severe SSP 2-4.5 scenario, low-frequency variability eventually showed slight yet statistically significant decreasing trends as the severity of the scenario increased. For the most severe scenario, almost all stations were affected by decreasing low-frequency variability. Finally, groundwater levels seemed, in most instances, slightly higher in the future than in the historical period, without any significant differences between emission scenarios.

5.1. Introduction

Groundwater is a vital freshwater resource for sustaining domestic, agricultural, and industrial activities. However, the sustainability of groundwater resources is under threat due to the impacts of climate change and human interventions (Jasechko et al., 2023; Mishra et al., 2024; Scanlon et al., 2023; Taylor et al., 2013). Assessing the potential effects of climate change on groundwater levels is of paramount importance, particularly in regions that heavily rely on groundwater resources and are vulnerable to changes in climate, such as northern France (Habets et al., 2013; Vergnes et al., 2023). Robust projections of future groundwater

levels under different climate change scenarios are essential for informed water resource management and the development of effective adaptation strategies. The Coupled Model Intercomparison Project Phase 6 (CMIP6) provides state-of-the-art information on plausible global to regional climate changes in the past, present, and future (Eyring et al., 2016). However, climate models show large uncertainty due to model physics differences, emission scenarios sensitivity, and internal climate variability (Atawneh et al., 2021; Hawkins & Sutton, 2009; Lehner et al., 2023). In addition, when analysing climate change impacts on hydrological variations and trends, additional uncertainties arise from the hydrological model structures and parameterisations (Clark et al., 2016; Melsen et al., 2018; Wu et al., 2024; Yuan et al., 2023). Nevertheless, only 20% of groundwater impact studies considered the influence on climate model uncertainties (Atawneh et al., 2021). Characterising such uncertainties is crucial for enhancing the reliability of climate change impact scenarios for groundwater resources.

Most of the time, physically-based hydrological models have been employed to assess the impacts of climate change on groundwater resources (Costantini et al., 2023; Halloran et al., 2023; Mishra et al., 2024; Vergnes et al., 2023). However, due to the substantial computational time and specific data requirements associated with these models, data-driven methods have become increasingly popular complements or sometimes alternatives (Bhasme et al., 2022; Chidepudi et al., 2023a; Hauswirth et al., 2021; Rehana & Rajesh, 2023; Wunsch et al., 2022). In recent years, some studies have employed artificial intelligence (AI), machine learning (ML), and Deep Learning (DL) models in conjunction with CMIP5 and CMIP6 climate projections to assess the impacts of climate change on groundwater levels (Chakraborty et al., 2021; Kayhomayoon et al., 2023; Nozari et al., 2022; Roshani & Hamidi, 2022; Xiong et al., 2022; Wunsch et al., 2022; Nourani et al., 2023; Secci et al., 2023). Most of these approaches used neural networks and at least one deep neural network architecture (i.e., DL). Many different architectures exist that can be mostly suited for specific tasks in handling time series data. Recently, Secci et al. (2023) compared different types of DL models (NARX, LSTM, and CNN) and found that long short-term memory neural networks (LSTM) outperformed the others due to their ability to handle long-range dependencies. Long et al. (2024) reached similar conclusions and demonstrated the ability to capture complex spatiotemporal patterns and nonlinear relationships between climate variables and hydrological processes.

Most of the studies dealing with groundwater level simulation using DL approaches either mainly considered aquifers dominated by seasonal variability or develop forecasts on quite short forecasting horizons, up to a few days or weeks (Boo et al., 2024; Rajaee et al., 2019; Tao et al., 2022; Uc-Castillo et al., 2023). However, it is now well-recognised that interannual to decadal variability affecting groundwater level originates from large-scale climate variability (Baulon et al., 2022a; El Janyani et al., 2012; Hanson et al., 2006; Holman et al.,2011; Liesch & Wunsch, 2019; Massei et al.,2010; Neves et al., 2019; Rust et al., 2019), and can significantly impact decadal trends at the regional scale in climate change projections. Such interannual to decadal variability is represented differently by different climate models and individual simulations from the same climate model (i.e., different ensemble members; Deser et al., 2012; Deser & Phillips, 2023). Emission scenarios can also modify low-frequency variability as climate change impacts large-scale modes of variability and teleconnections (Klavans et al., 2022; Mahmood et al., 2022; Terray, 2012). Therefore, It is also crucial to consider aquifers that behave on more low-frequency dynamics, develop and apply DL tools that can appropriately describe such variability in groundwater systems, and assess how systems subject to low-frequency variability would behave under climate change compared to those dominated by seasonal variations, which seemed to have received more attention so far.

The study presented here uses DL techniques and CMIP6 climate change scenarios to provide an overview of the potential impacts of climate change on groundwater levels for different types of aquifers dominated by seasonal variability or low-frequency variability (or a mix of the two). These three examples of contrasted behaviours are observed in northern France, which is here used as a case study. The research assesses potential alterations in groundwater level trends and variability due to future climate conditions. Employing a multi-station deep learning approach, we generated groundwater level projections for the region, incorporating three different climate change scenarios and socioeconomic pathways. This approach aims to capture the spatial patterns and temporal evolution of projected groundwater level changes across northern France according to the hydrological systems' characteristic behaviour (annual, mixed, inertial) to assess their sensitivity to different climate change scenarios. Furthermore, the study evaluates the performance, uncertainties, and limitations of the deep learning methodology and the climate model projections utilised in the groundwater level projection analysis.

The rest of the paper is structured as follows: Section 2 presents the data and study area. Section 3 presents the methodology of models and the trend and variability assessment of projections. Section 4 presents the dispersion of GWL projections under different scenarios. Section 5 discusses the time evolution of GWL projections. Section 6 presents our comparison with other relevant studies and conclusions.

5.2. Study Area and Data

The study focuses on the northern region of France, primarily encompassing the Paris Basin and its surrounding areas (Figure 5.1). This region was selected due to long-term groundwater level (GWL) data availability, which is crucial for accurate projections. A significant advantage of this area is the presence of three distinct GWL behaviours despite its relatively limited spatial coverage: 1- reactive aquifers dominated by seasonal variability ("annual" type), 2- aquifers with marked seasonal variations along with significant low-frequency variability ("mixed" type), and 3- aquifers dominated by low-frequency variability ("inertial" type).

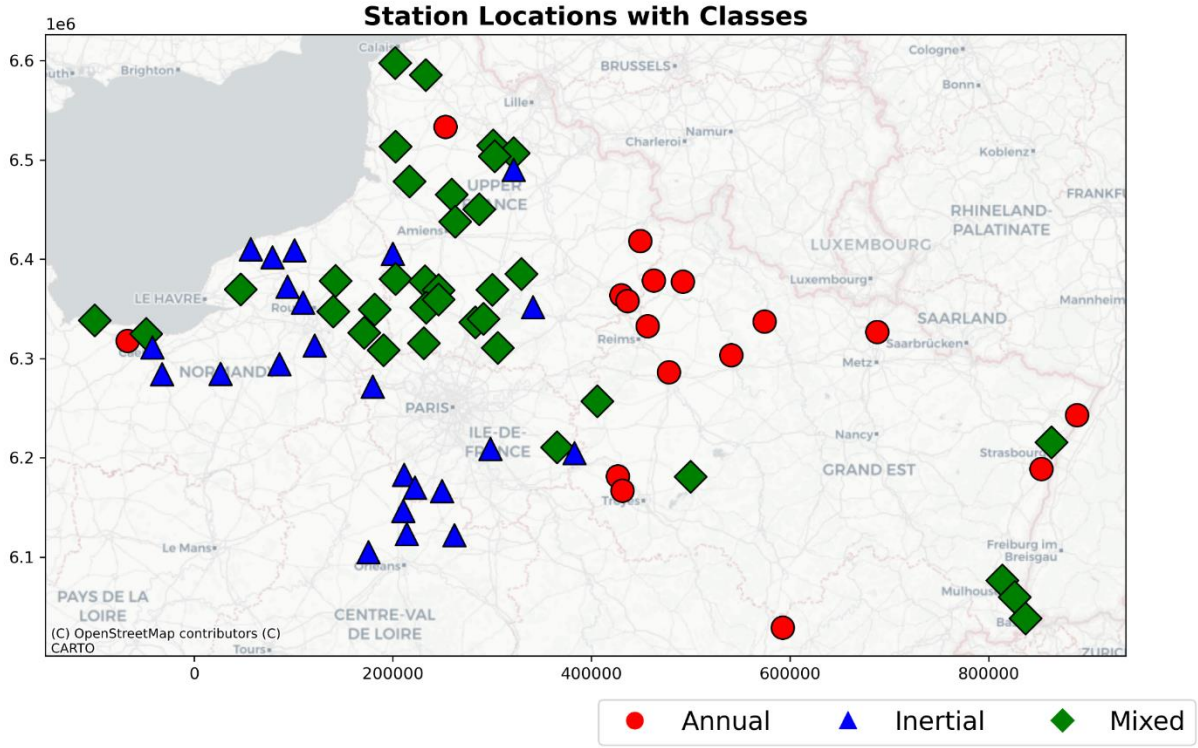

Figure 5.1: Study area and location of stations with details of GWL types: Annual (red dots), Inertial (blue triangles), Mixed (green lozenges)

Example time series of these three types are provided in Figure 5.4 - Figure 5.6 (observational data, left panel). This diversity in groundwater level patterns provides a valuable opportunity to assess the performance of the deep learning models in capturing various hydrogeological conditions and responses to climate variability.

The dynamic climate variables (precipitation and temperature) were obtained from the ERA5 reanalysis dataset (Hersbach et al., 2020), which provides forcing data at a high spatial resolution of 0.25 degrees. The selection of these two variables was made for several reasons. First, precipitation and temperature are available across all 16 selected climate models and three scenarios, ensuring consistency and reliability in the data used for analysis. Second, using these variables maintains coherence with other studies within the same research framework, allowing for better comparability and integration of results. Third, precipitation and temperature are fundamental drivers of hydrological processes, making them the most relevant basic variables for projections. Finally, focusing on these two key variables keeps the approach parsimonious regarding data availability and processing requirements, enhancing the efficiency and reproducibility of the analysis while still capturing the essential climate dynamics for groundwater projections. The groundwater level data were sourced from the ADES (Accès aux Données sur les Eaux Souterraines) database (https://ades.eaufrance.fr/; Winckel et al., 2022), specifically focusing on climate-sensitive wells minimally influenced by human activities and exhibiting strong sensitivity to climate variability (Baulon et al., 2022a).

To generate future GWL projections, climate data from 16 CMIP6 models were used as inputs in deep learning models trained on data from ERA5. Three Shared Socioeconomic Pathway (SSP) scenarios were considered: SSP2-4.5 (moderate emissions), SSP3-7.0 (severe emissions), and SSP5-8.5 (extreme emissions). These scenarios represent different future pathways of greenhouse gas emissions and socioeconomic factors, allowing for a comprehensive assessment of potential climate change impacts on groundwater resources. We chose to utilise the bias-corrected datasets from the NEX-GDDP-CMIP6 dataset, which has data for only one variant for each CMIP6 model. Hence, the uncertainty related to internal climate variability is not considered (Deser et al., 2012).

5.3. Methodology

5.3.1. DL models and neural network architectures

This study employs a deep learning approach to project future groundwater levels in northern France using climate projections from CMIP6 models under different emission scenarios. The methodology builds upon Chidepudi et al.'s (2024b) findings, which demonstrated the effectiveness of models trained with a multi-station approach on clustered data combined with wavelet preprocessing.

While the efficiency of LSTMs has long been demonstrated, other recurrent-based architectures have emerged more recently, like the Gated Recurrent Unit (GRU), as Cho et al. (2014) described. GRUs have a simpler architecture than LSTMs, which can lead to faster training. Additionally, GRUs have shown comparable or sometimes superior performance to LSTMs in various sequence modelling tasks, making them an attractive alternative for groundwater level modelling (Chidepudi et al., 2023a). The core of our approach then utilises GRU neural networks according to their ability to capture long-term dependencies in sequential data and their computational efficiency.

Our neural network models are trained with inputs from ERA5 data pre-processed using the Boundary Corrected Maximal Overlap Discrete Wavelet Transform (BC-MODWT), implemented with a 'la8' (least asymmetric) wavelet and four decomposition levels, as in Chidepudi et al. (2023a). This technique effectively separates input signals into different frequency bands while preserving time information and mitigating boundary effects (Quilty & Adamowski, 2018; Chidepudi et al., 2023a). It is important to note that for each input time series or feature, BC-MODWT produces five new time series: four wavelet detail levels corresponding to the decomposition levels and one smooth approximation. For more details about discrete wavelet transform and MODWT, the reader can refer to the very rich literature on the application of wavelet methodology to hydrology, such as Labat et al. (2000), Percival & Walden (2000) and Massei et al. (2017). This pre-processing step is applied to climate reanalysis data, which serves as input for the GRU models during training and validation. Specifically, if the original input consisted of N features, the BC-MODWT pre-processing would result in a 5N time series as input to the model, each maintaining the original temporal resolution but capturing different frequency components of the original signals. The observed groundwater levels are used as the target variable.

This study adopts a multi-station approach, where GRU models are trained within each GWL cluster (annual, mixed, or inertial) using aggregated data from multiple stations. Most recent studies have focused on single-station approaches, which consist of training DL models based

on a single target time series (i.e., the DL model learns from known values of the one-time series to be eventually simulated). For the last couple of years, recent research has suggested that models trained with more diverse data can result in more reliable hydrological projections (Wi & Steinschneider, 2022), as this would enhance the capabilities of the models for generalisability and transferability. The approach consists of training DL models for time series regression on multiple time series (in our case, GWL time series) simultaneously, hence leveraging a wider range of relevant values or hydrological events than available in a single station time series. The approach has already been called the "global models" or "multi-well" approach (Heudorfer et al., 2024) or "multi-station" (Chidepudi et al., 2023b & 2024b). This key finding highlights the potential advantages of multi-station approaches over single-station methods. Typically, the multi-station approach used herein leverages collective information from stations with at least 42 years of data (1970-2022), enhancing the model's ability to capture spatial patterns and increasing the generalisability of the projections. Only dynamic variables, such as precipitation and temperature, are used as input features.

Data pre-processing includes normalising each input variable individually to be in the 0-1 range and reshaping it into a 3D format suitable for GRU models. A sequence length of 48 months is used to capture long-term patterns in the data. To enhance robustness and mitigate the effects of random weight initialisation, multiple GRU models (10) are trained with different initialisations, creating an ensemble approach. As described in Chidepudi et al. (2024b), hyperparameter tuning was performed using Bayesian optimisation. The models are trained using early stopping and model checkpointing techniques to prevent overfitting and save the best-performing model. A 20% validation split is used to monitor performance during training. Importantly, models are trained with collective data from all stations in each cluster with at least 42 years of data (1970-2022), ensuring a comprehensive historical context.

5.3.2. Assessing trends and variability in projected GWL under three SSP scenarios

The trained ensemble GRU models are applied to the wavelet-transformed CMIP6 climate projection data to generate projections. As a reminder, the total number of features would equal 10 (two variables, i.e., precipitation and temperature, each decomposed into five wavelet components). The median of the ensemble predictions is calculated for each CMIP6 model, as shown in Figure 5.2. Then, the median across all models is computed to produce a robust projection for each scenario. For each GWL station, up to 160 projected time series

were generated, corresponding to 16 climate models and 10 initialisations of the DL model. For each climate scenario, the results are presented either: i) for each climate model individually (i.e., for each climate scenario, 10 GWL projected time series corresponding to 10 DM model initialisations; cf. Fig. 2); ii) 2- for all climate models together (i.e., 160 GWL projected time series corresponding to 16 climate models with ten initialisations of the DL model). A median time series of the 10 or 160 projections is derived in these two cases. It is then used to assess the ensemble trend and explore a possible change in the amplitude of the variability of GWL. In other words, the ensemble median time series (ensemble of either only 10 or 160 projected time series) is tested for stationarity in the weak sense (i.e. change in mean and variance) throughout the period 2030-2100 to assess whether a change in GWL is to be expected on average, and if the overall variability would also tend to change. Figure 5.3 illustrates such changes using one randomly selected GWL projection and is described later in the text. Ultimately, it comes to exploring whether water levels and their amplitude of variation (difference between high and low levels) will be expected to increase, decrease, or remain unchanged.

Trend analysis is performed using the Correlated Seasonal Mann-Kendall (CSMK) test, an extension of the classical Mann-Kendall trend test proposed by Hipel & McLeod (1994). The CSMK test is a non-parametric statistical method that detects monotonic trends in seasonal time series data with serial correlation. It is particularly suitable for hydrological and climate variables (Hussain et al., 2019). This test accounts for both seasonality and the correlation between observations in consecutive months or seasons, addressing the limitations of the classical MK test. The CSMK test does not require the data to follow a specific distribution and can handle missing values and outliers. It operates by separating the time series into seasonal groups, calculating the MK test statistic for each season, adjusting for serial correlation, and then combining the results to obtain an overall trend assessment. A standardised test statistic is computed and compared to a critical value (using a P-value of 0.05) to determine if a statistically significant trend exists. The CSMK test is robust against non-normality and censored data, making it particularly valuable for analysing trends in groundwater levels, precipitation, temperature, and other variables relevant to groundwater projections using

deep learning models, especially when these variables exhibit strong seasonal patterns and serial correlation. Sen's slope is also computed as part of the CSMK test.

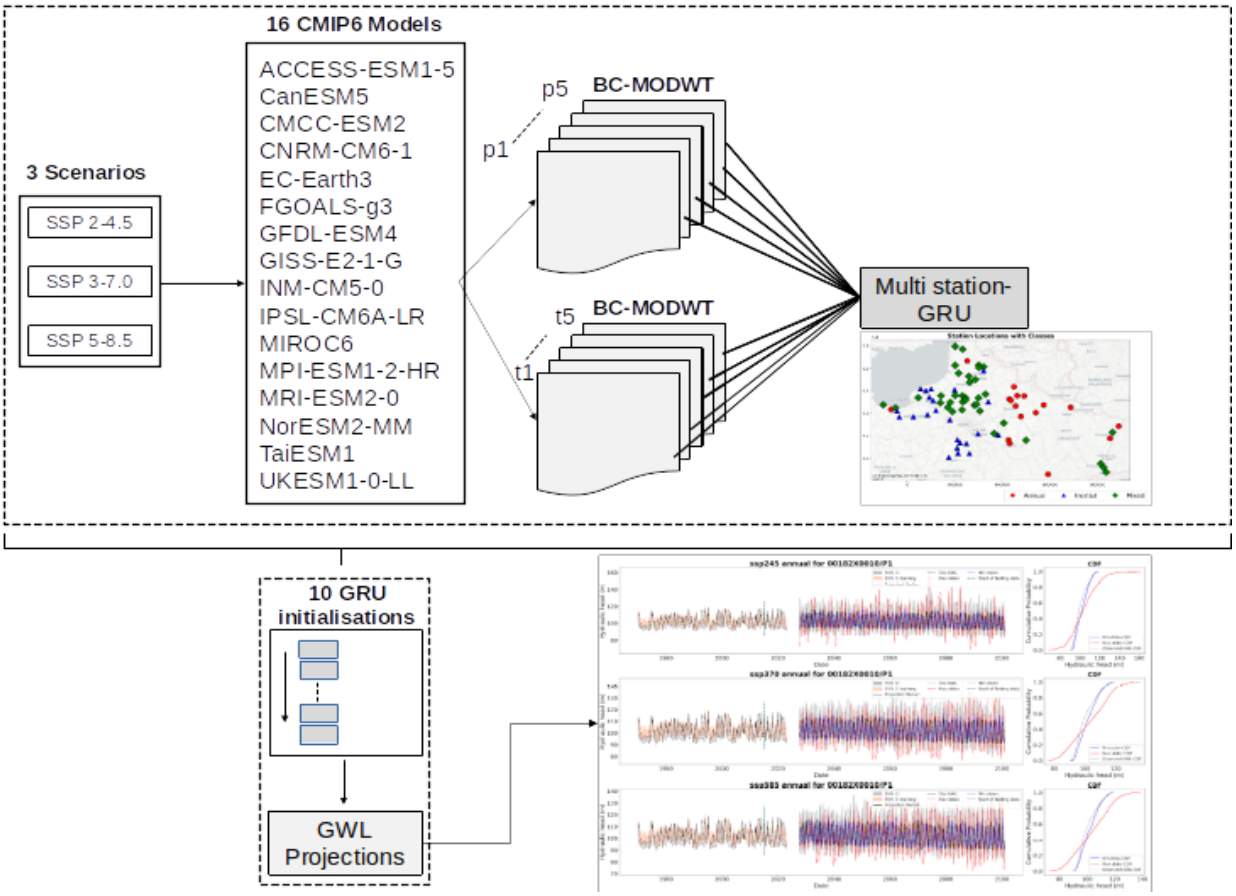

Figure 5.2: Structural workflow of the methodological approach for projections.

This analysis is conducted for each groundwater station, climate scenario, and GWL type. For assessing trends in GWL, the time series used for the MK trend test is the median of the ensemble. The ensemble would consist of either 10 or 160 projected time series:

- 10 projected time series when climate models are taken separately, i.e. for each climate model, ten projections corresponding to the ten different initialisations of the DL model are obtained;

- 160 projected time series when all climate models are grouped, leading to 10 DL initialisations * 16 climate models.

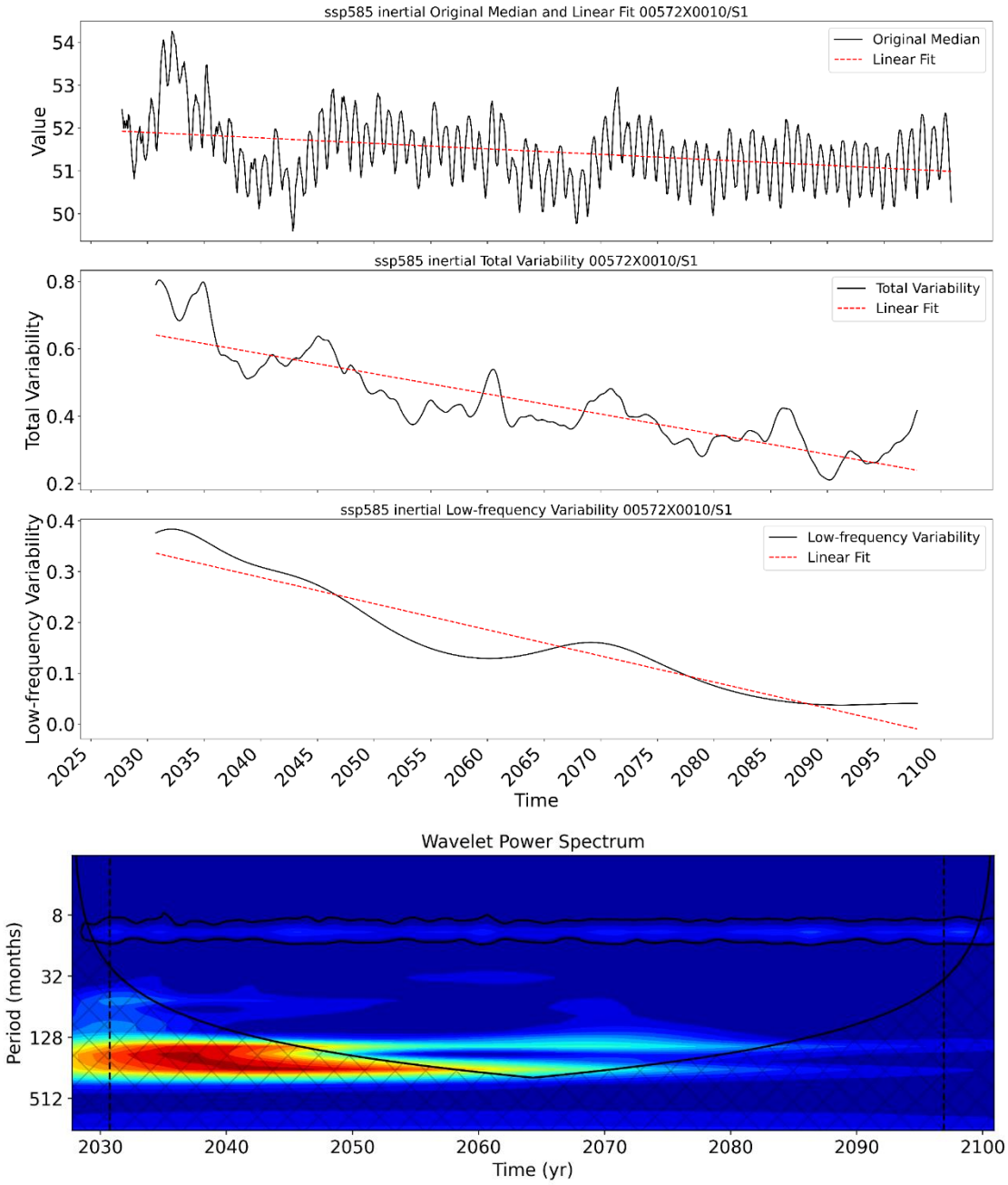

Figure 5.3: a) median of projections and fit line b) Scale averaged wavelet power for total variability c) Scale averaged wavelet power for low-frequency variability d) Continuous wavelet spectrum (scalogram) of GWL time series: red indicates high variance, and blue indicates low variance.

For assessing a change in the variability at each station, the time series used for the MK trend test is the estimated variance across time of the median of several projections. The times series' variance is estimated using so-called scale-averaged wavelet power, following the procedure described in Torrence and Compo (1998). At each time point, the scale-averaged wavelet power measures the variance in GWL for either all-time scales or a range of time scales. In the first case (all time scales), it measures the total variance of the time series across time (i.e. for all time steps). In contrast, the second case (selected range of time scales) measures the variance associated with one particular frequency band across time. The reader is referred to Torrence and Compo (1998) for a detailed and comprehensive explanation of wavelet scale-averaging's calculation and main interest.

Figure 5.3 illustrates this for one randomly selected GWL projection, representing the variance through time obtained for one GWL projection for a time scale greater than five years (i.e., low-frequency fluctuations of GWL at this station). In this example, scale-averaged wavelet power shows that total variability (Fig. 5.3b) and >5-yr low-frequency variability (Fig. 5.3c) tend to decrease over time. In order to prevent the results from being too much affected by edge effects, we removed the first and last 36 months from these scale-averaged power time series (Fig. 5.3b and c), which corresponded to removing as many wavelet coefficients as possible falling into the so-called cone of influence (cross-hatched area in Figure 5.3d). Such coefficients can be identified on the continuous wavelet spectrum of the time series and are located before the first. After the second vertical dashed line (Fig. 5.3d). The continuous wavelet spectrum also clearly shows that low-frequency variability (i.e. variance for periods higher than ~5 years / ~60 months on the plot) tends to decrease through time.

5.4. Dispersion of climate change impact projections on various GWL types under contrasting emission scenarios

Here, all 16 downscaled GCM were used as input to the DL models with ten different parameter initialisations (i.e. initialisation of the neural network weights) to create an ensemble of 160 projections per SSP scenario at each time point.

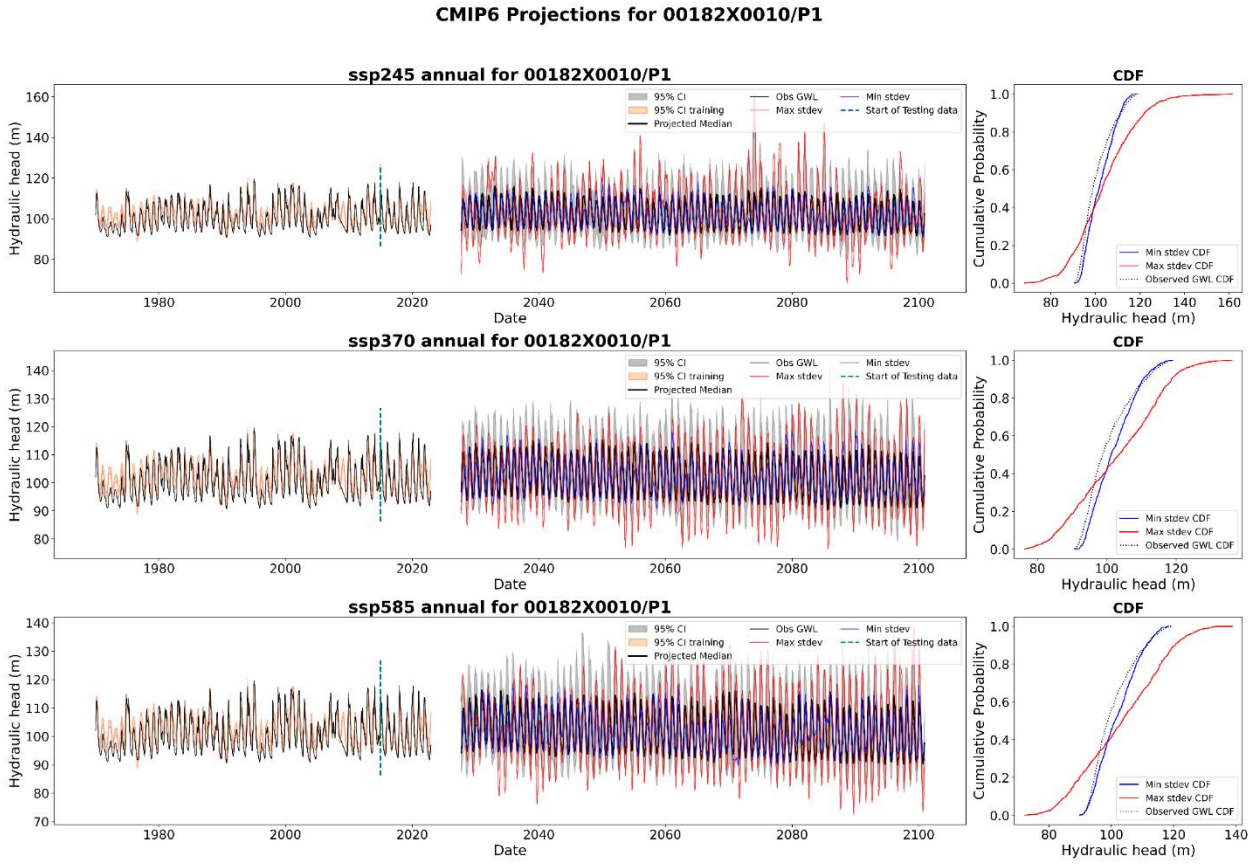

Figure 5.4: Annual type groundwater projections for three SSP scenarios (Top-bottom): a)(left panel)Training and testing results with confidence intervals, b)(middle panel) Projections for SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios show median (black), high variance (HV, red), low variance (LV, blue) projections, and confidence intervals (grey). c) (right panel) Cumulative Distribution Functions (CDFs) comparing historical observations (dotted black) with HV and LV projections (resp. as red and blue CDFs).

While projections are usually presented and quantified using the ensemble mean or median, relying on these metrics alone could suppress the information on the variations and uncertainty.

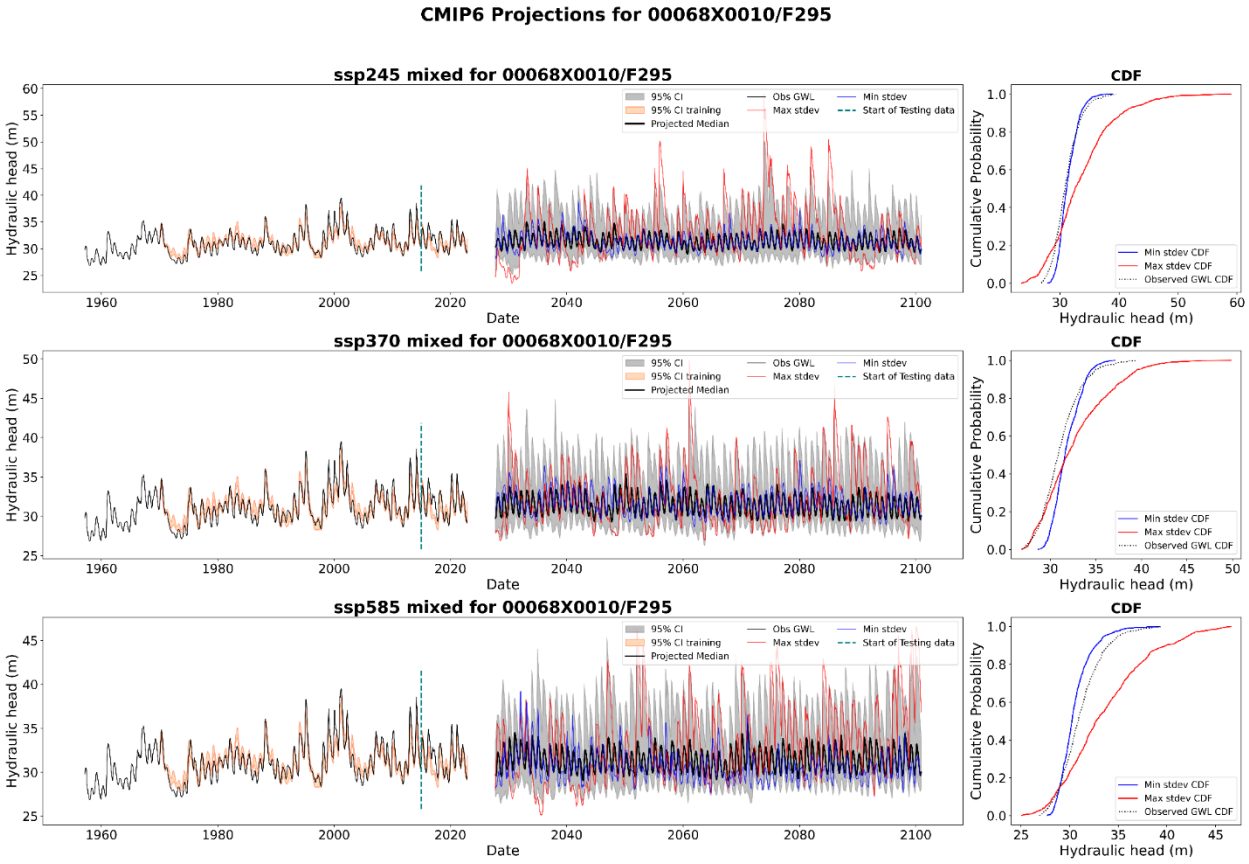

Figure 5.5: Mixed-type groundwater projections for three SSP scenarios (Top-bottom): a)(left panel)Training and testing results with confidence intervals, b)(middle panel) Projections for SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios show median (black), high variance (HV, red), low variance (LV, blue) projections, and confidence intervals (grey). c) (right panel) Cumulative Distribution Functions (CDFs) comparing historical observations (dotted black) with HV and LV projections (resp. as red and blue CDFs).

Here, we chose to represent our results in percentile rather than box plots, as recently suggested by Müller & Döll (2024), who found such a representation more suited to support participatory climate change adaptation processes and uncertain local climate hazards.

Figure 5.4-Figure 5.6 show the groundwater projections of three types (annual, mixed and inertial) for three SSPs. The left panel shows the training and test results of the historical period. The central panel shows the projected median groundwater level and the 95%

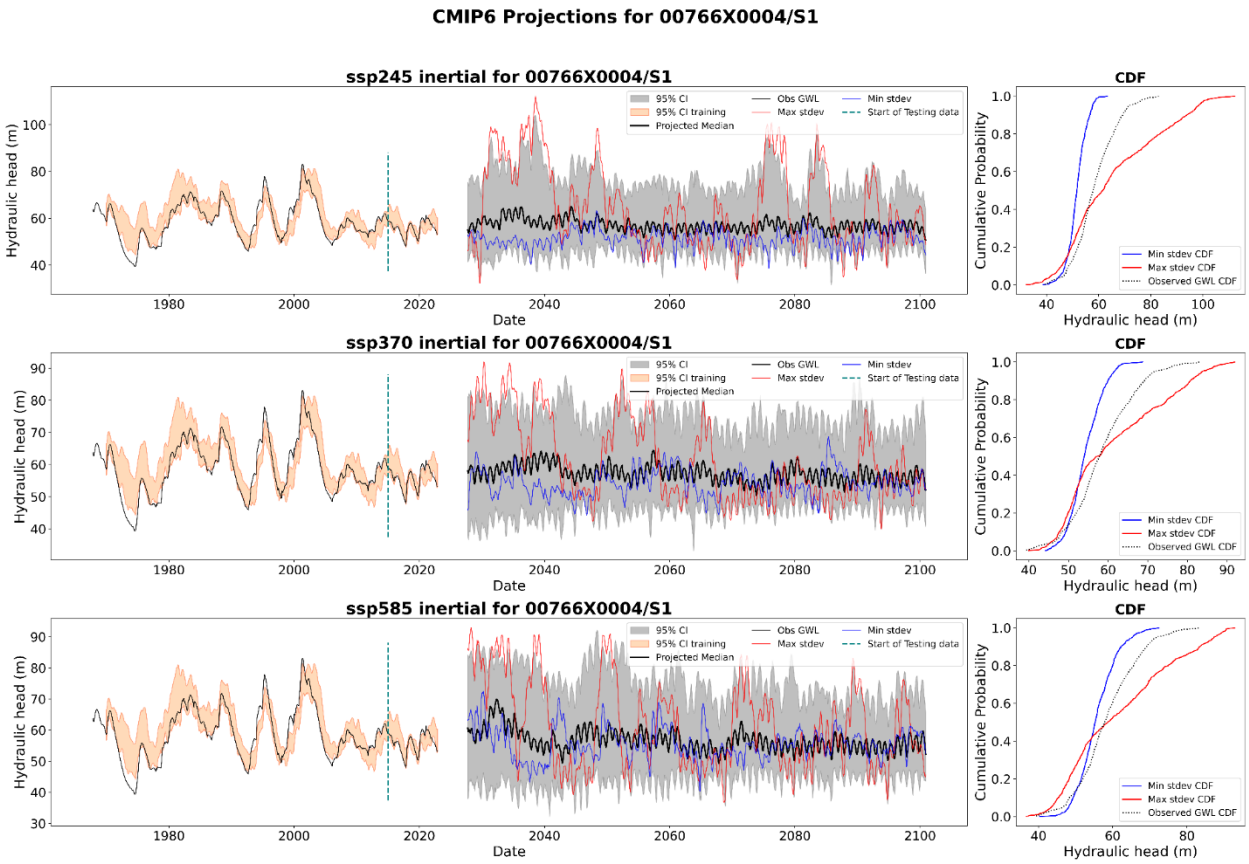

Figure 5.6: Inertial type groundwater projections for three SSP scenarios (Top-bottom): a)(left panel)Training and testing results with confidence intervals, b)(middle panel) Projections for SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios show median (black), high variance (HV, red), low variance (LV, blue) projections, and confidence intervals (grey). c) (right panel) Cumulative Distribution Functions (CDFs) comparing historical observations (dotted black) with HV and LV projections (resp. as red and blue CDFs).

confidence interval. We highlighted the projections with the highest and lowest variability (from now on referred to as HV and LV, resp. represented as red and blue lines in Figures 5.4-5.6), as these naturally correspond to either more pronounced or more dampened extremes resp, for HV and LV projections. In addition, comparisons between GWL types, i.e. with different amplitudes of low-frequency variability, provide information on how various representations of climate variability may impact the projections released.

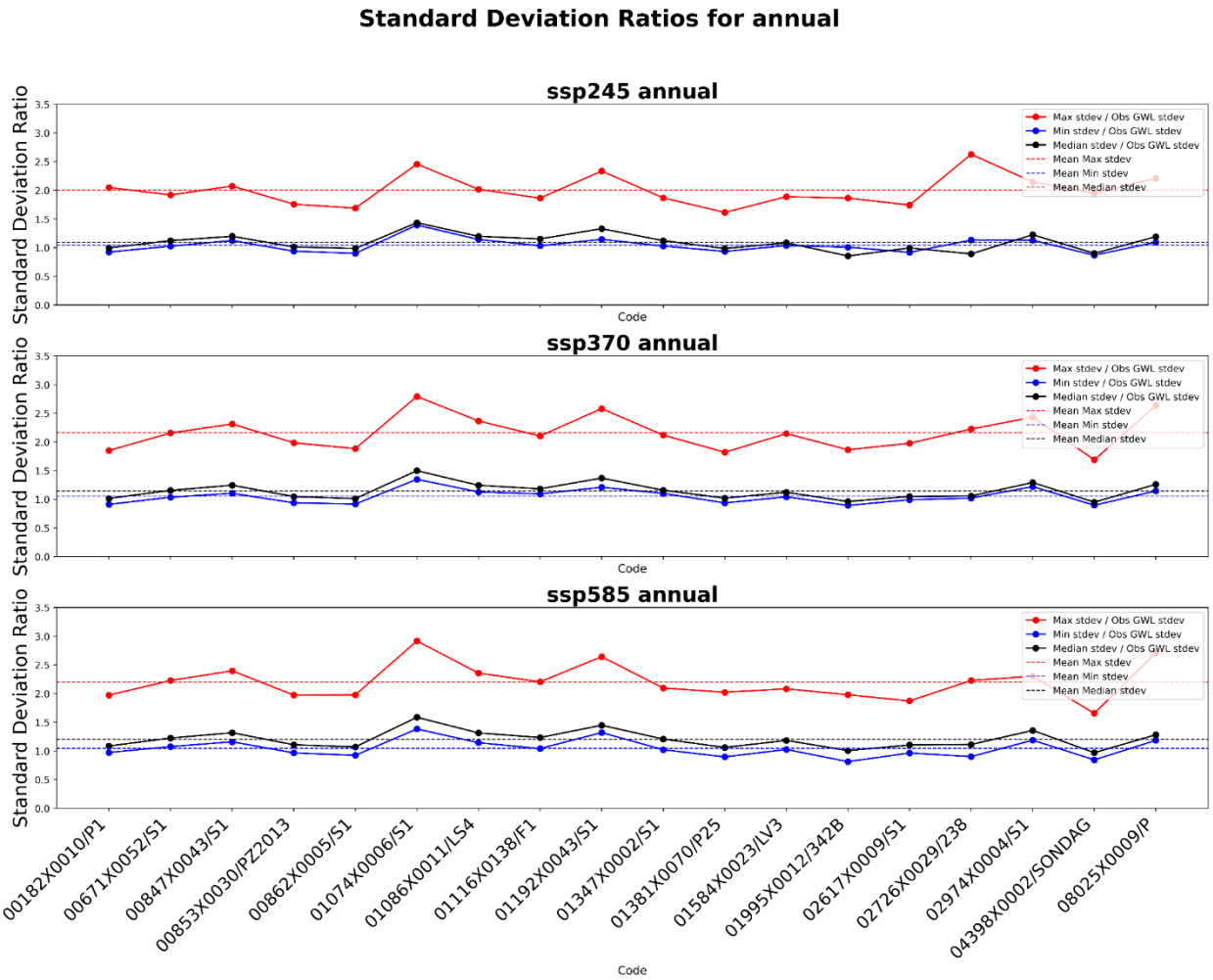

Figure 5.7: Variability comparison of all stations in annual-type GWL: Ratios of Median(black), HV (red) and LV (blue) projection standard deviations to observed standard deviations across all stations(X-axis) and emission scenarios (top-bottom).

The results for the training and testing stage are always quite satisfactory, with small confidence intervals (Figures 5.4a-5.6a). On the other hand, big differences may exist between the different projections for any given GWL type (annual, mixed and inertial), hence translating the uncertainty linked to the various climate models (Figures 5.4b-5.6b). While LV projections are relatively close to the median, HV ones would deviate substantially from the median and LV time series. The median time series always displayed a much more pronounced seasonal variability and a much lower low-frequency variability (the median time series appears somewhat "shrunk") compared to historical observations - noticeable for all GWL types - and even more particularly for the inertial. This increased seasonal variability is due

to the stochastic nature of low-frequency variability (stochastic noise eventually cancels out on average). In contrast, the annual cyclicity is, by definition, almost entirely deterministic:

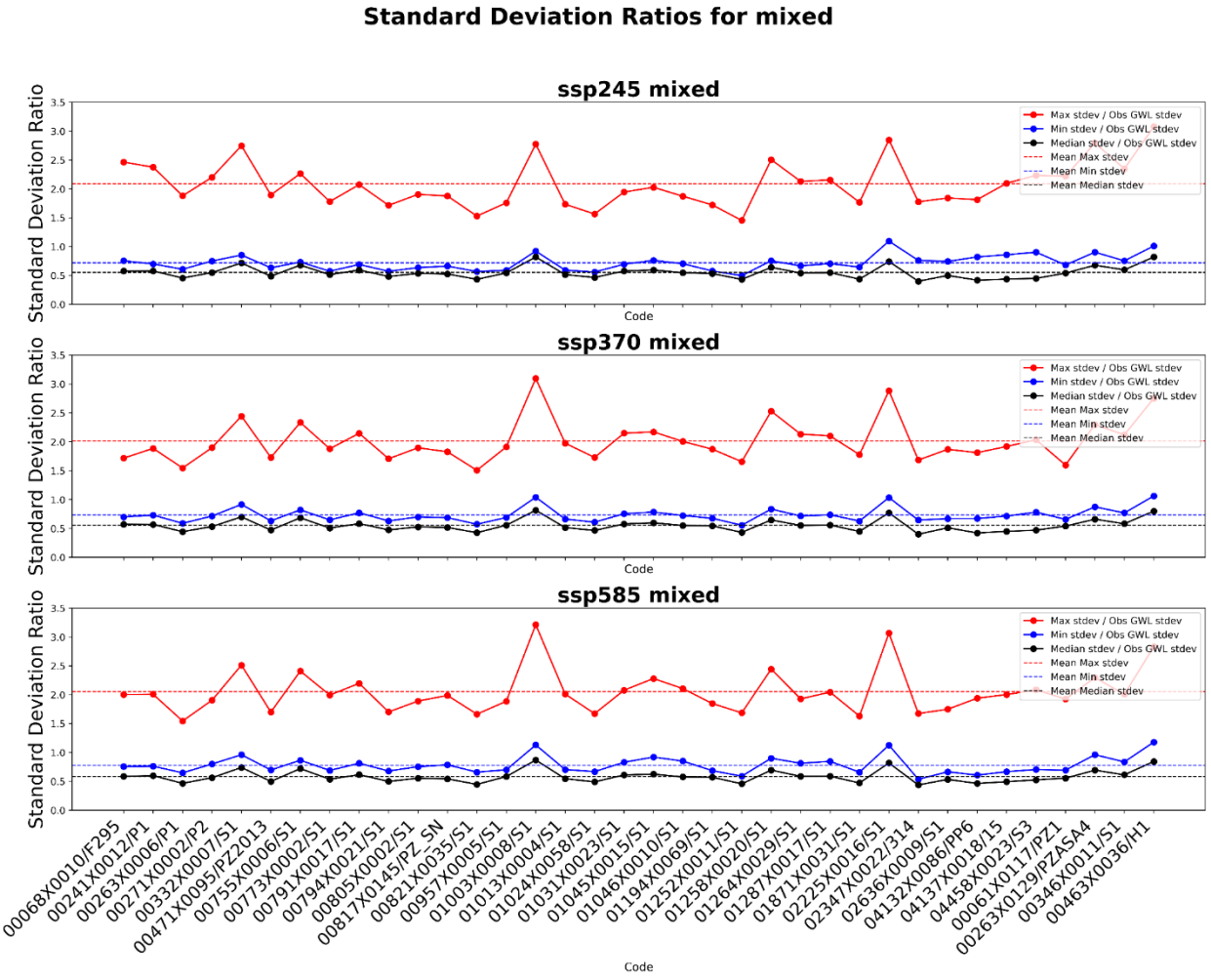

Figure 5.8: Variability comparison of all stations in mixed-type GWL: Ratios of Median(black), HV (red) and LV (blue) projection standard deviations to observed standard deviations across all stations(X-axis) and emission scenarios (top-bottom).

median computation from all 160 projections at each time point ultimately results in a low amplitude of any other variability than the annual cycle, which, in contrast, is over-expressed.

On the other hand, comparisons of the CDFs (Figures 5.4c-5.6c) of the LV and HV projections with the historical observed time series show that for the annual and mixed types, LV projections seem to approximately fit the historical observations for all 3 SSP scenarios

(Figures 5.4-5.5). which is not the case for the inertial type (Figure 5.6), for which the variability of the LV projection is always lower than for the historical period.

To check whether this observation drawn from the three randomly selected annual, mixed and inertial GWL stations presented in these figures can be generalised, we compared the variability of HV and LV projections to that of the historical GWL time series at each station.

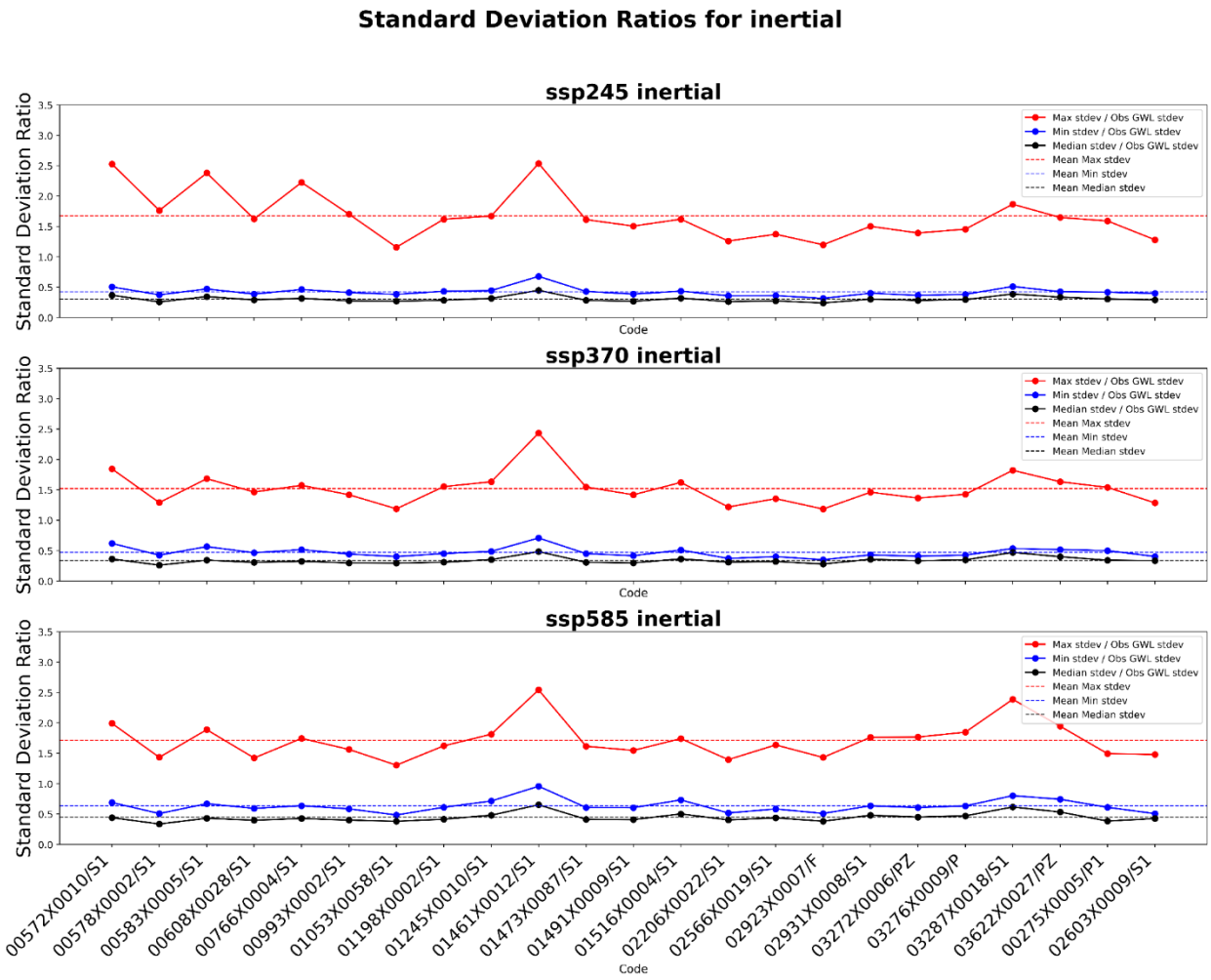

Figure 5.9: Variability comparison of all stations in inertial-type GWL: Ratios of Median(black), HV (red) and LV (blue) projection standard deviations to observed standard deviations across all stations(X-axis) and emission scenarios (top-bottom).

To this aim, we simply computed the standard deviation of HV, LV and observed GWL time series at each station and represented the ratios of HV standard deviation to observed standard deviation and of HV standard deviation to observed standard deviation (Figures 5.7-5.9) for all 3 GWL types and all emission scenarios.

The results confirmed the conclusions drawn from Figures 5.4 to 5.6 presented above. For all stations of the annual type and all emission scenarios (Fig.5.7), the LV to observation ratio is always close to 1, meaning that the overall variability of LV and observed historical GWL are always close.

In contrast, the HV projection always has a much higher standard deviation than the observation (~1.5 to 2 times), which is also true for the mixed-type GWL, albeit to a - slightly - lesser extent but does not hold for GWL projections of inertial types (Fig.5.9). Finally, as already mentioned above, one can notice here that except for the annual GWL type, the variability of the median projection time series is always lower than the LV projection for any given GWL station; the same explanation as provided above can be given here. No obvious differences among the different emission scenarios could be observed for any GWL type (Fig. 5.7 to 5.9).

5.5. Time evolution of GWL: Future trends and variability for annual, mixed and inertial types

5.5.1. Trends based on the ensemble approach

In the previous section, we described the different projections that could be obtained owing to the use of 16 different models for different GWL types and three emission scenarios. Here, we focus the analysis on the time evolution of GWL. For clarity, we evaluated such changes using the median time series only. Although we showed that the total amplitude of median projections is artificially lower than that of the observed historical time series, the aim here is not to assess the change between the historical and the future periods but only the change during the future period. We use the median as in most other works (Martel et al., 2022; Wunsch et al., 2022). In particular, we examined three different aspects of GWL time evolution: 1- we first assessed the potential changes in water levels using the MK trend test

performed on the median GWL time series, 2- we tested for changes in the variability of GWL using the MK trend test on scale-averaged power of the median GWL series, as described in the methodology section, 3- we repeated step 2- with a specific focus on the low-frequency variability. Here, low-frequency variability corresponds to time scales above five years, which have been recognised as responsible for several extreme events in groundwater levels in northwestern Europe (Massei et al., 2010; Baulon et al., 2022b; Rust et al., 2019; Neves et al., 2019; Liesch & Wunsch, 2019).

It is clear from Figure 5.10 that almost all GWL would decrease, regardless of the emission scenario and the GWL type. On the other hand, the overall variability of GWL through time showed more contrasted results (Figure 5.11). In all three SSP scenarios, the total variance of annual-type GWL is expected to increase while decreasing for almost all inertial and mixed types. No clear difference shows up between the scenarios. GWL types strongly influence the trends in total variance. Specifically, Annual type aquifers predominantly show increasing variance (blue circles). Inertial and mixed-type aquifers mostly exhibit decreasing variance (red triangles and diamonds). The increasing variance in annual type aquifers suggests these areas may face more extreme fluctuations in groundwater levels, i.e., be more prone to extremely high and low groundwater levels. The spatial patterns of increasing and decreasing variance trends persist across all three SSP scenarios (SSP2-4.5, SSP3-7.0, SSP5-8.5),

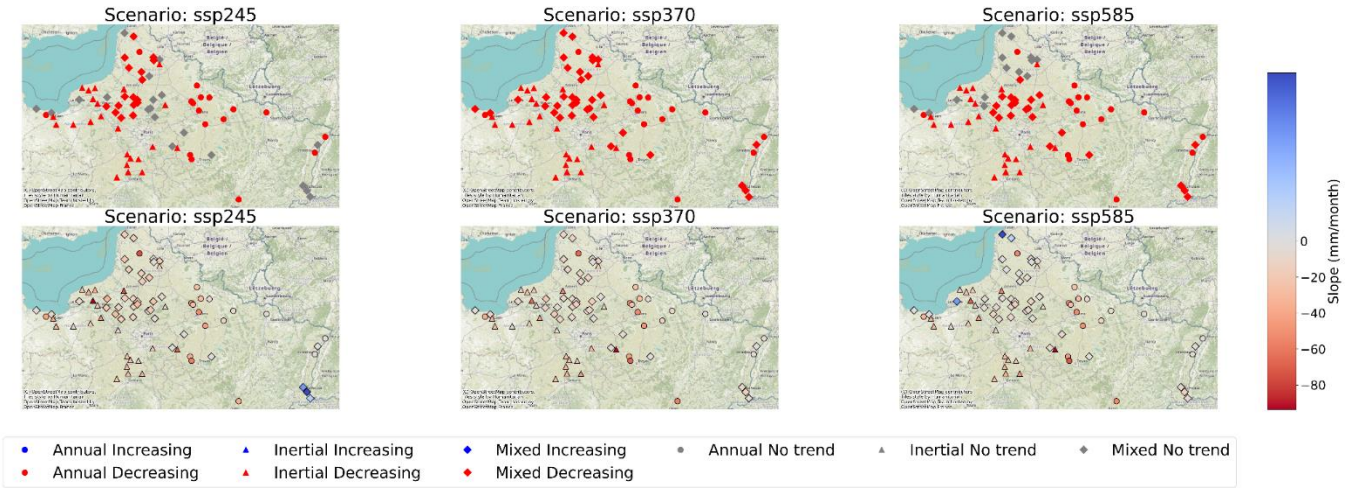

Figure 5.10: GWL trend on multi-model ensemble median projections from 16 CMIP6 models and 10 DL models for each scenario: Blue (Increasing) and Red (Decreasing)

suggesting that the overall pattern of change is relatively robust to different emission scenarios.

The trend in total variability (Figure 5.11) of annual type is always increasing. In contrast, for mixed and inertial types, it either decreases or shows no significant trends (central and South-central parts of the area).

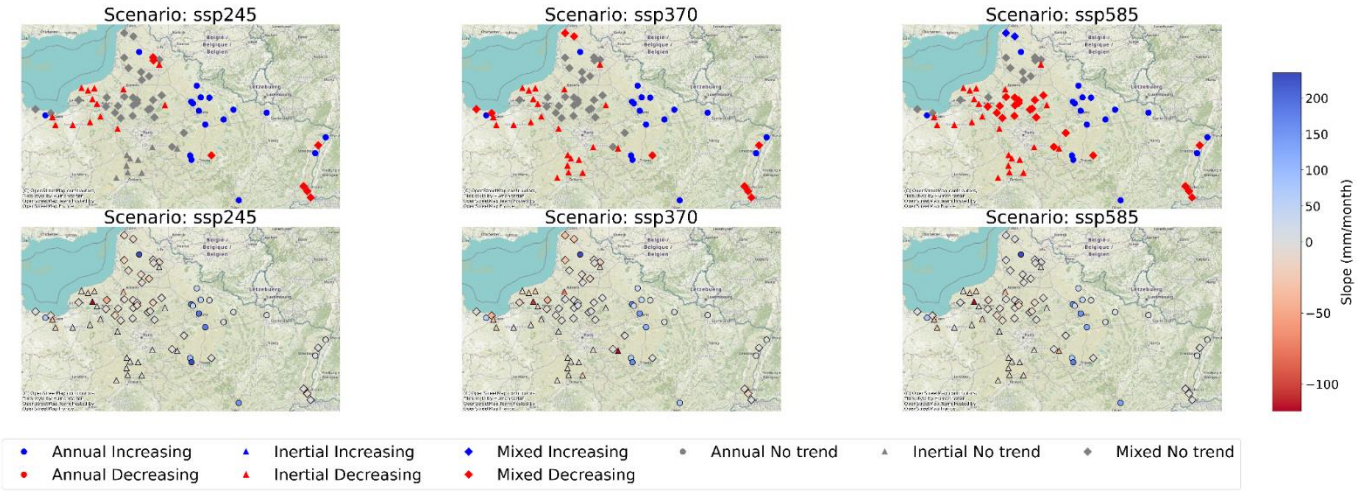

Figure 5.11: Trends and slopes in the total variability of projected GWL

We then focused on low-frequency, i.e. interannual-to-decadal variability only, defined as fluctuations over time scales exceeding five years. Such time scales are usually employed to compare hydro-climatic variability and large-scale climate variations (often along with their moving average) as depicted by climate indices and teleconnections mentioned in the introduction. Figure 5.12 shows the trends and slopes of low-frequency variance obtained from CWT. Changes in low-frequency variance could impact the occurrence and intensity of multi-year drought or wet periods. Our analysis (from Figure 5.12) revealed distinct trends across different emission scenarios. Under the SSP 2-4.5 scenario, many stations showed increasing trends. In contrast, the number of stations with increasing trends significantly decreased to very few for the SSP 3-7.0 scenario, while all stations displayed decreasing trends under the SSP 5-8.5 scenario.

To summarise (Figure 5.11-5.12), as emission scenarios worsen, the overall variability increases for annual-type stations, whereas interannual to decadal variability decreases for all

stations. Generally, more pessimistic emission scenarios correlate with a reduction in long-term groundwater level variability. While the magnitude of variability increase is typically low, it appears more pronounced in less pessimistic scenarios. These findings raise an important question: Could this be an effect of climate change reducing the amplitude of natural climate variability? The results suggest that low-frequency natural climate variability may diminish under more pessimistic emission scenarios, directly impacting interannual to decadal water level variations. This observation warrants further investigation into the complex interactions between anthropogenic climate change and natural climate variability that we did not explore in the current study.

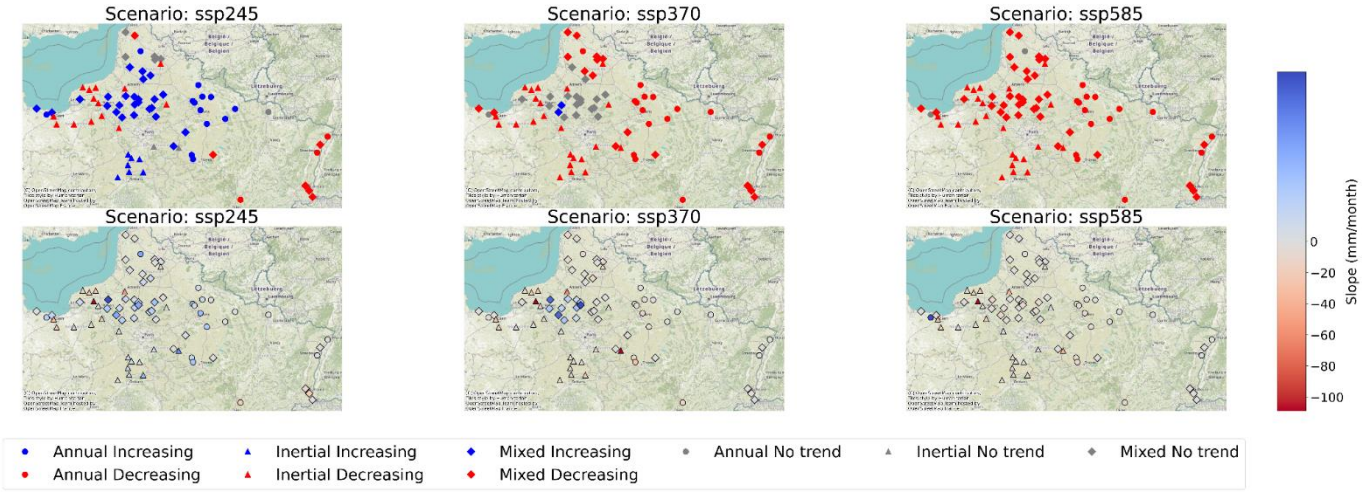

Figure 5.12: Trends and slopes in low-frequency variability from CWT

5.5.2. Differences in GWL projections according to different climate models

In the previous section, we explored trends in GWL and temporal changes in GWL variability using the median time series computed from all projections generated with different climate models (16 models) and different initialisations (10 initialisations). It is well known that estimating the differences in hydrological projections from climate change models and scenarios is crucial for understanding the sources of uncertainty and communicating contrasted but equally plausible outcomes to stakeholders. This is why we extracted and studied the highest and lowest variability of all projections in section 4 (Figures 5.4 to 5.6). The results of section 4 showed that large differences between projections would most likely

be due to the use of 16 climate model inputs rather than from DL models. Although we did not specifically assess their respective parts in the total uncertainty, comparing the range of uncertainty related to DL model initialisation on observed data (Figures 5.4a to 5.6a) and that obtained for projections (Figures 5.4b to 5.6b) showed quite clearly that climate model inputs would result in a much higher difference in GWL projections. In this section, we then explored the differences in trends and changes in variability of projections owing to the 16 climate models taken separately.

Similarly to what was done for the ensemble approach in section 5.1, we examined the direction (increasing or decreasing) and slope magnitudes of trends (level and variability) at all stations. As explained in the methodology section, in this case, the time series used for calculating GWL trends corresponded to the median of the 10 DL model initialisations for each climate model and each scenario. For assessing trends (i.e. potential change) in GWL variability (either total or considering low-frequency variations only), the time series used was the scale-average wavelet spectrum of the same median of the 10 DL model initialisations.

Figures 5.13a and 5.13b show, respectively, the trend direction and Sen's slopes of GWL projections from 16 CMIP6 models (as lines) and three scenarios each (as columns). Although the results seem rather contrasted, three main outcomes could be distinguished. First, for all SSP scenarios, decreasing or non-significant trends (with very slight decreasing or increasing slopes less than 10 mm/month) appeared to dominate. Amongst all models, most of the statistically significant increasing trends were observed for only SSP 2-4.5 with 4 out of 16 models (CMCC, GFDL, MRI and, to a lower extent, FGOALS). These models showed increasing trends with magnitudes up to more than 25 mm/month. Second, for SSP 3-7.0 and SSP 585 (2nd and 3rd column), all models showed mostly decreasing GWLs and some non-significant trends, except the CanESM5 and MPI-ESM1 models. The magnitude of negative slopes for SSP 3-7.0 and SSP 5-8.5 showed decreasing GWLs at rates as low as, or lower than, -50 mm/month. Although slightly increasing slopes could be identified in many models for those 2 SSP scenarios (e.g., EC-Earth3 SSP 3-7.0, CNRM SSP 3-7.0 and SSP 5-8.5), they were never statistically significant, except for the CanESM5 and MPI-ESM1 models. Third, it was noticeable that for the worse scenario, SSP 5-8.5, all stations located in the most eastern part of the area were affected by decreasing trends (or with no significant trends) for all 16 climate models; these stations consisted mainly of annual-type GWLs.

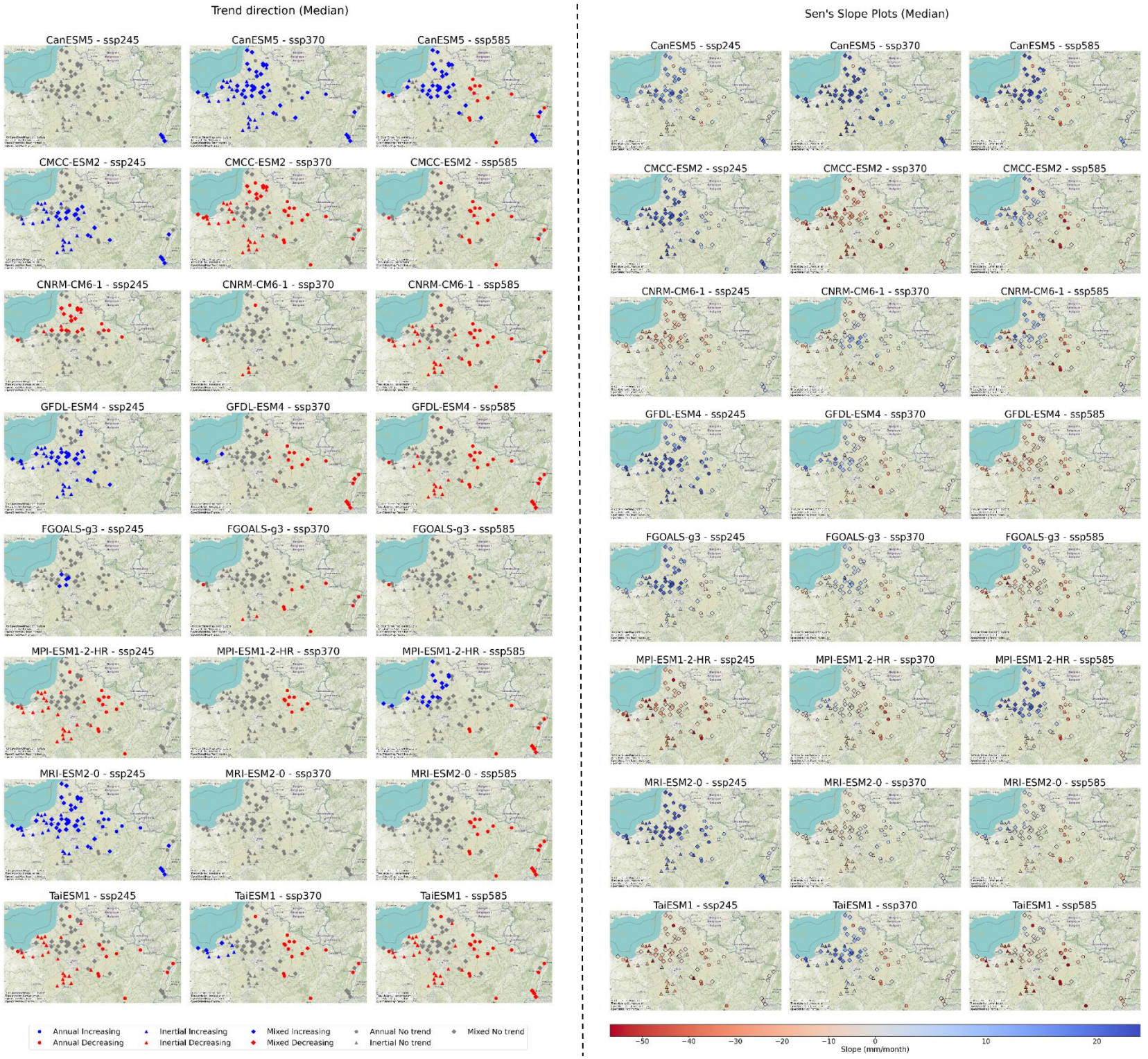

Figure continued….

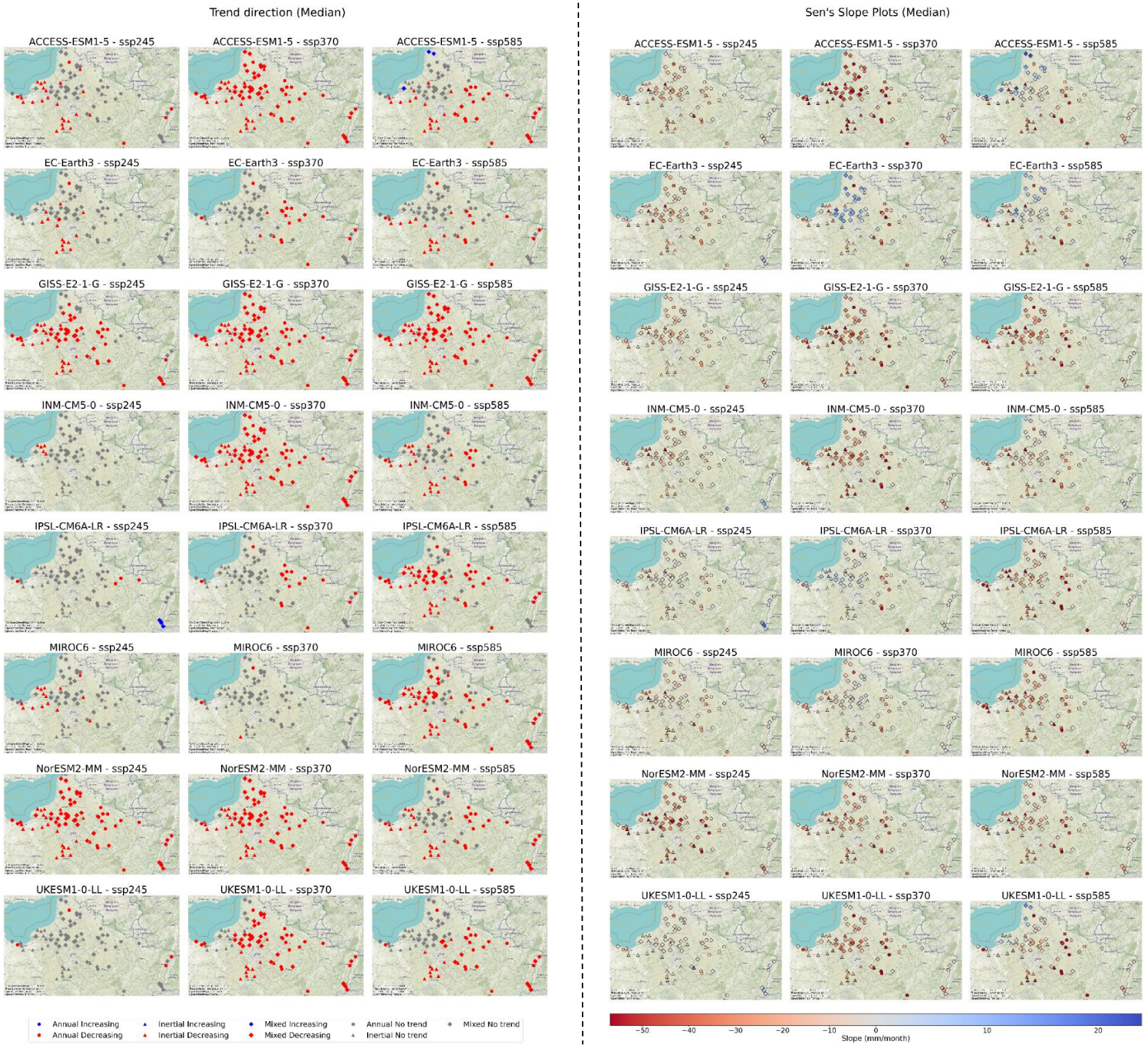

Figure 5.13: Individual groundwater level trends and slope magnitudes from 16 CMIP6 models under SSP2-4.5 (left), SSP3-7.0 (middle), and SSP5-8.5 (right) scenarios for 2100: Blue indicates increasing trends(a) or positive slopes(b), and Red indicates decreasing trends(a) and negative slopes. Grey indicates no significant trend.

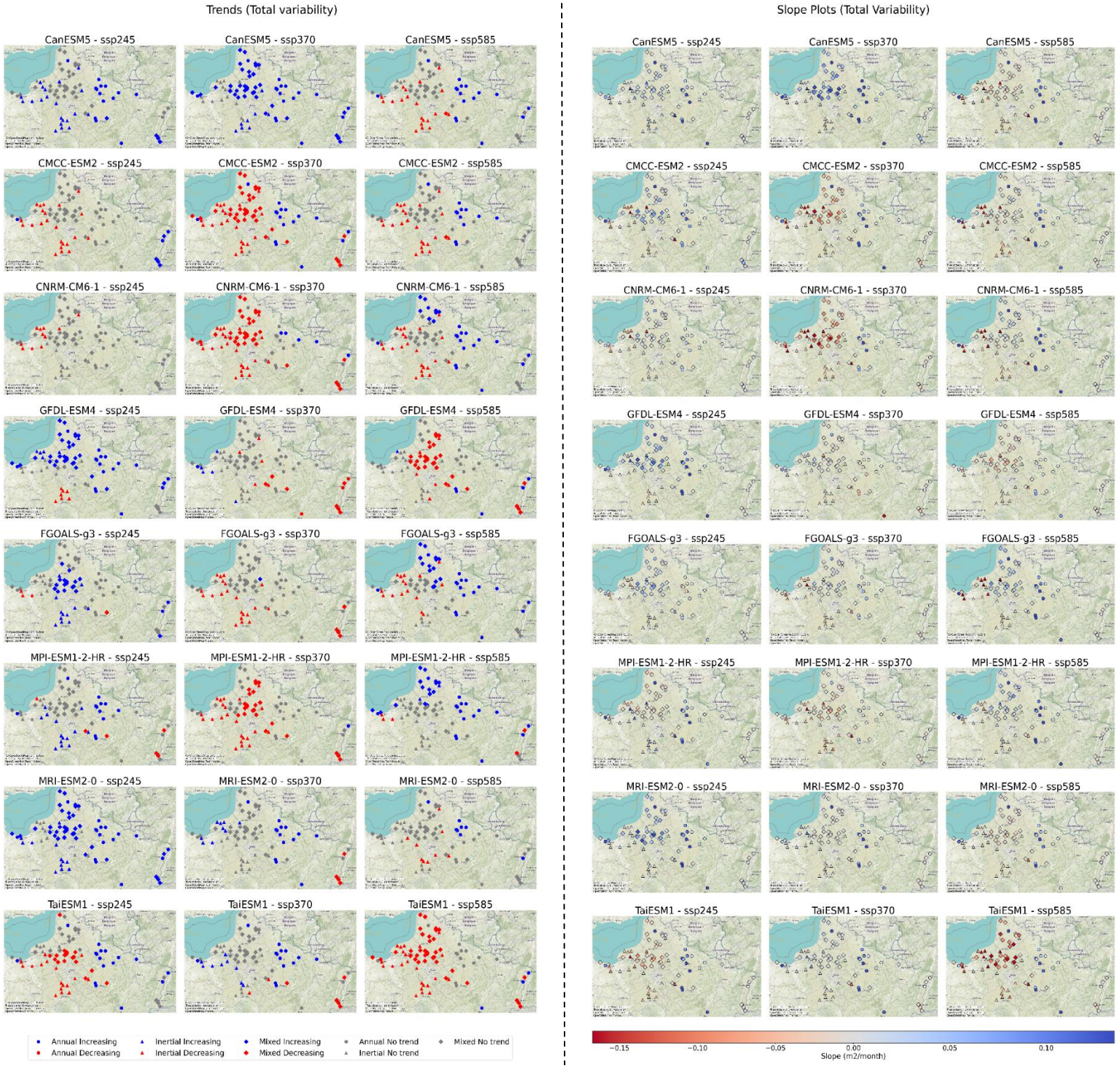

Figure Continued……

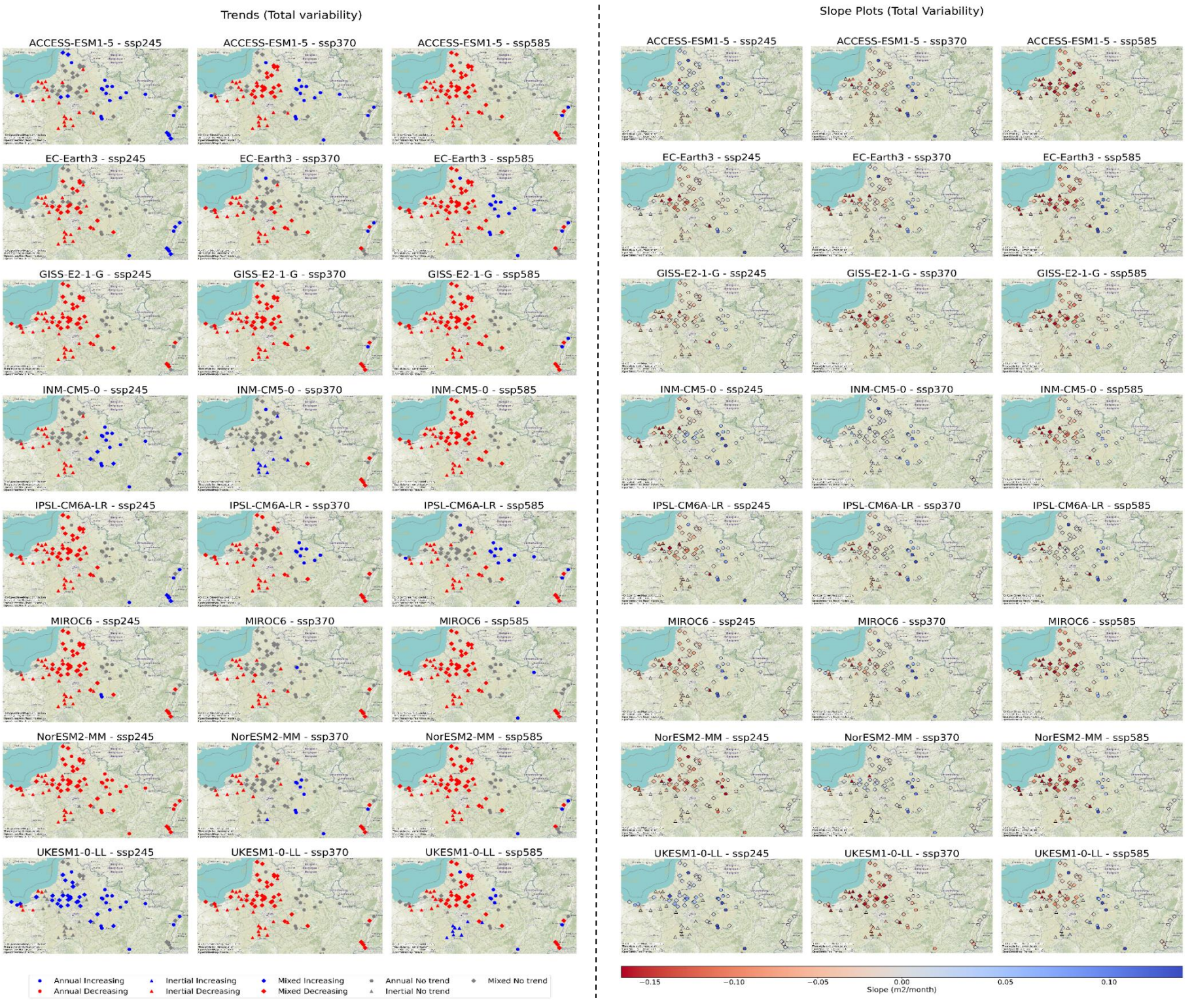

Figure 5.14: Trend and slope analysis on total variability from 16 CMIP6 models under SSP2-4.5 (left), SSP3-7.0 (middle), and SSP5-8.5 (right) scenarios for 2100: Blue indicates increasing trends(a) or positive slopes(b), and Red indicates decreasing trends(a) and negative slopes. Grey indicates no significant trend.

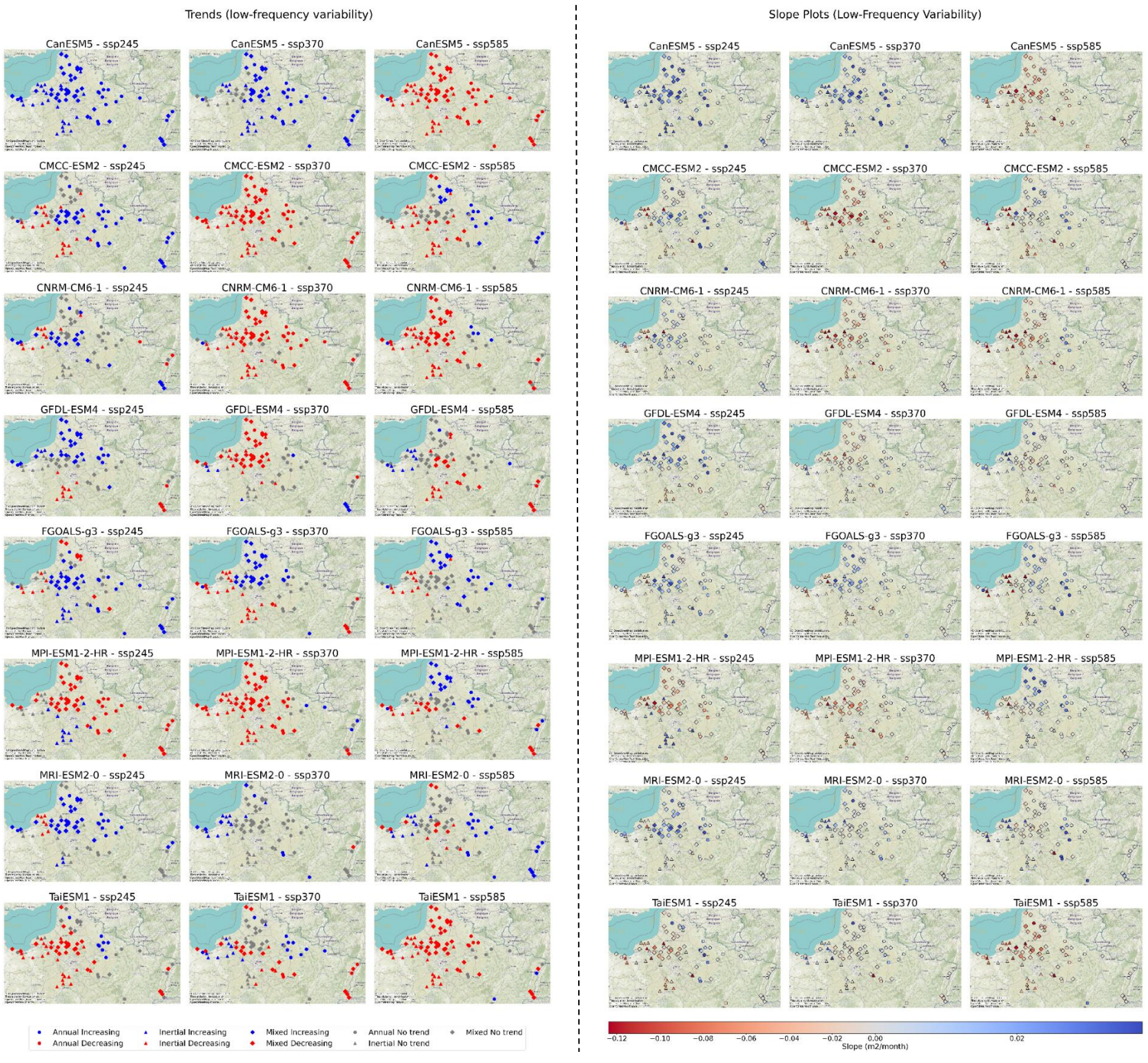

Figure continued…

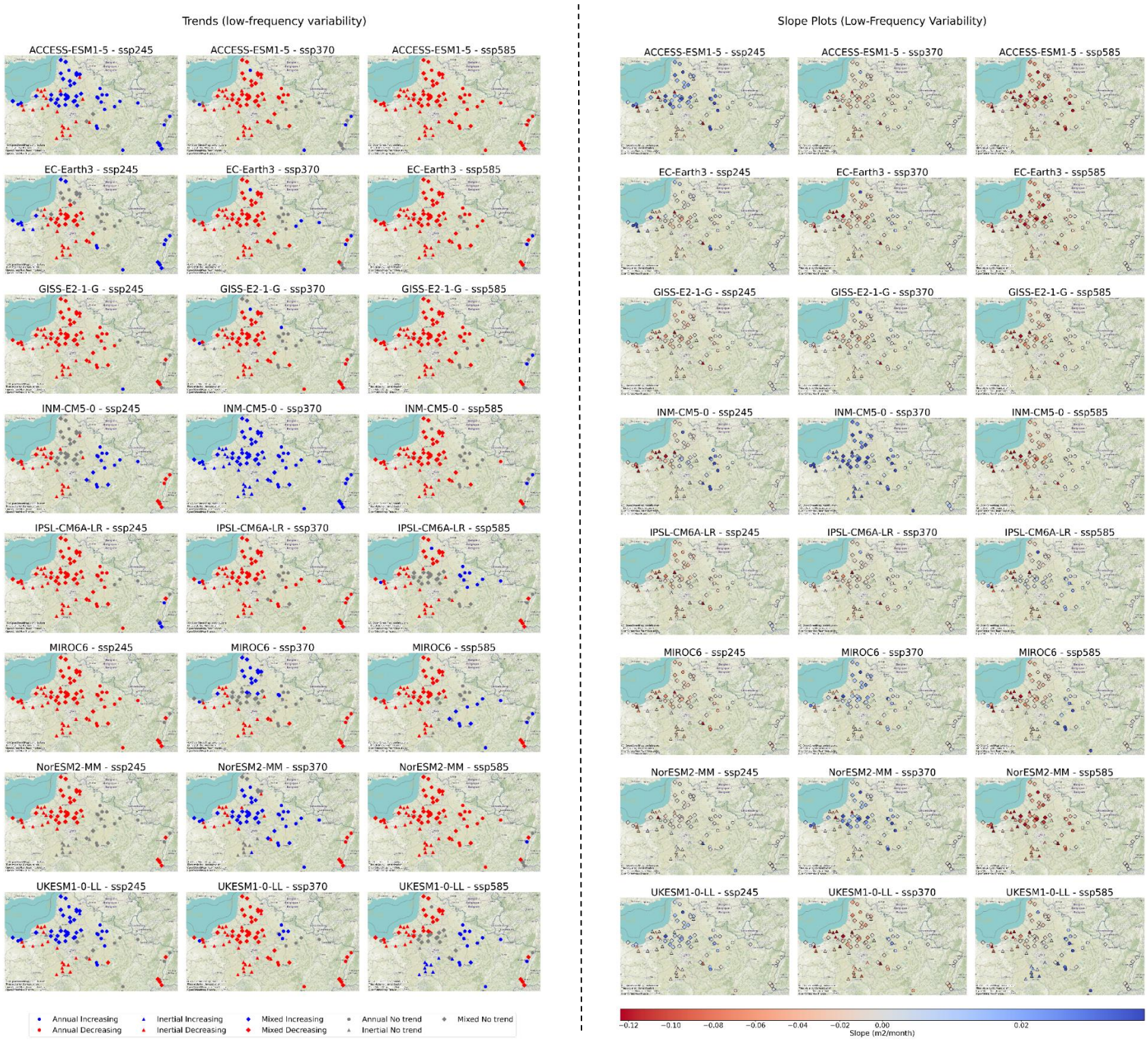

Figure 5.15:Trend and slope analysis on low-frequency variability from 16 CMIP6 models under SSP2-4.5 (left), SSP3-7.0 (middle), and SSP5-8.5 (right) scenarios for 2100: Blue indicates increasing trends(a) or positive slopes(b), and Red indicates decreasing trends(a) and negative slopes. Grey indicates no significant trend

As done with the ensemble approach, we then investigated how the different climate models led to potential changes in both total GWL variability and low–frequency variability. Figures 5.14a and 5.14b illustrate the results obtained for the total variability of the projected (median) time series. Unlike the results about changes in GWL, it seemed less easy to distinguish any general tendency in terms of changes in time series total variability according to the emission scenario for any model. As a first step, for the sake of simplicity, examination of trend direction only (Fig. 5.14) showed the same consistent pattern as previously noticed with the ensemble approach, albeit less clearly, where an obvious distinction between the eastern part (mainly annual-type aquifers) and the western part of the area was quite readily visible (Fig. 5.10). The magnitude of the slopes confirms this finding: for most models, one could observe either rather strong variability trends separating the eastern (increasing variability, blue-labelled stations) and western regions (decreasing variability, red-labelled stations), or only low-magnitude trends (most of the time not statistically significant). In brief, although differences between climate models exist, the inertial and mixed types (i.e., those with strong low-frequency variability) were mostly affected by decreasing variability through time over the period 2030 to 2100, whereas the annual-type aquifers would be characterised by increasing variability over the same period.

It then seemed that the amplitude low-frequency variability would decrease over time during the 2030 to 2100 period. The results of the ensemble approach showed that apart from SSP 2-4.5, the more pessimistic the emission scenario, the higher the number of stations (and the larger the region) affected (Fig. 5.12): there is a clear evolution from SSP 2-4.5 to SSP 5-8.5 for low-frequency variability. However, considered individually (Fig. 5.15), the 16 different models could show noticeable discrepancies in some cases: for instance, INM-CM5, NorESM2 or CanESM5 shows increasing low-frequency variability in terms of trend direction, contrary to many other models for SSP 3-7.0. Despite such disagreements, it seems rather clear that many models still display significantly increasing low-frequency variability for all station types (inertial, mixed or annual) with Sen's slopes of rather high magnitude for SSP 2-4.5, and that conversely, many models would display decreasing variability with strong Sen's slope magnitudes. Decreasing variability is also even more pronounced for SSP 5-8.5 than SSP 3-7.0. These results are consistent with the conclusions of the ensemble approach (Fig. 5.12). However, they also confirm that using the median time series in the ensemble approach seems well suited to identify clear general trends properly.

5.6. Discussion and Conclusion

In this study, we aimed to develop projections of GWL under three different climate change scenarios, focusing on three different GWL types: annual, mixed, and inertial. While some aquifers can be rather reactive to seasonal changes, dominated mainly by annual cyclicity, others are sometimes largely dominated by interannual to decadal variations that are driven by large-scale climate (Hanson et al.,2006; El Janyani et al., 2012; Rust et al.,2018; Baulon et al., 2022a). However, most studies dealing with groundwater level simulation or forecasting mainly considered time series represented by seasonal variations with a strong annual cycle governing water level variability. It was then desirable to study more complex aquifer dynamics mainly controlled by internal climate oscillations and assess how these different types of aquifers might be affected by climate change. In this framework, potential changes in variabilities under different SSP scenarios for such aquifers were explored by using deep learning GRU with wavelet pre-processing and CMIP6 bias-corrected precipitation and temperature data as input from 16 climate models. We analysed trends in groundwater levels and changes in variability (i.e. the amplitude of GWL variations) over time across the 2030-2100 period. Ten different initialisations of DL models and 16 climate model inputs resulted in an ensemble of 160 projections for each of the 3 SSP scenarios (SSP 2-4.5, 3-7.0, 5-8.5).

The analysis of the ensemble of 160 projections revealed that the lowest-variability (LV) GWL projections closely aligned with the range of variation of historical observations for annual and mixed types. In other words, only projections with the lowest possible variability would allow for maintaining the same range of variability encountered during the last ~60 decades. This was not the case for inertial-type GWL, for which the LV projection still has a significantly lower variability than previously observed in the historical period. The highest-variability (HV) projected GWL time series exhibited a much greater variability than all observed time series, with standard deviations approximately 1.5 to 2 times higher than for observed GWL. The annual-type GWL showed increasing total variability, while mixed and inertial types indicated decreasing across all scenarios. Additionally, distinct patterns for total variability showed up separating the eastern and western parts of the area (resp. increasing and decreasing variability): it seems like this is because the eastern part comprises almost all annual-type GWL, which are the ones exhibiting increasing variability. On the other hand, low-frequency variability seemed to be decreasing for almost all stations except for the less severe SSP 2-4.5 scenario; as well, more stations tended to exhibit decreasing low-frequency variability as emission scenarios became more severe (with all stations in this case for the most severe scenario).

Our CSMK test results on the median time series of the 160 projection ensemble indicated decreasing trends in groundwater levels for all scenarios and GWL types in northern France. This finding aligns with Habets et al. (2013), who projected a significant decrease in water resources for rivers and aquifers in two northern French basins despite model differences and uncertainties.

Conversely, Vergnes et al. (2023) projected higher groundwater levels on average over France in the future, including northern France, with mean annual GWL increasing by up to approximately 2 m. Vergnes et al. (2023) noted that their results did not match those of Wunsch et al. (2022) and were in contradiction with previous studies from Caballero et al. (2007) or Dayon et al. (2018). However, Wunsch et al. (2022) computed trends and relative changes essentially for the future period (2014-2100) from the projected levels alone, whereas Vergnes et al. (2023) compared the future period (2070-2099) to the historical period (1976-2005). In Germany, Wunsch et al. (2022) projected a median relative decrease in groundwater levels between 2014 and 2100, which seems to agree with our findings on trends in projected levels. It is also interesting to note that these authors showed increased variability in the annual cycle towards 2100, while our results highlighted an increase in the total variability of annual-type GWL in the eastern part of northern France (i.e., closest to Germany).

To facilitate comparison with these studies, we computed the mean difference (expressed as the relative change in %) between future periods and two historical reference periods: 1. Near future (2030-2050), 2. Middle future (2051-2070), 3. Far future (2071-2100), 4. Whole future period (2030-2100). We used 1970-2022 as our primary reference historical period (Figure. 5.16) and 1976-2005 as a secondary reference period (Figure. 5.17) to compare with Vergnes et al. (2023) directly. Figures 5.16 and 5.17 are used to support this analysis, from which several main points could be underlined. Typically, changes are generally small for all periods and scenarios, ranging from +8% to -6% (1970-2022 reference) and +8% to -10% (1976-2005 reference). The magnitude of changes is similar to Wunsch et al. (2022), although slightly less pronounced. There is a distinct pattern between the western and eastern parts, with future levels higher in the west and lower in the east compared to the historical reference as noticed for projected change in total variability (decreasing variability to the West, increasing to the East). No significant differences were observed between scenarios. Yet, a noticeable difference exists between the periods. Relative change is higher in mid and far-future periods than in the near future. Moreover, in the near future period (2030-2050), more stations indicated positive changes

in mean, followed by stabilisation in later periods. These increasing mean changes in some areas align with results from Vergnes et al. (2023).

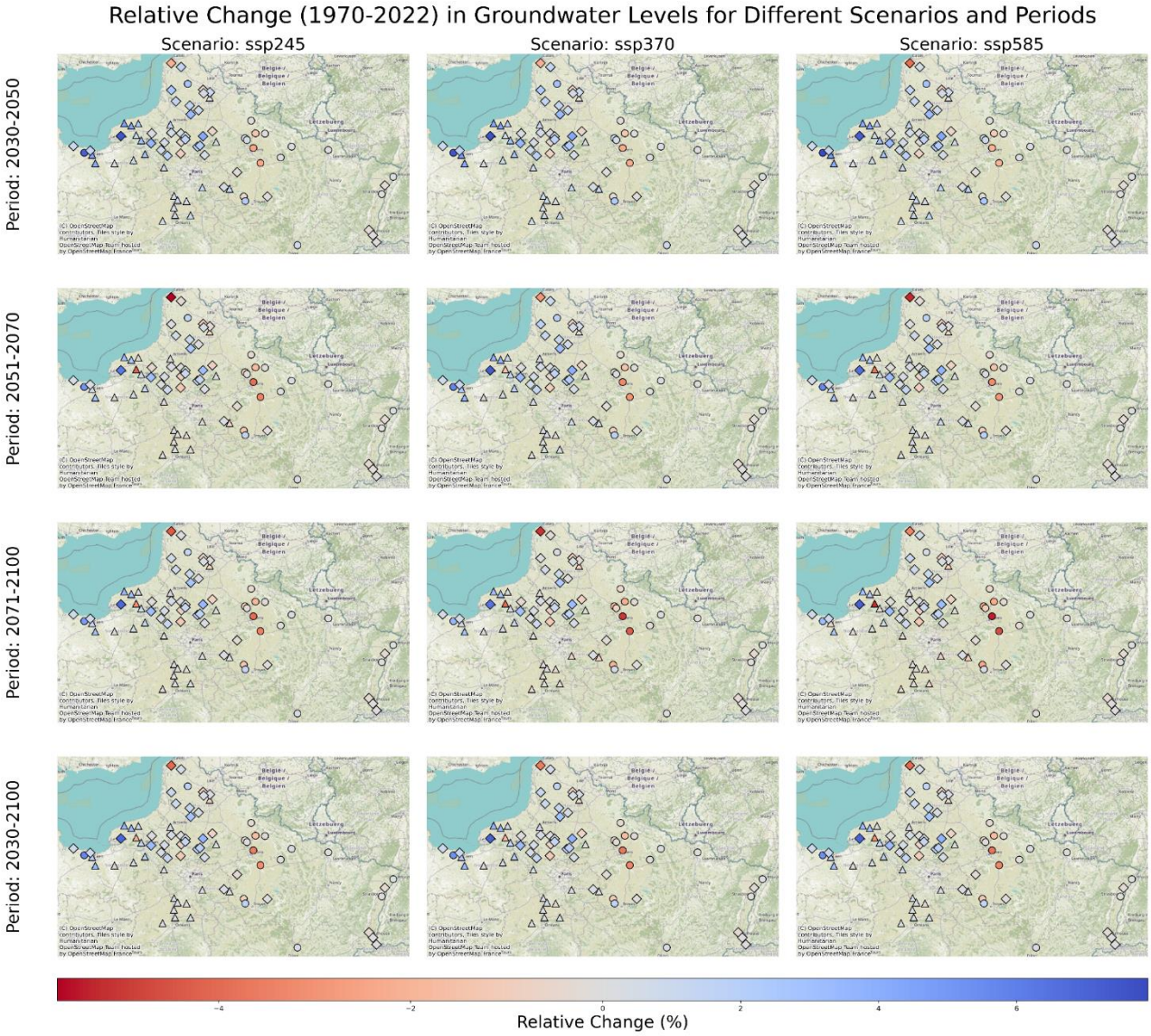

Figure 5.16: Relative change for future periods with respect to historical mean from 1970-2022

Our results reconcile those from Vergnes et al. (2023) and Wunsch et al. (2022), although according to Vergnes et al. (2023), they were apparently contradictory. Indeed, while projected GWL is expected to decrease over time (our study and Wunsch et al., 2022), changes in the annual mean projected levels remained slightly higher than historical levels (our study and Vergnes et al., 2023). This intriguing result warrants further investigation

and could be an interesting research question for future studies. It is worth noting that the relative positive change is mainly observed for inertial and mixed GWL types and much lower for annual types. Additionally, the differences between periods seemed consistent with the expected decreasing trends: for the far future, the difference in mean compared with the historical reference period seemed a little less than that of the near future, although the changes are so slight they are barely visible on the maps in figures 5.13 and 5.14.

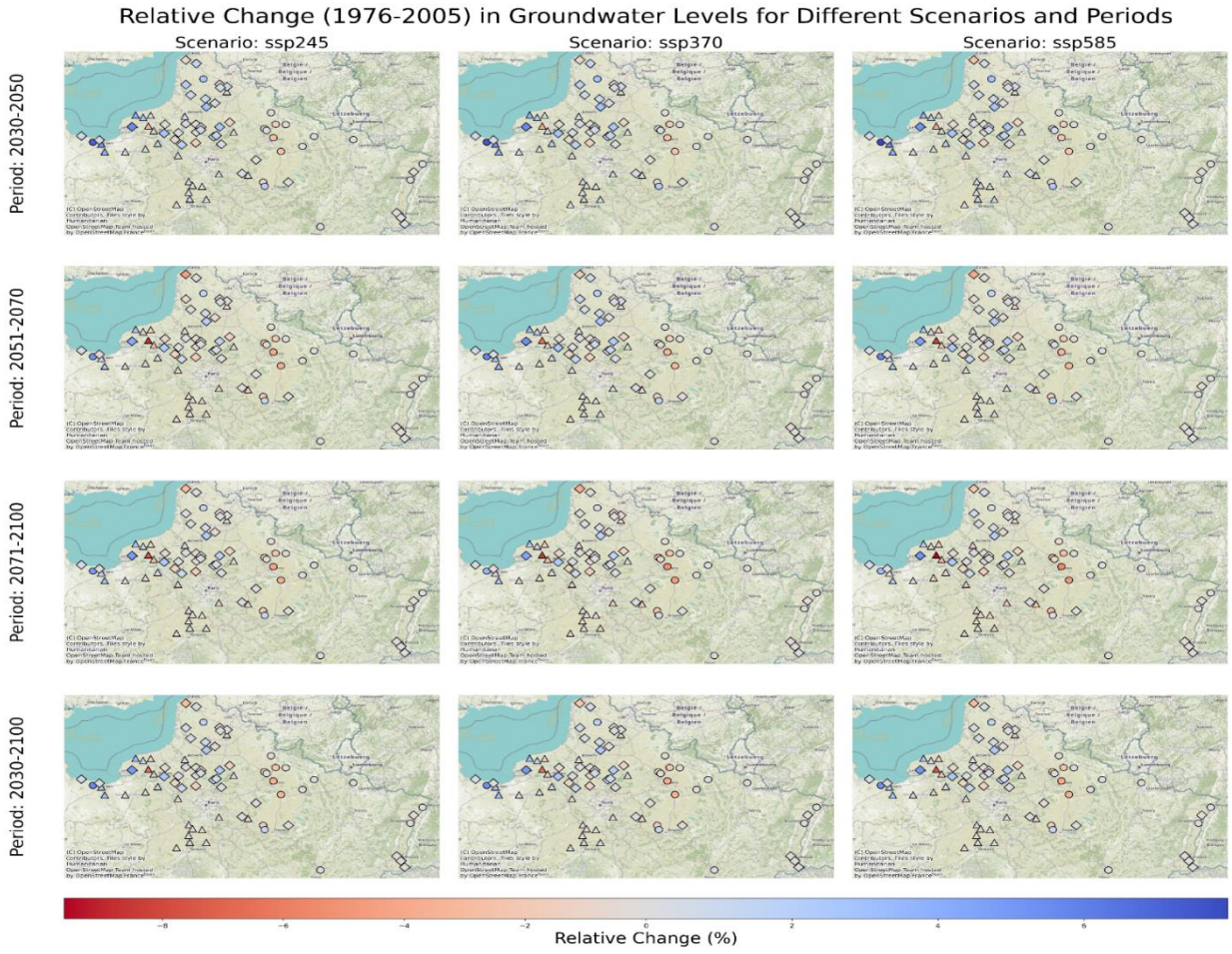

Figure 5.17: Relative change for future periods with respect to historical mean from 1976-2005 (similar to Vergnes et al. 2023)

Finally, it is also important to note that all these studies used different types of projected climate inputs (CMIP5 RCP or CMIP6 SSP scenarios, CMIP5 Euro-CORDEX regionalised climate projections for France, different ensembles or number of members), which makes direct comparison challenging. This highlights the need for a comprehensive, community-wide benchmarking experiment to understand better and reconcile these differences in future groundwater level projections.

While our study employed advanced techniques like utilising multiple CMIP6 climate models and scenarios, as well as training different initialisations for the GRU deep learning models, there are inherent strengths and limitations associated with these approaches that contribute to the total uncertainty affecting the results. The uncertainty associated with hydrological impact projections arises from multiple origins: it is partly linked to the climate model used (physics, initialisation), to the internal or "natural" variability, to the different emission scenarios, and ultimately to the hydrological model used. In particular, as mentioned in the introduction, low-frequency/long-time scale, natural variability can play a crucial role in modulating the effects of climate change by amplifying or attenuating (masking) hydroclimatic trends and associated extremes, e.g. as shown by Boé and Habets (2014) and emphasised by Massei et al. (2020). On interannual to decadal scales, climate oscillations and teleconnections such as the North Atlantic Oscillation (NAO) or the El Nino-Southern Oscillation (ENSO) were identified as a significant forcing factor of groundwater resources variations (Holman et al., 2011; Liesch & Wunsch, 2019; Massei et al., 2007, 2010; Rust et al., 2019). In many instances, such fluctuations may correspond to significant hydrological events, as shown in Baulon et al. (2022b) and describe multi-year periods of successive dry or wet years, to which human activities are

particularly vulnerable. This led Blösch et al. (2019) to classify the understanding of these phenomena as one of the "23 unsolved hydrological problems". The 16 model variants used in the work presented herein still represented the stochastic low-frequency climate variability differently, which allowed us to appreciate its potential impacts on GWL projections. However, using only one variant for precipitation and temperature of each of the 16 climate models prevented us from analysing the contribution of natural variability to the total uncertainty in the GWL projections released. A larger-ensemble approach would be needed to properly assess the contribution of natural climate variability to the uncertainty of GWL projections.

Finally, from a more technical standpoint, the DL models used herein have shown strong performance in identifying complex patterns in large datasets and capturing GWL variability, as also emphasised in our previous studies (Chidepudi et al., 2023a, 2024a), making them efficient for long-term simulations. They have proven to be a relevant alternative or complement to more complex modelling frameworks, like physically-based models, that are often more difficult to develop and implement. Their efficiency could be leveraged to facilitate benchmarking of hydrological and hydrogeological projections under climate change by conducting more similar studies globally

Conclusion and Pespectives

The thesis mainly focused on using DL for three aspects of groundwater modelling: simulation, reconstruction, and projection. We evaluated different DL models, including or not different pre-processing techniques, such as wavelet pre-processing or pre-clustering the training datasets, in both single-station and multi-station training modes. Three different types of GWLs were considered that would allow the DL models to encompass the main types of variability that can be generally encountered: "simple", almost periodic variability with a certain amount of infra-annual variations, low-frequency variability (rather smooth, long-term varying water tables), and mixed-type variability which corresponded to the superimposition of the two previous types. We then tested the relevance of the models to simulate the long-term evolution of groundwater levels in the past or the future. The northern part of metropolitan France was eventually selected as a case study because it fulfilled two major requirements: availability of the longest time series possible and the presence of the three contrasted above-mentioned types.

Our analysis of the current state-of-the-art, in which the most recent developments in using deep learning for groundwater were presented, highlighted the existence of clear potential opportunities and the need for advancing the knowledge about the application of DL for groundwater level simulation beyond the mainstream "forecasting" approach. While the application of DL for surface watersheds seemed to be quite advanced, a clear lack of investigation regarding groundwater systems appeared. We then first developed single-station DL models using recurrent-based architectures (LSTM, Bi-LSTM, GRU) and evaluated them for the three main GWL behaviours (annual, inertial and mixed) to test their capability to capture all types of frequency information. Results showed that wavelet pre-processing was a necessary step to make the models capable of representing the low-frequency part of GWL variability. Indeed, while all the necessary information for model training was readily available in precipitation and temperature for annual-type GWL, the models encountered difficulties in extracting low-frequency information from precipitation or temperature signals, which are commonly dominated by either random noise or annual cyclicity. The application of wavelet MODWT pre-processing enabled the retrieval of this information that was previously "hidden" within the input features, thereby significantly enhancing the learning process. MODWT-assisted models then significantly improved GWL simulations, especially for inertial and mixed types, with GRU models performing the best. Model outputs were interpreted using SHAP to understand the importance of different input features. These findings confirmed that wavelet-based pre-processing enhanced the ability

of DL models to capture low-frequency variability in GWL simulations. Subsequently, the single-station approach was extended all over northern France to achieve GWL reconstruction using ERA5 and ERA20C reanalysis datasets as input. The results highlighted the models' ability to achieve consistent reconstructions on the basis of the only one-century mixed-type GWL time series available. The model efficiently captured multi-decadal variability in all reconstructions: the influence of the Atlantic Multi-decadal Oscillation (AMO) on GWLs, already documented in a number of previous works, could be identified in all reconstructions, emphasising the potential of deep learning models to provide reliable long-term GWL simulations. The correlated seasonal Mann-Kendall trend test was applied to this database of reconstructed GWL time series and showed that no significant trend existed across the study area over the past century. However, this DL reconstruction experiment also emphasised the impact of a limited length of training GWL time series for achieving consistent simulations, especially in the case of low-frequency dominated GWL. The question then arose of the most appropriate way to maximise the amount of training information. We then investigated the possibility of retrieving information from all available adjacent time series usable for training but also from physical "static" features related to basins' properties. The previously developed DL models were then adapted to use regionalised information for large-scale simulation: several multi-station approaches were tested and compared with the specialised single-station approach. Results highlighted the effectiveness of clustering-based modelling in reducing data complexity and efficiently targeting relevant information. Additionally, it demonstrated the significant performance improvements achievable through the combination of wavelet pre-processing and prior clustering, particularly for GWLs dominated by low-frequency variations. The models developed allow quite a consistent capability for long-term simulation. We finally used a multi-station MODWT-GRU architecture trained for each GWL type to generate future groundwater projections under varying climate scenarios and discuss the possible changes in GWL in northern France up to 2100. Results from projected changes indicated decreasing trends in both groundwater levels (GWL mean) and GWL variability (GWL standard deviation), intensifying from SSP2-4.5 to SSP5-8.5. The variability of annual-type aquifers increased for all emission scenarios. Paradoxically, groundwater levels seemed higher in the future compared to the historical period for all scenarios, although the trends of GWL seemed to decrease throughout the future period consistently. Indeed, it seemed that on average, levels would tend to be higher in the near-future period, until approximately 2050, before decreasing to the end of the century. It would then explain overall higher future GWL compared to the historical period on average, although such future GWL would still tend to decrease until 2100.

Overall, several takeaways emerged from this thesis: 1) wavelet pre-processing consistently improved model performance across different GWL types and modelling approaches; 2) reconstructed GWLs were able to capture key multi-decadal variability related large-scale climate variations (AMO), with no significant noticeable trends affecting GWL during the historical period; 3) multi-station approaches, especially when combined with clustering based on spectral properties, showed promises in improving model accuracy; 4) climate change projections indicated decreasing trends in GWLs across northern France, with slightly increasing variability for annual types for all scenarios but decreasing variability for mixed and inertial types.

At the beginning of this dissertation, several research questions were raised, which we tried to answer in this work briefly, as summarised below:

1. *How can we develop deep learning models that would account for various types of groundwater level variations beyond solely annual cyclicities and not designed only for forecasting for a limited number of time steps (and not based on previous values of target variables)?*

For the models to be able to provide long-term simulations (Reconstruction and projections), it is necessary to use only external climate variables that are widely available from reanalysis (Chapter 3) or CMIP6 (Chapter 5) datasets. To better capture variations in low-frequency dominated GWL, BC-MODWT pre-processing is preferred for both DL modelling approaches (Single station (Chapter2) and Multi-station (Chapter4)). Multi-station approaches are preferred for very inertial GWL types. However, although both clustering and wavelet pre-processing improved the simulations in the multi-station framework, we still recommend the priority be placed on pre-processing rather than clustering wherever possible while still allowing for leveraging the same type of information (i.e. detecting the low-frequency). As it would prevent separating the dataset and reducing its size, which is not desirable for optimal training.

2. *How does spatial resolution of explanatory input climate variables taken from reanalysis datasets impact simulations or reconstruction? How accurate are these in capturing different variabilities?*

Though high-resolution SAFRAN reanalysis provides accurate simulations (Chapter 2), there is a need to look into global reanalysis datasets for better generalisation. Hence, ERA5 is an imminent alternative and gave comparable accuracy (Chapters 2 & 3), and for

very long-term reconstruction up to 1900, ERA20C is suitable, though ERA5 gave slightly better accuracy.

3. *3. Can such DL models consistently capture trends, long-term oscillations, and variability? What relevant strategies are needed to ensure this?*

BC-MODWT-DL models (wavelet-assisted DL models) allowed for extracting low-frequency information and improved simulations in all cases (Chapters 2-4). They could also capture multi-decadal variability systematically in all reconstructed GWLs (Chapter 3). Despite the low-frequency component in input features (precipitation and temperature) projections being almost non-existent, DL models trained for inertial or mixed behaviours were successful in generating projected GWLs that would comprise some amount of low-frequency variability: the models effectively contain the necessary information (related to physical basin properties) to represent such behaviours.

4. *What does DL simulated GWL time series tell about how different types of aquifers (e.g. annual dominated or low-frequency dominated) respond to climate change scenarios?*

While there were no significant trends on past reconstructed GWL (Chapter3), future projected changes indicated decreasing trends in groundwater levels and variability, intensifying from SSP2-4.5 to SSP5-8.5. The variability of annual-type aquifers increased for all emission scenarios (Chapter 5).

The results from this work are mainly focused on the monitored locations where the data is available for longer periods. After validating the results at wider scales, the approaches could be extended to unmonitored locations. Thereby investigating the potential of transfer learning and frugal AI approaches.

Future research could explore the simulation of groundwater levels, considering the magnitude of the interactions with surface hydrology (i.e., from high streamflow-groundwater interactions to almost nil interactions for confined aquifers). In all cases, the geological characteristics of the basins, as well as human influences such as land use and volumes of water abstracted, would need to be investigated. This means that all piezometric data, regardless of human impact or connectivity to the surface, and surface hydrological data in cases of significant surface-ground interactions could be utilised. In such complex situations where significant knowledge about different influences is required, data scarcity may be a challenge (e.g., knowledge about water abstraction, pumping data etc.). In such a case, physics-informed neural network approaches may help compensating

the lack of data by increasing the constraint on the neural network optimisation based on physical knowledge. Further, the spatialisation of groundwater information could be another open challenge. Currently developed models simulate groundwater point by point (i.e., only one piezometric station is simulated, whatever the learning conditions), but a major issue and challenge lies in producing spatialised simulation results in the form of a groundwater map.

BC-MODWT consistently improved the performance of the DL models for aquifers affected by low-frequency variability (i.e., mixed and inertial) in all approaches. Its impact on annual-dominated groundwater is relatively minimal. While all three DL models (LSTM, GRU and BiLSTM) were suitable for groundwater simulations with slight differences in their performance on unseen data, GRU is computationally the most efficient, taking less time to train models and to tune hyperparameter values. The usual approach of training models on a single station is widely adapted in building specialised models. For better generalisation and regionalisation efforts, a multi-station approach could become necessary, and the performance of these approaches could potentially depend on bigger datasets with relevant hydrogeological information. This would also require exploring more advanced encoding strategies for static attributes.

More research focusing on benchmarking these models is needed. Further integration of physical constraints and domain knowledge into DL models (e.g. physics-informed neural networks) or complementary with existing operational physically based groundwater models would be required with coordinated efforts from the global groundwater community. Similarly, for assessing trends and variability, standardising the approaches for different climate projections and future periods considered for computations will be needed for better generalisation and future comparisons with different modelling approaches. Regarding projections, uncertainty related to internal climate variability was not considered in this study, and this could be tested in future studies by exploring improved uncertainty quantification methods.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.

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

Afan, H. A., Ibrahem Ahmed Osman, A., Essam, Y., Ahmed, A. N., Huang, Y. F., Kisi, O., Sherif, M., Sefelnasr, A., Chau, K. wing, & El-Shafie, A. (2021). Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques. Engineering Applications of Computational Fluid Mechanics, 15(1), 1420–1439. https://doi.org/10.1080/19942060.2021.1974093

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

Althoff, D., Bazame, H. C., & Nascimento, J. G. (2021). Untangling hybrid hydrological models with explainable artificial intelligence. H2Open Journal, 4(1), 13–28. https://doi.org/10.2166/h2oj.2021.066

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2623–2631. https://doi.org/10.1145/3292500.3330701

Alizadeh, B., Ghaderi Bafti, A., Kamangir, H., Zhang, Y., Wright, D. B., & Franz, K. J. (2021). A Novel Attention-Based LSTM Cell Post-Processor Coupled with Bayesian Optimization for Streamow Prediction. Journal of Hydrology, 601(February), 126526. https://doi.org/10.1016/j.jhydrol.2021.126526

Anderson, S., & Radić, V. (2022). Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling. Hydrology and Earth System Sciences, 26(3), 795–825. https://doi.org/10.5194/hess-26-795-2022

Arsenault, R., Martel, J.-L., Brunet, F., Brissette, F., and Mai, J. (2023). Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models, Hydrol. Earth Syst. Sci., 27, 139–157, https://doi.org/10.5194/hess-27-139-2023

Ascott, M. J., Macdonald, D. M. J., Black, E., Verhoef, A., Nakohoun, 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 Resources Research, 56(12), e2020WR028056. https://doi.org/10.1029/2020WR028056

Atawneh, D. A., Cartwright, N., & Bertone, E. (2021). Climate change and its impact on the projected values of groundwater recharge: A review. Journal of Hydrology, 601, 126602. https://doi.org/10.1016/j.jhydrol.2021.126602

Bahmani, R., & Ouarda, T. B. M. J. (2021). Groundwater level modeling with hybrid artificial intelligence techniques. Journal of Hydrology, 595, 125659. https://doi.org/10.1016/j.jhydrol.2020.125659

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

Barclay, J. R., Topp, S. N., Koenig, L. E., Sleckman, M. J., & Appling, A. P. (2023). Train, inform, borrow, or combine? Approaches to process-guided deep learning for groundwater-influenced stream temperature prediction. Water Resources Research, 59, e2023WR035327. https://doi.org/10.1029/2023WR035327

Barthel, R., Banzhaf, S. Groundwater and Surface Water Interaction at the Regional-scale – A Review with Focus on Regional Integrated Models. Water Resour Manage 30, 1–32 (2016). https://doi.org/10.1007/s11269-015-1163-z

Barthel, R., Haaf, E., Giese, M., Nygren, M., Heudorfer, B., & Stahl, K. (2021). Similarity-based approaches in hydrogeology: proposal of a new concept for data-scarce groundwater resource characterization and prediction. Hydrogeology Journal, 29(5), 1693-1709. https://doi.org/10.1007/s10040-021-02358-4

Barzegar, R., Aalami, M. T., & Adamowski, J. (2021). Coupling a hybrid CNN-LSTM deep learning model with a Boundary Corrected Maximal Overlap Discrete Wavelet Transform for multiscale Lake water level forecasting.

Journal of Hydrology, 598, 126196. https://doi.org/10.1016/j.jhydrol.2021.126196

Bauer, P., Dueben, P., Chantry, M., Doblas-Reyes, F., Hoefler, T., McGovern, A., & Stevens, B. (2023). Deep learning and a changing economy in weather and climate prediction. Nature Reviews Earth and Environment, 4(8), 507–509. https://doi.org/10.1038/s43017-023-00468-z

Bauer, P., Hoefler, T., Stevens, B., & Hazeleger, W. (2024). Digital twins of Earth and the computing challenge of human interaction. Nature Computational Science 2024 4:3, 4(3), 154–157. https://doi.org/10.1038/s43588-024-00599-3

Baulon, L., Allier, D., Massei, N., Bessiere, H., Fournier, M., & Bault, V. (2022a). 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. (2022b). 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

Baulon, L. Déterminisme climatique et hydrogéologique de l'évolution à long terme des niveaux piézométriques. Sciences de la Terre. Normandie Université, 2023. Français. ⟨NNT : 2023NORMR015⟩. ⟨tel-04123371⟩

Berghuijs, W. R., Luijendijk, E., Moeck, C., van der Velde, Y., & Allen, S. T. (2022). Global recharge data set indicates strengthened groundwater connection to surface fluxes. Geophysical Research Letters, 49, e2022GL099010. https://doi.org/10.1029/2022GL099010

Beven, K., & Young, P. (2013). A guide to good practice in modeling semantics for authors and referees. Water Resources Research, 49(8), 5092–5098. https://doi.org/10.1002/wrcr.20393

Bhasme, P., Vagadiya, J., & Bhatia, U. (2022). Enhancing predictive skills in physically-consistent way: Physics Informed Machine Learning for hydrological processes. Journal of Hydrology, 615, 128618. https://doi.org/10.1016/j.jhydrol.2022.128618

Boé, J., & Habets, F. (2014). Multi-decadal river flow variations in France. Hydrol. Earth Syst. Sci, 18, 691–708. https://doi.org/10.5194/hess-18-691-2014

Bonnet, R., Boé, J., Dayon, G., & Martin, E. (2017). Twentieth-Century Hydrometeorological Reconstructions to Study the Multidecadal Variations of the Water Cycle Over France. Water Resources Research, 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. Hydrology and Earth System Sciences, 24(4), 1611–1631. https://doi.org/10.5194/hess-24-1611-2020

Bonnet, R., Boucher, O., Vrac, M., & Jin, X. (2022). Sensitivity of bias adjustment methods to low-frequency internal climate variability over the reference period: an ideal model study. Environmental Research: Climate, 1(1), 011001. https://doi.org/10.1088/2752-5295/ac6adc

Boo, K. B. W., El-Shafie, A., Othman, F., Khan, M. M. H., Birima, A. H., & Ahmed, A. N. (2024). Groundwater level forecasting with machine learning models: A review. Water Research, 252, 121249. https://doi.org/10.1016/j.watres.2024.121249

Bowes, B., Goodall, J., Sadler, J., Morsy, M., & Behl, M. (2019). Toward Forecasting Groundwater Table in Flood Prone Coastal Cities Using Long Short-term Memory and Recurrent Neural Networks. Earth and Space Science Open Archive.

Brocca, L., Barbetta, S., Camici, S., Ciabatta, L., Dari, J., Filippucci, P., Massari, C., Modanesi, S., Tarpanelli, A., Bonaccorsi, B., Mosaffa, H., Wagner, W., Vreugdenhil, M., Quast, R., Alfieri, L., Gabellani, S., Avanzi, F., Rains, D., Miralles, D. G., … Fernandez, D. (2024). A Digital Twin of the terrestrial water cycle: a glimpse into the future through high-resolution Earth observations. Frontiers in Science, 1, 1190191. https://doi.org/10.3389/fsci.2023.1190191

Caballero, Y., Voirin-Morel, S., Habets, F., Noilhan, J., LeMoigne, P., Lehenaff, A., & Boone, A. (2007). Hydrological sensitivity of the Adour-Garonne river basin to climate change. Water Resources Research, 43(7). https://doi.org/10.1029/2005WR004192

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

Chakraborty, D., Başağaoğlu, H., Gutierrez, L., & Mirchi, A. (2021). Explainable AI reveals new hydroclimatic insights for ecosystem-centric

groundwater management. Environmental Research Letters, 16(11), 114024. https://doi.org/10.1088/1748-9326/ac2fde

Chen, M., Qian, Z., Boers, N., Jakeman, A. J., Kettner, A. J., Brandt, M., Kwan, M., Batty, M., Li, W., Zhu, R., Luo, W., Ames, D. P., Barton, C. M., Cuddy, S. M., Koirala, S., Zhang, F., Ratti, C., Liu, J., Zhong, T., . . . Lü, G. (2023). Iterative integration of deep learning in hybrid Earth surface system modelling. Nature Reviews Earth & Environment, 4(8), 568-581. https://doi.org/10.1038/s43017-023-00452-7

Cheng, S., Cheng, L., Qin, S., Zhang, L., Liu, P., Liu, L., et al. (2022). Improved understanding of how catchment properties control hydrological partitioning through machine learning. Water Resources Research, 58, e2021WR031412. https://doi.org/10.1029/2021WR031412

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

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

Chidepudi, S. K. R., Massei, N., Jardani, A., & Henriot, A. (2024a). 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., Dieppois, B., Henriot, A., and Fournier, M. (2024b). 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?, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-794

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

Chollet, F. (2015). Keras. GitHub. https://github.com/fchollet/keras

Chuphal, D. S., & Mishra, V. (2023). Hydrological model-based streamflow reconstruction for Indian sub-continental river basins, 1951–2021. Scientific Data, 10(1), 1-11. https://doi.org/10.1038/s41597-023-02618-w

Clark, M.P., Wilby, R.L., Gutmann, E.D. et al. Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. Curr Clim Change Rep 2, 55–64 (2016). https://doi.org/10.1007/s40641-016-0034-x

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

Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., Fransen, H. J. H., Verhoef, A., Van Loon, A. F., Sulis, M., & Abesser, C. (2021). Global Groundwater Modeling and Monitoring: Opportunities and Challenges. Water Resources Research, 57(12). https://doi.org/10.1029/2020WR029500

Costantini, M., Colin, J., & Decharme, B. (2023). Projected Climate-Driven Changes of Water Table Depth in the World's Major Groundwater Basins. Earth's Future, 11(3), e2022EF003068. https://doi.org/10.1029/2022EF003068

Cornish, C. R., Bretherton, C. S., & Percival, D. B. (2006). Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. Boundary-Layer Meteorology, 119(2), 339–374. https://doi.org/10.1007/s10546-005-9011-y

Coulibaly, P., Anctil, F., Aravena, R., & Bobée, B. (2001). Artificial neural network modeling of water table depth fluctuations. Water Resources Research, 37(4), 885–896. https://doi.org/10.1029/2000WR900368

Dayon, G., Boé, J., Martin, É., & Gailhard, J. (2018). Impacts of climate change on the hydrological cycle over France and associated uncertainties. Comptes Rendus Geoscience, 350(4), 141-153. https://doi.org/10.1016/j.crte.2018.03.001

de Graaf, I. E., Marinelli, B., & Liu, S. (2024). Global analysis of groundwater pumping from increased river capture. Environmental Research Letters, 19(4), 044064. https://doi.org/10.1088/1748-9326/ad383d

Deser, C., Phillips, A., Bourdette, V. et al. Uncertainty in climate change projections: the role of internal variability. Clim Dyn 38, 527–546 (2012). https://doi.org/10.1007/s00382-010-0977-x

Deser, C. and Phillips, A. S. (2023). A range of outcomes: the combined effects of internal variability and anthropogenic forcing on regional climate trends over Europe, Nonlin. Processes Geophys., 30, 63–84, https://doi.org/10.5194/npg-30-63-2023

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 of the Past, 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. International Journal of Climatology, 36(15), 4679–4696. https://doi.org/10.1002/joc.4660

Dorigo, W., Dietrich, S., Aires, F., Brocca, L., Carter, S., Cretaux, J., Dunkerley, D., Enomoto, H., Forsberg, R., Güntner, A., Hegglin, M. I., Hollmann, R., Hurst, D. F., Johannessen, J. A., Kummerow, C., Lee, T., Luojus, K., Looser, U., Miralles, D. G., Pellet, V., Recknagel, T., Vargas, C. R., Schneider, U., Schoeneich, P., Schröder, M., Tapper, N., Vuglinsky, V., Wagner, W., Yu, L., Zappa, L., Zemp, M., & Aich, V. (2021). Closing the Water Cycle from Observations across Scales: Where Do We Stand?. Bulletin of the American Meteorological Society, 102(10), E1897-E1935. https://doi.org/10.1175/BAMS-D-19-0316.1

Edijatno, & Michel, C. (1989). Un modèle pluie-débit journalier à trois paramètres. La Houille Blanche, 75(2), 113–122. https://doi.org/10.1051/lhb/1989007

El Janyani, S., Massei, N., Dupont, J., Fournier, M., & Dörfliger, N. (2012). Hydrological responses of the chalk aquifer to the regional climatic signal. Journal of Hydrology, 464-465, 485-493. https://doi.org/10.1016/j.jhydrol.2012.07.040

El Janyani, S., Dupont, JP., Massei, N. et al. Hydrological role of karst in the Chalk aquifer of Upper Normandy, France. Hydrogeol J 22, 663–677 (2014). https://doi.org/10.1007/s10040-013-1083-z

Espinoza, E. A., Loritz, R., Chaves, M. Á., Bäuerle, N., & Ehret, U. (2024). To bucket or not to bucket? Analyzing the performance and interpretability

of hybrid hydrological models with dynamic parameterization. Hydrol. Earth Syst. Sci, 28, 2705–2719. https://doi.org/10.5194/hess-28-2705-2024

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016

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

Feng, D., Beck, H., Lawson, K., & Shen, C. (2023). The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. Hydrology and Earth System Sciences, 27(12), 2357–2373. https://doi.org/10.5194/hess-27-2357-2023

Fossa, M., Dieppois, B., Massei, N., Fournier, M., Laignel, B., & Vidal, J. P. (2021). Spatiotemporal and cross-scale interactions in hydroclimate variability: a case-study in France. Hydrology and Earth System Sciences, 25(11), 5683–5702. https://doi.org/10.5194/hess-25-5683-2021

Ghazi, B., Jeihouni, E., Kouzehgar, K., & Haghighi, A. T. (2021). Assessment of probable groundwater changes under representative concentration pathway (RCP) scenarios through the wavelet–GEP model. Environmental Earth Sciences, 80(12), 446. https://doi.org/10.1007/s12665-021-09746-9

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2019). GRUN: An observation-based global gridded runoff dataset from 1902 to 2014. Earth System Science Data, 11(4), 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Ghobadi, F., Yaseen, Z. M., & Kang, D. (2024). Long-term streamflow forecasting in data-scarce regions: Insightful investigation for leveraging satellite-derived data, Informer architecture, and concurrent fine-tuning transfer learning. Journal of Hydrology, 631, 130772. https://doi.org/10.1016/j.jhydrol.2024.130772

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

Giese, M., Haaf, E., Heudorfer, B., & Barthel, R. (2020). Comparative hydrogeology – reference analysis of groundwater dynamics from neighbouring observation wells. Hydrological Sciences Journal, 65(10), 1685–1706. https://doi.org/10.1080/02626667.2020.1762888

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural networks, 18(5-6), 602-610.

Gualtieri, G. (2022). Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review. Renewable and Sustainable Energy Reviews, 167, 112741. https://doi.org/10.1016/j.rser.2022.112741

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. Journal of Hydrology, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haaf, E., Giese, M., Heudorfer, B., Stahl, K., & Barthel, R. (2020). Physiographic and climatic controls on regional groundwater dynamics. Water Resources Research, 56, e2019WR026545. https://doi.org/10.1029/2019WR026545

Haaf, E., Giese, M., Reimann, T., & Barthel, R. (2023). Data-driven estimation of groundwater level time-series at unmonitored sites using comparative regional analysis. Water Resources Research, 59, e2022WR033470. https://doi.org/10.1029/2022WR033470

Habets, F., Boé, J., Déqué, M. et al. Impact of climate change on the hydrogeology of two basins in northern France. Climatic Change 121, 771–785 (2013). https://doi.org/10.1007/s10584-013-0934-x

Habets, F., Ackerer, P., Amraoui, N., Augeard, B., Besson, F., Caballero, Y., ... & Viennot, P. (2015). Aqui-FR, un système multi-modèle hydrogéologique à l'échelle nationale. Géologues, 187, 105-109.

Hagen, J. S., Hasibi, R., Leblois, 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. Hydrological Sciences Journal, 68(3), 412–431. https://doi.org/10.1080/02626667.2023.2165927

Halloran, L. J., Millwater, J., Hunkeler, D., & Arnoux, M. (2023). Climate change impacts on groundwater discharge-dependent streamflow in an alpine headwater catchment. Science of The Total Environment, 902, 166009. https://doi.org/10.1016/j.scitotenv.2023.166009

Hanson, R.T., Dettinger, M.D. & Newhouse, M.W. Relations between climatic variability and hydrologic time series from four alluvial basins across the southwestern United States. Hydrogeol J 14, 1122–1146 (2006). https://doi.org/10.1007/s10040-006-0067-7

Harris, C. R., Millman, K. J., J., S., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P. Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array (2020). Array programming with NumPy. Nature, 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

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? Hydrology and Earth System Sciences, 26(22), 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

Hawkins, E., & Sutton, R. (2009). The Potential to Narrow Uncertainty in Regional Climate Predictions. Bulletin of the American Meteorological Society, 90(8), 1095-1108. https://doi.org/10.1175/2009BAMS2607.1

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. Hydrology and Earth System Sciences, 28(3), 525–543. https://doi.org/10.5194/hess-28-525-2024

Heudorfer, B., Haaf, E., Stahl, K., & Barthel, R. (2019). Index-based characterization and quantification of groundwater dynamics. Water Resources Research, 55, 5575–5592. https://doi.org/10.1029/2018WR024418

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

Holman, I.P., Rivas-Casado, M., Bloomfield, J.P. et al. Identifying non-stationary groundwater level response to North Atlantic ocean-atmosphere teleconnection patterns using wavelet coherence. Hydrogeol J 19, 1269–1278 (2011). https://doi.org/10.1007/s10040-011-0755-9

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science and Engineering, 9(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hussain et al., (2019). pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. Journal of Open Source Software, 4(39), 1556, https://doi.org/10.21105/joss.01556

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. Hydrological Processes, 30(18), 3107–3125. https://doi.org/10.1002/hyp.10850

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

Jahangir, M. S., & Quilty, J. (2024). Generative deep learning for probabilistic streamflow forecasting: Conditional variational auto-encoder. Journal of Hydrology, 629, 130498. https://doi.org/10.1016/j.jhydrol.2023.130498

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R. G., Fallatah, O., & Kirchner, J. W. (2023). Rapid groundwater decline and some cases of recovery in aquifers globally. Nature, 625(7996), 715-721. https://doi.org/10.1038/s41586-023-06879-8

Jiang, S., Zheng, Y., & Solomatine, D. (2020). Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning. Geophysical Research Letters, 47(13). https://doi.org/10.1029/2020GL088229

Jing, H., He, X., Tian, Y., Lancia, M., Cao, G., Crivellari, A., Guo, Z., & Zheng, C. (2022). Comparison and interpretation of data-driven models for simulating site-specific human-impacted groundwater dynamics in the North China Plain. Journal of Hydrology, 616, 128751. https://doi.org/10.1016/j.jhydrol.2022.128751

Jones, A., Kuehnert, J., Fraccaro, P., Meuriot, O., Ishikawa, T., Edwards, B., Stoyanov, N., Remy, S. L., Weldemariam, K., & Assefa, S. (2023). AI for climate impacts: applications in flood risk. Npj Climate and Atmospheric Science, 6(1), 1–8. https://doi.org/10.1038/s41612-023-00388-1

Jung, H., Saynisch-Wagner, J., & Schulz, S. (2024). Can eXplainable AI offer a new perspective for groundwater recharge estimation?—Global-scale modeling using neural network. Water Resources Research, 60, e2023WR036360. https://doi.org/10.1029/2023WR036360

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. Journal of Hydrology, 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. Journal of Hydrology: Regional Studies, 43, 101185. https://doi.org/10.1016/j.ejrh.2022.101185

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? Computers and Electronics in Agriculture, 203, 107482. https://doi.org/10.1016/j.compag.2022.107482

Kayhomayoon, Z., Jamnani, M. R., Rashidi, S., Ghordoyee Milan, S., Arya Azar, N., & Berndtsson, R. (2023). Soft computing assessment of current and future groundwater resources under CMIP6 scenarios in northwestern Iran. Agricultural Water Management, 285, 108369. https://doi.org/10.1016/j.agwat.2023.108369

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

Klavans, J. M., Clement, A. C., Cane, M. A., & Murphy, L. N. (2022). The Evolving Role of External Forcing in North Atlantic SST Variability over the Last Millennium. Journal of Climate, 35(9), 2741-2754. https://doi.org/10.1175/JCLI-D-21-0338.1

Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., & Nearing, G. (2022). Uncertainty estimation with deep learning for rainfall-runoff modeling. Hydrology and Earth System Sciences, 26(6), 1673–1693. https://doi.org/10.5194/hess-26-1673-2022

Kraft, B., Jung, M., Körner, M., Koirala, S., & Reichstein, M. (2022). Towards hybrid modeling of the global hydrological cycle. Hydrol. Earth Syst. Sci, 26, 1579–1614. https://doi.org/10.5194/hess-26-1579-2022

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

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., 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., 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

Labat, D., Ababou, R., & Mangin, A. (2000). Rainfall–runoff relations for karstic springs. Part II: Continuous wavelet and discrete orthogonal multiresolution analyses. Journal of Hydrology, 238(3-4), 149-178. https://doi.org/10.1016/S0022-1694(00)00322-X

Lavers, D. A., Simmons, A., Vamborg, F., & Rodwell, M. J. (2022). An evaluation of ERA5 precipitation for climate monitoring. Quarterly Journal of the Royal Meteorological Society, 148(748), 3152-3165. https://doi.org/10.1002/qj.4351

Lee, S., Lee, KK. & Yoon, H. Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. Hydrogeol J 27, 567–579 (2019). https://doi.org/10.1007/s10040-018-1866-3

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., and Dadson, S. J. (2021). Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models, Hydrol. Earth Syst. Sci., 25, 5517–5534, https://doi.org/10.5194/hess-25-5517-2021

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth Syst. Dynam., 11, 491–508, https://doi.org/10.5194/esd-11-491-2020

Liang, Z., Liu, Y., Hu, H., Li, H., Ma, Y., & Khan, M. Y. A. (2021). Combined Wavelet Transform With Long Short-Term Memory Neural Network for Water Table Depth Prediction in Baoding City, North China Plain. Frontiers in Environmental Science, 9. https://doi.org/10.3389/fenvs.2021.780434

Liesch, T., & Wunsch, A. (2019). Aquifer responses to long-term climatic periodicities. Journal of Hydrology, 572, 226-242. https://doi.org/10.1016/j.jhydrol.2019.02.060

Liu, J., Koch, J., Stisen, S., Troldborg, L., & Schneider, R. J. M. (2024). A national-scale hybrid model for enhanced streamflow estimation-consolidating a physically based hydrological model with long short-term memory (LSTM) networks. Hydrol. Earth Syst. Sci, 28, 2871–2893. https://doi.org/10.5194/hess-28-2871-2024

Liu, Q., Gui, D., Zhang, L., Niu, J., Dai, H., Wei, G., & Hu, B. X. (2022). Simulation of regional groundwater levels in arid regions using interpretable

machine learning models. Science of The Total Environment, 831, 154902. https://doi.org/10.1016/J.SCITOTENV.2022.154902

Liu, W., Yu, H., Yang, L., Yin, Z., Zhu, M., & Wen, X. (2021). Deep learning-based predictive framework for groundwater level forecast in arid irrigated areas. Water (Switzerland), 13(18), 2558. https://doi.org/10.3390/w13182558

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

Long, J., Wang, L., Chen, D., Li, N., Zhou, J., Li, X., Guo, X., Liu, H., Chai, C., & Fan, X. (2024). Hydrological Projections in the Third Pole Using Artificial Intelligence and an Observation-Constrained Cryosphere-Hydrology Model. Earth's Future, 12(4), e2023EF004222. https://doi.org/10.1029/2023EF004222

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

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., Delgado-Torres, C., Samsó, M., and Bretonnière, P.-A. (2022). Constraining low-frequency variability in climate projections to predict climate on decadal to multi-decadal timescales – a poor man's initialized prediction system, Earth Syst. Dynam., 13, 1437–1450, https://doi.org/10.5194/esd-13-1437-2022

Maria Clerc-Schwarzenbach, F., Selleri, G., Neri, M., Toth, E., van Meerveld, I., & Seibert, J. (2024). HESS Opinions: A few camels or a whole caravan? https://doi.org/10.5194/egusphere-2024-864

Martel, J.-L., Brissette, F., Troin, M., Arsenault, R., Chen, J., Su, T., & Lucas-Picher, P. (2022). CMIP5 and CMIP6 model projection comparison for hydrological impacts over North America. Geophysical Research Letters, 49, e2022GL098364. https://doi.org/10.1029/2022GL098364

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. International Journal of Climatology, 30(14), 2146–2154. https://doi.org/10.1002/joc.2022

Massei, N., & Fournier, M. (2012). Assessing the expression of large-scale climatic fluctuations in the hydrological variability of daily Seine river flow (France) between 1950 and 2008 using Hilbert-Huang Transform. Journal of Hydrology, 448–449, 119–128. https://doi.org/10.1016/j.jhydrol.2012.04.052

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. Journal of Hydrology, 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. Proceedings of the International Association of Hydrological Sciences, 383, 141–149. https://doi.org/10.5194/piahs-383-141-2020

Maxwell, R. M., Condon, L. E., & Kollet, S. J. (2015). A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. Geosci. Model Dev, 8, 923–937. https://doi.org/10.5194/gmd-8-923-2015

Mcdonald, M. G., Harbaugh, A. W., Clark, W. P., & Peck, D. L. (1988). A modular three-dimensional finite-difference ground-water flow model. United States Department of the Interior.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 1(Scipy), 56–61. https://doi.org/10.25080/majora-92bf1922-00a

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J. J. F., Clark, M. P., Uijlenhoet, R., and Teuling, A. J. (2018). Mapping (dis)agreement in hydrologic projections, Hydrol. Earth Syst. Sci., 22, 1775–1791, https://doi.org/10.5194/hess-22-1775-2018

Mishra, V., Shah, R., Azhar, S., Shah, H., Modi, P., & Kumar, R. (2018). Reconstruction of droughts in India using multiple land-surface models (1951-2015). Hydrology and Earth System Sciences, 22(4), 2269–2284. https://doi.org/10.5194/hess-22-2269-2018

Mishra, V., Dangar, S., Tiwari, V. M., Lall, U., & Wada, Y. (2024). Summer monsoon drying accelerates India's groundwater depletion under climate change. Earth's Future, 12, e2024EF004516. https://doi.org/10.1029/2024EF004516

Müller, L. and Döll, P. (2024). Quantifying and communicating uncertain climate change hazards in participatory climate change adaptation processes, Geosci. Commun., 7, 121–144, https://doi.org/10.5194/gc-7-121-2024

Mohanty, S., Jha, M. K., Kumar, A., & Panda, D. K. (2013). Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi-Surua Inter-basin of Odisha, India. Journal of Hydrology, 495, 38–51. https://doi.org/10.1016/j.jhydrol.2013.04.041

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

Mouatadid, S., Adamowski, J. F., Tiwari, M. K., & Quilty, J. M. (2019). Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting. Agricultural Water Management, 219(March), 72–85. https://doi.org/10.1016/j.agwat.2019.03.045

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

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 System Science Data, 14(9), 4035–4056. https://doi.org/10.5194/essd-14-4035-2022

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., & Matias, Y. (2024). Global prediction of extreme floods in ungauged watersheds. Nature, 627(8004), 559–563. https://doi.org/10.1038/s41586-024-07145-1

Neves, M. C., Jerez, S., & Trigo, R. M. (2018). The response of piezometric levels in Portugal to NAO, EA, and SCAND climate patterns. Journal of Hydrology, 568, 1105-1117. https://doi.org/10.1016/j.jhydrol.2018.11.054

Niu, X., Lu, C., Zhang, Y., Zhang, Y., Wu, C., Saidy, E., Liu, B., & Shu, L. (2023). Hysteresis response of groundwater depth on the influencing factors using an explainable learning model framework with Shapley values.

Science of The Total Environment, 904, 166662. https://doi.org/10.1016/j.scitotenv.2023.166662

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., 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., 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., Alami, M. T., & Vousoughi, F. D. (2016). Hybrid of SOM-Clustering Method and Wavelet-ANFIS Approach to Model and Infill Missing Groundwater Level Data. Journal of Hydrologic Engineering, 21(9). https://doi.org/10.1061/(asce)he.1943-5584.0001398

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., 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

Nozari, S., Bailey, R. T., Haacker, E. M., Zambreski, Z. T., Xiang, Z., & Lin, X. (2022). Employing machine learning to quantify long-term climatological and regulatory impacts on groundwater availability in intensively irrigated regions. Journal of Hydrology, 614, 128511. https://doi.org/10.1016/j.jhydrol.2022.128511

O'Reilly, C. H., Patterson, M., Robson, J., Monerie, P. A., Hodson, D., & Ruprich-Robert, Y. (2023). Challenges with interpreting the impact of

Atlantic Multidecadal Variability using SST-restoring experiments. npj Climate and Atmospheric Science, 6(1), 1-12. https://doi.org/10.1038/s41612-023-00335-0

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

Pedregosa, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot andÉdouardand, M., Duchesnay, A., & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. In Journal of Machine Learning Research (Vol. 12).

Percival, D. B., & Walden, A. T. (2000). Wavelet Methods for Time SeriesAnalysis. In Wavelet Methods for Time SeriesAnalysis. Cambridge University Press. https://doi.org/10.1017/cbo9780511841040

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, L., & Fisher, M. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. Journal of Climate, 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. Journal of Hydrology, 563(February), 336–353. https://doi.org/10.1016/j.jhydrol.2018.05.003

Quilty, J. M., Sikorska-Senoner, A. E., & Hah, D. (2022). A stochastic conceptual-data-driven approach for improved hydrological simulations. Environmental Modelling and Software, 149, 105326. https://doi.org/10.1016/j.envsoft.2022.105326

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, 336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037

Rahmani, F., Appling, A., Feng, D., Lawson, K., & Shen, C. (2023). Identifying Structural Priors in a Hybrid Differentiable Model for Stream Water Temperature Modeling. Water Resources Research, 59(12). https://doi.org/10.1029/2023WR034420

Rehana, S., & Rajesh, M. (2023). Assessment of impacts of climate change on Indian riverine thermal regimes using hybrid deep learning methods. Water Resources Research, 59, e2021WR031347. https://doi.org/10.1029/2021WR031347

Roshani, A., Hamidi, M. (2022). Groundwater Level Fluctuations in Coastal Aquifer: Using Artificial Neural Networks to Predict the Impacts of Climatical CMIP6 Scenarios. Water Resour Manage 36, 3981–4001 https://doi.org/10.1007/s11269-022-03204-2

Rust, W., Holman, I., Corstanje, R., Bloomfield, J., & Cuthbert, M. (2018). A conceptual model for climatic teleconnection signal control on groundwater variability in Europe. Earth-Science Reviews, 177, 164-174. https://doi.org/10.1016/j.earscirev.2017.09.017

Rust, W., Holman, I., Bloomfield, J., Cuthbert, M., & Corstanje, R. (2019). Understanding the potential of climate teleconnections to project future groundwater drought. Hydrology and Earth System Sciences, 23(8), 3233–3245. https://doi.org/10.5194/hess-23-3233-2019

Saeed, A., Li, C., Danish, M., Rubaiee, S., Tang, G., Gan, Z., & Ahmed, A. (2020). Hybrid bidirectional lstm model for short-term wind speed interval prediction. IEEE Access, 8, 182283–182294. https://doi.org/10.1109/ACCESS.2020.3027977

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. Science of the Total Environment, 857(February 2022), 159289. https://doi.org/10.1016/j.scitotenv.2022.159289

Scanlon, B. R., Reedy, R. C., Faunt, C. C., Pool, D., & Uhlman, K. (2016). Enhancing drought resilience with conjunctive use and managed aquifer recharge in California and Arizona. Environmental Research Letters, 11(4). https://doi.org/10.1088/1748-9326/11/4/049501

Scanlon, B. R., Fakhreddine, S., Rateb, A., De Graaf, I., Famiglietti, J., Gleeson, T., Grafton, R. Q., Jobbagy, E., Kebede, S., Kolusu, S. R., Konikow, L. F., Long, D., Mekonnen, M., Schmied, H. M., Mukherjee, A., MacDonald, A., Reedy, R. C., Shamsudduha, M., Simmons, C. T., . . . Zheng, C. (2023). Global water resources and the role of groundwater in a resilient water future. Nature Reviews Earth & Environment, 4(2), 87-101. https://doi.org/10.1038/s43017-022-00378-6

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

Secci, D., Giovanna Tanda, M., D'Oria, M., & Todaro, V. (2023). Artificial intelligence models to evaluate the impact of climate change on groundwater resources. Journal of Hydrology, 627, 130359. https://doi.org/10.1016/j.jhydrol.2023.130359

Sharma, S., & Mujumdar, P. P. (2024). Baseflow significantly contributes to river floods in Peninsular India. Scientific Reports, 14(1), 1-10. https://doi.org/10.1038/s41598-024-51850-w

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., … Lawson, K. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. Nature Reviews Earth and Environment, 4(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019, 3285–3292. https://doi.org/10.1109/BigData47090.2019.9005997

Slimani, S., Massei, N., Mesquita, J. et al. Combined climatic and geological forcings on the spatio-temporal variability of piezometric levels in the chalk aquifer of Upper Normandy (France) at pluridecennal scale. Hydrogeol J 17, 1823–1832 (2009). https://doi.org/10.1007/s10040-009-0488-1

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

Slater, L. J., Arnal, L., Boucher, M. A., Chang, A. Y. Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., & Zappa, M. (2023). Hybrid forecasting: blending climate predictions with AI models. Hydrology and Earth System Sciences, 27(9), 1865–1889. https://doi.org/10.5194/hess-27-1865-2023

Slater, L., Coxon, G., Brunner, M., McMillan, H., Yu, L., Zheng, Y., et al. (2024). Spatial sensitivity of river flooding to changes in climate and land cover through explainable AI. Earth's Future, 12, e2023EF004035. https://doi.org/10.1029/2023EF004035

Sun, A. Y., Jiang, P., Mudunuru, M. K., & Chen, X. (2021). Explore spatio-temporal learning of large sample hydrology using graph neural networks. Water Resources Research, 57, e2021WR030394. https://doi.org/10.1029/2021WR030394

Tahmasebi, P., Kamrava, S., Bai, T., & Sahimi, M. (2020). Machine learning in geo- and environmental sciences: From small to large scale. Advances in Water Resources, 142, 103619. https://doi.org/10.1016/j.advwatres.2020.103619

Tao, H., Majeed Hameed, M., Abdulameer Marhoon, H., Zounemat-Kermani, M., Salim, H., Sungwon, K., Oleiwi Sulaiman, S., Leong Tan, M., Sa'adi, Z., Danandeh Mehr, A., Falah Allawi, M., Abba, S. I., Mohamad Zain, J., Falah, M. W., Jamei, M., Dhanraj Bokde, N., Bayatvarkeshi, M., Al-Mukhtar, M., Kumar Bhagat, S., … Mundher Yaseen, Z. (2022). Groundwater Level Prediction using Machine Learning Models: A Comprehensive Review. Neurocomputing, 489, 271–308. https://doi.org/10.1016/j.neucom.2022.03.014

Tarek, M., Brissette, F. P., and Arsenault, R. (2020) Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, Hydrol. Earth Syst. Sci., 24, 2527–2544, https://doi.org/10.5194/hess-24-2527-2020

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., . . . Treidel, H. (2013). Ground water and climate change. Nature Climate Change, 3(4), 322-329. https://doi.org/10.1038/nclimate1744

Terray, L. (2012). Evidence for multiple drivers of North Atlantic multi-decadal climate variability. Geophysical Research Letters, 39(19). https://doi.org/10.1029/2012GL053046

Thébault, C., Perrin, C., Andréassian, V., Thirel, G., Legrand, S., and Delaigue, O. (2024). Multi-model approach in a variable spatial framework for streamflow simulation, Hydrol. Earth Syst. Sci., 28, 1539–1566, https://doi.org/10.5194/hess-28-1539-2024

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., Hrast Essenfelder, A., Isabellon, M., Jonas, T., Maetens, W., Magni, D., Masante, D., Mazzeschi, M., … Spinoni, J. (2023). Drought in Europe March 2023. GDO Analytical Report, 1, 1–31. https://doi.org/10.2760/998985

Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. Bulletin of the American Meteorological Society, 79(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Trichakis, I., Burek, P., de Roo, A., & Pistocchi, A. (2017). Towards a pan-european integrated groundwater and surface water model: Development and applications. Environmental Processes, 4, S81–S93. https://doi.org/10.1007/s40710-017-0216-0

Tripathy, K. P., & Mishra, A. K. (2023). Deep learning in hydrology and water resources disciplines: Concepts, methods, applications, and research directions. Journal of Hydrology, 628, 130458. https://doi.org/10.1016/j.jhydrol.2023.130458

Tsai, W. P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., & Shen, C. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. Nature Communications, 12(1), 1–13. https://doi.org/10.1038/s41467-021-26107-z

Tursun, A., Xie, X., Wang, Y., Liu, Y., Peng, D., Rusuli, Y., & Zheng, B. (2024). Reconstruction of missing streamflow series in human-regulated catchments using a data integration LSTM model. Journal of Hydrology: Regional Studies, 52, 101744. https://doi.org/10.1016/j.ejrh.2024.101744

Uc-Castillo, J. L., Marín-Celestino, A. E., Martínez-Cruz, D. A., Tuxpan-Vargas, J., & Ramos-Leal, J. A. (2023). A systematic review and meta-analysis of groundwater level forecasting with machine learning techniques: Current status and future directions. Environmental Modelling & Software, 168, 105788. https://doi.org/10.1016/j.envsoft.2023.105788

Uz, M., Atman, K. G., Akyilmaz, O., Shum, C. K., Keleş, M., Ay, T., Tandoğdu, B., Zhang, Y., & Mercan, H. (2022). Bridging the gap between GRACE and GRACE-FO missions with deep learning aided water storage

simulations. Science of the Total Environment, 830. https://doi.org/10.1016/j.scitotenv.2022.154701

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. Computing in Science and Engineering, 13(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Vergnes, J. P., Caballero, Y., & Lanini, S. (2023). Assessing climate change impact on French groundwater resources using a spatially distributed hydrogeological model. Hydrological Sciences Journal, 68(2), 209–227. https://doi.org/10.1080/02626667.2022.2150553

Vidal, P., Martin, E., Franchistéguy, L., Baillon, M., & Soubeyroux, M. (2010). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. International Journal of Climatology, 30(11), 1627-1644. https://doi.org/10.1002/joc.2003

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

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

Wi, S., & Steinschneider, S. (2022). Assessing the physical realism of deep learning hydrologic model projections under climate change. Water Resources Research, 58, e2022WR032123. https://doi.org/10.1029/2022WR032123

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

Wu, C., Zhang, X., Wang, W., Lu, C., Zhang, Y., Qin, W., Tick, G. R., Liu, B., & Shu, L. (2021). Groundwater level modeling framework by combining the wavelet transform with a long short-term memory data-driven model. Science of the Total Environment, 783, 146948. https://doi.org/10.1016/j.scitotenv.2021.146948

Wu, Y., Miao, C., Slater, L., Fan, X., Chai, Y., & Sorooshian, S. (2024). Hydrological Projections under CMIP5 and CMIP6: Sources and Magnitudes

of Uncertainty. Bulletin of the American Meteorological Society, 105(1), E59-E74. https://doi.org/10.1175/BAMS-D-23-0104.1

Wu, M., Feng, Q., Wen, X., Yin, Z., Yang, L., & Sheng, D. (2021). Deterministic analysis and uncertainty analysis of ensemble forecasting model based on variational mode decomposition for estimation of monthly groundwater level. Water (Switzerland), 13(2). https://doi.org/10.3390/w13020139

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

Xiong, J., Abhishek, Guo, S., & Kinouchi, T. (2022). Leveraging machine learning methods to quantify 50 years of dwindling groundwater in India. Science of the Total Environment, 835(March), 155474. https://doi.org/10.1016/j.scitotenv.2022.155474

Xu, W., Chen, J., Corzo, G., Xu, C. Y., Zhang, X. J., Xiong, L., Liu, D., & Xia, J. (2024). Coupling Deep Learning and Physically Based Hydrological Models for Monthly Streamflow Predictions. Water Resources Research, 60(2), e2023WR035618. https://doi.org/10.1029/2023WR035618

Yadav, B., Gupta, P. K., Patidar, N., & Himanshu, S. K. (2020). Ensemble modelling framework for groundwater level prediction in urban areas of India. Science of the Total Environment, 712, 135539. https://doi.org/10.1016/j.scitotenv.2019.135539

Yao, Y., Zhao, Y., Li, X., Feng, D., Shen, C., Liu, C., Kuang, X., & Zheng, C. (2023). Can transfer learning improve hydrological predictions in the alpine regions? Journal of Hydrology, 625, 130038. https://doi.org/10.1016/j.jhydrol.2023.130038

Yuan, Q., Thorarinsdottir, T. L., Beldring, S., Wong, W. K., & Xu, C. (2023). Assessing uncertainty in hydrological projections arising from local-scale internal variability of climate. Journal of Hydrology, 620, 129415. https://doi.org/10.1016/j.jhydrol.2023.129415

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

Zhang, J., Zhang, X., Niu, J., Hu, B. X., Soltanian, M. R., Qiu, H., & Yang, L. (2019). Prediction of groundwater level in seashore reclaimed land using wavelet and artificial neural network-based hybrid model. Journal of Hydrology, 577, 123948. https://doi.org/10.1016/j.jhydrol.2019.123948

Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. Journal of Hydrology, 561, 918–929. https://doi.org/10.1016/j.jhydrol.2018.04.065

Zhao, X., Fang, K., Chen, F., Martín, H., & Roig, F. A. (2023). Reconstructed Jing River streamflow from western China: A 399-year perspective for hydrological changes in the Loess Plateau. Journal of Hydrology, 621, 129573. https://doi.org/10.1016/j.jhydrol.2023.129573

Supplementary information

List of Tables:

Table S1: List of GWL stations and corresponding class and static attributes

Well ID	Class	longitude	latitude	Geological Context	Medium: Type of Porosity	Lithology
00068X0010/F295	mixed	2.09	50.80	Magmatic and metamorphic	Matrix / cracks	Chalk
00182X0010/P1	annual	2.27	50.50	Sedimentary	Matrix / cracks	Chalk
00241X0012/P1	mixed	1.82	50.39	Sedimentary	Matrix / cracks	Chalk
00263X0006/P1	mixed	2.70	50.39	Magmatic and metamorphic	Matrix / cracks	Chalk
00271X0002/P2	mixed	2.89	50.35	Magmatic and metamorphic	Matrix / cracks	Chalk
00332X0007/S1	mixed	1.95	50.19	Sedimentary	Matrix / cracks	Chalk
00471X0095/PZ2013	mixed	2.58	50.02	Sedimentary	Matrix / cracks	Chalk
00572X0010/S1	inertial	0.51	49.79	Sedimentary	Matrix/fracture/karst	Chalk
00578X0002/S1	inertial	0.71	49.74	Sedimentary	Matrix/fracture/karst	Chalk
00583X0005/S1	inertial	0.91	49.78	Sedimentary	Matrix/fracture/karst	Chalk
00608X0028/S1	inertial	1.80	49.76	Sedimentary	Matrix/fracture/karst	Chalk

00671X0052/S1	annual	4.04	49.84	Sedimentary	Matrix / cracks	Chalk
00755X0006/S1	mixed	0.42	49.56	Sedimentary	Matrix/fracture/karst	Chalk
00766X0004/S1	inertial	0.84	49.57	Sedimentary	Matrix/fracture/karst	Chalk
00773X0002/S1	mixed	1.28	49.60	Sedimentary	Matrix/fracture/karst	Chalk
00791X0017/S1	mixed	1.82	49.62	Sedimentary	Matrix / cracks	Chalk
00794X0021/S1	mixed	2.09	49.60	Sedimentary	Matrix / cracks	Chalk
00805X0002/S1	mixed	2.21	49.55	Sedimentary	Matrix / cracks	Chalk
00817X0145/PZ_SN	mixed	2.70	49.55	Sedimentary	Matrix / cracks	Chalk
00821X0035/S1	mixed	2.96	49.65	Sedimentary	Matrix / cracks	Chalk
00847X0043/S1	annual	3.86	49.52	Sedimentary	Matrix / cracks	Chalk
00853X0030/PZ2013	annual	4.16	49.61	Sedimentary	Matrix / cracks	Chalk
00862X0005/S1	annual	4.42	49.60	Magmatic and metamorphic	Porous	Sand
00957X0005/S1	mixed	-0.90	49.37	Sedimentary	Karst / fissures	Limestone

00993X0002/S1	inertial	0.98	49.48	Sedimentary	Matrix/fracture/karst	Chalk
01003X0008/S1	mixed	1.26	49.43	Sedimentary	Matrix/fracture/karst	Chalk
01013X0004/S1	mixed	1.63	49.44	Sedimentary	Matrix/fracture/karst	Chalk
01024X0058/S1	mixed	2.10	49.45	Sedimentary	Matrix / cracks	Chalk
01031X0023/S1	mixed	2.21	49.50	Sedimentary	Matrix / cracks	Chalk
01045X0015/S1	mixed	2.54	49.36	Sedimentary	Matrix / cracks	Chalk
01046X0010/S1	mixed	2.61	49.38	Sedimentary	Matrix / cracks	Chalk
01053X0058/S1	inertial	3.07	49.45	Alluvial	Porous	Sand
01074X0006/S1	annual	3.92	49.49	Sedimentary	Matrix / cracks	Chalk
01086X0011/LS4	annual	4.10	49.34	Sedimentary	Matrix / cracks	Chalk
01116X0138/F1	annual	5.16	49.37	Magmatic and metamorphic	Karst / fissures	Limestone
01192X0043/S1	annual	-0.60	49.25	Sedimentary	Karst / fissures	Limestone
01194X0069/S1	mixed	-0.43	49.29	Sedimentary	Karst / fissures	Limestone

01198X0002/S1	inertial	-0.38	49.21	Sedimentary	Karst / fissures	Limestone
01245X0010/S1	inertial	1.09	49.22	Sedimentary	Matrix/fracture/karst	Chalk
01252X0011/S1	mixed	1.54	49.30	Sedimentary	Matrix/fracture/karst	Chalk
01258X0020/S1	mixed	1.71	49.20	Sedimentary	Matrix/fracture/karst	Chalk
01264X0029/S1	mixed	2.08	49.24	Sedimentary	Matrix/fracture/karst	Chalk
01287X0017/S1	mixed	2.75	49.21	Magmatic and metamorphic	Matrix / cracks	Sand
01347X0002/S1	annual	4.86	49.17	Magmatic and metamorphic	Porous	Gaize
01381X0070/P25	annual	6.18	49.30	Sedimentary	Porous	Alluvium
01461X0012/S1	inertial	-0.29	49.06	Sedimentary	Karst / fissures	Limestone
01473X0087/S1	inertial	0.24	49.06	Sedimentary	Matrix/fracture/karst	Chalk
01491X0009/S1	inertial	0.77	49.11	Sedimentary	Matrix/fracture/karst	Chalk
01516X0004/S1	inertial	1.61	48.98	Sedimentary	Matrix/fracture/karst	Chalk
01584X0023/LV3	annual	4.29	49.07	Sedimentary	Matrix / cracks	Chalk

01871X0031/S1	mixed	3.65	48.89	Not available	Porous	Marl
01995X0012/342B	annual	7.98	48.81	Sedimentary	Porous	Alluvium
02206X0022/S1	inertial	2.68	48.61	Magmatic and metamorphic	Karst / fissures	Limestone
02225X0016/S1	mixed	3.28	48.62	Sedimentary	Karst / fissures	Limestone
02347X0022/314	mixed	7.75	48.65	Sedimentary	Porous	Alluvium
02566X0019/S1	inertial	1.90	48.46	Magmatic and metamorphic	Karst / fissures	Limestone
02617X0009/S1	annual	3.83	48.44	Sedimentary	Matrix / cracks	Chalk
02636X0009/S1	mixed	4.49	48.44	Sedimentary	Porous	Alluvium
02726X0029/238	annual	7.66	48.49	Sedimentary	Porous	Alluvium
02923X0007/F	inertial	1.99	48.38	Magmatic and metamorphic	Porous	Sand
02931X0008/S1	inertial	2.24	48.36	Magmatic and metamorphic	Matrix / cracks	Limestone
02974X0004/S1	annual	3.87	48.36	Sedimentary	Matrix / cracks	Chalk

03272X0006/PZ	inertial	1.89	48.24	Magmatic and metamorphic	Karst / fissures	Limestone
03276X0009/P	inertial	1.92	48.10	Magmatic and metamorphic	Karst / fissures	Limestone
03287X0018/S1	inertial	2.35	48.09	Magmatic and metamorphic	Karst / fissures	Limestone
03622X0027/PZ	inertial	1.58	47.99	Magmatic and metamorphic	Karst / fissures	Limestone
04132X0086/PP6	mixed	7.31	47.82	Sedimentary	Porous	Alluvium
04137X0018/15	mixed	7.42	47.72	Sedimentary	Porous	Alluvium
04398X0002/SONDAG	annual	5.32	47.53	Sedimentary	Karst	Limestone
04458X0023/S3	mixed	7.52	47.59	Sedimentary	Porous	Alluvium
00061X0117/PZ1	mixed	1.82	50.87	Magmatic and metamorphic	Matrix / cracks	Chalk
00263X0129/PZASA4	mixed	2.72	50.33	Magmatic and metamorphic	Matrix / cracks	Chalk
00275X0005/P1	inertial	2.89	50.25	Magmatic and metamorphic	Matrix / cracks	Chalk
00346X0011/S1	mixed	2.33	50.11	Sedimentary	Matrix / cracks	Chalk

00463X0036/H1	mixed	2.36	49.95	Sedimentary	Matrix / cracks	Chalk
02603X0009/S1	inertial	3.44	48.58	Sedimentary	Karst / fissures	Limestone

Table S2: Station wise training and testing dates and period. The first four years in the training period were used to make sequences hence only ERA5 data was used for this period.

Code	Training Period (Dates)	Training Period (Years)	Testing Period (Dates)	Testing Period (Years)
00068X0010/F295	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00182X0010/P1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00241X0012/P1	1968-04-01 to 2014-12-01	47	2015-01-01 to 2022-12-01	8
00263X0006/P1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00271X0002/P2	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00332X0007/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00471X0095/PZ2013	1966-06-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8

00572X0010/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00578X0002/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00583X0005/S1	1967-03-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
00608X0028/S1	1966-10-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
00671X0052/S1	1970-03-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
00755X0006/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00766X0004/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00773X0002/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00791X0017/S1	1970-04-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
00794X0021/S1	1966-10-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
00805X0002/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00817X0145/PZ_SN	1966-11-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
00821X0035/S1	1970-04-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
00847X0043/S1	1966-10-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8

00853X0030/PZ2013	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00862X0005/S1	1967-02-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
00957X0005/S1	1970-02-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
00993X0002/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01003X0008/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01013X0004/S1	1967-03-01 to 2014-12-01	48	2015-01-01 to 2022-08-01	8
01024X0058/S1	1970-02-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
01031X0023/S1	1966-10-01 to 2014-12-01	48	2015-01-01 to 2020-03-01	5
01045X0015/S1	1966-10-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
01046X0010/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01053X0058/S1	1966-10-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
01074X0006/S1	1966-10-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
01086X0011/LS4	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01116X0138/F1	1971-05-01 to 2014-12-01	44	2015-01-01 to 2022-12-01	8

01192X0043/S1	1970-02-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
01194X0069/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01198X0002/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01245X0010/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01252X0011/S1	1967-03-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
01258X0020/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01264X0029/S1	1970-04-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
01287X0017/S1	1970-04-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
01347X0002/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01381X0070/P25	1967-09-01 to 2014-12-01	47	2015-01-01 to 2022-12-01	8
01461X0012/S1	1966-06-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01473X0087/S1	1970-02-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
01491X0009/S1	1968-12-01 to 2014-12-01	46	2015-01-01 to 2022-12-01	8
01516X0004/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8

01584X0023/LV3	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01871X0031/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
01995X0012/342B	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
02206X0022/S1	1969-06-01 to 2014-12-01	46	2015-01-01 to 2022-12-01	8
02225X0016/S1	1967-02-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
02347X0022/314	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
02566X0019/S1	1970-03-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
02617X0009/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
02636X0009/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
02726X0029/238	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
02923X0007/F	1970-12-01 to 2014-12-01	44	2015-01-01 to 2022-12-01	8
02931X0008/S1	1966-09-01 to 2014-12-01	48	2015-01-01 to 2022-12-01	8
02974X0004/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
03272X0006/PZ	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8

03276X0009/P	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
03287X0018/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
03622X0027/PZ	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
04132X0086/PP6	1970-03-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
04137X0018/15	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-01-01	7
04398X0002/SONDAG	1970-02-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
04458X0023/S3	1971-12-01 to 2014-12-01	43	2015-01-01 to 2022-12-01	8
00061X0117/PZ1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00263X0129/PZASA4	1970-02-01 to 2014-12-01	45	2015-01-01 to 2022-12-01	8
00275X0005/P1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00346X0011/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
00463X0036/H1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8
02603X0009/S1	1966-02-01 to 2014-12-01	49	2015-01-01 to 2022-12-01	8

Table S3: Best hyperparameters for all standalone models of single station approach

Code	Best Params GRU	Best Params LSTM	Best Params BILSTM
00068X0010/F295	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.004285640735795779, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00182X0010/P1	{'learning_rate': 0.007291305160743269, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.009707480295841904, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.0034631671584833116, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00241X0012/P1	{'learning_rate': 0.006938969099074053, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.008993463169584801, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00263X0006/P1	{'learning_rate': 0.006821777398460312, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}

00271X0002/P2	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.005143094276731783, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00332X0007/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001440785178275633, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
00471X0095/PZ2013	{'learning_rate': 0.009605417373868429, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001940640704224857 2, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00572X0010/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00578X0002/S1	{'learning_rate': 0.003729090520557473, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1,	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1,	{'learning_rate': 0.001946438853175019 8, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 224, 'n_layers': 1,

	'n_units_l0': 70, 'dropout_l0': 0.2}	'n_units_l0': 80, 'dropout_l0': 0.2}	'n_units_l0': 80, 'dropout_l0': 0.2}
00583X0005/S1	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.006821777398460312, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00608X0028/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00671X0052/S1	{'learning_rate': 0.001986907762154767 7, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.002765878983990436 7, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
00755X0006/S1	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.002765878983990436 7, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.001471958750607762, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00766X0004/S1	{'learning_rate': 0.003151425264040375, 'optimizer':	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.001346013249503339, 'optimizer': 'adam', 'epochs':

	'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	450, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
00773X0002/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00791X0017/S1	{'learning_rate': 0.006562911757711773, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00794X0021/S1	{'learning_rate': 0.008819753706964301, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.003487742134799628, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00805X0002/S1	{'learning_rate': 0.009773436531085809, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001582159258248788 3, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}

00817X0145/PZ_SN	{'learning_rate': 0.0022009248329412116, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00821X0035/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00847X0043/S1	{'learning_rate': 0.00998126342923339, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.007630901178166286, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00853X0030/PZ2013	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00862X0005/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1,	{'learning_rate': 0.008329460349014989, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1,	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1,

	'n_units_l0': 80, 'dropout_l0': 0.2}	'n_units_l0': 20, 'dropout_l0': 0.2}	'n_units_l0': 40, 'dropout_l0': 0.2}
00957X0005/S1	{'learning_rate': 0.005262638466167993, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.005741434770268582, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.004228752765456005, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00993X0002/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
01003X0008/S1	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
01013X0004/S1	{'learning_rate': 0.009720173077787352, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 128, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.003487742134799628, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001763895960635345 3, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 10, 'dropout_l0': 0.2}
01024X0058/S1	{'learning_rate': 0.009707480295841904, 'optimizer':	{'learning_rate': 0.001300122762188 1476, 'optimizer': 'adam',	{'learning_rate': 0.002419619449357 1197, 'optimizer': 'adam', 'epochs':

	'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	400, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
01031X0023/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0014886622754444637, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
01045X0015/S1	{'learning_rate': 0.008819753706964301, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0027658789839904367, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
01046X0010/S1	{'learning_rate': 0.004871081986357781, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.009242062120765548, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01053X0058/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0010091326698400099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}

01074X0006/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01086X0011/LS4	{'learning_rate': 0.0035650145243888115, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.008819753706964301, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.0081030567555325962, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
01116X0138/F1	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.008625283569644069, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0018484463533347947, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
01192X0043/S1	{'learning_rate': 0.008819753706964301, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.006508409660582904, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
01194X0069/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1,	{'learning_rate': 0.003886552995763 6045, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 48, 'n_layers': 1,	{'learning_rate': 0.0019905093688178995, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 1,

	'n_units_l0': 80, 'dropout_l0': 0.2}	'n_units_l0': 50, 'dropout_l0': 0.2}	'n_units_l0': 100, 'dropout_l0': 0.2}
01198X0002/S1	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.002765878983990436 7, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
01245X0010/S1	{'learning_rate': 0.007735205021322065, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01252X0011/S1	{'learning_rate': 0.007865894872388346, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.0010091326698 40099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0014886622754444637, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
01258X0020/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0034877421347 99628, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01264X0029/S1	{'learning_rate': 0.003151425264040375, 'optimizer':	{'learning_rate': 0.0010091326698 40099, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.0021556933852825582, 'optimizer': 'adam', 'epochs':

	'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	500, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}
01287X0017/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0081030567555532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01347X0002/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.008863787581270316, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01381X0070/P25	{'learning_rate': 0.009774776777496903, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.008759599404809302, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01461X0012/S1	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0019869077621547677, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}

01473X0087/S1	{'learning_rate': 0.007630901178166286, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.005906369914893648, 'optimizer': 'adam', 'epochs': 50, 'batch_size': 176, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001986907762154767677, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}
01491X0009/S1	{'learning_rate': 0.0059131273523082, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.002155693385282558 2, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}
01516X0004/S1	{'learning_rate': 0.007476627885185544, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.006742941264953408, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}
01584X0023/LV3	{'learning_rate': 0.009707480295841904, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.008091344208943023, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}
01871X0031/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1,	{'learning_rate': 0.002742831255412809 7, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 128, 'n_layers': 1,	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1,

	'n_units_l0': 80, 'dropout_l0': 0.2}	'n_units_l0': 40, 'dropout_l0': 0.2}	'n_units_l0': 70, 'dropout_l0': 0.2}
01995X0012/342B	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0019464388531750198, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
02206X0022/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
02225X0016/S1	{'learning_rate': 0.001300122762 1881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
02347X0022/314	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001986907762154767677, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}
02566X0019/S1	{'learning_rate': 0.008103056755532596, 'optimizer':	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs':

	'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
02617X0009/S1	{'learning_rate': 0.009282156083736918, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.0025030353278011654, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
02636X0009/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
02726X0029/238	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
02923X0007/F	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0019362365596948302, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}

02931X0008/S1	{'learning_rate': 0.0059131273523082, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0027940569610945824, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 176, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
02974X0004/S1	{'learning_rate': 0.0081030567555 32596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0086193397178 05171, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}
03272X0006/PZ	{'learning_rate': 0.0065955504050 89572, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
03276X0009/P	{'learning_rate': 0.0081030567555 32596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0010091326698 40099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
03287X0018/S1	{'learning_rate': 0.0081030567555 32596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1,	{'learning_rate': 0.0087595994048 09302, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1,	{'learning_rate': 0.0015596967176583343, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1,

	'n_units_l0': 40, 'dropout_l0': 0.2}	'n_units_l0': 20, 'dropout_l0': 0.2}	'n_units_l0': 70, 'dropout_l0': 0.2}
03622X0027/PZ	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
04132X0086/PP6	{'learning_rate': 0.009774776777496903, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.005143094276731783, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.002155693385282558, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}
04137X0018/15	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
04398X0002/SONDAG	{'learning_rate': 0.008819753706964301, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.002864586050269265 4, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.007227642053055 42, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
04458X0023/S3	{'learning_rate': 0.002794056961 0945824, 'optimizer':	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs':

	'adam', 'epochs': 150, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00061X0117/PZ1	{'learning_rate': 0.009704538365201885, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.0013001227621881476, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00263X0129/PZASA4	{'learning_rate': 0.007192474099753512, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.001997768076037276 3, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0015635240773203178, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00275X0005/P1	{'learning_rate': 0.009347592099550283, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0010091326698400099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00346X0011/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.004557599941262 12, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.0038243884220289666, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}

00463X0036/H1	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.002221836274003990 5, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}
02603X0009/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}

Table S4: Best hyperparameters for all wavelet models in single station approach

Code	Best Params GRU	Best Params LSTM	Best Params BILSTM
00068X0010/F295	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
00182X0010/P1	{'learning_rate': 0.009534403218946923, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.006595199234314437, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}

00241X0012/P1	{'learning_rate': 0.007096620660484497, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.006875829206466048, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00263X0006/P1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.006389829936455146, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00271X0002/P2	{'learning_rate': 0.009534403218946923, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.009469968207464949, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 128, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
00332X0007/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.007224044366448904, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00471X0095/PZ2013	{'learning_rate': 0.008795055777615794, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 64, 'n_layers': 1,	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1,	{'learning_rate': 0.006040973157822555, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1,

	'n_units_l0': 90, 'dropout_l0': 0.2}	'n_units_l0': 80, 'dropout_l0': 0.2}	'n_units_l0': 80, 'dropout_l0': 0.2}
00572X0010/S1	{'learning_rate': 0.007495659935295373, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00578X0002/S1	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.007045962219824817, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
00583X0005/S1	{'learning_rate': 0.009952652202455213, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
00608X0028/S1	{'learning_rate': 0.009783853471961944, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
00671X0052/S1	{'learning_rate': 0.009873018120325887, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.001300122762181476, 'optimizer': 'adam', 'epochs': 450, 'batch_size':	{'learning_rate': 0.004338726506305758, 'optimizer': 'adam', 'epochs': 350, 'batch_size':

	200, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	208, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00755X0006/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.007764723178214306 4, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00766X0004/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
00773X0002/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
00791X0017/S1	{'learning_rate': 0.009952652202245213, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.008625283569644069, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00794X0021/S1	{'learning_rate': 0.004470922913421275,	{'learning_rate': 0.003146923594339205 7, 'optimizer':	{'learning_rate': 0.001009132669840099, 'optimizer':

	'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00805X0002/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.00455759994126212, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
00817X0145/PZ_SN	{'learning_rate': 0.007096620660484497, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.008596635502376394, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.006086547336704553, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00821X0035/S1	{'learning_rate': 0.009498471575558025, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.006935572081797071, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00847X0043/S1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.009751298348816691, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}

00853X0030/PZ2013	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.004518702848821970 5, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 128, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.008644529555669741, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00862X0005/S1	{'learning_rate': 0.004969445949093749, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.007906626615390122, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.006565484501923348, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
00957X0005/S1	{'learning_rate': 0.001877203523325561, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.007214354848482611, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
00993X0002/S1	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
01003X0008/S1	{'learning_rate': 0.009245886088352598, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 64, 'n_layers': 1,	{'learning_rate': 0.006689047421574244, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 224, 'n_layers': 1,	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1,

	'n_units_l0': 30, 'dropout_l0': 0.2}	'n_units_l0': 90, 'dropout_l0': 0.2}	'n_units_l0': 70, 'dropout_l0': 0.2}
01013X0004/S1	{'learning_rate': 0.0091564675275 1386, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.00447092291342 1275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.00683453354251 3314, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
01024X0058/S1	{'learning_rate': 0.0081030567555 32596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.00447092291342 1275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.00722404436644 8904, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
01031X0023/S1	{'learning_rate': 0.0051614275817 75894, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.00382691428780 26804, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.00608654733670 4553, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
01045X0015/S1	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.00741548593668 1335, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.00100913266984 0099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01046X0010/S1	{'learning_rate': 0.0095344032189 46923, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.00315142526404 0375, 'optimizer': 'adam', 'epochs': 150, 'batch_size':	{'learning_rate': 0.00608654733670 4553, 'optimizer': 'adam', 'epochs': 250, 'batch_size':

	150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	208, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
01053X0058/S1	{'learning_rate': 0.009282156083736918, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.002155693385282558 2, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.002500314424817318, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}
01074X0006/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.007630901178166286, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01086X0011/LS4	{'learning_rate': 0.009704538365201885, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
01116X0138/F1	{'learning_rate': 0.009707480295841904, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.009472945579003304, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}

01192X0043/S1	{'learning_rate': 0.008759599404809302, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.005264284135366972, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.009907298163988354, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
01194X0069/S1	{'learning_rate': 0.004969445949093749, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.006086547336704553, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
01198X0002/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
01245X0010/S1	{'learning_rate': 0.009534403218946923, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.001986907762154767677, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
01252X0011/S1	{'learning_rate': 0.009282156083736918, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 80, 'n_layers': 1,	{'learning_rate': 0.001300122762186147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1,	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1,

	'n_units_l0': 100, 'dropout_l0': 0.2}	'n_units_l0': 70, 'dropout_l0': 0.2}	'n_units_l0': 80, 'dropout_l0': 0.2}
01258X0020/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.007210036192517294, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
01264X0029/S1	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.002419619449357119 7, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
01287X0017/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.009681298865877503, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 192, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
01347X0002/S1	{'learning_rate': 0.009981300982599807, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01381X0070/P25	{'learning_rate': 0.004969445949093749, 'optimizer':	{'learning_rate': 0.007742174649008 56, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs':

	'adam', 'epochs': 250, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	250, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
01461X0012/S1	{'learning_rate': 0.009704538365201885, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.009707480295841904, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.008590069145513327, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}
01473X0087/S1	{'learning_rate': 0.009952652202452213, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.001967305713793621 4, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.007950991288447225, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
01491X0009/S1	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.003826890237851838, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
01516X0004/S1	{'learning_rate': 0.0099891338549917, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.007917876446858305, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
01584X0023/LV3	{'learning_rate': 0.004284251906762034,	{'learning_rate': 0.001300122762188147 6, 'optimizer':	{'learning_rate': 0.008103056755532596, 'optimizer':

	'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
01871X0031/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
01995X0012/342B	{'learning_rate': 0.004358974703068952, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.006806686454144779, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
02206X0022/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
02225X0016/S1	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001300122762188147 6, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.004875095617603179, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}

02347X0022/314	{'learning_rate': 0.008759599404809302, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.002765878983990437, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
02566X0019/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.006794402113345505, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
02617X0009/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.002765878983990437, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
02636X0009/S1	{'learning_rate': 0.009774776777496903, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.007495659935295373, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
02726X0029/238	{'learning_rate': 0.009774776777496903, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 80, 'n_layers': 1,	{'learning_rate': 0.008759599404809302, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1,	{'learning_rate': 0.008819753706964301, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1,

	'n_units_l0': 100, 'dropout_l0': 0.2}	'n_units_l0': 20, 'dropout_l0': 0.2}	'n_units_l0': 20, 'dropout_l0': 0.2}
02923X0007/F	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
02931X0008/S1	{'learning_rate': 0.005143094276731783, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
02974X0004/S1	{'learning_rate': 0.009282156083736918, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.009282156083736918, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
03272X0006/PZ	{'learning_rate': 0.009469968207464649, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
03276X0009/P	{'learning_rate': 0.004228752765456056, 'optimizer': 'adam', 'epochs':	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size':	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size':

	350, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
03287X0018/S1	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0029741064884176733, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.006864998400835032, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
03622X0027/PZ	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0027658789839904367, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.0027658789839904367, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 160, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}
04132X0086/PP6	{'learning_rate': 0.0020822598979197784, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.007214354848482611, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}
04137X0018/15	{'learning_rate': 0.008173382490197835, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 80, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.009707480295841904, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}

04398X0002/SONDAG	{'learning_rate': 0.008759599404809302, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.008819753706964301, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 20, 'dropout_l0': 0.2}	{'learning_rate': 0.009660529800880094, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
04458X0023/S3	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.004122917646627026 5, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 48, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.009707480295841904, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 30, 'dropout_l0': 0.2}
00061X0117/PZ1	{'learning_rate': 0.009534403218946923, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.006875829206466048, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}
00263X0129/PZASA4	{'learning_rate': 0.002204680597211 8027, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 16, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.009751298348816691, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}
00275X0005/P1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1,	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1,	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1,

	'n_units_l0': 40, 'dropout_l0': 0.2}	'n_units_l0': 60, 'dropout_l0': 0.2}	'n_units_l0': 80, 'dropout_l0': 0.2}
00346X0011/S1	{'learning_rate': 0.008103056755532596, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 1, 'n_units_l0': 40, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
00463X0036/H1	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
02603X0009/S1	{'learning_rate': 0.007495659935295373, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.004557599941262 12, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 112, 'n_layers': 1, 'n_units_l0': 100, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}

Table S5: Input shape of different multi-station approaches

Approach	Shape of input data for training (Standalone)	Wavelet
NO	(39952, 48, 7)	(39952, 48, 35)
OHE	(39952, 48, 83)	(39952, 48, 111)

STAT	(39952, 48, 24)	(39952, 48, 52)
STAT_OHE	(39952, 48, 100)	(39952, 48, 128)

Table S6: Hyperparameters of Multi-station Standalone Models

Approach	GRU	LSTM	BiLSTM
NO	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0010091326698400099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
OHE	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}

STAT	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.003487742134799628, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 96, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0029166390295633435, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
STAT_OHE	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0010091326698400099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}

Table S7: Hyperparameters of Multi-station Wavelet models

Approach	GRU	LSTM	BiLSTM
NO	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam',	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam',	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam',

	'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
OHE	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
STAT	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
STAT_OHE	{'learning_rate': 0.001009132669840099,	{'learning_rate': 0.001009132669840099,	{'learning_rate': 0.005269329024971373,

	'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}

Table S8: Hyperparameters of cluster based multi-station standalone models

Approach	GWL Type	GRU	LSTM	BILSTM
Cluster_NO	annual	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.0027940569610945824, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 192, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}

	mixed	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
Cluster_OHE	annual	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	mixed	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
Cluster_Static	annual	{'learning_rate': 0.005269329024971373,	{'learning_rate': 0.001009132669840099,	{'learning_rate': 0.001009132669840099,

		'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.005028014857866192, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 256, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	mixed	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
Cluster_Static_OHE	annual	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size':	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size':	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size':

		240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	mixed	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}

Table S9: Hyperparameters of cluster based multi-station wavelet models

Approach	GWL Type	GRU	LSTM	BILSTM
Cluster_NO	annual	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.004470922913421275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.003151425264040375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}

	mixed	{'learning_rate': 0.0010091326698 40099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0068217773984 60312, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 144, 'n_layers': 1, 'n_units_l0': 50, 'dropout_l0': 0.2}	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
Cluster_OHE	annual	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	mixed	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0010091326698 40099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
Cluster_Static	annual	{'learning_rate': 0.0031514252640 40375,	{'learning_rate': 0.0044709229134 21275,	{'learning_rate': 0.0052693290249 71373,

		'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0050280148578 66192, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 256, 'n_layers': 1, 'n_units_l0': 90, 'dropout_l0': 0.2}	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	mixed	{'learning_rate': 0.0010091326698 40099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
Cluster_Static_OHE	annual	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.0044709229134 21275, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 208, 'n_layers': 1, 'n_units_l0': 60, 'dropout_l0': 0.2}	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	inertial	{'learning_rate': 0.0052693290249 71373, 'optimizer': 'adam', 'epochs': 150, 'batch_size':	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size':	{'learning_rate': 0.0031514252640 40375, 'optimizer': 'adam', 'epochs': 150, 'batch_size':

		240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	64, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}
	mixed	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.005269329024971373, 'optimizer': 'adam', 'epochs': 150, 'batch_size': 240, 'n_layers': 1, 'n_units_l0': 80, 'dropout_l0': 0.2}	{'learning_rate': 0.001009132669840099, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 224, 'n_layers': 1, 'n_units_l0': 70, 'dropout_l0': 0.2}

Table S10: Optimal parameters for standard and MODWT-assisted GRU Models for three types of GWLs (Mixed, inertial and annual) and two different input types (Effective precipitation (PE) or Precipitation and air temperature (PT))

GRU	PT	PTWT	PE	PEWT
Mixed	{'learning_rate': 0.006867727616302 92, 'optimizer': 'adam', 'epochs': 300, 'batch_size': 64, 'n_layers': 2, 'n_units_l0': 80, 'dropout_l0': 0.2,	{'learning_rate': 0.003990606939187562, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 144, 'n_layers': 6, 'n_units_l0': 40, 'dropout_l0': 0.2,	{'learning_rate': 0.006495625254590144, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 192, 'n_layers': 2, 'n_units_l0': 100, 'dropout_l0': 0.2,	{'learning_rate': 0.001240110471 0986497, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 48, 'n_layers': 3, 'n_units_l0': 100, 'dropout_l0': 0.2, 'n_units_l1': 100,

	'n_units_l1': 40, 'dropout_l1': 0.2}	'n_units_l1': 10, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 90, 'dropout_l4': 0.2, 'n_units_l5': 10, 'dropout_l5': 0.2}	'n_units_l1': 80, 'dropout_l1': 0.2	'dropout_l1': 0.2, 'n_units_l2': 50, 'dropout_l2': 0.2}
Inertial	{'learning_rate': 0.0021546271126850858, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 16, 'n_layers': 3, 'n_units_l0': 40,	{'learning_rate': 0.0087837581354567 27, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 32, 'n_layers': 3, 'n_units_l0': 30,	{'learning_rate': 0.0011098849592139203, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 32, 'n_layers': 6, 'n_units_l0': 60,	{'learning_rate': 0.0034877421347996 28, 'optimizer': 'adam', 'epochs': 100, 'batch_size': 96, 'n_layers': 5, 'n_units_l0': 50, 'dropout_l0': 0.2,

	'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2}	'dropout_l0': 0.2, 'n_units_l1': 100, 'dropout_l1': 0.2, 'n_units_l2': 10, 'dropout_l2': 0.2}	'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 90, 'dropout_l2': 0.2, 'n_units_l3': 70, 'dropout_l3': 0.2, 'n_units_l4': 50, 'dropout_l4': 0.2, 'n_units_l5': 100, 'dropout_l5': 0.2}	'n_units_l1': 50, 'dropout_l1': 0.2, 'n_units_l2': 70, 'dropout_l2': 0.2, 'n_units_l3': 60, 'dropout_l3': 0.2, 'n_units_l4': 70, 'dropout_l4': 0.2}
Annual	{'learning_rate': 0.004281617773380174, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 80, 'n_layers': 5,	{'learning_rate': 0.009240879941149954, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 32, 'n_layers': 5,	{'learning_rate': 0.003990606939187562, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 144, 'n_layers': 6,	{'learning_rate': 0.003966724265631709, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 80, 'n_layers': 3, 'n_units_l0': 70,

	'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 30, 'dropout_l1': 0.2, 'n_units_l2': 50, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 80, 'dropout_l4': 0.2}	'n_units_l0': 60, 'dropout_l0': 0.2, 'n_units_l1': 50, 'dropout_l1': 0.2, 'n_units_l2': 10, 'dropout_l2': 0.2, 'n_units_l3': 50, 'dropout_l3': 0.2, 'n_units_l4': 80, 'dropout_l4': 0.2}	'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 10, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 90, 'dropout_l4': 0.2, 'n_units_l5': 10, 'dropout_l5': 0.2}	'dropout_l0': 0.2, 'n_units_l1': 50, 'dropout_l1': 0.2, 'n_units_l2': 10, 'dropout_l2': 0.2}

Table S11: Optimal parameters for standard and MODWT-assisted LSTM Models for three types of GWLs (Mixed, inertial and annual) and two different input types (Effective precipitation (PE) or Precipitation and air temperature (PT))

LSTM	PT	PTWT	PE	PEWT

Mixed	{'learning_rate': 0.003990606939187562, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 144, 'n_layers': 6, 'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 10, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 90, 'dropout_l4': 0.2, 'n_units_l5': 10,	{'learning_rate': 0.001642602063722058, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 16, 'n_layers': 3, 'n_units_l0': 100, 'dropout_l0': 0.2, 'n_units_l1': 100, 'dropout_l1': 0.2, 'n_units_l2': 30, 'dropout_l2': 0.2}	{'learning_rate': 0.0017243790893987354, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 4, 'n_units_l0': 30, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 10, 'dropout_l3': 0.2}	{'learning_rate': 0.0011286448692694375, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 16, 'n_layers': 2, 'n_units_l0': 30, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2}

	'dropout_l5': 0.2}			
Inertial	{'learning_rate': 0.0019279681079150224, 'optimizer': 'adam', 'epochs': 250, 'batch_size': 80, 'n_layers': 4, 'n_units_l0': 70, 'dropout_l0': 0.2, 'n_units_l1': 80, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 10, 'dropout_l3': 0.2}	{'learning_rate': 0.0036725622777068682, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 32, 'n_layers': 5, 'n_units_l0': 60, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 10, 'dropout_l2': 0.2, 'n_units_l3': 60, 'dropout_l3': 0.2, 'n_units_l4': 10, 'dropout_l4': 0.2}	{'learning_rate': 0.0062532462360794 6, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 80, 'n_layers': 3, 'n_units_l0': 30, 'dropout_l0': 0.2, 'n_units_l1': 40, 'dropout_l1': 0.2, 'n_units_l2': 20, 'dropout_l2': 0.2}	{'learning_rate': 0.0039906069391875 62, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 144, 'n_layers': 6, 'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 10, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 90, 'dropout_l4': 0.2, 'n_units_l5': 10, 'dropout_l5': 0.2}

Annual	{'learning_rate': 0.009737236666114474, 'optimizer': 'adam', 'epochs': 500, 'batch_size': 64, 'n_layers': 5, 'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 30, 'dropout_l2': 0.2, 'n_units_l3': 100, 'dropout_l3': 0.2, 'n_units_l4': 100, 'dropout_l4': 0.2}	{'learning_rate': 0.006495625254590144, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 192, 'n_layers': 2, 'n_units_l0': 100, 'dropout_l0': 0.2, 'n_units_l1': 80, 'dropout_l1': 0.2}	{'learning_rate': 0.004312167719228594, 'optimizer': 'adam', 'epochs': 400, 'batch_size': 16, 'n_layers': 2, 'n_units_l0': 50, 'dropout_l0': 0.2, 'n_units_l1': 60, 'dropout_l1': 0.2}	{'learning_rate': 0.001562653587605288, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 3, 'n_units_l0': 90, 'dropout_l0': 0.2, 'n_units_l1': 100, 'dropout_l1': 0.2, 'n_units_l2': 50, 'dropout_l2': 0.2}

Table S12: Optimal parameters for standard and MODWT-assisted BiLSTM Models for three types of GWLs (Mixed, inertial and annual) and two different input types (Effective precipitation (PE) or Precipitation and air temperature (PT))

BiLSTM	PT	PTWT	PE	PEWT
Mixed	{'learning_rate': 0.004212772944331677, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 80, 'n_layers': 3, 'n_units_l0': 70, 'dropout_l0': 0.2, 'n_units_l1': 80, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2}	{'learning_rate': 0.0034309766901050947, 'optimizer': 'adam', 'epochs': 350, 'batch_size': 16, 'n_layers': 4, 'n_units_l0': 70, 'dropout_l0': 0.2, 'n_units_l1': 30, 'dropout_l1': 0.2, 'n_units_l2': 80, 'dropout_l2': 0.2, 'n_units_l3': 50, 'dropout_l3': 0.2}	{'learning_rate': 0.002966383208485377, 'optimizer': 'adam', 'epochs': 200, 'batch_size': 16, 'n_layers': 3, 'n_units_l0': 80, 'dropout_l0': 0.2, 'n_units_l1': 90, 'dropout_l1': 0.2, 'n_units_l2': 70, 'dropout_l2': 0.2}	{'learning_rate': 0.00218172101811288, 'optimizer': 'adam', 'epochs': 450, 'batch_size': 16, 'n_layers': 2, 'n_units_l0': 100, 'dropout_l0': 0.2, 'n_units_l1': 100, 'dropout_l1': 0.2}
Inertial	{'learning_rate': 0.003990606939187562,	{'learning_rate': 0.001314944556129716, 'optimizer': 'adam',	{'learning_rate': 0.002141506081144842,	{'learning_rate': 0.0039906069391875 62, 'optimizer': 'adam',

	'optimizer': 'adam', 'epochs': 450, 'batch_size': 144, 'n_layers': 6, 'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 10, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 90, 'dropout_l4': 0.2, 'n_units_l5': 10, 'dropout_l5': 0.2}	'epochs': 500, 'batch_size': 16, 'n_layers': 6, 'n_units_l0': 60, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 20, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 30, 'dropout_l4': 0.2, 'n_units_l5': 30, 'dropout_l5': 0.2}	'optimizer': 'adam', 'epochs': 400, 'batch_size': 16, 'n_layers': 3, 'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2}	'epochs': 450, 'batch_size': 144, 'n_layers': 6, 'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 10, 'dropout_l1': 0.2, 'n_units_l2': 40, 'dropout_l2': 0.2, 'n_units_l3': 20, 'dropout_l3': 0.2, 'n_units_l4': 90, 'dropout_l4': 0.2, 'n_units_l5': 10, 'dropout_l5': 0.2}
Annual	{'learning_rate': 0.004409980567805262, 'optimizer': 'adam',	{'learning_rate': 0.008515913194855923, 'optimizer': 'adam',	{'learning_rate': 0.004849787847878661,	{'learning_rate': 0.0020512170635433725, 'optimizer': 'adam',

	'epochs': 250, 'batch_size': 208, 'n_layers': 2, 'n_units_l0': 80, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2}	'epochs': 200, 'batch_size': 16, 'n_layers': 2, 'n_units_l0': 100, 'dropout_l0': 0.2, 'n_units_l1': 40, 'dropout_l1': 0.2}	'optimizer': 'adam', 'epochs': 350, 'batch_size': 48, 'n_layers': 3, 'n_units_l0': 10, 'dropout_l0': 0.2, 'n_units_l1': 60, 'dropout_l1': 0.2, 'n_units_l2': 60, 'dropout_l2': 0.2}	'epochs': 400, 'batch_size': 16, 'n_layers': 3, 'n_units_l0': 40, 'dropout_l0': 0.2, 'n_units_l1': 70, 'dropout_l1': 0.2, 'n_units_l2': 20, 'dropout_l2': 0.2}

Deep learning for the simulation, reconstruction & projection of groundwater level variations

Keywords: MODWT, Groundwater, LSTM, GRU, BiLSTM

Résumé

Des simulations précises du niveau des eaux souterraines (GWL) sont indispensables pour générer les reconstructions et les projections servant à analyser les tendances et la variabilité historiques et futures des eaux souterraines à long terme. Dans cette thèse, nous étudions l'utilisation d'approches d'apprentissage profond (DL) pour les simulations, reconstructions et projections du niveau des eaux souterraines, en mettant l'accent sur les questions liées à la représentation de la variabilité à basse fréquence interannuelle à décennale, et en utilisant divers produits de réanalyses climatiques et sorties de GCM. Une approche de pré-traitement par ondelettes assistant les modèles DL a été développée, en particulier à partir de transformée en ondelettes discrète à chevauchement maximal (MODWT) en une étape de décomposition les signaux d'entrée. Les modèles récurrents à mémoire long- et court- terme (LSTM) et leurs développements plus récents (unité récurrente à porte GRU et LSTM bidirectionnels BiLSTM) ont été plus spécifiquement utilisés et évalués, pour développer des approches d'apprentissage à station unique et à stations multiples. Les résultats de l'approche à station unique ont indiqué que les modèles GRU assistés par MODWT permettaient d'extraire des informations à basse fréquence et surpassaient considérablement les modèles « simples » (i.e. sans pré-traitement) dans la simulation des GWL, en particulier pour les GWL de type inertiel. La méthode SHAP a été utilisée pour appréhender l'interprétabilité des résultats des modèles et le fonctionnement des modèles eux-mêmes, mettant ainsi notamment en évidence les caractéristiques d'entrée les plus importantes. Pour les reconstructions GWL à long terme, les modèles DL ont été construits en utilisant les ensembles de données de réanalyse climatique ERA5 et ERA20C du centre européen de prévisions météorologiques à moyen terme (ECMWF), permettant des reconstructions jusqu'en 1940 et 1900, respectivement. Ces modèles basés ont pu capturer avec succès la variabilité multidécennale dans tous les niveaux de nappe reconstruits, un enjeu important en contexte de changement climatique dans la mesure où la variabilité multidécennale peut fortement interférer avec les effets du changement climatique. Plusieurs approches d'apprentissage multi-stations et de clustering ont été utilisées pour les simulations GWL à grande échelle, intégrant des variables climatiques dynamiques et des caractéristiques statiques des aquifères. Les modèles spécifiquement entraînés sur différents types de GWL, regroupés sur la base de leurs propriétés spectrales, ont obtenu des résultats significativement meilleurs que ceux entraînés sur l'ensemble des données. Enfin, un modèle GRU multi-stations entraîné pour chaque type de GWL avec un prétraitement MODWT avec correction des effets de bord (BC-MODWT) a été utilisé pour générer des projections jusqu'en 2100. Les changements futurs indiquent des tendances à la baisse des niveaux et de la variabilité des eaux souterraines, s'intensifiant de SSP2-4.5 à SSP5-8.5, malgré des niveaux des eaux souterraines projetés plus élevés en moyenne par rapport à la période historique dans tous les scénarios. Nous expliquons ce résultat apparemment contre-intuitif par le fait que les niveaux projetés sont systématiquement bien plus élevés en début de période future (jusqu'à ~2050) par rapport à la période historique. Nos résultats indiquent enfin que la variabilité des aquifères de type annuel a augmenté pour tous les scénarios d'émission.

Abstract

Accurate groundwater level (GWL) simulations facilitate reconstructions and projections for analysing historical and future groundwater trends and variability at the decadal scale. In this thesis, we investigate the use of deep learning (DL) approaches for GWL simulations, reconstructions, and projections, with a focus on capturing low-frequency variability and leveraging climate reanalysis and GCM model outputs. A wavelet-assisted DL framework was developed, using the Maximal Overlap Discrete Wavelet Transform (MODWT) as a pre-processing step to decompose input signals. We specifically evaluated advanced DL models, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM), for single-station and multi-station approaches. The single station approach results indicated that MODWT-assisted GRU models allowed for extracting low-frequency information and significantly outperformed standalone models in simulating GWLs, particularly for inertial-type GWL. The Shapley Additive Explanations (SHAP) technique was used to interpret model outputs and highlight important input features. For long-term GWL reconstructions, DL models were trained on ERA5 and ERA20C climate reanalysis datasets, enabling reconstructions up to 1900 and 1940, respectively. These DL-based models were able to capture multi-decadal variability in all reconstructed GWLs. Several multi-station training approaches and clustering were used for large-scale GWL simulations, incorporating dynamic climatic variables and static aquifer characteristics. Models specifically trained on different GWL types, clustered by spectral properties, performed significantly better than those trained on the whole dataset. Finally, A multi-station GRU model trained for each GWL type with boundary-corrected MODWT (BC-MODWT) pre-processing was used to generate projections until 2100. Future changes show decreasing trends in groundwater levels and variability, intensifying from SSP2-4.5 to SSP5-8.5, despite projected groundwater levels being higher on average compared to the historical period in all scenarios. We explain this seemingly counter-intuitive result by the fact that projected levels are systematically much higher at the beginning of the future period (up to ~2050) compared to the historical period. Finally, our results indicate that the variability of annual-type aquifers has increased for all emission scenarios.