CLINT

CLIMATE INTELLIGENCE

Preliminary report on AI-enhanced Climate Services for Extreme Impacts

Deliverable D6.2

May 2024

This project is part of the H2020 Programme supported by the European Union, having received funding from it under grant agreement No 101003876

Programme call:	Building a low-carbon, climate resilient future: climate action in support of the Paris Agreement (H2020-LC-CLA-2018-2019-2020)
Grant Agreement ID:	101003876
Project Title:	CLINT
Partners:	POLIMI (Project Coordinator), CMCC, HEREON, CSIC, SMHI, HKV, E3M, TCDF, DKRZ, IHE, ECMWF, UAH, JLU, OGC, UCM
Work-Package:	WP6: AI-enhanced climate services to foster adaptation at the pan-European scale
Deliverable #:	D6.2
Deliverable Type:	Report
Contractual Date of Delivery:	31 May 2024
Actual Date of Delivery:	31 May 2024
Title of Document:	Preliminary report on AI-enhanced Climate Services for extreme impacts
Responsible partner:	E3M
Author(s):	Kostas Kavvaidas, Kristina Govorukha, Evita Kourtza, Ilias Pechlivanidis, Du Yiheng, Ilaria Cemenzi, Odysseas Vlachopoulos, Elena Xoplaki, Niklas Luther
Content of this report:	The report integrates milestone MS15 and illustrates the preliminary analysis of AI-enhanced CS for European EE impacts for the water, energy, and food sectors.
Availability:	This report is public.

Document revisions		
Author	Revision content	Date
Kristina Govorukha	D6.2 – v00 Initial draft of the deliverable.	16/10/2023
Ilias Pechlivanidis, Du Yiheng, Ilaria Cemenzi, Kostas Kavvaidas, Kristina Govorukha	D6.2 – v01 Revision 01	03/04/2024
Kostas Kavvaidas, Kristina Govorukha, Evita Kourtza	D6.2 – v02 Revision 02	21/04/2024
Kristina Govorukha, Odysseas Vlachopoulos, Niklas Luther, Elena Xoplaki	D6.2 – v03 Revision 03	29/04/2024
Kristina Govorukha	D6.2 – v04 Final formatting and submission for internal revision.	30/04/2024
Du Yiheng, Odysseas Vlachopoulos, Kristina Govorukha	D6.2 – v05 Revision after after the internal review by Étienne Plésiat and Linus Magnusson.	23/05/2024
Kristina Govorukha	D6.2 – v06 Final formatting	28/05/2024
Guido Ascenso, Andrea Castelletti	D6.2 – vF Final check	29/05/2024

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AI	Artificial Intelligence
ANN	Artificial Neural Network
BNN	Bayesian Neural Networks
C3S	Copernicus Climate Change Services
CCS/U	Carbon Capture Storage / Utilization
CRO-SL	Coral Reef Optimization with Substrate Layers
CS	Climate Services
CV	Cross-Validation
DL	Deep Learning
EE	Extreme Events
ELM	Extreme Learning Machines
ETS	Emissions Trading System
GLM	Generalised Linear Model
LSTM	Long-Short-Term-Memory
MARS	Multivariate Adaptive Regression Splines
ML	Machine Learning
MSE	Mean Squared Error
PCA	Principal Components Analysis
QM	Quantile Mapping
ReLU	Rectified Linear Unit Activation Function
RF	Random Forests
RFECV	Recursive Feature Elimination with Cross-Validation
RFNBO	Renewable Fuels of Non-biological Origin
RNN	Recurrent Neural Network
RoR	Run of River
SNN	Stochastic Neural Networks

EXECUTIVE SUMMARY

This deliverable lays out the preliminary results on Artificial Intelligence and Machine Learning (AI/ML) enhanced climate services at the European scale. The analysis focuses on historical, forecasted, and projected climate conditions and discusses the implications for extreme impacts.

In the dynamic realm of climate research, the integration of AI/ML into climate services is transforming our understanding and response strategies to climatic changes impacting the water, energy, and food sectors. This transformation is heavily reliant on the accessibility and integrity of data, as demonstrated by the application of methods and case studies in this deliverable, spanning a broad array of environmental parameters including temperature, river discharge, and precipitation.

Progressing from preliminary findings, the methods and applications detailed in this deliverable underscore the evolving understanding of specific data and climate service needs across vital sectors. Each sector strives to pinpoint the most pertinent climate parameters to integrate into impact models effectively. Moreover, there is a focus on selecting suitable AI/ML methodologies to address the climate data complexity and enhance its integration into modeling frameworks.

Water Sector

For AI-enhanced hydrological modeling, we applied four post-processing methods to runoff simulations produced by the E-HYPE model. Moreover, to better understand the climate service for future projections, we analyzed how changes in climatic factors influence runoff variations in present and future conditions.

As a result of post-processing, the model was significantly improved in terms of both total runoff volumes and extreme flow conditions. Spatial patterns of the model enhancement have been observed, with more accurate predictions occurring in Central Europe, especially under conditions of extreme events. These enhancements in hydrological modeling will be operationalized to provide climate services with more reliable tools for predicting water availability and managing water resources efficiently. In particular, the ability to capture extreme events with higher accuracy is crucial for developing resilient water management strategies in response to climate change.

On the other hand, results of attribution analysis for runoff variations have provided us knowledge of leading drivers for future water resources under different climate change scenarios, therefore deepening our understanding of the outputs from climate projection service.

Energy Sector

To ensure the link between climate-induced data and power demand and hydropower generation profiles, climate data have been processed to feed into the PRIMES-IEM's

modeling framework. The focus of work under this deliverable has been on variations in air temperatures and river discharge under historic and projected climate conditions. This approach employs machine learning techniques to correlate climate data with energy production and consumption metrics.

In preliminary results, we demonstrate an application that links variations in the river discharge and air temperature into the energy modeling framework. The power demand is linked with changes in air temperature, with emphasis on electric-driven heating and cooling demand in European countries. Moreover, river discharge derived by the AI-enhanced E-HYPE model, is linked with hydropower generation. Its availability is projected for future climate scenarios, indicating the variations in available hydropower potential among scenarios and regions. In both cases this will enhance the understanding of how extreme weather events can affect the energy sector.

Improved data representation and modeling capabilities enable better projections for energy demand and supply and ensure resilience and more stable energy supply in response to climate variability.

Food Sector

In this deliverable, we present the preliminary results from the development of an AI/ML-enhanced crop model ECroPS, which includes downscaled and calibrated climate forecasts to assess the impacts of climate extremes on crop yields.

The emulator has shown that concurrent climate extremes have a profound impact on crop yields. The integration of seasonal climate forecasts into the emulator allows improved crop performance prediction under various climate scenarios.

These advancements provide crucial insights into food systems' vulnerability to climate extremes, aiding in the development of more effective agricultural strategies and policies. This is vital for ensuring food security and optimizing crop productivity in the face of global climate change.

The methods and applications outlined in the deliverable represent preliminary findings that lead to a deeper comprehension of the specific needs and the extent of data required across the water, energy, and food sectors. Each sector has endeavored to identify the most relevant climate parameters necessary for effectively integrating climate data into their respective impact models. Additionally, these sectors have focused on selecting appropriate methods, enhanced with AI/ML techniques where possible, to manage the complexity of the available data and improve the accuracy of the overall modelling framework.

1. INTRODUCTION

1.1. CLIMATE SERVICES

Climate services (CS) are increasingly acknowledged as pivotal in facilitating adaptation, mitigation, and disaster risk management in the face of escalating climate change challenges. These services play a crucial role in preparing for and responding to extreme events (EE), which are predicted to grow in frequency and intensity. CS have emerged as a critical tool in the global response to climate change, particularly following the urgency highlighted by the Paris Agreement in 2015. These services transform climate-related data into customized products and tools, essential for informed decision-making in areas like adaptation and mitigation strategies and enable climate-informed decision-making and climate-smart policy and planning.

Climate services cover a broad range of climate-related data across different sectors and across different user groups. Transformation of data and enriching it with user-relevant information into customized products is the main task of CS. Delivering the CS through projections, forecasts, trends, economic analyses, and assessments facilitates climate-informed decision-making and counseling (see the definition of CS, according to the European Commission's Roadmap for Climate Services (2015)). The value of CS lies in its accessibility, understanding of user needs, and provision of scientifically based and credible information to diverse user groups across society and aid to cope with climate variability, climate-related disasters and related social and economic damages. CS contributes significantly by engaging society and institutions directly in information and communication about climate-driven impacts on the agriculture, water, and energy sectors, and helping to develop appropriate adaptation strategies (see goals of the Climate Europe Project 2021 and 2022)[1]. Development and design of CS must take into account that the data and tools provided are not only scientifically robust but also practically usable and tailored to the specific needs of diverse user groups. The scope of users varies by the requirements to the CS, as time resolution and variables, see Figure 1-1.

[1] Climateurope2 is coordinated by the Barcelona Supercomputing Center. The goal of the project is to develop standardization procedures and recommendations for climate services. Project web page: [https://climateurope2.eu/]

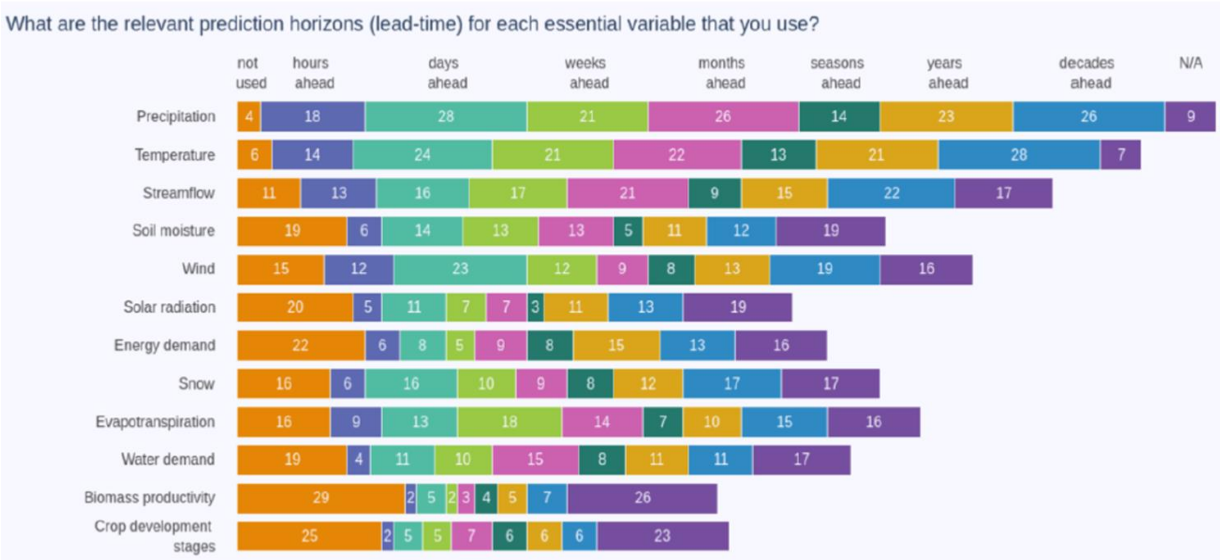


Figure 1-1 Different needs and requirements to the time resolution provided by end-users of CS, based on the CLINT User Survey results, presented in Deliverable 6.1.

The evolution of CS is driven by the integration of advanced climate data and machine learning techniques. The implementation of CS involves sophisticated AI frameworks for analyzing climatological data, thereby enhancing the detection, prediction, and attribution of EEs and their impacts. This synergy enables the provision of accessible, timely, and decision-relevant information, vital for policymakers and end-users across various socio-economic sectors.

The challenge lies in ensuring the accessibility and usability of climate information. Recent advances in modeling capabilities, combined with improved observation tools, have expanded the range of available climate information. Yet, the emphasis is on tailoring this information to fit the specific needs of climate-sensitive sectors and overcoming barriers that limit societal uptake allowing for more standardization, conceptualization, operationalization and evaluation of CS information and data (Weichselgartner & Arheimer, 2019).

CS cover the needs of urban planners, financial investors, and communities to effectively anticipate and adapt to projected climatic trends and variations. Essential to this process is high-quality, reliable data, including but not limited to climate variables like temperature, precipitation, wind patterns, soil moisture levels. This climate data provided to the users of climate services streams into the direct applications as well as to the professional users that need access to long-term historical averages, maps, climate projections and scenarios to elaborate risk and vulnerability analyses, impact assessment studies dedicated for different sectors, see Figure 1-2. Climate data and information products combined with impact assessment models, enrich CS with non-climate information on agricultural production, health trends, distributional effects across affected population, energy and other socio-economic variables.

Figure *1-2 Overview of CS and end-users, flows of information (specific for the work accomplished within WP6)*

Advancements in the availability of climate data, and development of AI/ML techniques gives a positive impact on CS, as stated in Haupt et al. 2021: "ML post-processing can act as a bridge between the physical representation of the atmosphere provided by numerical weather prediction and the decision-making requirements of end-users."

Recent observations of extreme weather events, such as tropical cyclones, tropical nights, heatwaves, droughts, and floods, have heightened the attention of policymakers and the public. The increasing frequency and intensity of these events necessitate an enhanced focus on the forecasting and warning of such extreme weather phenomena and contributing to CS.

Extreme Events (EE) are defined as significant occurrences with substantial societal and economic impacts, spanning critical sectors including water, food, and energy. The identification and forecasting of these events over multiple regions or countries, across trans-boundary river systems demands attention not just at the regional level but also at national and pan-European scales. The IPCC special report defines extreme weather events as "risks/impacts to human health, livelihoods, assets and ecosystems from extreme weather events such as heatwaves, heavy rain, drought and associated wildfires, and coastal flooding" (IPCC 2018, p. 11).

Power and food sectors are exposed to EE and climate variability at all temporal scales and at various spatial scales. The report gives a demonstration of how the enhancements of data availability at different time and spatial scales can improve the impact assessments.

Until recently, applications in the power sector only partially included the impacts of climate variability and EE due to issues such as limited data coverage or spatial resolution (Dubus et al, 2023). Recent dataset developments, notably European Meteorological derived High Resolution RES generation time series for present and future scenarios (EMHIRES) and Renewables Ninja[2], provide time series for wind and solar power in Europe, but lack comprehensive coverage on hydropower and electricity demand. The most recent services, as C3S Energy operational service (C3S-E) (Dubus et al, 2023), provide high resolution and large temporal coverage of data that encompasses all key renewable energy variables (wind, solar PV, hydro, and demand) across a range of timescales. The Destination Earth (DestinE)[3] the undertaking, supported by the European Commission, aims to develop and maintain accurate, interactive and dynamic simulations of the Earth system to better prepare for natural disasters, adapt to climate change and predict the socioeconomic impacts. Empowered by AI and maintaining a pool of observational data and simulation, this initiative develops tools for climate-informed decision making. These improvements in data availability positively contribute to the development of long-term planning and investment strategies across Europe. Power sector will benefit from further enhancement of CS to also include and improve detection and forecasting of EE.

The impacts of weather and climate on the food sector and its relevant spheres, from production to distribution chains and food security, are well discussed in the scientific domain and within stakeholders' ecosystems (Jägermeyr et al., 2021). There is a multitude of studies and reviews around the impacts of climate and climate change especially during the unprecedented recent years of compound events and EEs such as droughts (e.g., Zscheischler & Fischer (2020), Lesk, et al., (2022), Brás, et. al., (2021)). However, the challenge remains around the wider access of interested parties on affordable, accurate, timely and highly resolved characterization, assessment and prediction of food related impacts following EE due to their most severe and mitigation-demanding nature in the socioeconomic sense.

CS that can provide these attributes to its served products can be of great value for the food sector for short- to long-term assessments, utilizing seasonal forecasts, historical climate observations and future projections such as CMIP6. This is greatly reinforced by the available open data streams of Copernicus Climate Change Service (C3S) seasonal forecast service which provide standardized data of ensemble forecasts for probabilistic analyses and the Coupled Model Intercomparison Project, currently in Phase 6 (CMIP6) by WCRP, which allows

[2] EMHIRES (https://data.jrc.ec.europa.eu/collection/id-0055); Ninja (https://www.renewables.ninja/)

[3] Destination Earth (destination-earth.eu)

studying the impacts of future weather and climate to the crop production, under different assumptions and scenarios.

By leveraging advancements in climate data and AI/ML techniques, CS provides a comprehensive understanding of EE, facilitating effective adaptation and mitigation strategies. Thus, the growing value of CS lies in their ability to quantify the impacts of extreme weather events, aiding decision-makers in navigating the challenges posed by such events and shaping resilient responses at both local and wider geographical levels. The current deliverable focuses on improvements to current CS available in water, energy and food sectors. The deliverable demonstrates AI/ML methods and their implementation in the impact assessment models. The preliminary results highlight the achieved improvements for better CS and climate-informed decision making in the respective sectors.

1.2. STRUCTURE AND OBJECTIVE OF THIS DOCUMENT

Objective of this report is to enhance methods and data that can be used in CS. The Deliverable focuses on the European scale and makes use of the global scale climate modeling results. By using selected examples, the report will illustrate how climate data can be used by the integrated models for water, energy and food sectors. We also discuss what enhancements of sector specific models can be implemented to take advantage of combining the available climate data and AI/ML techniques. These illustrations can bridge the gap between the abundance of information provided by CS and informed policy and planning decision making.

The deliverable is structured in 6 chapters. Chapter 1 (current chapter) introduces CS, discusses the main challenges, and outlines the deliverable. Throughout Chapter 2, we provide the description of the impact models for the water, energy, and food sectors and discuss model-specific enhancements. In the following chapters, we will present the preliminary results for AI/ML applications in impact models. Dedicated chapters present AI enhanced climate services for each sector represented by the model: Chapter 3 is dedicated for the water sector, Chapter 4 - the energy sector, and Chapter 5 - the food sector. Following the methodology description, each chapter presents the results and outlines the next steps. Chapter 6 presents the summary of work done in the deliverable, conclusions and moving forward for water, energy, and food sectors.

1.3. CONNECTION TO OTHER WORK PACKAGES

This deliverable is complementary to the findings and methodologies described in the WP2, WP3, but does not necessarily reinforce the ML/AI methods developed in these work packages. The methodologies presented in this deliverable at Pan-European Scale can be streamlined to activities under WP7 for local case studies.

2. IMPACT MODELS DESCRIPTION AND ENHANCEMENTS

2.1. WATER SECTOR – E-HYPE HYDROLOGICAL MODEL

2.1.1. The hydrological model structure

The Hydrological Predictions for the Environment, HYPE, model (Lindström et al., 2010) is an advanced semi-distributed, process-based hydrological model designed for basin-scale analysis. It incorporates conceptual routines to simulate key land and subsurface hydrological processes, including snow and ice dynamics, evaporation, flow mechanisms, soil moisture variations, groundwater movements, and water distribution through various water bodies, driven by parameters related to the landscape's physical characteristics (Figure 2-1, left). The model features the degree-day method for snow processes and accounts for spatial differences through Hydrological Response Units (HRUs) and layers the soil into three sections for detailed water path simulation. Rainfall and snowmelt infiltration are modelled based on soil type, with excess water leading to macropore flow, surface runoff, or overland flow, depending on soil moisture levels. Potential evaporation rates are calculated using the modified Jensen-Haise model (Oudin et al., 2005), with actual evaporation dependent on soil moisture. Runoff and water percolation dynamics are triggered by soil moisture thresholds, with groundwater levels reflecting soil zone saturation.

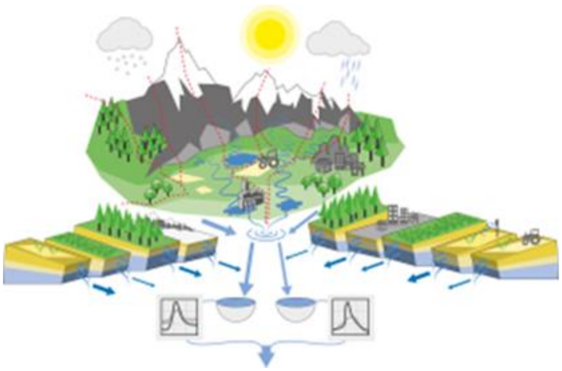



Figure 2-1 (Left) Schematic representation of the processes described in the HYPE model structure. (Right) Domain of the pan-European E-HYPE hydrological model.

The HYPE model routes generated discharge through subbasins, using a river routing routine that simulates flow attenuation and delay. If lakes or reservoirs exist within a subbasin, their flow is managed using rating curves. Lakes, classified by area, accumulate runoff and upstream river flow, with direct precipitation and evaporation accounted for until depletion. Lake outflow is determined by rating curves, with specific regulation for managed lakes and reservoirs, which can include constant or seasonal outflow adjustments. Irrigation is modelled based on crop water demands, using either the FAO-56 method (Allen et al., 1998) or reference levels for submerged crops (e.g. rice), with water sourced from local water bodies or groundwater, adjusted for conveyance losses. Water needed is sourced from rivers, lakes, reservoirs, and groundwater, both within and outside the originating sub-basin, subject to the availability at these sources. After accounting for conveyance losses, the withdrawn water is then used to increase infiltration in the soils requiring irrigation.

2.1.2. The pan-European E-HYPE hydrological setup

The HYPE model is configured on a continental scale, covering the entire pan-European area of 8.8 million km² (Figure 2-1 right). This configuration is known as E-HYPE and divides the region into approximately 35,400 sub-basins, averaging 215 km² each, while it operates on a daily time-step (Hundecha et al., 2016). This setup (version 3.0) utilizes open data, incorporating 8 soil types and 15 land use categories to define up to 75 Hydrological Response Units (HRUs). The model's construction leveraged a variety of open data sources for continental and global analysis, and details can be found in Hundecha et al., (2016). River networks and sub-catchments were outlined using WWF's Hydrosheds data, while HRUs were established from land use and soil information from several databases. CORINE provided land use data, while lakes and reservoirs data were sourced from the GLWD (Global Lakes and Wetlands Database) and GranD (Global Reservoir and Dam) databases. Irrigated areas were identified using GMIA (Global Map of Irrigation Areas) and MIRCA (Monthly Irrigated and Rainfed Crop Areas) datasets, and soil types were derived from the Harmonized World Soil Database (HWSD).

2.1.3. Meteorological forcing input

The E-HYPE hydrological model relies on the HydroGFD v3.0 meteorological dataset (Berg et al., 2021) to simulate soil, lake, and river conditions during both the spin-up phase and historical simulations. Due to the scarcity of daily meteorological data at large scales, this challenge has been addressed by adjusting reanalysis data with monthly global observations to ensure the monthly water balance aligns with observed data, while short-term processes are predicted by meteorological models (Weedon et al., 2014). However, such datasets are typically static, covering only historical periods until updated.

The HydroGFD method, developed at the Swedish Meteorological and Hydrological Institute (SMHI), merges various global data sources to enable near real-time updates of meteorological data (Figure 2-2). This operational system at SMHI integrates global models from the ECMWF and observations for precipitation from the GPCC and temperature from the CRU and CPC. The composition of these data sources varies over time, producing different datasets. The historical HydroGFD dataset (1979-2016) corrects the ERA5 reanalysis with GPCC precipitation data and CRU temperature data. For more recent periods up to three months ago, the HydroGFD-Extended dataset updates observations with the latest from GPCC, CRU, and CPC. The most recent data, from the last two to three months, is provided by the HydroGFD-Near Real-time product, utilizing the ERA5t forecast with a five-day delay. The system outputs data with a spatial resolution of 0.25 degrees (approximately 25 km). It provides data every 3 hours for the HydroGFD-Historical and HydroGFD-Extended products, and every 6 hours for the HydroGFD-NearRealtime product. From this information, daily totals for precipitation and daily averages, as well as minimum and maximum temperatures, are derived.

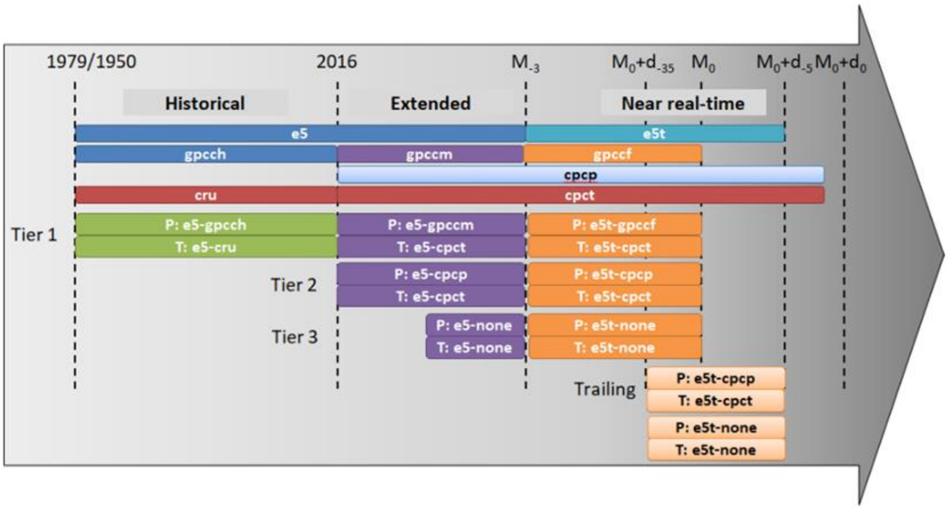


Figure *2-2 The HydroGFD system that creates meteorological data for hydrological modeling, tracking both current and historical weather conditions.*

2.1.4. The hydrological model performance across Europe

The E-HYPE model has been calibrated using in-situ and satellite data for different variables in order to ensure reliable hydrological simulations. This calibration approach following a multi-period, a multi-variable and a multi-criteria framework (Figure 2-3), allows parameter transferability across catchments, facilitating modeling in both gauged and ungauged areas. Calibration and validation of the model employed data from 115 gauging discharge stations and 538 independent stations, respectively. The model shows good performance having streamflow as a target variable across different catchment types with a median Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) around 0.54 and volume errors within ±2%, indicating robustness across varied hydrological settings (see Hundecha et al., 2016). Variation exists in model performance among catchment types, influenced by factors like rainfall patterns, baseflow contributions and land use, with overall tendencies to underestimate streamflow in calibration catchments and overestimate in validation ones. Despite some variability, especially in arid regions where performance drops, the model's general capability is affirmed, with median NSE and volume errors remaining stable across calibration and validation phases. The model's performance could be linked to the geographical gradient, with better accuracy in northern forested catchments compared to southern agricultural ones.

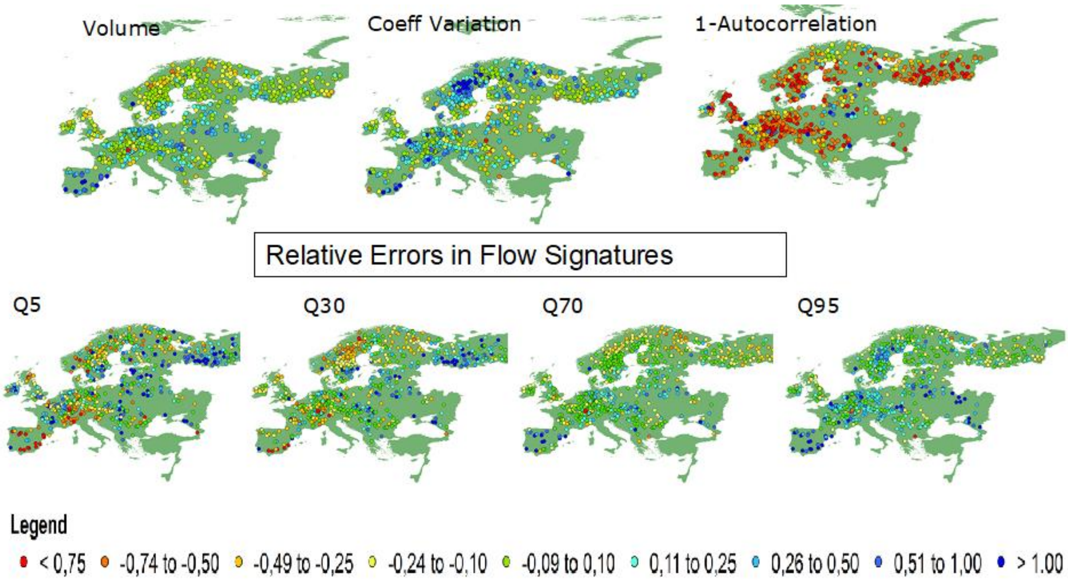


Figure 2-3 E-HYPE model performance using different streamflow signatures (accounting for volume, coefficient of variation, autocorrelation, and 5th, 30th, 70th and 95th percentiles).

Earth observation products were also used in the model evaluation, with particular focus on modelled monthly actual evaporation and the corresponding MODIS dataset in the period 2000-2012 (Figure 2-4; see also Hundecha et al., 2020). The evaluation of E-HYPE against the MODIS actual evaporation indicates that over the entire pan-European domain, the model is capable of representing the long-term actual evaporation and its temporal dynamics. Overall, the results emphasize the E-HYPE adaptability to diverse environmental conditions, while highlighting areas for potential improvement.

Figure 2-4

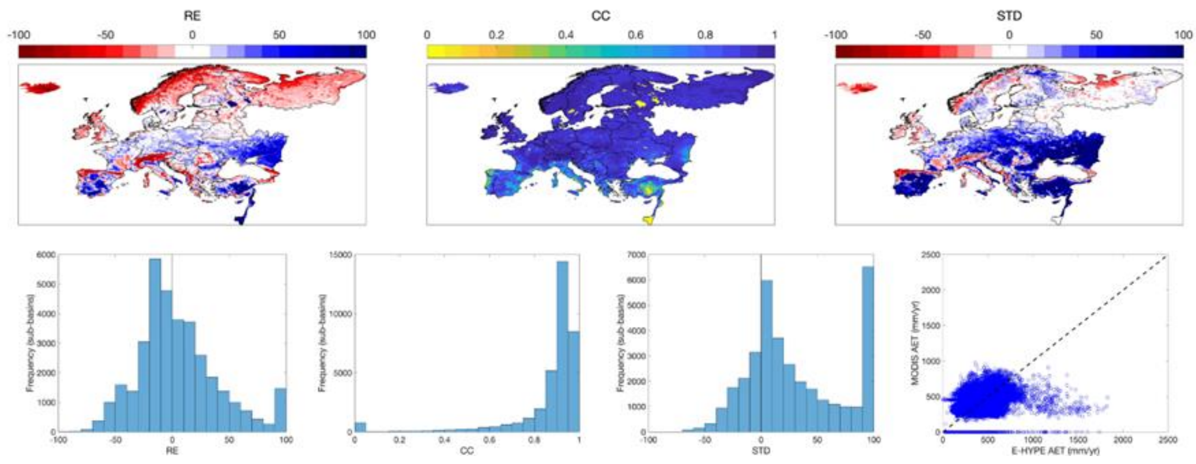


Figure 2-4 E-HYPE performance using MODIS monthly actual evaporation as reference in the period 2000-2012. The performance is assessed in terms of volume (RE; relative error), timing (CC; correlation coefficient) and variability (STD; standard deviation) including also their histograms. Finally, the bottom right subplot compares E-HYPE and MODIS actual evaporation for all 35408 sub-basins.

2.1.5. Hydro-climatic data under current and future conditions

Daily precipitation and air temperature data for the pan-European domain at the catchment scale were sourced from seven bias-adjusted regional climate models, which are members of the Euro-CORDEX model ensemble (Berg et al., 2021). Three representative concentration pathways (RCPs) are used accounting for a low, medium and high emission scenario (RCP2.6, 4.5 and 8.5 respectively). The ensemble includes 3 different emission scenarios, 3 global circulation models (GCMs), 5 regional climate models (RCMs), and two spatial resolutions (Table 2-1). The dataset included a reference, 1971-2000, and three future periods 2011-2040, 2041-2070, and 2071-2100 (Berg et al., 2021), named early, mid and late centuries. Note that the ensemble does not comprehensively cover all sources of uncertainty within the modeling chain, with for instance the MPI-ESM-LR GCM and the CCLM4-8-17 RCM over- and under-represented within the ensemble, respectively. Moreover, there are instances where the same model combinations are utilized, yet the projections are produced using different initial Earth system states of the GCM scenarios (projection ID 6 and 7). Nevertheless, we hypothesize that all available climate models are equally probable to project the future conditions, meaning also that they are free of systematic errors and have acceptable historical performance.

Table 2-1 The ensemble of Euro-CORDEX projections used to produce hydrological impacts. All members are available for the RCP2.6, 4.5 and 8.5 emission scenario.

ID	GCM	RCM	Abbreviation
1	EC-EARTH	CCLM4-8-17	EC-EARTH+CCLM4-8-17
2		RACMO22E	EC-EARTH+RACMO22E
3	HadGEM2-ES	RACMO22E	HadGEM2-ES+RACMO22E
4		RCA4	HadGEM2-ES+RCA4
5	MPI-ESM-LR	RCA4	MPI-ESM-LR+RCA4
6		REMO2009	MPI-ESM-LR+REMO2009 r1
7		REMO2009	MPI-ESM-LR+REMO2009 r2

Note*:* RCMs have a 0.11 degree (about 12.5 Km) horizontal spatial resolution. The GCM model MPI-ESM-LR has two realizations, r1i1p1 (r1) and r2i1p1 (r2), indicating a different initial Earth system's state of the GCM scenarios.

2.2 ENERGY SECTOR – PRIMES MODEL

2.2.1. PRIMES and PRIMES-IEM models for European energy system

The PRIMES model is a large-scale applied energy system model designed to provide long-term energy system projections and system restructuring, both on the demand and supply side on a country-by-country basis and the entire European energy system. The modeling suite is based on interlinked models, including the detailed model of the power generation sector and dedicated models for industrial, transportation, residential, and services sectors to assess changes in the demand for energy. The PRIMES framework incorporates behavioral aspects as well as formulations of discrete choice theory to capture the idiosyncratic behavior of individual actors. The detailed model description is given in E3-Modeling (2018).

The model is formulated based on microeconomic foundations with explicit representation of engineering constraints, technologies (existing and perspective) and vintage of infrastructure characteristic to each sector. The model includes non-linear formulation of potentials by type (e.g., for renewable energy resources, fossil fuel availability and trade, acceptability, compliance with policies etc.) and technology learning. Microeconomic principles (utility maximization, cost minimization and market equilibrium) are combined with detailed engineering constraints and integrate EU and national energy and climate policies.

A series of research programs funded by the European Commission led to the development of the PRIMES model. The PRIMES model has been peer reviewed (European Commission, 2011). The model is used for the EU impact assessments and development of the EU reference scenario. Recent applications include the EU Reference scenario 2020 to project changes in energy, transport and GHG emissions trends to 2050 (European Commission, 2021), which updates the previous version published in 2016 (European Commission, 2016). The PRIMES model is used in the impact assessment of EU climate and energy policies such as the Fit-for-55 policy package, focusing on reaching ambitious targets in 2030 and building the roadmap for climate neutrality in 2050. The European Commission (2020) impact assessment relies on PRIMES scenarios that outline the effects of macroeconomic changes, fuel price changes, and technological changes on EU energy, transport systems, and greenhouse gas (GHG) emissions. The PRIMES model was applied to the assessment of the recent energy crisis caused by the Russian war in Ukraine, developed to assess the necessary measures to reduce the dependence on Russian fossil fuels in the EU energy system, the RepowerEU plan, (European Commission 2022). The latest application in 2024 includes the analysis of the 2040 roadmap, an assessment of the Europe's 2040 climate target plan (European Commission 2024).

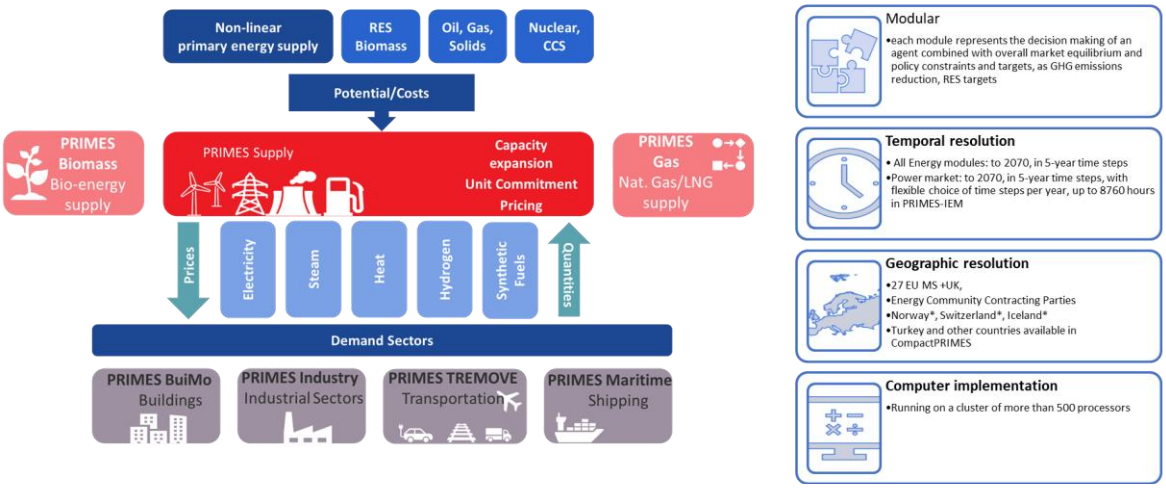


Figure 2-5 PRIMES model and its components.

2.2.2. The pan-European model setup PRIMES-IEM

The PRIMES model covers all energy sectors (Figure 2-5) and ensures the continuity between the available Eurostat statistics for historic periods and projections. The model covers the EU27+UK as well as 10 non-EU countries including the EFTA countries Norway, Switzerland and Iceland. With 5-year steps, the model covers horizons until 2100.

PRIMES - Biomass. The model analyzes the supply of biomass and waste for energy needs. The model quantifies the required transformation capacity to produce bioenergy commodities, associated production costs and end user prices. The Biomass module links the energy system model and the available projections from acknowledged European models such as CAPRI, GLOBIOM, GAINS for the projections on agriculture, forestry and non-CO2 emissions.

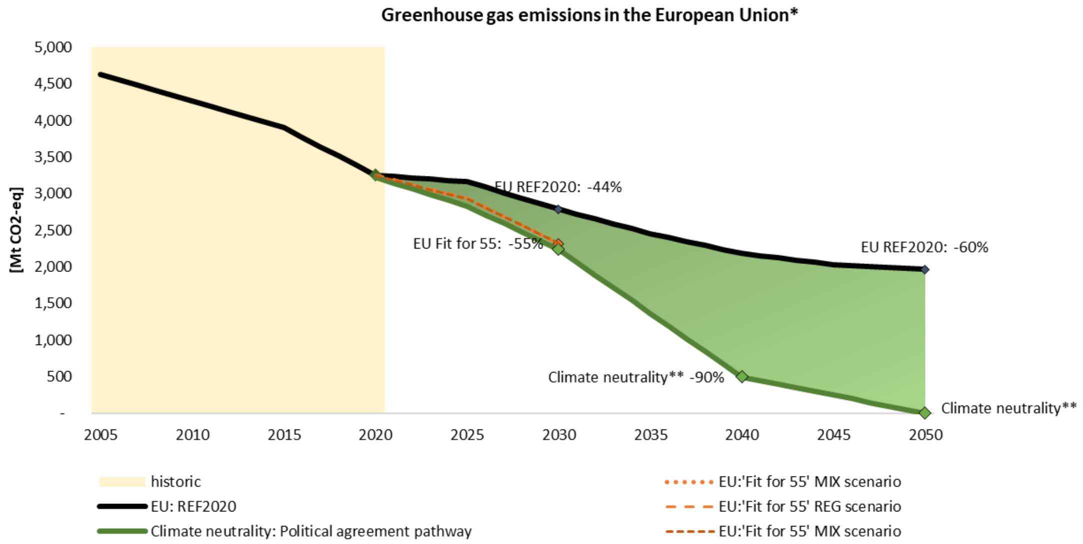
PRIMES - TREMOVE. The dedicated model of the transport system, equipped to analyze changes in the demand for passengers and freight transport. The model includes a specter of transport modes and vehicles to produce the projections of activity levels, vehicle fleet, energy consumption by fuel type and transport mode, emissions. The model produces detailed long-term outlooks per country and EU overall, taking in consideration EU and national policies, standards and characteristics of the vehicle fleet.

PRIMES - Maritime. Within the energy-economy-environment modeling nexus, this model represents the maritime sector. For each of the EU Member-States, the model produces long-term energy and emission projections, as well as integrated projections with PRIMES-TREMOVE and PRIMES-PRIMES.

The current deliverable uses the **PRIMES-IEM** model (see Figure 2-5 on the right). This model has a flexible choice of time-steps per year with up to 8760 single-hour steps. This model has a detailed representation of the European power and heat generation. PRIMES-IEM has a stylized representation of the network for supply and exchanges of power, heat and steam simultaneously. The model produces long-term projections of the power mix needed to satisfy electricity and heat demand in demand sectors: industry, residential, services, and transportation. The model is based on the microeconomic foundations and captures the competition between power generation technologies, combined heat and power (CHP) plants and boilers, self-production and grid-based supply of energy. The model includes a detailed representation of the power generation technologies and their development, including introduction of prospective technologies (e.g., carbon capture storage and use CCS/U) and fuels, including prospective use of synthetic fuels and renewable fuels of non-biological origin (RFNBO). The long-term projections take into account changes in the power mix driven by the policy, economy and technological framework. The PRIMES-IEM model inherits the key characteristics of the PRIMES model (as implementation of policy instruments and climate targets) and is dedicated to the analysis of the electricity market under different climate conditions.

2.2.3. EU policy roadmap to climate neutrality under future climate conditions

Recent applications of PRIMES include the official evaluation of the Fit-for-55 legislative package, aiming to reduce greenhouse gas emissions to -55% below 1990 levels by 2030 European Commission (2021c). The long-term assessment focuses on the energy efficiency measures, pathways for renewable energy development in the demand and supply sectors. The EU long-term climate plan is driven by the overall climate neutrality target in 2050 (Figure 2-6). Improving the long-term strategies for the energy sector with the impact from climate change, contributes to our preparedness and development of adaptation measures necessary to increase the resilience of the European energy system.



Notes:

*Greenhouse gas emissions (Domestic & Intra-EU Maritime and Aviation)

**2030 -55% GHG emission reduction target and 2050 climate neutrality target outlined in the European Green Deal; for 2040 -90% GHG emission reduction target in the political agreement announced in the communication of the EC (COM/2024/63)

Figure 2-6 EU emissions pathways to 2050.

Climate change introduces specific challenges for technological and socio-economic frameworks, particularly highlighted during the summer of 2018 in Europe and globally. This period evidenced heatwaves in North America, Western Europe, and the Caspian Sea region, and extreme rainfall in Southeast Europe and Japan, showcasing the direct impacts of climate variability (Kornhuber et al. 2019). In Europe, prolonged droughts and heatwaves disrupted electricity production, notably in France and Germany, due to reduced output from nuclear and coal power plants. This situation led to spikes in electricity prices, especially in Italy and Spain, in anticipation of rising temperatures. The energy transition towards renewable sources faces tests from extreme weather patterns, such as the February–March 2018 cold spells and the July 2018 heatwaves (Platts 2018, EDO 2018). These incidents underscore the importance of flexibility and balancing capacity in power systems heavily reliant on variable renewable energy. Thermal plants, essential for such capacity, depend on the availability and temperature of cooling water, rendering them vulnerable during hot periods. Thermoelectric power plants could suffer significant capacity reductions, up to 84–86%, by 2040–2069 due to cooling water limitations, emphasizing the need for enhanced risk preparedness in the power sector (van Vliet et al. 2016). While carbon capture technologies offer a pathway to decarbonize electricity generation, they significantly increase water demand for cooling, highlighting the risk of water scarcity during droughts and the need for investment in efficient cooling technologies (Byers et al. 2016).

There are several dimensions of the power generation and power demand that are vulnerable to changing climate conditions, in particular to the increased occurrence of extreme weather events, such as heatwaves, droughts, and cold spells. Some of these dimensions are: changing demand for power for cooling and heating; efficiency losses and cooling needs of the thermal power generation; changing availability of hydropower generation.

Within the ongoing CLINT project, the PRIMES-IEM modeling framework is being extended to include the risks of extreme events on the energy production in Europe, to improve the long-term projections and understand the risks associated with climate change. The work will contribute to policy dialog for adaptation measures necessary to increase the resilience of the European energy system. The PRIMES-IEM modeling framework can be improved with ML techniques to facilitate the use of available climate data to assess impacts of changing climate and rising risks of extreme events over Europe.

2.3 FOOD SECTOR – ECROPS MECHANISTIC CROP GROWTH MODEL

2.3.1. The crop growth model

Crop growth modeling has been a cornerstone of advancing and establishing a spatiotemporal knowledge on yield prediction and impact assessment of climate and human practices on crop production.

Crop growth models like ECroPS (Engine for Crop Parallelizable Simulations) and WOFOST (World Food Studies) simulate and predict crop growth under varying environmental conditions, integrating complex biological, physiological, and environmental processes to simulate the growth and development of crops over time (Toreti et al., 2019).

At the core of models such as ECroPS lies a comprehensive understanding of the physiological mechanisms governing crop growth. By simulating processes such as photosynthesis, respiration, transpiration, and nutrient uptake, these models offer insights into how crops respond to environmental factors such as temperature, rainfall, soil moisture, and nutrient availability.

One of the primary applications of crop growth models like ECroPS is yield prediction. By simulating crop growth across different agro-climatic regions and management practices, these models can forecast potential yields under various scenarios. This capability is important to farmers, policymakers, and agribusinesses, enabling them to make informed decisions regarding crop selection, planting dates, irrigation scheduling, fertilizer application, and pest management among others. Furthermore, yield predictions facilitate risk assessment and management, helping stakeholders mitigate the impacts of adverse weather events or other environmental stressors on agricultural production.

In the context of climate change, crop growth models play a critical role in assessing the vulnerability and adaptive capacity of agricultural systems. By simulating future climate scenarios and their potential impacts on crop growth and yield, these models provide essential information for climate change adaptation and mitigation strategies. In this sense we can evaluate the resilience of different crop varieties, cropping systems, and management practices to projected changes in temperature, precipitation, and extreme weather events. This information empowers stakeholders to develop tailored adaptation measures, such as shifting planting dates, adopting heat-tolerant cultivars, implementing water-saving technologies, and adjusting agronomic practices to optimize resource use efficiency and minimize environmental impacts.

More specifically, ECroPS is in principle a dynamical process based crop model developed to simulate crop growth for a wide range of crops and crop varieties anywhere on the globe, but focuses its functionality and final products by the EC-JRC on the European domain. ECroPS is a new modeling framework developed to deal with the high computational demand of high-resolution regional climate model simulations. The core of ECroPS is the mechanistic crop growth model WOFOST that explains crop growth on the basis of the underlying processes, such as photosynthesis, respiration and how these processes are influenced by environmental conditions. Internally, various modules include phenological development, light interception, gross CO_2 assimilation, growth and maintenance respiration, dry matter partitioning, source and sink limited leaf area development, soil water balance and soil nutrition balance.

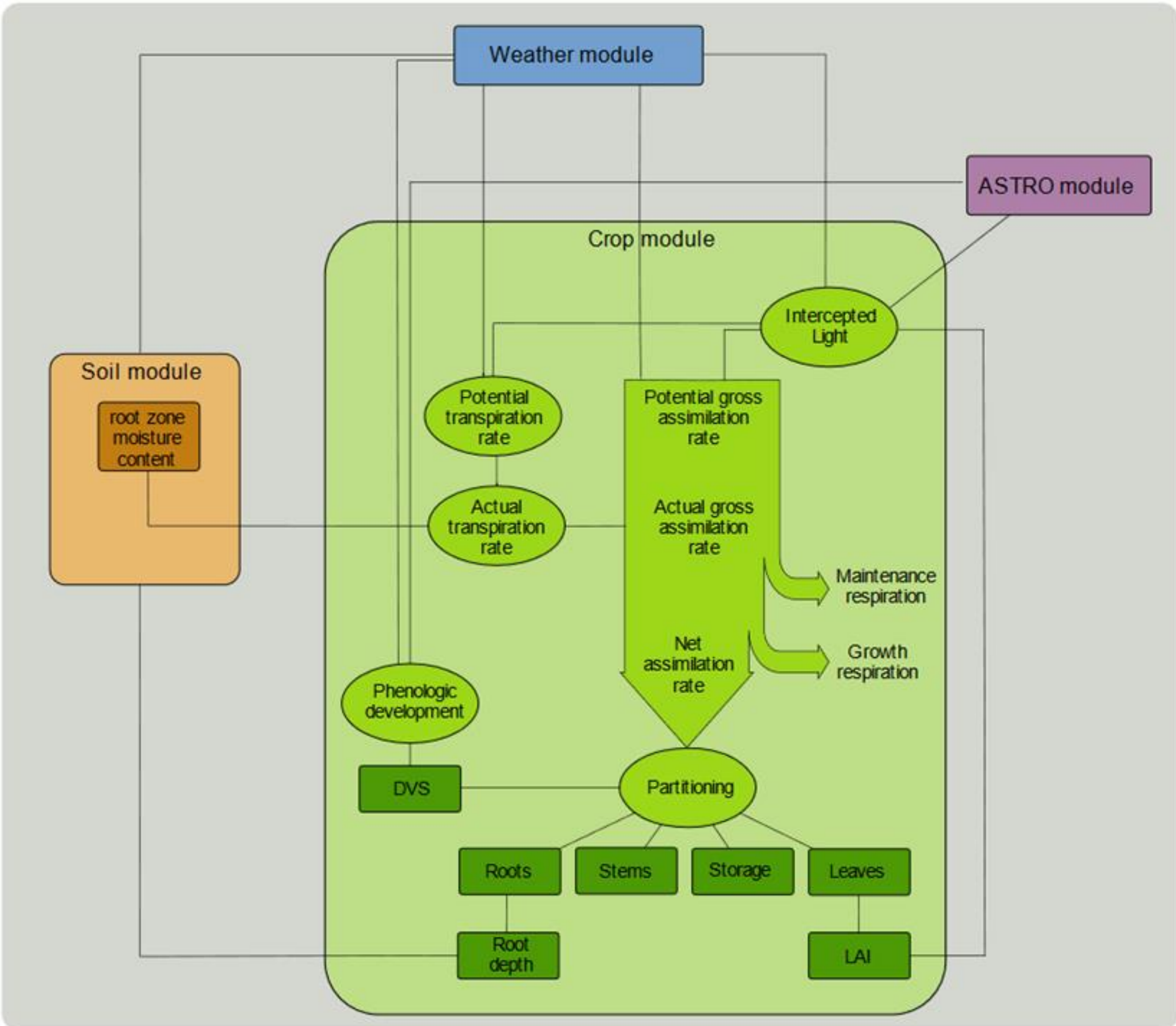


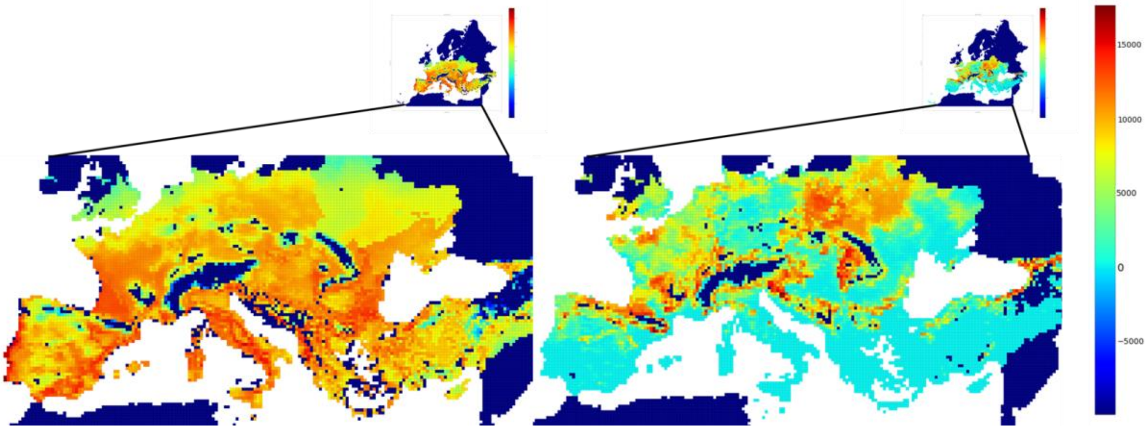
Figure 2-7 Simplified general structure of the dynamic, explanatory crop growth model ECroPS (based on the WOFOST core, see de Wit et al., 2019).

2.3.2. ECroPS inputs and output data streams

As a summary of its key mechanisms, ECroPS integrates:

- Biophysical Processes: ECroPS simulates the biophysical processes involved in crop growth, including photosynthesis, respiration, transpiration, and biomass accumulation. These processes are influenced by environmental factors such as temperature, solar radiation, humidity, and soil moisture.

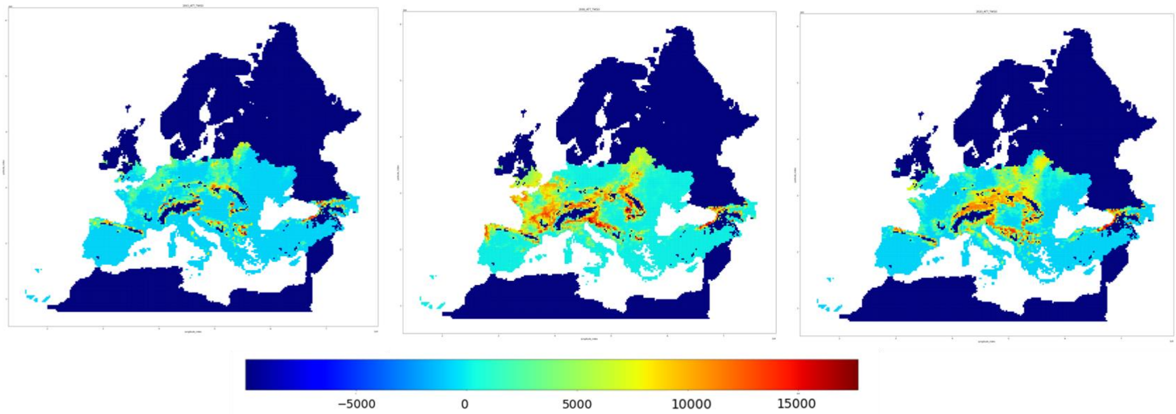
- Crop Development: The model tracks the growth stages of the crop based on thermal time, which is the accumulation of heat units required for crop development. Different crops have specific phenological stages (e.g., emergence, flowering, maturity), and ECroPS simulates these stages based on empirical relationships.
- Crop Management: ECroPS allows for the simulation of various crop management practices, such as planting dates, irrigation, fertilization, and crop residues. These management inputs influence crop growth and yield potential by affecting factors such as water and nutrient availability.
- Water Balance: The model simulates the water balance of the crop system, accounting for rainfall, irrigation, evaporation, and transpiration. Soil water availability is a critical factor affecting crop growth, and ECroPS calculates soil moisture dynamics based on water inputs and losses.
- Nutrient Dynamics: ECroPS considers the dynamics of nutrient uptake and cycling within the soil-plant system. It simulates the availability of essential nutrients (e.g., nitrogen, phosphorus) in the soil and their uptake by the crop, which influences biomass accumulation and yield formation.
- Environmental Stress: The model incorporates factors that can cause environmental stress to crops, such as drought, heat stress, and nutrient deficiencies. These stresses can affect crop growth and yield, and ECroPS simulates their impact based on physiological responses.
- Yield Formation: Based on the simulated crop growth processes and environmental conditions, ECroPS predicts crop yield potential under different scenarios. Yield formation is influenced by factors such as biomass accumulation, crop development, and environmental stress.



Note: Graph's legend is Total Weight of Storage Organs (TWSO) that represents the harvestable product (the yield) of the crop in kg/ha.

Figure 2-8 Example simulations of potential and water-limited configurations for grain maize.

Overall, ECroPS is a comprehensive crop growth model that integrates biophysical processes, crop development, management practices, and environmental factors to simulate crop growth and yield potential. It is used worldwide for agricultural research, decision support, and policy analysis. Furthermore, the impact of highly relevant climate extremes can be simulated, including heat stress and droughts.



Note: Graph's legend is Total Weight of Storage Organs (TWSO) that represents the harvestable product (the yield) of the crop in kg/ha.

Figure 2-9 From left to right: The 2003, 2008 and 2020 heatwaves and droughts are visible through the ECroPS simulations for grain maize.

The crop model is parameterized to account for the spatial variability in crop model parameters across Europe, thus considering different spatial variety distribution for the main crops (Ceglar et al., 2019). The latter is based on pan-European spatial calibration of several crop model parameters relating to variety prevalence in different European growing regions. ECroPS distinguishes three levels of crop production: potential production (determined by crop variety, radiation and temperature), water limited production (water availability limits potential production) and nutrient limited production (in which nutrient availability limits water limited production).

2.3.3. Crop growth modeling challenges and their surrogate potential

Today, within the canvas of crop growth modeling, AI is emerging as a compelling alternative to traditional crop growth models, offering novel approaches for simulating and emulating complex agricultural systems with accuracy and efficiency. While conventional mechanistic models like ECroPS and WOFOST indeed serve as valuable tools for understanding crop-environment interactions, they often struggle for visibility, wider utilization and are rarely encompassed to operational schemes.

AI-based techniques, including ML and deep learning (DL), present a disruptive paradigm for crop modeling by leveraging various streams of data to infer complex relationships and patterns. Unlike mechanistic models that rely on predefined equations and assumptions, AI algorithms learn directly from data, enabling them to uncover nonlinear and dynamic relationships between environmental variables and crop responses.

One of the most important traits of AI-based models is that they offer superior scalability and computational efficiency compared to traditional simulation approaches, making them well-suited for large-scale applications and real-time decision support. By leveraging parallel computing architectures and distributed processing frameworks, AI algorithms can analyze massive datasets and perform complex computations in a fraction of the time required by conventional models, enabling rapid scenario analysis and sensitivity testing.

AI techniques such as emulation and surrogate modeling have emerged as powerful tools for accelerating the calibration, validation, and uncertainty quantification of complex crop models. By training ML/DL algorithms to mimic the behavior of computationally expensive simulation models, researchers can create lightweight and interpretable emulators that replicate model outputs with high fidelity while drastically reducing computational costs. These emulators enable efficient sensitivity analysis, optimization, and ensemble modeling, facilitating robust decision-making under uncertainty.

ECroPS and its internal modular components are highly complex and resource demanding. This is due to the nature of ECroPS, which is a sophisticated mechanistic crop growth model that integrates all the abovementioned modeled biophysical processes related to plants, with multiple weather, soil and crop parametrization inputs and calibration requirements. This particular aspect of the crop growth model poses an important limitation towards a scalable and widely adopted solution. Overall, the usage complexity, special data demanding and the overall configuration of ECroPS, make the integration into existing decision support pipelines outside research environments very difficult, if possible at all. In summary, we aim at operationalizing a crop growth modeling process while lowering the inherent uncertainty that emerges from the multiple weather variables that ECroPS uses. This is achieved by the current AI-based surrogate modeling efforts, a model that will also allow the realizations of a vast number of runs (ensemble runs) of the surrogate model, thus achieving more statistically robust mean outputs and probabilistic interpretations.

2 INVESTIGATION ON AI-ENHANCED CLIMATE SERVICES FOR EXTREME IMPACTS - WATER SECTOR

3.1. HYDROLOGICAL KNOWLEDGE ENHANCEMENT THROUGH AI - BACKGROUND AND OBJECTIVES

3.1.1. Hybrid hydrological modeling to enhance performance at local scale

Hydrological modeling has significantly advanced our understanding of the water cycle, offering insights into the flow, allocation, and quality of water resources (Guse et al., 2021; Yang et al., 2021). This progress has enabled both scientists and policy-makers with an improved understanding of water's dynamics and its interaction with both climate and environmental parameters (Barendrecht et al., 2019; Botzen et al., 2009; Di Baldassarre et al., 2019). Such an improved understanding is important for effective water resource management, optimizing water usage, ensuring supplies, and enhancing resilience against natural disasters, like floods and droughts. It enables the formulation of accurate risk assessments and the development of targeted mitigation strategies. The application of catchment-scale hydrological models has been widely focusing on understanding and predicting water movement, storage, and quality within specific, often small drainage areas, allowing for detailed analysis of local water resources and environmental impacts (Pechlivanidis et al., 2011). However, many river systems across the world cover a large domain, are transboundary, and even interconnected through water reallocation or groundwater interactions. In such cases, large-scale hydrological models (LSHM) have been preferred to extend the hydrological predictions over broader geographical areas, including entire river basins, continents, or the global scale, to address regional to global water cycle dynamics and interactions with climate change and human activities (Pechlivanidis and Arheimer, 2015).

Deploying LSHMs on national to global scales introduces a set of challenges. These include uncertainties in model structural and parameter identification, which can degrade model accuracy and lead to gaps in our understanding of water cycle dynamics. Variability in hydrological responses, driven by diverse climatic conditions, soil types, topographical features, and human activities, such as irrigation practices and reservoir management, complicates the modeling process further, especially in regions with insufficient gauging stations, where traditional data gathering methods fall short. This scarcity of data complicates the calibration and verification of LSHMs, which are crucial steps for ensuring model reliability and robustness. Moreover, the absence of a comprehensive meteorological dataset that can accurately reflect regional weather patterns, especially precipitation, introduces significant errors/uncertainties in hydrological predictions.

Machine learning (ML) and statistical methods can be used to assist hydrological modeling by analyzing complex patterns within hydrological data, thereby enhancing the accuracy and reliability of water cycle simulations (Bézenac et al., 2019; Geer, 2021; Kraft et al., 2022; Moradkhani et al., 2005; Xu and Liang, 2021). By employing advanced ML or statistical algorithms (i.e. neural networks, decision trees, ensemble learning, quantile mapping and regression models) hydrologists can now capture the nonlinear relationships between various hydrological processes and environmental factors more effectively (AlDahoul et al., 2023; Hauswirth et al., 2021; Papacharalampous et al., 2022). This has led to significant improvements in predicting water distribution, flow, and quality across different landscapes and climatic conditions. Particularly, ML methods are useful at handling large datasets, allowing for the extraction of data information, correlations between variables and the refinement of model outputs to closely align with observations. The integration of these ML and statistical methods into hydrological modeling and/or hybrid combination of these two model families allows for more advanced approaches in understanding and predicting the impacts of extreme events (Slater et al. 2023).

In line with the above challenges, post-processing techniques have emerged as important tools for refining hydrological and meteorological model outputs, including both statistical and ML-based techniques. Statistical methods, like quantile mapping, adjust and downscale outputs to align with observed data, while ML techniques, such as neural networks and ensemble learning, explore complex data patterns to improve prediction accuracy (Enayati et al., 2020; Liu et al., 2022; Papacharalampous and Tyralis, 2022). These advanced post-processing tools have shown their capability of reducing uncertainties and increasing forecast accuracy. However, the effectiveness of these post-processing ML methods can be compromised by issues such as overfitting and the "black box" nature of some ML models, which questions their applicability for decision-making. Similarly, traditional statistical methods may lack the flexibility to capture complex, nonlinear relationships within the data, leading to less accurate predictions. Additionally, both statistical and ML methods often require a large amount of high-quality historical data, which may not be available in all regions. Despite their limitations, there are potential ways of improving their performance, for example, to take account of unique characteristics of each local catchment by considering topography, soil type, vegetation, and climate (Du et al., 2023; Pechlivanidis et al., 2020). By integrating local regimes into post-processing strategies, hydrological modeling takes a significant step forward, into more refined, reliable, and regionally relevant predictions.

Within CLINT, the objective is to enhance the quality of streamflow simulations derived from large-scale hydrological models at the catchment and regional level, thereby enhancing their applicability for local decision-making processes. Specifically, we investigate two scientific questions: (1) *Can the outputs from process-based hydrological models be post-processed using statistical/machine learning models at the local scale across various hydrological regimes?* (2) *Can this hybrid hydrological modeling improve the performance at the local scale?* To address these questions, we apply a comprehensive post-processing approach to the streamflow outputs of the pan-European E-HYPE process-based hydrological model, using a combination of four statistical and machine learning techniques. This approach is tested against a substantial dataset covering approximately 2000 streamflow gauging stations, which represent a wide range of hydrological gradients in Europe.

3.1.2. Attributing runoff changes to climatic drivers in present and under future conditions

Various studies have investigated the hydro-meteorological changes across the European domain, reporting significant shifts in runoff patterns due to climatic change. The Intergovernmental Panel on Climate Change (IPCC, 2022) refers to pronounced streamflow variations due to rising temperatures and evolving precipitation regimes. The results specifically highlight the Mediterranean region's vulnerability, with the projections indicating drier conditions with diminished annual precipitation (Schneider et al., 2013; Yeste et al., 2021), whereas the boreal climate zone is anticipated to experience runoff increments due to increased precipitation and temperature (Donnelly et al., 2017). The temperate zone's flow regimes are expected to exhibit varying degrees of impact, from minimal in oceanic to substantial in continental areas. Furthermore, the incidence and severity of hydrological extremes, such as floods (Thober et al., 2018) and droughts (Gu et al., 2023), are likely to increase alongside seasonal streamflow changes.

Empirical analyses, including those by Stahl et al. (2012), have observed increased trends in annual streamflow within western and northern Europe, contrasting with downward trends in the south and parts of eastern Europe over the period 1962-2004, signifying seasonal adjustments in hydrological patterns. Gudmundsson et al. (2019) identified significant streamflow increases in Central Europe from 1960 to 2000, with notable annual streamflow variations in Northern Europe during the last three decades (1971-2010). Regional assessments in Central Europe have spotlighted divergent seasonal runoff changes, particularly influenced by infrastructure developments and snow cover alterations in recent decades (Rottler et al., 2020). Overall, the existing research has advanced our knowledge of the interplay between climatic fluctuations and hydrological responses, yet gaps remain in identifying the exact impacts of distinct climatic variables. The complexity of disentangling climate-induced changes from those driven by anthropogenic actions, such as land-use transformation and water management practices, remains a key challenge, highlighting the necessity for refined hydrological attribution techniques.

Various methodologies have been deployed for attributing observed hydrological changes to climate change, each with inherent strengths and limitations. Statistical approaches and hydrological modeling have been instrumental in correlating observed changes to climatic drivers, especially in catchments that are not affected by human activities (Hannaford et al., 2013; Hundecha and Merz, 2012). Nonetheless, the inherent uncertainties in hydrological simulations necessitate cautious interpretation when attributing streamflow changes to climate change or human disturbances (Tang et al., 2022).

The Budyko framework offers a conceptual approach for estimating the hydrological sensitivity to changes in precipitation and potential evapotranspiration within a catchment, facilitating the delineation of climate change contributions based on the water balance equation (Liu et al., 2021). The framework helps in explaining how climate and other factors influence long-term average runoff and evaporation in river basins. By delineating the relationship between climate variables and hydrological responses, the Budyko framework facilitates the attribution of changes in runoff to variations in precipitation and evapotranspiration. This has been instrumental in assessing the impacts of climate change on water resources, enabling hydrologists to predict how alterations in climate patterns may affect water availability and distribution. Its simplicity and the minimal data requirements have made it a foundational tool in hydrological studies, particularly in analyzing the effects of climatic variability and change on hydrological regimes across diverse geographical locations. Up to this date, the application of the framework was done on single or few river systems, and although such studies provide detailed insights, continental-scale analyses yield a broader perspective on transboundary climate and hydrological dynamics, albeit with a compromise on localized resolution, thereby enabling the identification of overarching hydrological trends through comparative analyses.

Within CLINT, the aim is to elucidate the connection between European runoff variations and climatic alterations, including associated uncertainties, both in the current and future conditions. We aim to answer three questions: (1) *How will runoff, precipitation and potential evapotranspiration change under future conditions?* (2) *What is the sensitivity of runoff changes to precipitation and potential evapotranspiration across Europe?* and (3) *How do precipitation and evapotranspiration changes relatively contribute to runoff changes under future conditions?* To address these questions, we take advantage of the Euro-CORDEX data and assess the runoff changes in the three future periods and the climatic elasticity (see Section 2.1.5). We then use the theoretical Budyko framework to quantify and attribute runoff changes to precipitation and evapotranspiration under different future periods and emission scenarios.

3.2. METHODOLOGY – HYBRID HYDROLOGICAL MODELING TO ENHANCE PERFORMANCE AT LOCAL SCALE

3.2.1. Description of the methods

To enhance hydrological model performance at local scales, several statistical and machine learning based methods are adopted to post-process the modeled streamflow to the local observations. Here we briefly explain the four methods, including two statistical (generalized linear model and quantile mapping) and two machine learning (random forest and long short-term memory) methods. Detailed information of the machine learning methods can be found in Deliverable 2.2.

Generalized Linear Model (GLM): GLM is a statistical method which extends linear regression to accommodate non-normal distributions of the error terms (Madsen and Thyregod, 2010). It allows for the inclusion of different types of predictor variables and the modeling of response variables that follow distributions other than the normal, such as Gaussian, to provide a flexible framework for understanding the relationships between variables.

Quantile Mapping (QM): QM is a statistical technique used for bias correction by adjusting the distribution of one variable to match the target variable distribution, therefore it effectively corrects systematic biases in model outputs (Gudmundsson et al., 2012). The technique involves comparing the quantiles of the input data with those of the target data and applying corrections to align them. This method is particularly useful for improving the accuracy of hydrological predictions, as it ensures that the corrected model output maintains the statistical properties of the observed data across the entire distribution. The tricubic spline method is adopted here to allow for a smooth adjustment of the cumulative distribution functions, thereby providing a refined approach to addressing biases in both the center and tails of the distribution.

Random Forest (RF): RF is an ensemble learning method that builds multiple decision trees and merges their predictions to improve accuracy and control over-fitting (Pham et al., 2021). By employing a multitude of decision trees, each trained on random subsets of the data and features, RF provides a robust predictive framework that can handle high-dimensional spaces and non-linear relationships without requiring extensive parameter tuning.

Long Short-Term Memory (LSTM) model: LSTM model is a type of recurrent neural network designed to capture long-term dependencies in sequential data, an essential feature for modeling hydrological processes (Kratzert et al., 2018). In our study, with a configuration that includes three layers with a varying number of cells (100-50-20), LSTM models are trained for learning from the temporal patterns in runoff data over a 3-day lookback period. Its mechanisms for adding or removing information to a cell state enables it to learn from the temporal dynamics in the data.

Next, in order to better understand the performance results after the post-processing and link the performance to the local physiographic characteristics, we apply the classification and regression trees ML-based method.

Classification And Regression Trees (CART): CART is a non-parametric decision tree learning technique that models the prediction of a target variable by recursively partitioning the data set and fitting a simple model within each partition (Breiman et al., 2017). In this study, CART is used to identify the most important predictors of model performance and to model the complex, non-linear relationships between them. The algorithm splits the data into subsets based on the values of the input features that result in the largest reduction in heterogeneity of the target variable. This process continues until further splitting does not significantly improve the model's accuracy or until predefined stopping criteria are met, such as a minimum number of observations in each leaf of the tree. To avoid overfitting, the technique of pruning is used by removing branches that have little to no contribution to the model's predictive power, aiming to find the optimal balance between the tree's complexity and its accuracy on a validation set.

3.2.2. Application: post-processing for hydrological services

A significant challenge in enhancing continental and global hydro-climate services lies in the insufficient integration of local knowledge and datasets from end-users, necessitating advanced post-processing techniques. Recent advancements have demonstrated the efficacy of statistical and machine learning (ML) methodologies in elevating the accuracy and reliability of hydrological models, ensuring outputs more accurately reflect specific local hydrological dynamics (Slater et al., 2023). This investigation, therefore, explores four post-processing strategies aimed at refining streamflow predictions, with the objective being to enhance the precision of volume estimates and the characterization of hydrological extremes, thereby bridging the gap between generic model outputs and localized hydrological realities.

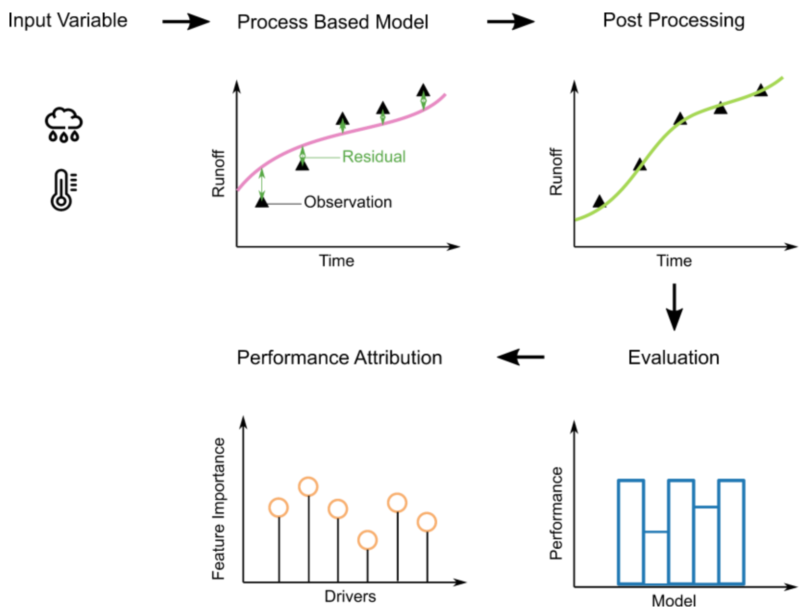


Figure *2-1 Framework for post-processing hydrological model outputs using local observations.*

In Figure 2-1, we present the architecture of our post-processing framework. This framework is benchmarked against a process-based model E-HYPE that is driven by forcing data, i.e. temperature and precipitation, to produce hydrological simulations across the pan European domain. Despite the efficiency of this model, discrepancies (referred to as residuals) persist between the simulated outputs and observational data. During the post-processing phase, specific algorithms (i.e. statistical or machine learning based) are employed to bridge this gap, thereby minimizing the residuals. In our experiment, two statistical methods (GLM and QM) and two machine learning based methods (RF and LSTM) are employed. The performance of these post-processors is evaluated from various perspectives using selected metrics which are introduced in 3.2.2. Furthermore, we investigate the significance of features by linking the performance of the post-processors with potential influencing factors, including climatology, topography, human impact, and hydrological regimes (Section 3.2.3).

Simulated runoff from E-HYPE was obtained for the period 1961-2023. Observations were collected in the pan European domain from various data sources, including GRDC, EWA, SMHI, BHDC, Spanish and Italian authorities. Detailed information can be found in Hundecha et al 2016. Around 2000 stations were selected based on data availability with at least 10 years of data (Figure 2-2), which is divided into training and testing periods with 80%/20% split. The figure illustrates comprehensive spatial coverage of the stations across the entire study domain, with a higher concentration in Central Europe and relatively fewer stations in the southern (e.g. Spain) and the eastern part of the region.

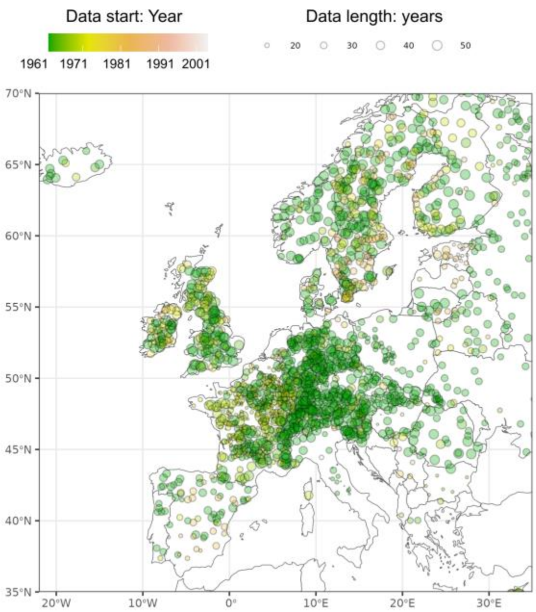


Figure 2-2 Data availability: stations and length of the observation data.

3.2.2. Evaluation: improvement of post processing framework

To evaluate the added value of the post processing algorithms, three evaluation metrics were used to assess the volume, high and low streamflows.

Mean Absolute Error (MAE): MAE is a measure of errors between paired observations expressing the same phenomenon. It calculates the average magnitude of errors in a set of predictions, without considering their direction. MAE is a linear score which means that all individual differences are weighted equally in the average. The Scaled Mean Absolute Error (SMAE) serves to adjust the MAE in relation to the average runoff observed at each station, thus facilitating the comparison of MAE values across stations that have different runoff magnitudes.

$$MAE = \frac{\sum_{t=1}^{T} \left| y_o{}^t - y_m{}^t \right|}{T},$$

$$SMAE = \frac{MAE}{\underline{y_o}},$$

Nash-Sutcliffe Efficiency (NSE): NSE is a normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("signal"). NSE values range from -∞ to 1, where 1 indicates perfect model prediction. NSE is especially used in hydrological modeling to assess how well a model can replicate observed outcomes.

$$NSE = 1 - \frac{\sum_{t=1}^{T} (y_o{}^t - y_m{}^t)^2}{\sum_{t=1}^{T} (y_o{}^t - \underline{y_o})^2},$$

Logarithmic Nash-Sutcliffe Efficiency (logNSE): The logNSE modifies the traditional NSE to emphasize the performance of a model in predicting low flow conditions. By applying the logarithm to both observed and predicted values before calculating efficiency, logNSE is particularly sensitive to differences in low flow predictions, making it valuable for evaluating hydrological models where capturing low flows is critical.

$$log\ NSE = 1 - \frac{\sum_{t=1}^{T} (log(y_o{}^t) - log(y_m{}^t))^2}{\sum_{t=1}^{T} (log(y_o{}^t) - log(\underline{y_o}))^2}$$

Improvement at each station is denoted by calculation of skills, which quantifies the performance of post-processing methods relative to raw simulations, where negative skill values indicate deterioration, and positive values denote improvements. A skill value approaching 1 signifies a greater enhancement in predictive performance, highlighting the effectiveness of the post-processing techniques in refining hydrological forecasts.

$$Skill = \frac{Score_{pp} - Score_{raw}}{Score_{perfect} - Score_{raw}}$$

3.2.3. Detection: drivers of post processing efficacy

To improve climate services through a deeper understanding of the hydrological model process, assessing model performance and its potential drivers is crucial. Machine learning techniques have shown their capability of uncovering relationships between influencing factors and the target variables, especially when these relationships are non-linear and complex. In our study, we use CART analysis to identify key drivers that affect the performance of the E-HYPE model across Europe, with a focus on general model performance on volume and both high and low flows.

CART is an algorithm that classifies the space defined by the input descriptors (i.e., physiographic, hydrological, and climatic) based on the output variable (e.g., performance metric of the E-HYPE simulation). By splitting data into subsets based on certain criteria, it builds a binary tree structure that represents decision paths. The method also provides information on the probabilities of different output groups at each “leaf”. In this study, we chose to divide the model performance into four groups by quartiles. A terminal “leaf” exists at the end of each branch of the “tree,” where the probability of belonging to any of the four output groups was inspected. The original trees were pruned to avoid overfitting. We next calculated the descriptors' importance by summing changes in the probability of splitting on every descriptor and dividing the sum by the number of branch nodes (Pechlivanidis et al 2020). This importance score was then standardized, spanning from 0 to 100 for comparability.

The association between model performance and potential drivers were investigated using the method above, by calculating the feature importance of each potential driver. The considered potential drivers were listed in Table 2-1, including topography, climate, human impact and hydrological regimes. Some drivers are highly interdependent and therefore introduce uncertainty to CART analysis. Here the highly interdependent drivers (Pearson correlation coefficient > 0.6) were removed, and therefore 8 potential drivers were kept for CART analysis, as shown by bond font in Table 2-1.

Table 2-1 Potential drivers considered in this study, including topography, climate, human impact and hydrological regimes.

No.	Name	Abbreviation	Unit
1	Precipitation	Prec	mm
2	Temperature	Temp	°C
3	Snow depth	Snow	cm
4	Actual evapotranspiration	AET	mm
5	Potential evapotranspiration	PET	mm
6	Dryness index	PET/Prec	--
7	Evaporative index	AET/Prec	--
8	Upstream Area	Area	km^2
9	Elevation	Elev	m
10	Relief ratio	Relief	--
11	Slope	Slope	%
12	Degree of Regulation	DoR	%
13	Hydrological Clusters	Cluster	--

By applying the concept of feature importance, a comprehensive ranking index, as defined by Jiang et al. (2015), enables the evaluation and comparison of potential drivers' influence across various models. This ranking index (RI) is mathematically expressed as,

$$RI = 1 - \frac{1}{nm}\sum_{i=1}^{n} rank_i,$$

m represents the total number of potential drivers, which in this study is eight, and n denotes the number of models, set at five (raw model and four post processing methods) for this analysis. $rank_i$ indicates the assigned rank of each potential driver, with 1 being the most critical and 8 the least. Thus, an RI value approaching 1 signals a more accurate and effective simulation outcome.

With this rank index, the analysis identifies the three most influential drivers across both the unprocessed model and the various post-processing methods. This approach can reveal the underlying drivers of the model performance and provide information on where post-processing methods can significantly refine the model's accuracy. These insights are invaluable for the strategic application of statistical and AI techniques to enhance model accuracy and contribute to the reliability of climate services.

3.3. METHODOLOGY – ATTRIBUTING RUNOFF CHANGES TO CLIMATIC DRIVERS IN PRESENT AND UNDER FUTURE CONDITIONS

3.3.1. Enhanced attribution of extreme events to climate change

The Budyko framework (Figure 2-3) was adopted to attribute future runoff changes to precipitation and potential evaporation as primary climatic factors of these changes. The Budyko framework explains the division of precipitation (P) into evapotranspiration (E) and runoff (Q) by balancing the water supply from the atmosphere (P) and the water demand by the atmosphere (potential evaporation, E0). In this framework the theoretical equation of the water-energy balance expressing the changes in E due to changes in P and E0 for a given catchment and deviations from the original Budyko relation is combined with the water balance equation governing runoff changes in long-term periods to obtain the runoff changes (ΔQ) as a function of the climate drivers. In this approach changes in catchment storage were assumed negligible compared to changes in P, E, and Q as the changes are considered to occur between the long-term period of 30-years typical of hydroclimatic applications. The attribution of runoff alterations to climate changes was calculated as the relative contribution of the influencing climate factor to runoff changes (δp, δE0), i.e. the ratio between the contribution of each climatic factor (ΔQ_p, ΔQ_{E0}) and the total, $\Delta Q_p+\Delta Q_{E0}$, climatic contribution (Liu et al., 2017):

$$\delta_p = \frac{\Delta Q_p}{\Delta Q_p + \Delta Q_{E0}};\ \delta_{E0} = \frac{\Delta Q_{E0}}{\Delta Q_p + \Delta Q_{E0}}$$

The contribution of each climate factor on runoff changes in the early (2011-2040), middle (2041-2070), and late (2071-2100) centuries was obtained by multiplying the elasticity coefficient (εp, εE0) by the relative change of the climatic factor (ΔP, ΔE0) to the reference period (1971-2000).

The elasticity coefficients express the sensitivity of runoff to changes in the climatic drivers and represent the changes (increase/decrease) in runoff response to a 1% increase in precipitation and potential evaporation (Roderick and Farquhar's, 2011).

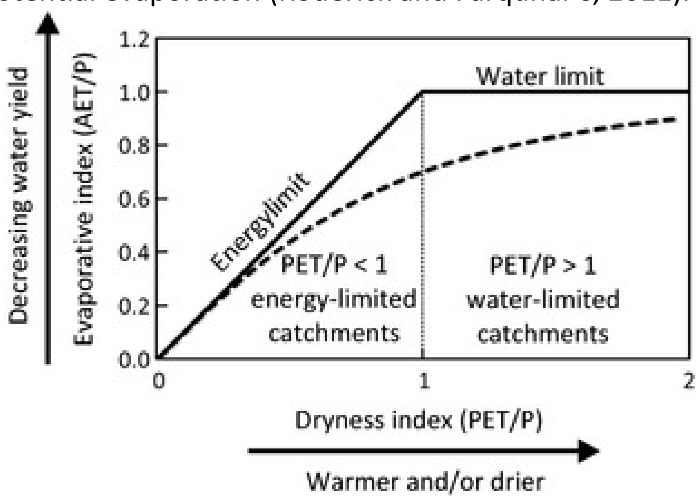


Figure *2-3 Conceptualized illustration of the Budyko diagram (here PET and AET are the potential and actual evapotranspiration, respectively).*

3.4. RESULTS – HYBRID HYDROLOGICAL MODELING TO ENHANCE PERFORMANCE AT LOCAL SCALE

3.4.1. Tailoring hydrological services to local conditions through post processors

With three metrics, we assess the performance of post-processing methods at different stations. Figure 2-4 presents results from an example station, where we show the time series of the raw and post processed simulations, and corresponding metrics. The analysis of four distinct post-processing methodologies reveals an overall enhancement of the initial simulation's performance, specifically regarding volume metrics and the accuracy of high flow extremes. This improvement is quantitatively supported by a reduction in MAE and an increase in NSE, indicating a more accurate representation of hydrological dynamics by the post-processed models compared to the raw simulations. However, the examination of low flow predictions reveals a special pattern for the GLM, where it achieved a marginally lower logNSE compared to the raw simulation data. This slight decrease in logNSE for GLM indicates a reduction in model performance when predicting low flow conditions. Such a discrepancy underscores the challenge of enhancing model accuracy across the full scale of hydrological conditions. This detailed analysis, on one hand, proves the general success of post-processing techniques in refining hydrological forecasts, and on the other hand, also emphasizes the importance of method selection based on the specific hydrological process being modelled.

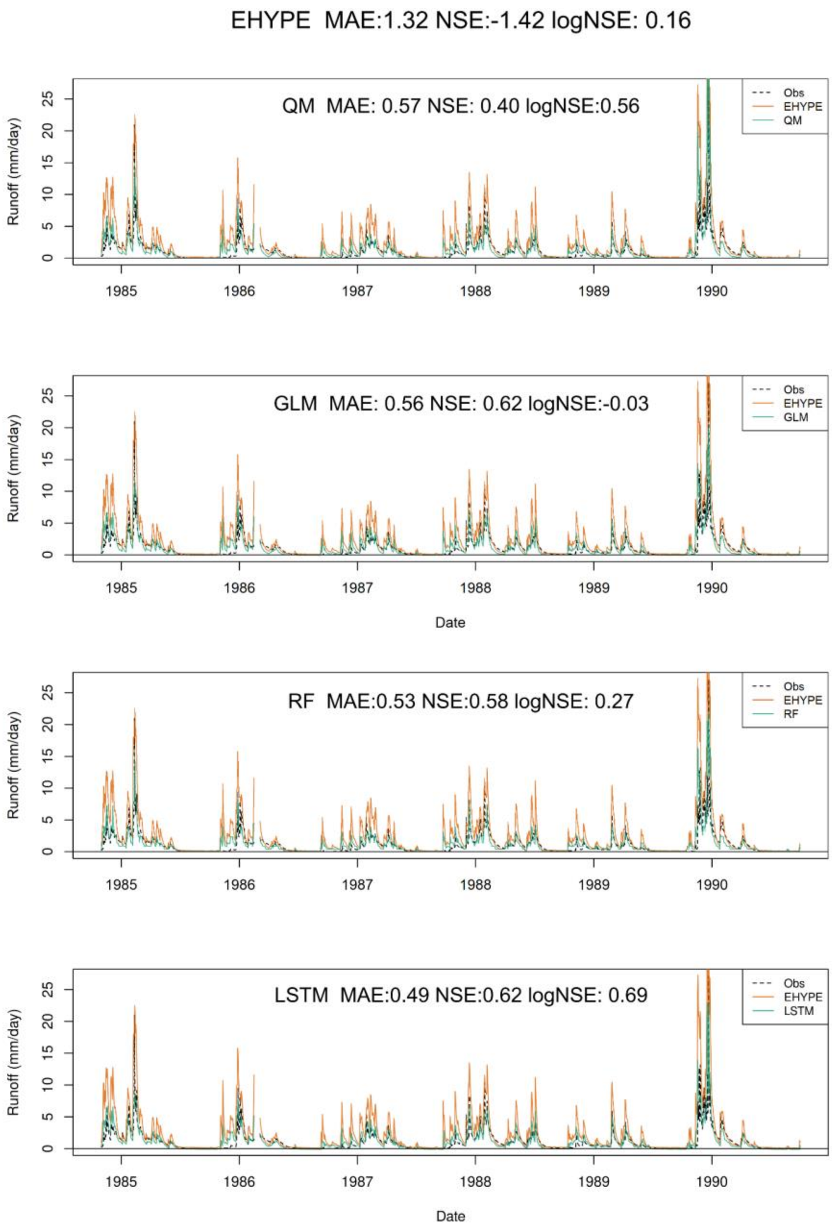


Figure *2-4 Example of post-processing (Spain).*

As illustrated in Figure 2-5, an extensive analysis that extends to all stations within the study domain. This figure plots the cumulative distribution of three evaluative metrics for both unprocessed and post-processed time series, allowing for a comparative analysis of their overall performance. From the visual representation, it is evident that all four post-processing approaches enhance model performance across the three key streamflow characteristics: total volume, and the extremes of high and low flows. This enhancement is indicated by the distributions' shifts towards optimal values, 0 for the SMAE and 1 for both the NSE and the logNSE. Such shifts in the cumulative distributions suggest that, in terms of total volume, the four methods show comparable adequacy across various performance levels. However, distinctions among the methods become more obvious when focusing on extreme flow conditions. Machine learning methods outperform statistical ones within the 'fair' performance category (ranging from 0.2 to 0.5), indicating a superior capability in handling both high and low flow extremes. The patterns are even more noticeable in the 'very poor' and 'unsatisfactory' performance categories (below 0.2), where the QM method shows relatively the weakest performance for both high and low extremes. This detailed analysis underscores the capabilities of post-processing techniques in enhancing hydrological model accuracy, particularly highlighting the strengths and limitations of ML versus statistical methods in predicting extreme flow conditions.

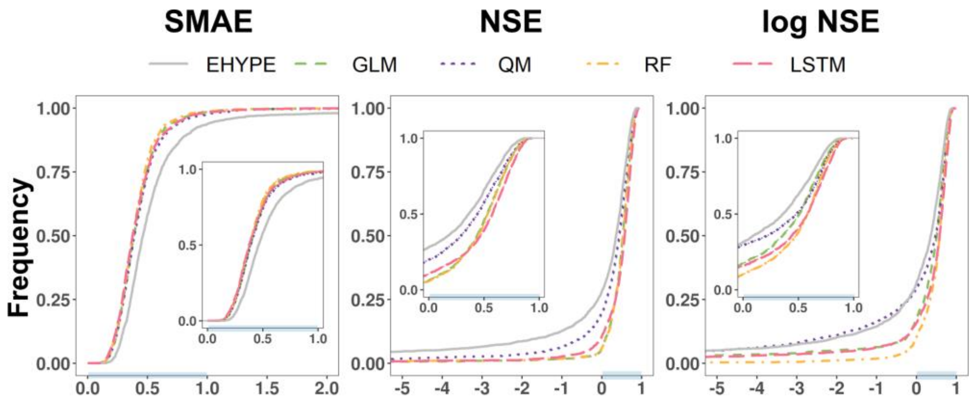


Figure 2-5 E-HYPE performance and post-processed performance for three metrics (SMAE, NSE and log NSE) accounting for errors in volume, and high and low extremes.

Based on the metrics presented in Figure 2-5, skill scores were calculated for each station to investigate the comparative added value provided by various post-processing methods over raw simulations. These calculations are depicted in Figure 2-6, where skills are color-coded as: green-blue shades indicate higher skills, yellow means lower skills, and grey denotes stations where no improvement in skill was found. A consistent pattern emerges across the different post-processing techniques, with stations in Central Europe generally showing higher skills. This trend is particularly notable for extreme values, where an increase in skill levels of both NSE and logNSE, relative to the SMAE, indicates that post-processing significantly improves model performance at both high and low extremes.

Spatial variations of the improvement across different post processing methods are evident, as shown by the performance in the UK, where the QM method shows limited improvement, with many stations showing no skill enhancement. In contrast, two machine learning methods reveal considerable skill improvements in this region. This distinction may be attributed to the machine learning algorithms' superior capability to detect complex and nonlinear relationships within the dataset, which is less pronounced in traditional statistical approaches.

When comparing these outcomes to the baseline performance of raw model simulations, it is observed that stations with suboptimal raw performance are precisely those where significant improvements are obtained through post-processing. This observation underscores the post-processing methods' effectiveness in enhancing model accuracy, particularly in areas initially having lower predictive quality.

Overall, the analysis suggests that there is no universally superior model; each post-processing method presents varying degrees of skill across different spatial locations and according to different evaluation metrics. This variability highlights the importance of selecting the appropriate post-processing technique based on specific regional characteristics and the particular aspects of hydrological behavior being modeled.

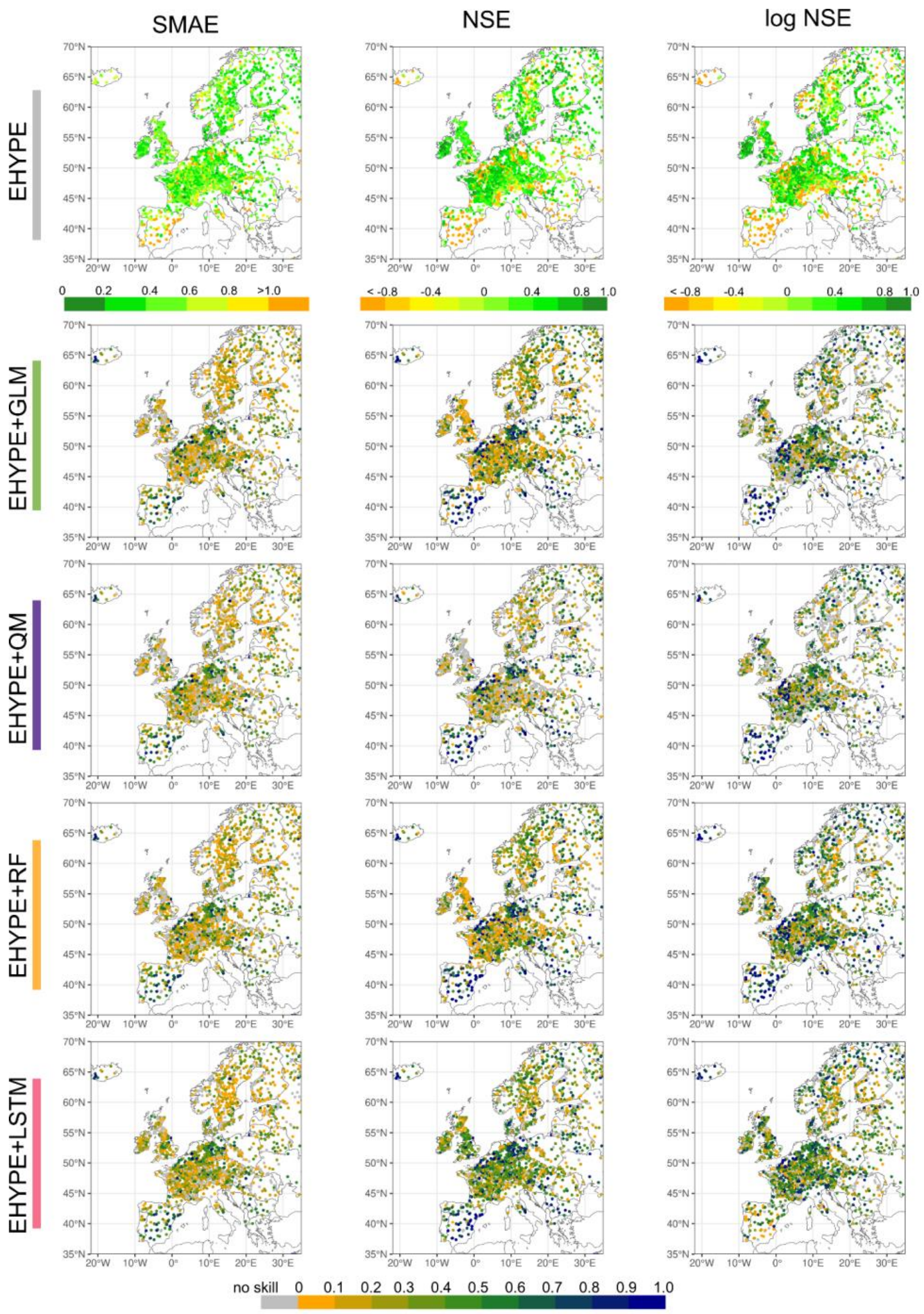


Figure 2-6 Skills achieved from post processing: EHYPE raw simulation metrics, and skills for four methods.

3.4.2. Detection of drivers of post processing performance

Building on the insights gained from evaluating performance, we further examined the potential drivers of the performance of hydrological services, with a particular focus on both volume and extremes. Through CART analysis, we explored the importance of various factors by considering them as inputs, with performance metrics and skill scores serving as the targets. Figure 2-7 illustrates the importance of each feature for the raw model and each post-processing approach, using SMAE as a representative metric, alongside the comprehensive rank index highlighting the three most influential drivers.

In Figure 2-8, we graphically represent the relationship between the performance metric/skill for a specific model and its most critical drivers. This analysis is exclusively conducted for the model where a given driver is identified as having the highest significance. For instance, the hydrological cluster gained highest significance in the raw E-HYPE simulation among the hybrid approaches, accounting for over 40% importance, then the SMAE of the raw E-HYPE model is decomposed and plotted according to these clusters. This allows us to observe the variations across different clusters, including their distribution and median values, offering a deeper understanding of how each cluster contributes to the overall performance. Similar plots are also drawn for mean temperature and mean precipitation, where a clear trend is observed between the skill and the driver. This approach not only highlights the critical drivers affecting hydrological service performance but also allows a targeted analysis of how these drivers influence specific models.

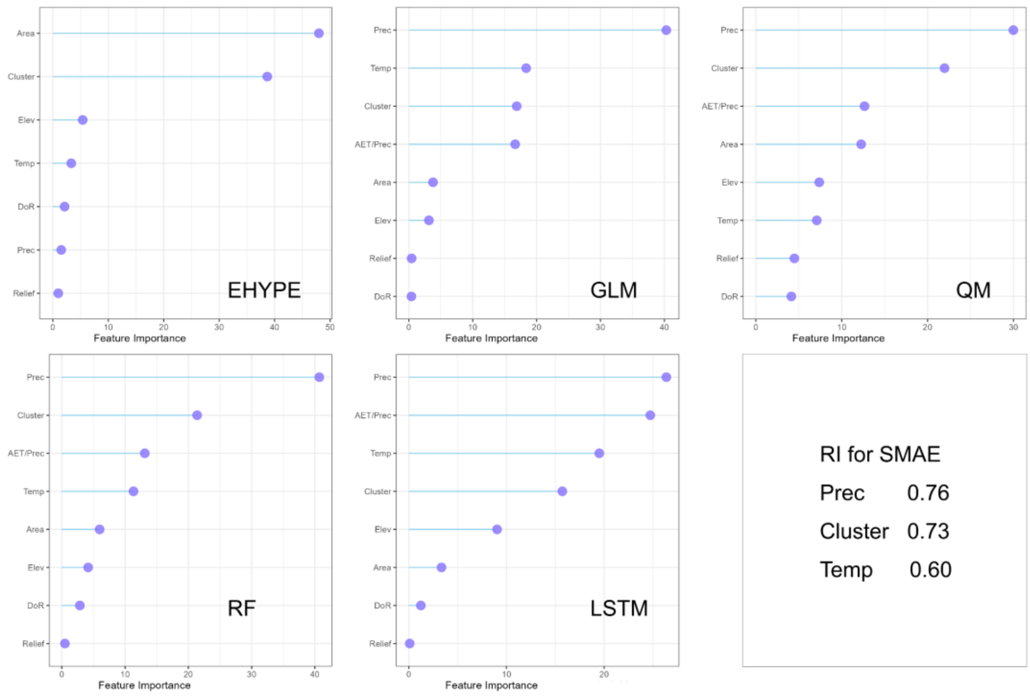


Figure *2-7 Feature importance of local physiographic drivers to the performance of hydrological simulation in terms of the volume (denoted by SMAE).*

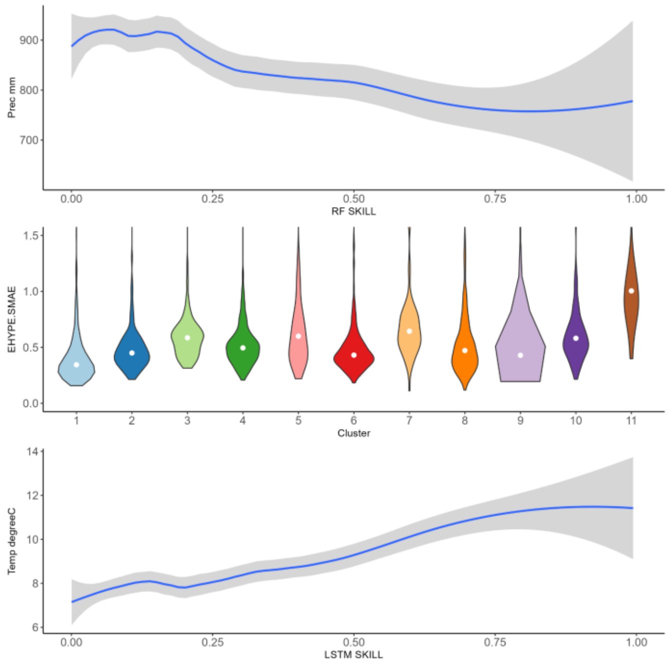


Figure 2-8 Examples of SMAE change corresponding to the leading drivers.

Similarly, the same analysis extends to include both NSE and logNSE, identifying the three most impactful drivers for each metric. For NSE, mean precipitation, mean temperature, and hydrological clusters emerge as the top drivers; while for logNSE, the leading factors are hydrological clusters, elevation, and mean precipitation, each with its unique ranking across the different metrics.

The recurring presence of the hydrological cluster as one of the leading drivers in both volume and extremes underscores its important role in understanding the model performance. Recognizing such key factors is essential for refining hydrological models, as it directs attention to the elements that most significantly impact the accuracy and reliability of water cycle simulations. Through this analysis, we gain deeper insights into the mechanisms driving model performance, allowing targeted improvements in hydrological services.

3.4.3. Summary of results

In this section, we applied four post-processing methods to runoff simulations generated by the E-HYPE model and evaluated their performance using three different metrics while also the spatial distribution of the skills. Moreover, we have extended the investigation in order to identify the primary factors that improve the performance for each post-processing method. The key findings are:

- The analysis reveals a notable improvement by post-processing the raw simulations, in terms of both total volume and high and low extremes. This is evidenced by a decrease in SMAE and an increase in NSE and logNSE, which suggests that post-processed models provide a more accurate representation of hydrological dynamics than the raw simulations.

- Across the different post-processing techniques, a similar spatial pattern of skill improvements is observed, showing higher skills in stations located in central Europe. This pattern is enhanced in the context of extreme events, which also indicates the added value from post processing methods on high and low streamflow extremes.
- Key drivers were identified for influencing the model performance after post-processing and these are: mean precipitation, mean temperature, hydrological clusters and elevation. Each driver ranked differently across the various metrics, indicating their different impacts on model performance. Notably, the recurrent identification of hydrological clusters as a significant factor for both volume and extremes emphasizes its importance in refining model accuracy, in terms of volume and extremes.

3.5. RESULTS - ATTRIBUTING RUNOFF CHANGES TO CLIMATIC DRIVERS IN PRESENT AND UNDER FUTURE CONDITIONS

3.5.1. Hydro-climatic changes under future conditions

Figure 2-9 illustrates the ensemble precipitation changes over Europe in the early, middle and late centuries for the low (RCP 2.6), medium (RCP 4.5) and high (RCP 8.5) emission scenarios. Northern Europe experienced an increase in precipitation under low emission scenarios, specifically in North Eastern Europe and Scandinavia, in contrast to the reference period (Figure 2-9a-c). On the other hand, in the early and middle century, Southern Europe encountered a decrease in precipitation, particularly in Portugal, Spain, Greece, and Cyprus, displaying a different spatial behavior. This difference was more pronounced under the medium (Figure 2-9d-f) and high emission scenarios (Figure 2-9h-g) in the present and future periods in comparison to the low emission scenario (Figure 2-9a-c).

Based on the findings presented in Figure 2-10, it is evident that there has been a notable rise in potential evaporation in Europe when compared to the reference period. The increase in potential evaporation was particularly prominent in the North European and Alps regions for all three periods under the low emission scenario (Figure 2-10a-c). Furthermore, the potential evaporation continued to increase in these areas under medium and high-emission scenarios. In Central and Southern Europe, potential evaporation showed a significant increase observed in the middle and late century under medium (Figure 2-10e,f) and high emission scenarios (Figure 2-10h,i).

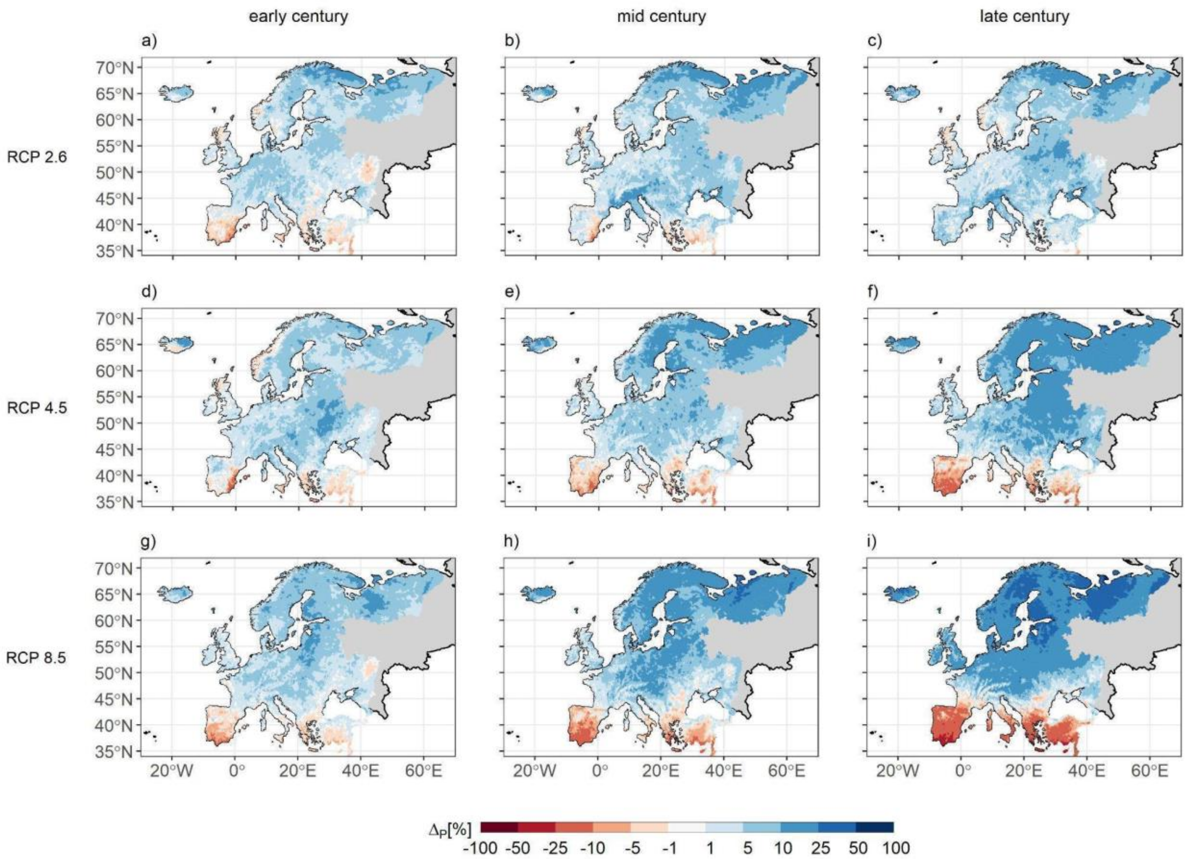


Figure *2-9 Model ensemble median of precipitation changes for the RCP 2.6 (a,b,c), RCP 4.5 (d,e,f) and RCP 8.5 (g,h,i) emission scenarios in the early (a,d,g), mid (b,e,h) and late (c,f,i) century relative to the historical period (1971-2000) across Europe.*

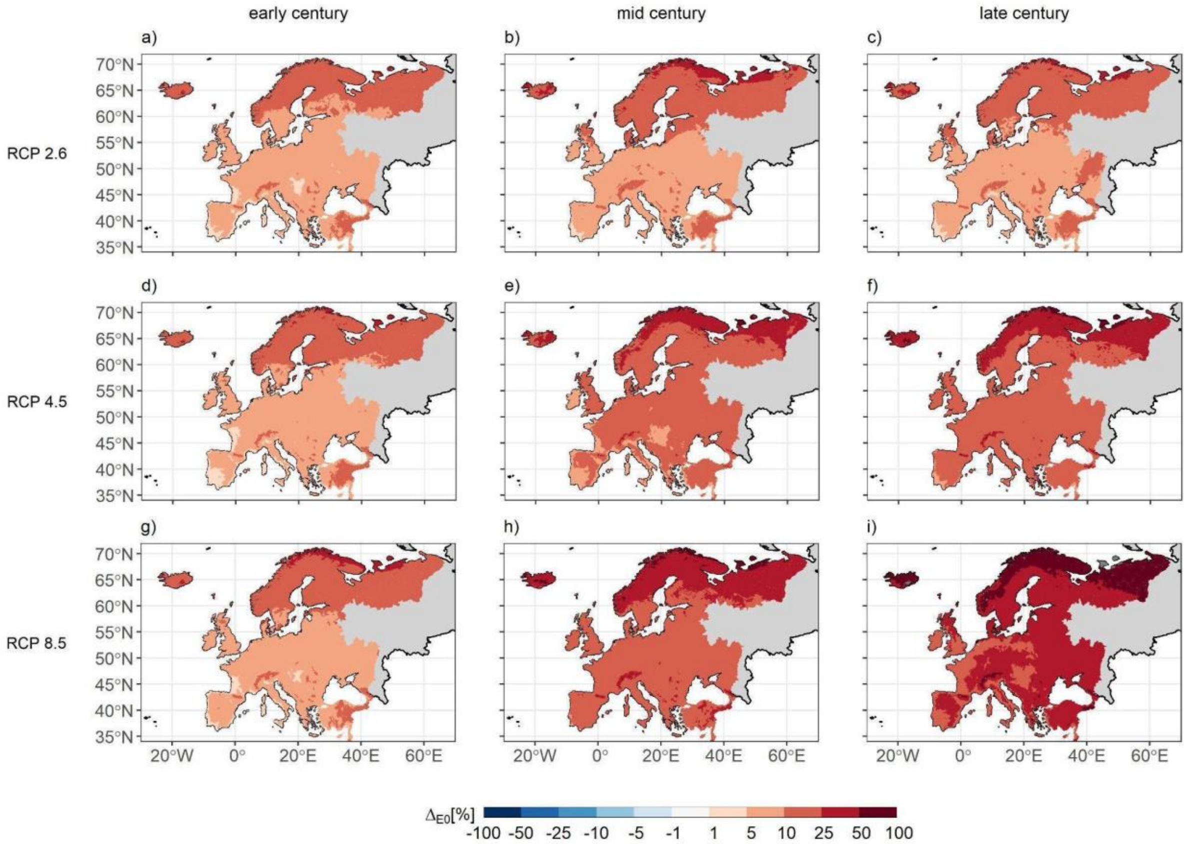


Figure 2-10 Model ensemble median of potential evapotranspiration changes for the RCP 2.6 (a,b,c), RCP 4.5 (d,e,f) and RCP 8.5 (g,h,i) emission scenarios in the early (a,d,g), mid (b,e,h) and late (c,f,i) century relative to the historical period (1971-2000) across Europe.

3.5.2. Sensitivity of runoff changes to precipitation and evaporation

Figure 2-11 displays the elasticity coefficients of runoff to precipitation (P) and potential evaporation (E0) in the European domain. These coefficients reveal the percentage changes in runoff, either increase or decrease, in response to a 1% increase in precipitation and potential evaporation. The results show that runoff tends to respond positively to precipitation changes and negatively to potential evaporation. The elasticity coefficient of precipitation between zero and five in 90% of European regions (Figure 2-11a), whereas for potential evaporation, it ranges between zero and three (Figure 2-11b). The median ensemble elasticities over Europe were 1.2 for precipitation and 0.5 for potential evaporation, respectively. This result means that on average, a 10% increase or decrease in precipitation would lead to a 12% increase or decrease in runoff, respectively. Similarly, a 10% increase in potential evaporation would result in a 5% decrease in runoff. Northern and central Europe exhibited a higher sensitivity to precipitation changes than to evaporation changes, while in some regions, such as Eastern Europe, the sensitivity to both climatic factors was comparable.

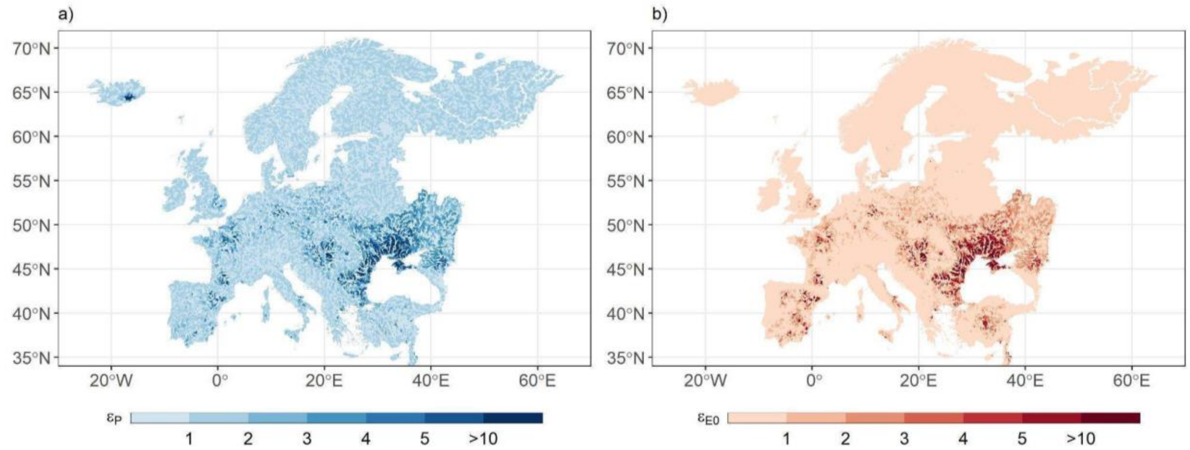


Figure 2-11 Elasticity coefficients of runoff changes to (a) precipitation and (b) potential evaporation (E0) across Europe.

3.5.3. Spatial variability of runoff changes under future conditions

Figure 2-12 presents the results of the ensemble runoff changes in the different periods under future conditions. The findings reveal that under low emission scenarios, there was an increase in runoff compared to the reference period in North-western Europe and the Alps, while a decrease was observed in Southern and Eastern Europe (Figure 2-12a-c). This spatial variability was more significant in the middle and late century under medium (Figure 2-12e,f) and high emission scenarios (Figure 2-12i,g) than under the low emission scenario (Figure 2-12b,c). Furthermore, runoff decreased in central Europe under medium and high emission scenarios, especially in the middle (Figure 2-12e-f) and late century (Figure 2-12h-i).

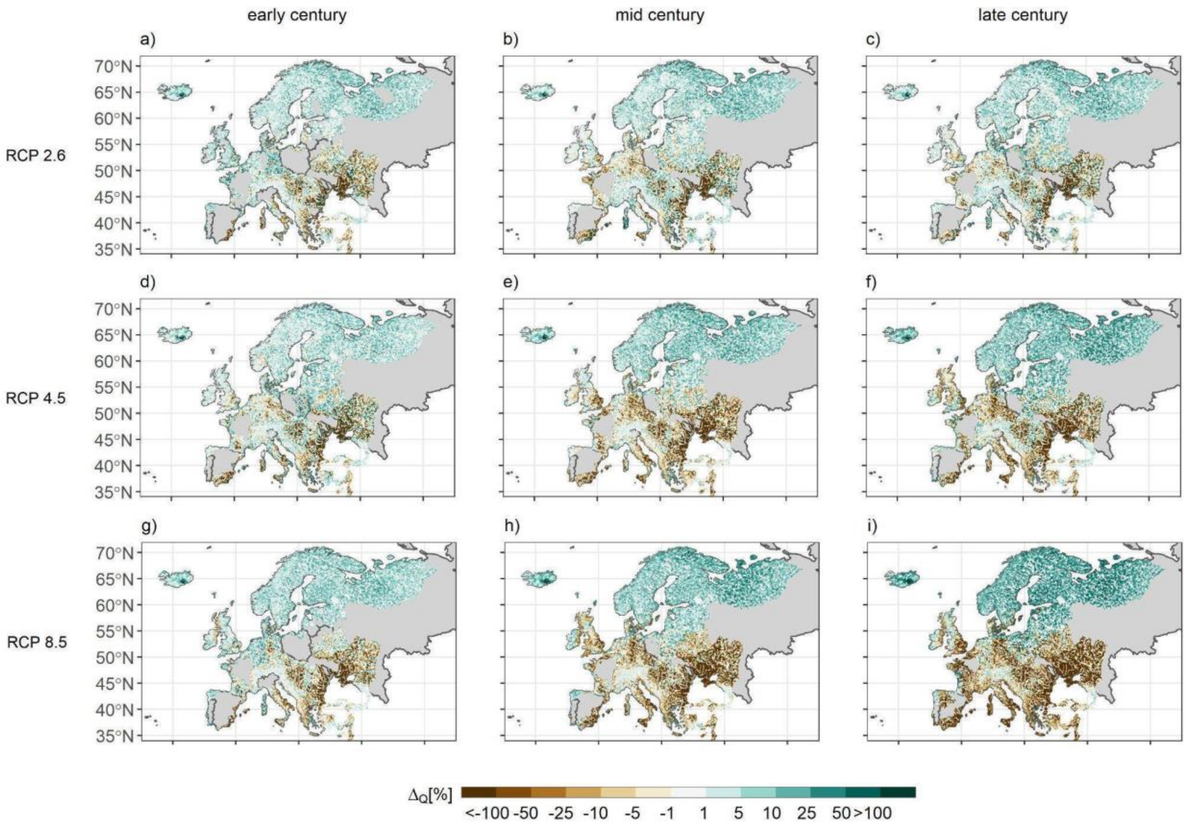


Figure 2-12 Ensemble median of projected runoff changes for the RCP 2.6 (a,b,c), RCP 4.5 (d,e,f) and RCP 8.5 (g,h,i) emission scenarios in the early (a,d,g), mid (b,e,h) and late (c,f,i) century relative to the historical period (1971-2000) across Europe.

3.5.4. Attribution of precipitation and evapotranspiration to runoff changes

Figure 2-13 displays the contribution of climatic factors such as precipitation and evaporation to the changes in runoff during the early, mid and late centuries under low, medium and high emission scenarios. Temporal and spatial differences in the contribution of precipitation and evapotranspiration were found across Europe with the severity of the emission scenarios. Precipitation mainly contributed to runoff changes in Scandinavia and North-eastern Europe, especially under medium (Figure 2-13d-f) and high (Figure 2-13g-i) emission scenarios. While precipitation contributed to runoff changes in central and eastern Europe under low-emission scenarios, evapotranspiration changes became the main factor of runoff changes in central Europe from the early to late century under both the medium (Figure 2-13d-f) and high (Figure 2-13g-i) emission scenarios. In southern Europe, precipitation was the climatic factor contributing to decreased runoff, especially under medium scenarios (Figure 2-13d-f). In contrast, evapotranspiration contributed more than precipitation (Figure 3.13g-i) to runoff under high-emission scenarios. The Budyko elasticity framework provided a quantitative understanding of how changes in climatic factors affect alterations in runoff, bringing new insights into the context of hydroclimatic attribution under future scenarios.

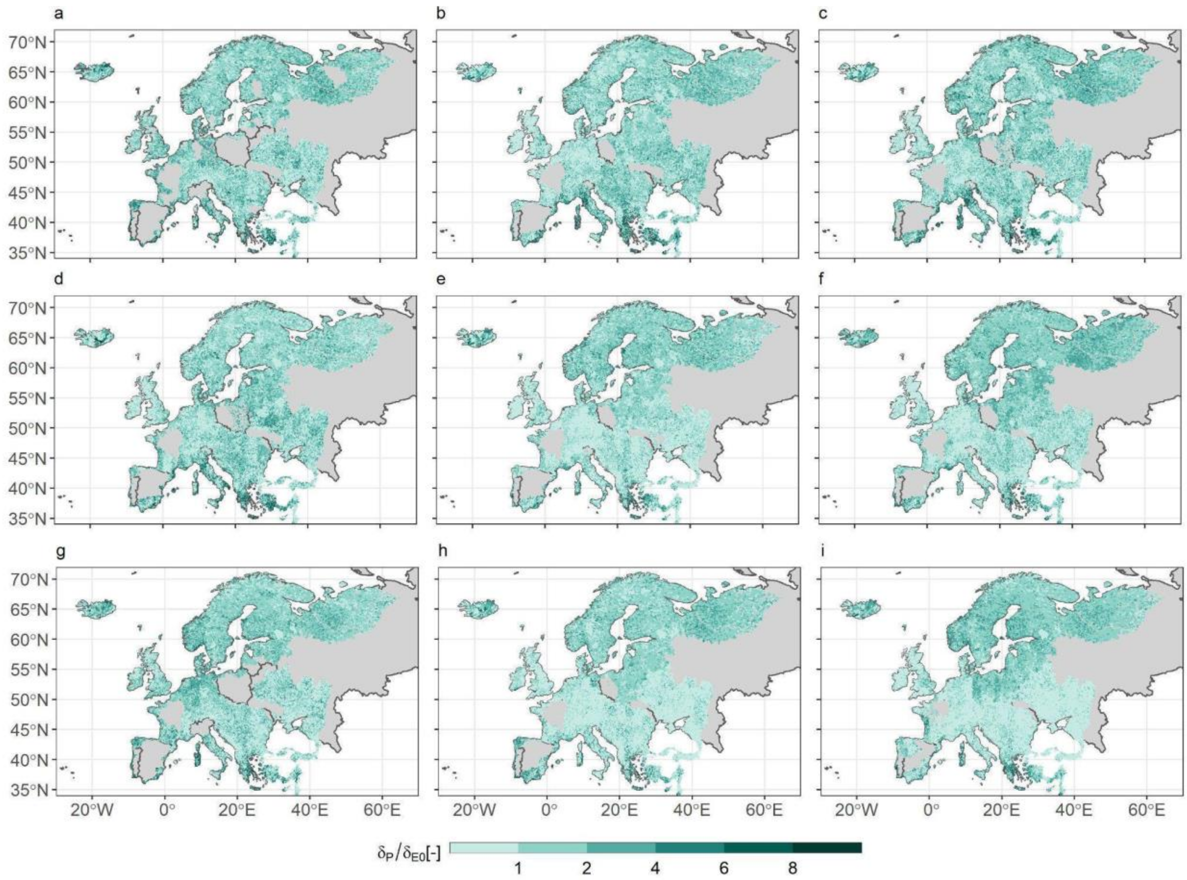


Figure 2-13 Ensemble median ratio of the relative precipitation (δp) and evapotranspiration (δE0) contributions to runoff changes for the RCP 2.6 (a,b,c), RCP 4.5 (c,d,e) and RCP 8.5 (e,f,g) emission scenarios in the early (a,d,g), mid (b,e,h) and late (c,f,i) century relative to the historical period (1971-2000) across Europe.

3.5.5. Summary of results

Here, we showed the application of the Budyko framework to provide a quantitative understanding of how changes in climatic factors affect alterations in runoff during the early, mid and late centuries under low, medium and high emission scenarios across Europe. The key findings include:

- Under future conditions, precipitation showed a contrasting spatial behavior with increasing precipitation in Northeastern Europe, specifically in Northeastern Europe and Scandinavia and a substantial decrease in Southern Europe towards high emission scenarios. Evaporation increased with the increasing severity of the emission scenario.
- On average, the sensitivity of runoff to precipitation changes was more substantial than to evaporation across Europe. Northern and central Europe showed a higher sensitivity to precipitation than evaporation changes, while in other regions, the sensitivity to both climatic factors was comparable.
- Precipitation was the main factor of increasing runoff in Scandinavia and North-eastern Europe, especially under medium and high emission scenarios, while it contributed to decrease runoff in southern Europe.
- Evaporation changes were the main factor of runoff alterations in central Europe from the early to late century and with the increasing severity of the emission scenario.

These findings provide a foundation for the next phase of our work, which involves using machine learning to investigate how the attribution of runoff extremes to climatic factors is connected to the local hydrological regimes to evolve climate services towards hydrological attribution in the context of centennial projections.

3 INVESTIGATION ON AI-ENHANCED CLIMATE SERVICES FOR EXTREME IMPACTS - ENERGY SECTOR

Recent literature has explored the topic of closer linking of weather and climate data with energy models. The increased dependency of the energy system on climate and weather variables requires the energy models to be able to accurately represent the effect of climate variability (De Felice M et al., 2023). The effects of extreme weather events on the energy sector can fall into two categories: either on operational impacts or on infrastructure destruction, e.g. a tropical cyclone or a flood permanently destroying a power plant or power grids (Xu et al.,2024). Although tropical cyclones are not pertinent to the European region, climate projections point to the higher risk of colder and stormier winters, heavier rainfalls, storms and summer heatwaves (Collins et al., 2019). This work focuses on the operational impacts to the energy sector.

There are several dimensions of the operational impacts that are vulnerable to changing climate conditions, as especially to the increased occurrence of extreme weather events, such as heatwaves, droughts, and cold spells. The following chapters focus on the two dimensions:

Power demand: heats and droughts cause an increase to the electricity demand for cooling in households and the food industry. Taking into account seasonal forecasts plays a great role in the planning (Orlov et al., 2020). The decarbonization of the industrial, buildings and transport sectors among other foresees higher electrification. Higher demand for power is magnified in the periods of high and low temperatures for heating and cooling needs. Despite regional differences, there is a general increase in electricity demand for cooling across various sectors (Rujiven et al., 2019).

Hydropower generation: changes in hydropower generation at the national or regional level induced by climate change. By considering the space-time dependencies of both increases and decreases of change in climate-induced runoff, recent advanced studies have significantly contributed to finding gains between aggregate hydro power potential (Yalew et el., 2020, Gernaat et al., 2021). Climate-driven fluctuations in hydropower availability across large spatial distances contributed to the understanding of the future decarbonized power sector needs for time flexibility (Wörman et al., 2020). Coordinating hydropower operations over extensive geographical areas can lead to a more consistent availability of power, thereby minimizing the reliance on energy storage systems to buffer periods of reduced energy generation.

The methodological approach comprises two parallel processes for an assessment of the impacts of extreme weather events on the European power generation system with the detailed unit commitment model PRIMES (PRIMES-IEM):

- an analysis of the impact of heatwaves and cold spells on cooling/heating demand, is accomplished with the simulation of demand load curves.
- an assessment of changes in power generation from run of river hydropower and pumped storages, modeled based on hydro inflows provided by SMHI.

By using both methods, we can better understand and model energy systems' reactions to extreme events like droughts and cold spells. Hydro inflows are mainly driven by such factors as precipitation, temperature, snow and ice melt, depending on the region (Quaranta et al., 2022). A change in the temperature also affects the amount of energy required for cooling and heating. Reaching the EU climate neutrality target in 2050 requires the transformation of the energy, industry, transport and buildings sectors. The target can be achieved through different decarbonization pathways, including increasing renewable energy uses, reaching higher energy efficiency, introducing carbon capture and storage. Other options include electrifying heating systems, introducing bioenergy, using hydrogen and synthetic fuels that increase demand for green electricity. To achieve carbon neutrality, the power generation system should be designed to be resilient to the effects of climate change, including the increasing probability of extreme weather events.

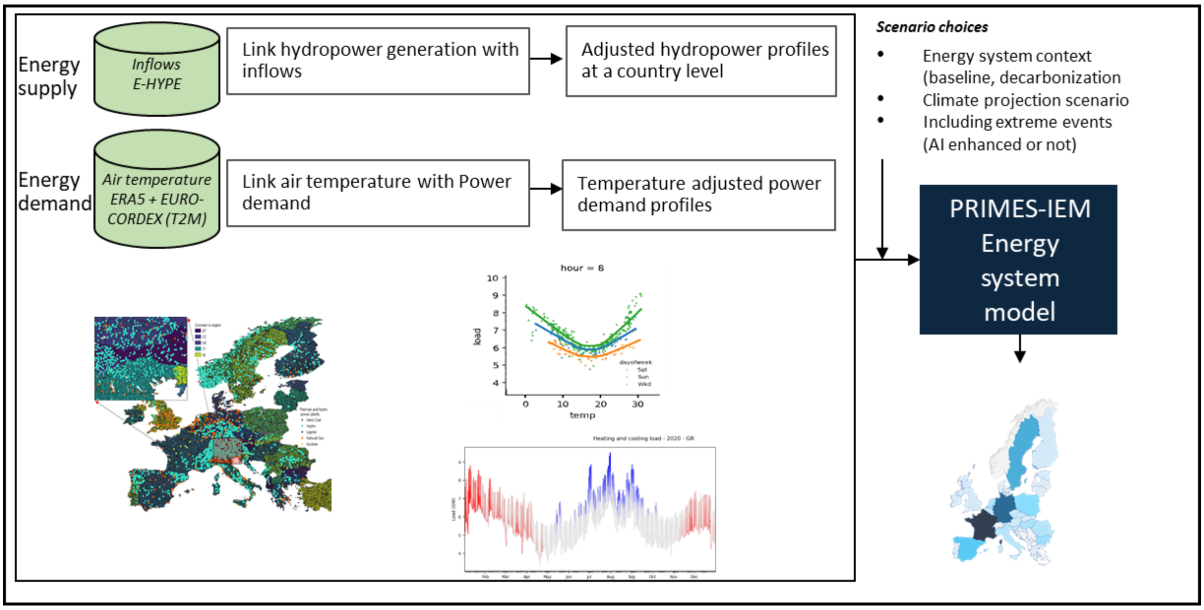


Figure *3-1 PRIMES model framework design to include climate services data for the energy sector*

The following chapters describe the process (Figure 3-1) of implementing the ML-techniques to take into account the climate data and hydrologic projections (E-HYPE) into the PRIMES-IEM model.

The long-term energy scenarios will benefit from the modification of the PRIMES IEM modeling suite to include EE effects on the energy system and especially power supply modules. The PRIMES IEM modeling suite can integrate both supply and demand side effects of climate change on the energy sector, considering current and planned EU climate and energy policies. The model applies constraints associated with water supply for hydropower production and feedstock supply curves for biomass production and can also be extended to taking into consideration such EE as tropical cyclones and extreme droughts. The PRIMES IEM demand module can reflect structural changes in the industrial and domestic demand for energy and electricity for heating and cooling services induced by EE, such as an increased occurrence of heatwaves and warm nights.

4.1. ADJUSTING POWER DEMAND TO TEMPERATURE CHANGES UNDER CLIMATIC DRIVERS IN PRESENT AND UNDER FUTURE CONDITIONS

Climate and weather fluctuations impact both the supply and demand for energy. High and low temperatures indicate significant energy requirements for heating and cooling. The relationship between temperature and electricity consumption is nonlinear: rising temperatures increase the amount of energy needed for cooling in summer or heating in winter. For example, significant heatwaves struck Europe in August 2003, July 2010, and June–August 2015. These events increased the need for cooling power. In the same way, lowering the wintertime thermal comfort threshold causes an abrupt increase in the amount of energy needed for heating. This work aims to identify the thresholds and the heating and cooling thermo-sensitivity of power demand.

4.1.1. Methodology - constructing adjusted load curves

There is general agreement that temperature is a key weather driver of electricity demand and the relationship between electricity demand and outside temperature follows a U-shaped curve (Bloomfield et al., 2020). The temperature-energy demand function (or temperature response function) describes the typical response of energy demand with temperature, and it comprises a thermosensitive part, which is temperature-dependent, and a non-thermosensitive part, which is considered as baseload. At low temperatures, when temperature T is already below the heating temperature point T_h ($T < T_h$), with later temperature increase the demand for heating decreases. At high temperatures, when temperature T is above the cooling temperature point T_c ($T > T_c$), the temperature increases and the cooling demand also increases. $P_{baseload}$ is the non-thermosensitive part of demand. T_h and T_c are the temperature threshold points that indicate the range of temperatures that are relevant for heating or cooling respectively. These are also the threshold points that are used to calculate heating degree days (HDD) and cooling degree days (CDD).

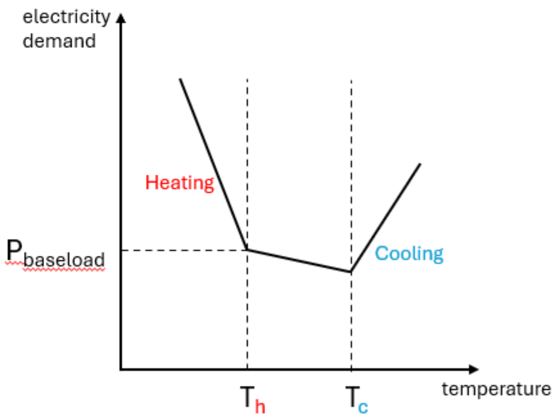


Figure *3-2 Temperature - demand response curve*

The temperature-demand response curve is schematically presented on Figure 3-2. It is typical for the temperature-demand response curve to be analyzed for different times of the day (day, night, individual hours of the day) or different days of the week (weekday, weekend, holiday). Common approaches for modeling the temperature-demand relationship include multiple linear regressions, building energy consumption simulation and machine learning methods such as artificial neural networks (ANNs) and random forests (RFs). More recently, models using multivariate adaptive regression splines (MARS) have become common in literature (Hiruta et al., 2022), primarily due to their ability to model non-linear interactions between the variables using hinge functions (Friedman, 1991). They are also preferred over machine learning methods like ANN or RFs due to their explainability and capacity to reflect the effect of each variable on the model construction and output.

In the current study, the goal is to identify the statistical relationship between temperature and electricity demand at a country level and use this to analyze the change in demand due to temperature changes in future climatic conditions. We are using the following hourly temperature and load data for each country:

- 2-meter temperature data from ERA5 reanalysis (Felice & Kavvadias, 2022),
- aggregate electricity demand from the ENTSO-E Transparency platform.[4]

The py-earth package[5] implements the MARS algorithm from (Friedman, 1991), and simplifies fitting a MARS model (or an Earth model using the package's terms) to load-temperature data for a country. Considering a max number of 2 coefficient terms (one for heating and one for cooling), we get a MARS model that follows:

$$E(T_t) = a * (T_t - hinge)^+ + b * (hinge - T_t)^+ + c$$

Where,

T_t the hourly mean temperature observation

$hinge$ the temperature threshold points for heating (T_h) and cooling (T_c) respectively

a, b the slope coefficients

c the intercept

The model shows reasonable results for countries where the temperature response function relies on one temperature hinge point, as for example the relationship between temperature and heating demand in Estonia in 2015, see Figure 3-3 below, on the right. However, when both demand for heating and cooling determine the relationship between temperature and power demand, the MARS model estimates day and night temperature hinge points that are skewed towards high temperatures. This leads to underestimating the cooling share of thermosensitive demand.

[4] https://transparency.entsoe.eu/

[5] https://contrib.scikit-learn.org/py-earth/

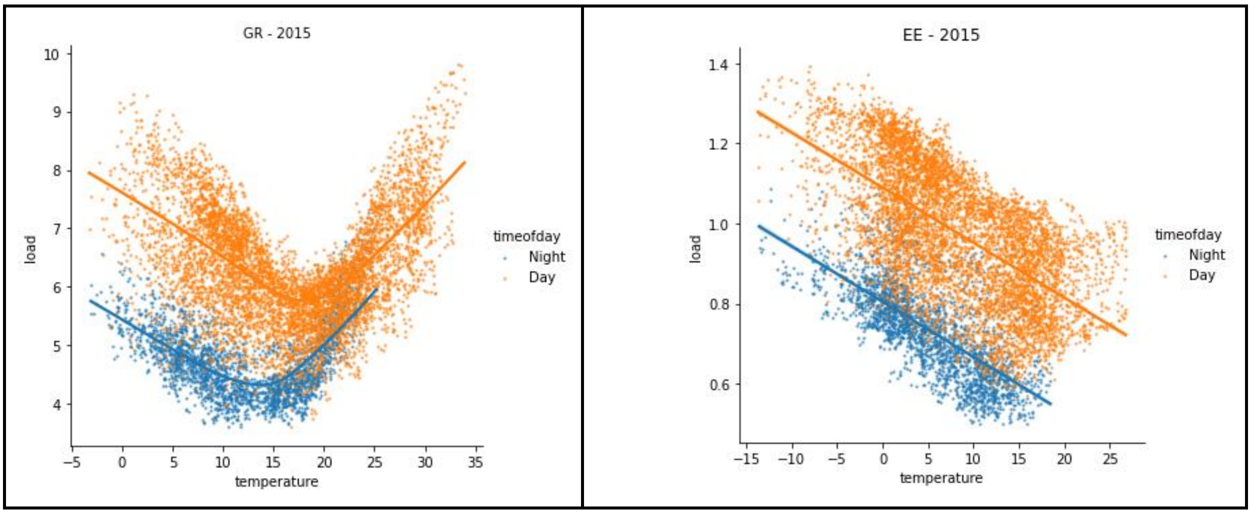


Figure *3-3 Temperature-load data for Greece-2015 and Estonia-2015*

Degree-day models usually apply fixed temperature thresholds to calculate degree days and then use linear functions to translate degree days into demand for heating or cooling. Eurostat's Energy Statistics unit uses fixed temperature thresholds across all countries to calculate degree days. The thresholds are 15°C for heating degree days (HDD) and 24°C for cooling degree days (CDD). However, this ignores differences between countries and leads to over- or under-estimation of demand (Wenz et al., 2017). Staffell et al., (2023) highlight the importance of having optimal parameters to derive HDDs and CDDs, including temperature thresholds.

Improved representation of temperature-load relationship

We implement an improvement over the MARS model that is based on the simplicity of linear regression methods but attempts to capture the climatic and thermal comfort differences between European countries. Below we illustrate our method in Figure 3-4. The steps in our method are as follows:

a. We assume that there are two temperature threshold / hinge points for each country. For heating we expect this will be in the range 10°C - 25°C and for cooling in the range 13°C - 29°C.
b. For each temperature in this range, we calculate Spearman's correlation for the relationship between HDD (or CDD) and load observations. We consider only positive Spearman correlation values with p-value over 0.05.
c. We consider the temperature point for which Spearman's correlation maximizes, as the optimal temperature threshold point. We identify one temperature threshold for heating T_h based on the Spearman correlation values of HDDs, and one temperature threshold for cooling T_c based on the Spearman correlation values of CDDs. For countries where only heating load is observed via a single negative gradient curve (e.g., Scandinavian countries), only T_h is calculated.
d. For selected cases, we verify the temperature threshold points by performing a visual inspection of the Spearman correlation against temperature.

e. We consider that temperature points outside the range of T_h to T_c, represent thermosensitive loads. We fit two separate linear models:
   - For($T < T_h$): $E(T_t) = h_{coeff} * T_t + h_{baseload}$
   - For ($T > T_c$): $E(T_t) = c_{coeff} * T_t + c_{baseload}$
   - Where h_{coeff} or c_{coeff} is the slope and $h_{baseload}$ or $c_{baseload}$ is the line intercept

f. We follow this process separately for combinations of (i) daytime and nighttime (ii) weekday and weekend observations as it was observed that these sets have different behavior.

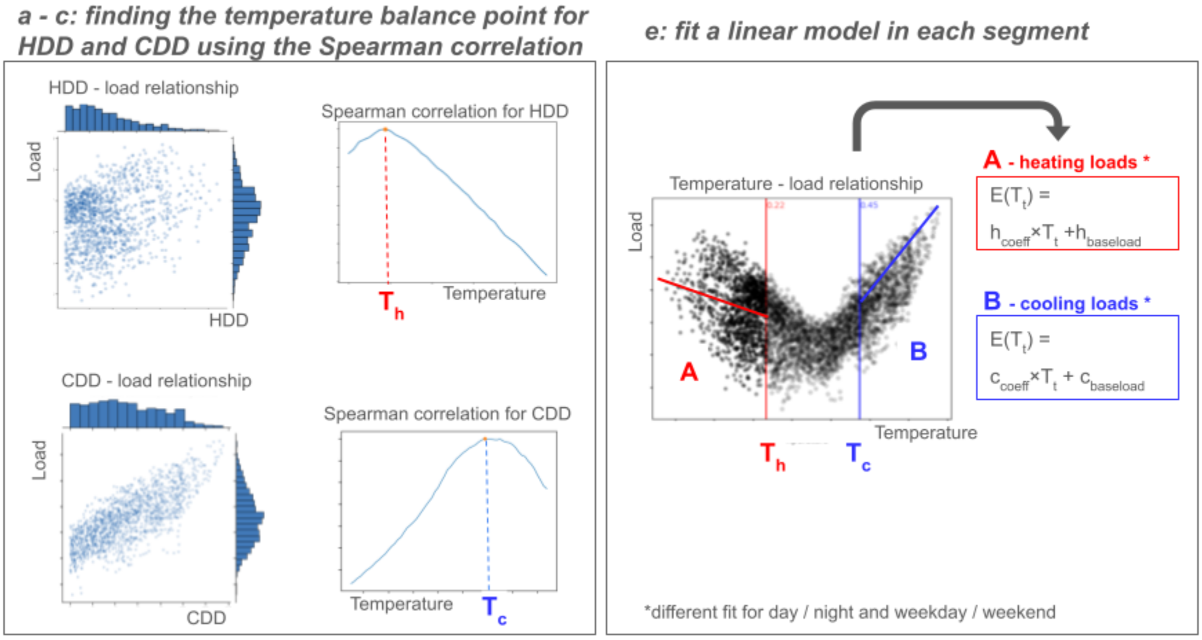


Figure 3-4 Steps (a) to (c), finding the temperature balance point for HDD and CDD using the Spearman correlation for a country and a particular year, time of day and day of week. Step (e), fitting a linear model for each segment.

Bias adjustment

Using the above model to describe how load varies with temperature in each country, we aim to build an adjusted demand time series for temperature predictions of different climatic scenarios. Our method is based on additive delta load correction enhanced by making use of temperature hinge points.

The uncertainty of climatic data has often been highlighted in climate science, especially in terms of bias and limitations found in observations (Bloomfield et al., 2021). There are two methods that are commonly found in literature for using climate model data: (1) the simple additive delta correction and (2) a seasonal quantile-based correction (Bloomfield et al., 2022). We perform the bias correction with delta method by associating the temperature difference between base and future year with a load difference, as in the following steps:

a. We consider a base year as reference for the graphical representation of the temperature-load relationship and derive the model parameters:

$h_{coeff,base}$, $c_{coeff,base}$ coefficient for heating and cooling in the base period

$h_{baseload,base}$, $c_{baseload,base}$ baseload for heating and cooling in the base period

$T_{h,base}$,$T_{c,base}$ temperature threshold (hinge point) for heating and cooling in the base period

b. For a future year with hourly temperatures $T_{t,future}$, we estimate the change in load dE for the associated change in temperature dT compared to the base year. We consider a change in heating or cooling load, depending on the base year temperature hinge points.

- For $T_{t,future} \leq T_{h,base} \rightarrow dE = h_{coeff,base} \cdot dT$
- For $T_{t,future} \geq T_{c,base} \rightarrow dE = c_{coeff,base} \cdot dT$
- For $T_{h,base} < T_{t,future} < T_{c,base} \rightarrow dE = 0$
- where $dT = T_{t,future} - T_{h,base}$

c. We consider that the future demand at each hourly step will be the base year load corrected by the calculated dE. So for each hourly $T_{t,\, future}$:

- If $T_{t,future} \leq T_{h,base} \rightarrow E_{t,future} = h_{baseload,base} + dE$
- If $T_{t,future} \geq T_{c,base} \rightarrow E_{t,future} = c_{baseload,base} + dE$

4.1.2. Results - constructing adjusted demand curves

Using 2020 as a reference year, we showcase the different temperature hinge points calculated by the degree-day delta correction method for each country. For Northern countries like Finland and Latvia, our method estimates HDD hinge points at just above 10°C, considerably lower than the Eurostat's 15°C baseline for HDD, see Figure 3-5. For CDD, our method finds temperature hinge points close or higher than Eurostat's baseline for CDD of 24°C, Figure 3-6.

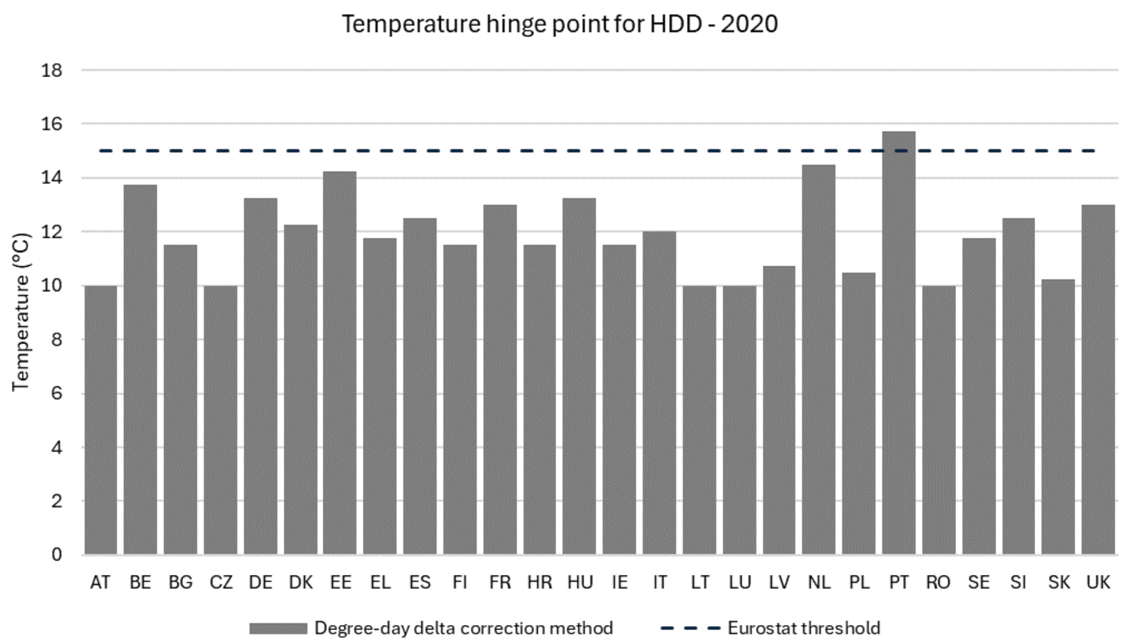


Figure *3-5 Temperature hinge points for HDD for 2020; comparison between degree-day delta correction method and Eurostat threshold. Results from the degree-day delta correction method are for daytime - weekday observations.*

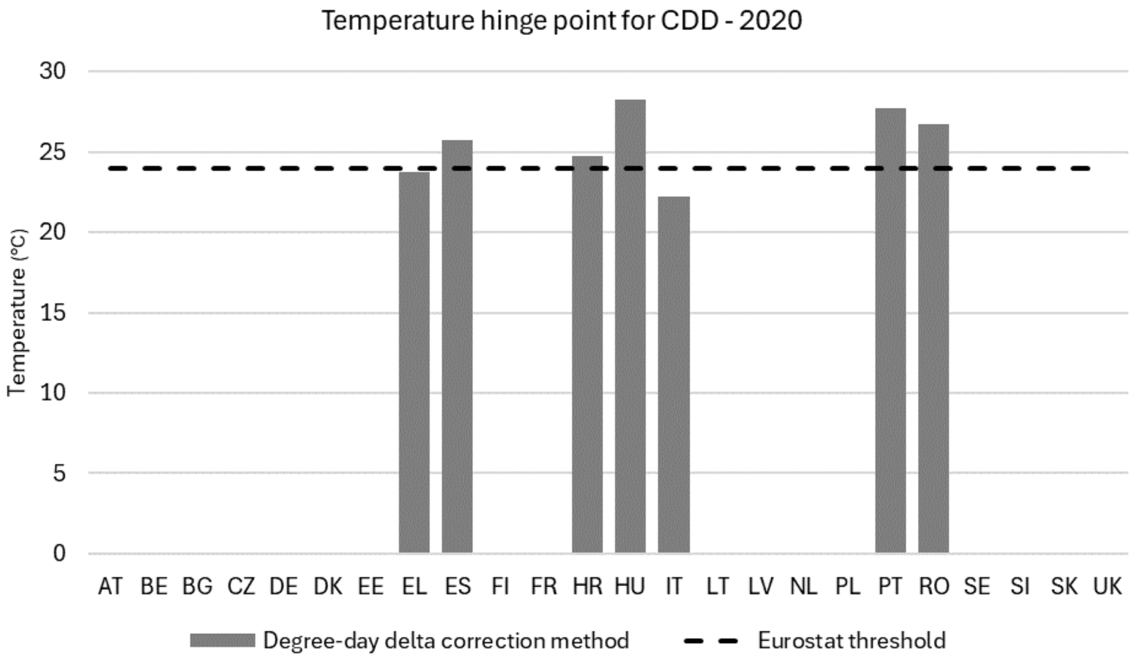


Figure 3-6 Temperature hinge points for CDD for 2020; comparison between degree-day delta correction method and Eurostat threshold. Results from the degree-day delta correction method are for daytime - weekday observations.

Based on the above described method we explore the relationship between temperature and heating or cooling demand for electricity for 2020. Figure 3-7 below shows the hourly power demand for Greece and the hinge points for HDD/CDD obtained from the degree-day delta correction method. Figure 3-7 shows the similar results for Finland. On both figures, heating and cooling demand are marked as red or blue depending on whether the mean hourly temperature is above or below the calculated HDD or CDD hinge points.

Figure 3-8 for Greece shows that the need for cooling is visible primarily on peaks of hourly demand, whereas heating is across the whole range of demand. As expected, the period December to February is the most common period for heating and the summer months are more prevalent for cooling. For both countries the demand time series shows variability across the year, although this is more obvious in the case of Finland. This is an indication that Finland is more driven by the demand for electricity for heating and therefore its heating demand is more temperature dependent. Temperature hinge points for nighttime are lower than daytime temperature thresholds, showing that thermal and cooling comfort thresholds might be stricter at nighttime.

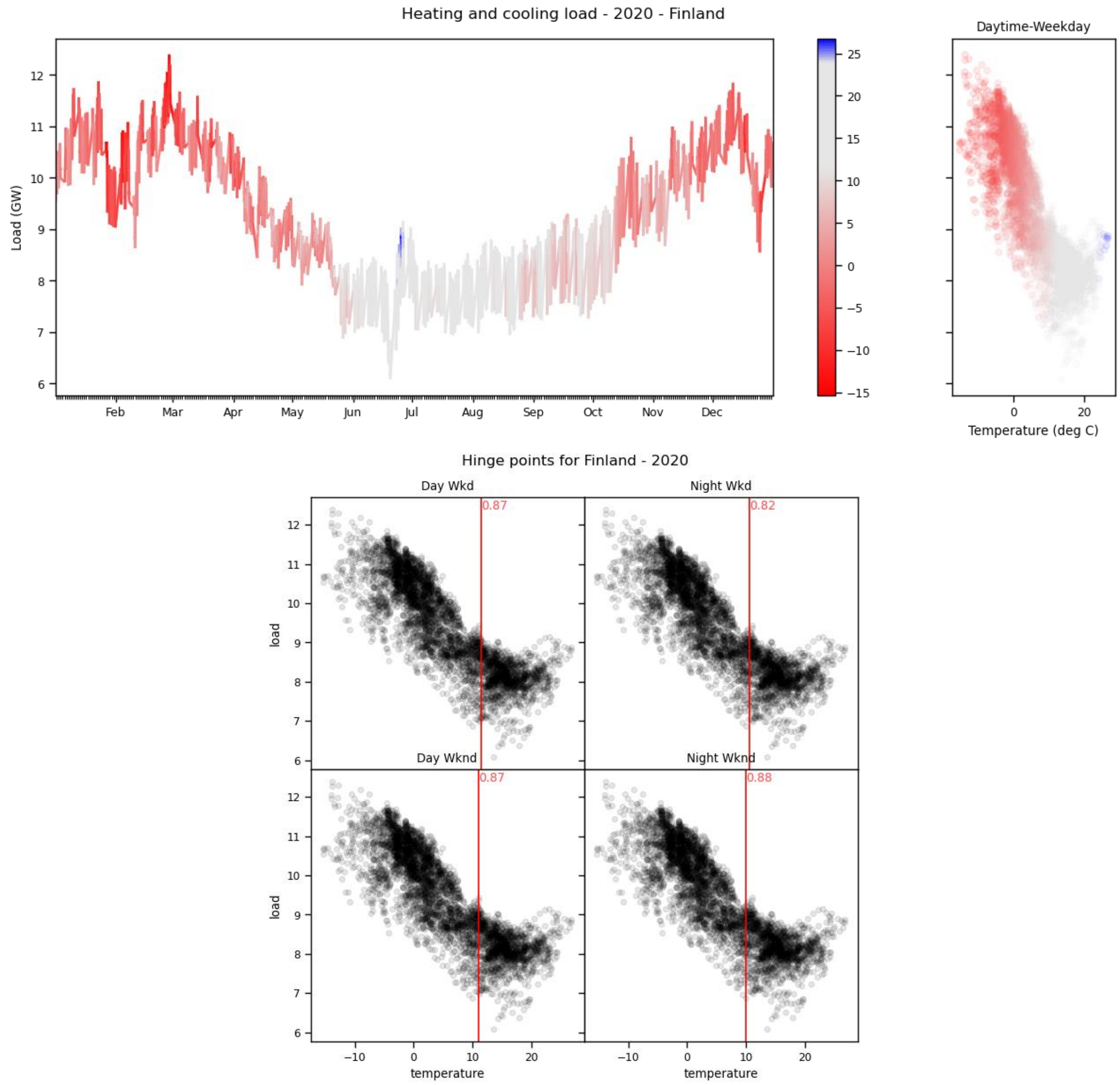


Figure *3-7 (i) Hourly demand for Finland 2020; blue indicates cooling and red indicates heating load based on the identified hinge points. (ii) Temperature hinge points for HDD and CDD for each set of daytime/nighttime, weekday/weekend temperature - load observations.*

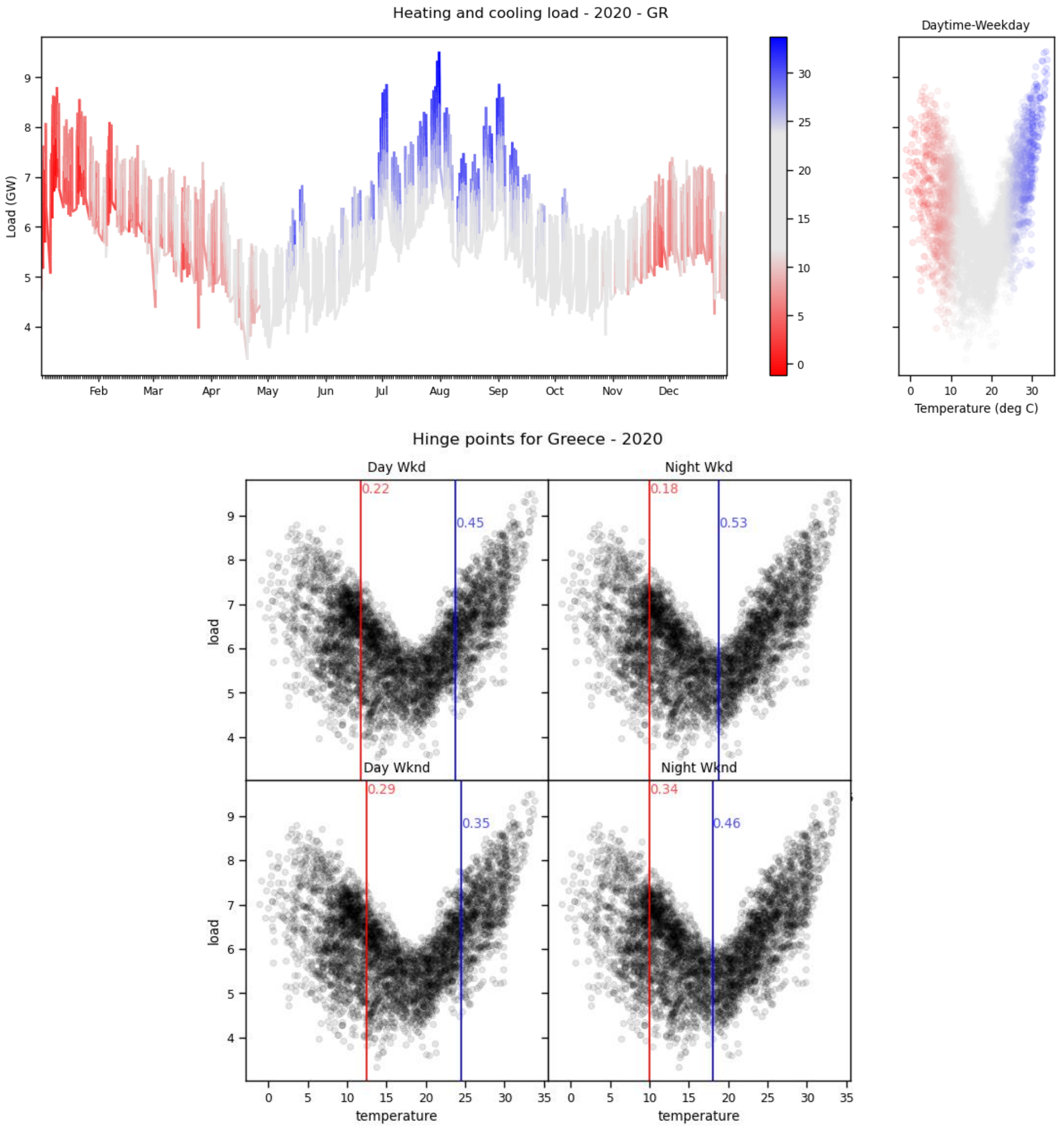


Figure 3-8 (i) Hourly demand for Greece in 2020; blue indicates cooling and red indicates heating load based on the identified hinge points. (ii) Temperature hinge points for HDD and CDD for each set of daytime/nighttime, weekday/weekend temperature - load observations.

We have used the degree-day delta correction method to construct an adjusted demand time series for historical temperature data from 1980 to 2021, based on the hinge points defined for the base year 2015. Figure 3-9 shows the adjustments in power demand curves, achieved with the above-described method for historic periods with known 2-meter temperature from ERA5 reanalysis data. The Figure 3-9 below demonstrates how we can use data on electricity demand from a base year to get adjusted load curves for historic periods, as well as for the different climatic scenarios in the future.

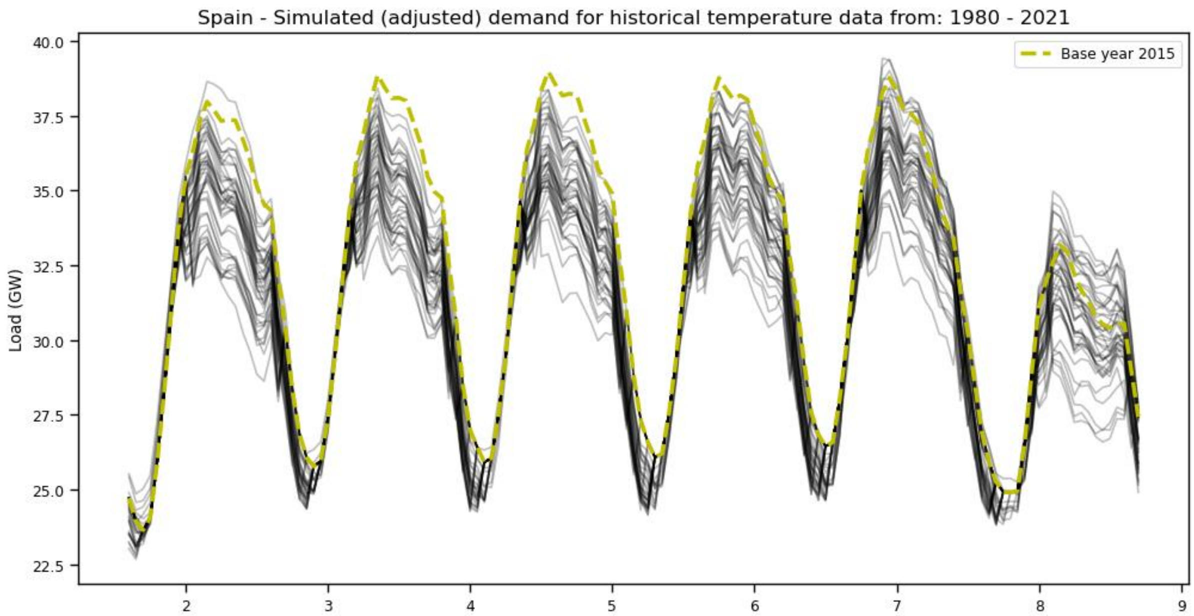


Figure *3-9 Simulated (adjusted) demand time series for historical temperature data. Example for Spain using 2015 as base year and historical temperature data for 1980 - 2021 for the period from July 13th to 19th for all years*

Table 3-1 demonstrates how annual electricity demand can vary for different climatic years for different countries. The annual adjusted demand for each country was simulated using 2015 as base year and temperature values for each climatic year from the period 1980 to 2020.

Table 3-1 Descriptive statistics of simulated (adjusted) annual demand for historical temperature years from 1980 to 2020, using 2015 as a base year (excluded from min / max / median statistics).

Country	Simulated (adjusted) annual demand							
	Median GW	CV	Min GW	Min Temp. Year	Min $T_{daily\ mean}$ (°C)	Max GW	Max Temp.Year	Max $T_{daily\ mean}$ (°C)
Austria (AT)	61.7	0.56	61	2014	8.14	62.3	1985	5.2
Belgium (BE)	75.6	0.7	74.7	2020	11.5	76.9	1985	8.3
Bulgaria (BG)	34.3	0.63	33.8	2019	12.7	34.6	1987	10.3
Czechia (CZ)	56.4	0.62	55.8	2014	9.79	57.25	1996	6.4
Germany (DE)	460.8	0.51	456.8	2020	10.6	466.5	1996	7.2
Denmark (DK)	29.7	0.66	29.4	2020	10.1	30.2	2010	7.1
Greece (GR)	44.7	0.75	44.1	1983	14	45.6	2012	15.5
Spain (ES)	217.3	0.36	215.9	1993	12.8	219.9	2017	15
Finland (FI)	74.6	1.51	72.3	2020	4.7	77.8	1985	-0.09
France (FR)	427.5	1.54	415.3	2020	12.6	444	1985	9.9
Hungary (HU)	35.7	0.32	35.5	1980	9.3	36	2012	11.9
Italy (IT)	273.3	0.79	269.8	1980	11.36	279	2003	13.4
Netherlands (NL)	100	0.22	99.6	1988	10.3	100.8	2018	11.3
Portugal (PL)	131.8	0.14	131.4	2020	10	132.1	1985	6.7
Romania (RO)	46.1	0.29	45.8	2019	11.4	46.4	1985	8.2
Sweden (SE)	123.1	1.87	118.4	2020	5.5	129.5	1985	1
United Kingdom (UK)	302.3	0.69	299.8	2007	9.8	308	2010	8.3

Note: Red indicates years when adjusted demand reflects a higher or lower demand for heating due to very low or very high temperatures during winter, respectively. Blue indicates years when adjusted demand reflects higher demand for both heating and cooling. This is due to lower temperatures during winter and higher temperatures during summer.

The Table 3-1 shows the minimum and maximum adjusted demand, the temperature year in which this was found and the daily mean temperature in that year. We excluded 2015 from finding the min, max and median values, as this was the year used as base year for the simulation. We derive the following key points:

- 1985 is the year with the maximum adjusted electricity demand for 7 out of the 17 countries used in the simulation. 1985 featured lower than average temperatures and for many countries it was a year with extremely low winter temperatures, reaching as low as -35°C. Very low temperatures led to increased need for heating and therefore higher simulated annual electricity demand.
- In contrast, 2020 is the year with the minimum adjusted demand for 7 countries. 2020 featured higher mean daily temperatures and, more specifically, higher temperatures during winter. The lowest daily temperature across the simulated countries was -2.1°C in 2020, compared to -8°C for the whole period 1980-2020. Higher temperatures led to lower need for heating and therefore lower simulated annual electricity demand.
- For Greece, Spain, Hungary and the Netherlands, the simulation showed that both heating and cooling demand can lead to increased annual demand. The maximum adjusted demand was for a year when daily temperatures varied from very low to very high, thus leading to increased demand for heating and cooling.
- The coefficient of variation (CV) is a unitless measure of the level of dispersion around the mean. A higher value (over 1), such as for Finland, France and Sweden shows higher variation of annual electricity demand.

Focus on extreme events – Case 1: Warm winter

We use the method described in section 4.1.1 to get the adjusted demand time series for two cases of extreme weather: the warm winter of 2020 and the heatwave in 2018. This allows us to explore the impact on demand if mean hourly temperatures changed from their base year values to the extreme event values. For both cases we are using 2015 as the reference year.

2020 was an especially warm winter for EU countries, with significant lower HDDs than the 20-year average. The Market Observatory for DG Energy reported that higher than usual temperatures during the first quarter of 2020 led to 247 HDDs below average, translating to 2.7°C higher temperature than usual per day (DG Energy 2020).

Germany saw the largest deviation in February 2020 with almost 100 HDDs lower than its long-term average. Less HDDs led to lower demand for heating. Looking at the first week of February 2020 (2nd to 9th February), Figure 3-10 shows that warmer temperatures in winter can lead to lower power demand load. Using 2015 as our base year, we estimate that for a mean hourly temperature increase of 5.6°C, the simulated (adjusted) load during that week was 4% lower than in 2015 (for the same period).

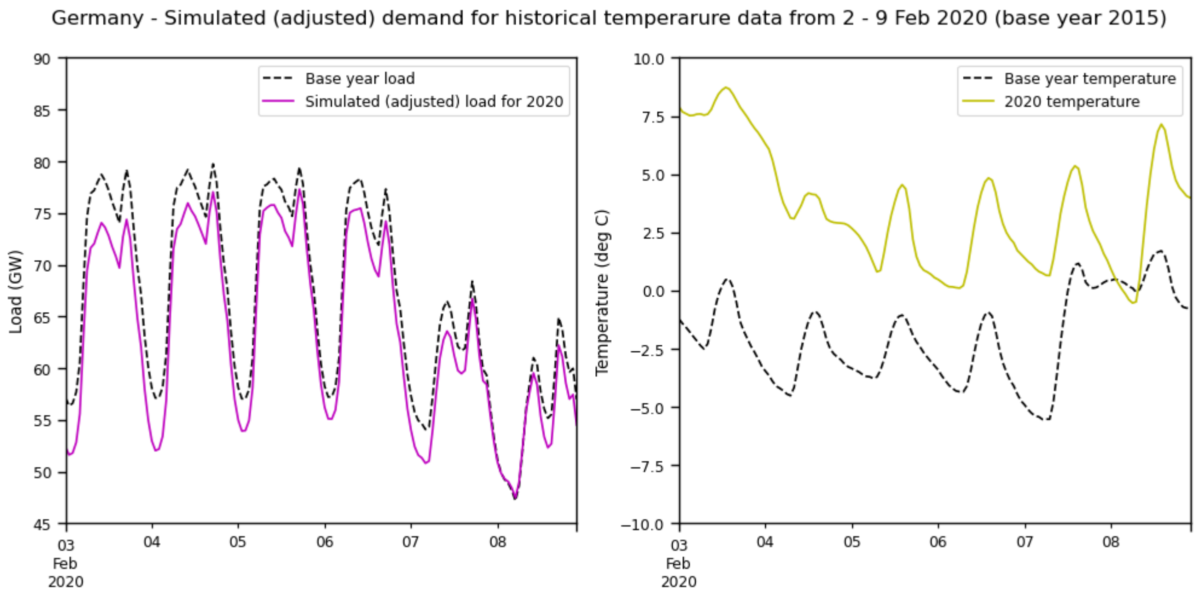


Figure 3-10 Simulated (adjusted) demand time series and temperature trend for 3 - 9 February 2020 using 2015 as base year for Germany.

Focus on extreme events – Case 2: Heatwave

Similarly, we examine the case of the 2018 heatwave. July and August featured particularly hot weather, which had an impact on cooling demand especially in Southern EU countries (DG Energy 2018).

During August 2018, Spain saw higher temperatures than the long-term average, reaching almost 30°C. We estimate that for the week 13 to 19 August 2018, the mean hourly temperature was 1.6°C higher than the base year (2015) for the same period. Higher temperature increases the need for cooling demand. As Figure 3-11 shows the simulated (adjusted) demand for that period is over the base year load and the overall demand for the week increased by 3.34%.

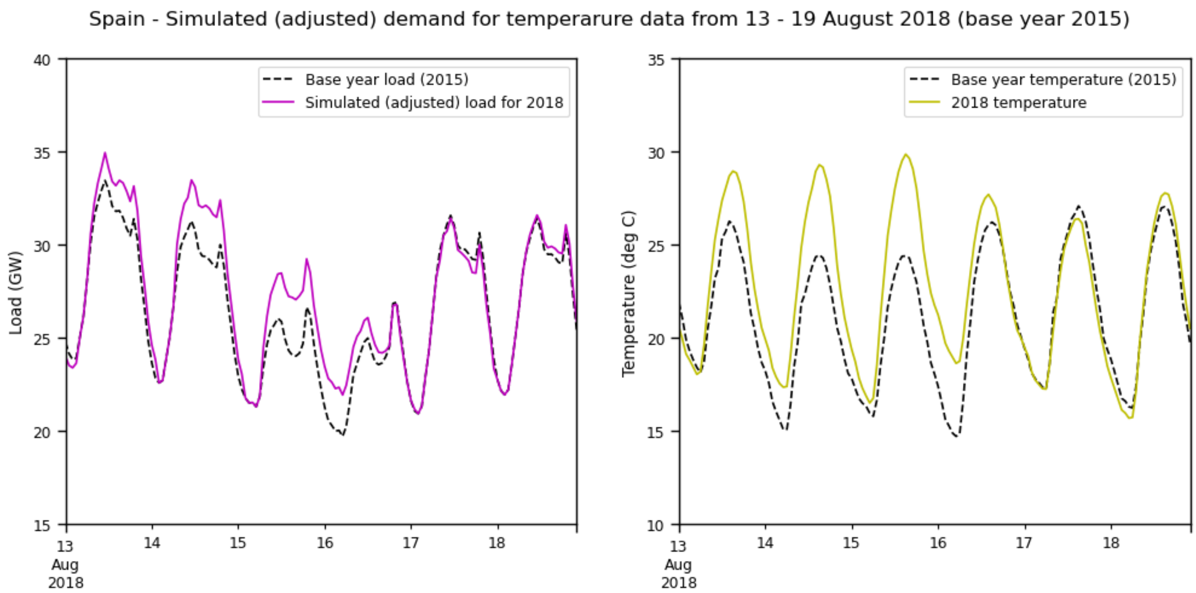
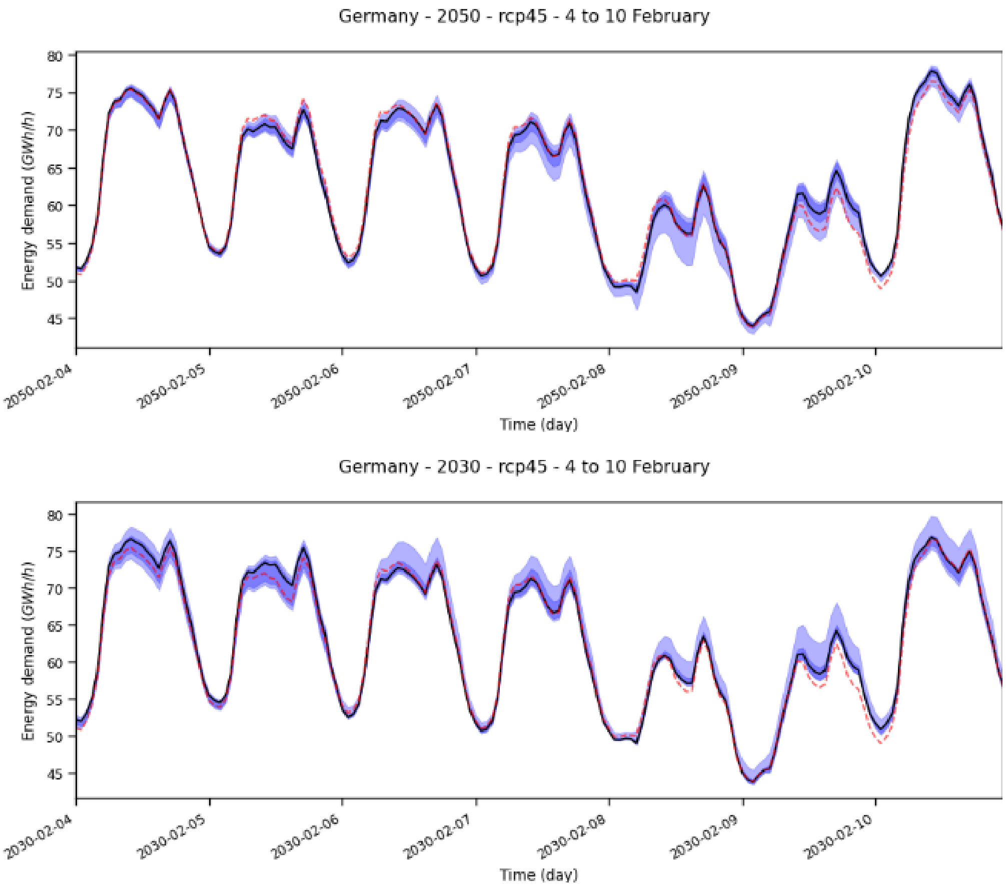


Figure 3-11 Figure: Adjusted demand time series and temperature trend for 13 - 19 August 2018 using 2015 as base year for Spain.

Focus on future climate – Scenario RCP 4.5

We used daily air temperature data from 9 climate models covering three representative concentration pathways. We focused on scenario RCP 4.5 and used temperature projections for the period 2025 – 2070 to simulate the impact on adjusted energy demand.

Temperature data from each climate model sub basin. We matched sub basins to countries and used the sub basin area values to obtain a weighted average of daily air temperature. Before constructing adjusted load curves, we performed bias correction with quantile mapping on the daily temperature values for the same historical period as the temperature observations in the ERA5 reanalysis dataset. We also downscaled the initial timeseries to an hourly step, using as reference the average hourly temperature profile of each country's ERA5 temperature observations for 2020, which we consider as base year.



Note: Black line shows the median adjusted hourly load based on the ensemble of 9 EURO-CORDEX models' simulation of energy demand with the degree-day delta correction method. Light blue area represents the 5th to 95th percentile range, and the 25th to 75th percentile is represented by the dark blue shaded region. Red dash line shows the hourly load for the base year 2020 for the same period.

Figure *3-12 Simulated (adjusted) load patterns for 2030 and 2050 years for Germany under RCP4.5 for the period 4 to 10 February.*

Figure 3-12 demonstrates the impact of the rise in average temperature to hourly electricity demand for Germany in two separate time periods. Our simulations indicated that there is little change in the average hourly electricity load. The biggest differences between the models' ensemble simulations of load occur during peak electricity demand during the daytime, at times of very high or very low air temperatures. The results show variation of the temperature projections in the model ensemble during the peak demand hours (see light and dark blue shaded areas on the Figure 3-12).

We compare changes in simulated (adjusted) demand due to (simulated) predicted changes in temperature compared to the base year 2020. During the week in February, hourly temperatures show up to 12.48°C higher by 2050, which leads to lower need for heating demand and therefore a decrease in hourly electricity load up to 4.86GW.

4.2. ADJUSTING HYDRO POWER GENERATION PROFILES TO CHANGES IN RIVER DISCHARGE UNDER CLIMATIC DRIVERS IN PRESENT AND UNDER FUTURE CONDITIONS

Run-of-river hydropower generators are weather-dependent generators that rely on the flow of water through rivers to drive their turbines. The application of machine learning can be used to find a non-parametric relationship between run-of-river hydro generation and river discharge. Recent literature has focused on the application of machine learning techniques to predict weather-dependent electricity generation (De Felice, 2020; Ho, Dubus, De Felice, & Troccoli, 2020). This paper explores using machine learning techniques to predict run-of-river hydro generation based on hydrological data for Europe. It trains a machine learning model for each relevant EU Member State. It then uses the model to predict run-of-river generation up to year 2100 using river discharge projections for a given climate scenario.

There is some literature on using machine learning techniques to predict run-of-river hydro. De Felice (2020) used four years of data on river discharge from the JRC-EFAS-Hydropower database and run-of-river generation data from ENTSO-E to fit a machine learning model. Using linear regression, the paper explored ridge regression and random forest models and evaluated the performance using root mean squared error (RMSE). Felice (2020) found that ridge regression and random forest yielded the best results. Ho et al. (2020) uses a random forest regression model to predict hydro generation in 12 European countries based on temperature and precipitation data. Lagged values were included to capture the delay between temperature and precipitation’s effect on hydro generation. For example, using lags allowed the dataset to capture a proxy for snow depth. The paper found that the random forest regression model performs better at predicting run-of-river generation than reservoir hydro, and attributes this to “human intervention and management” of reservoir-based hydro (Ho, Dubus, De Felice, & Troccoli, 2020).

In the following chapters we introduce methods used to find the relationship between river discharge for each subbasin and run-of-river hydro generation at European level. Two machine learning methods were explored to find the best fit: XGBoost and Artificial Neural Networks (ANN). The goal of the research was to prepare the country specific run-of-river hydro generation time series for the input in the PRIMES-IEM model. The purpose of this modification is to improve the scenarios for the European energy sector based on the ML/AI enhanced datasets on river discharge from the dedicated impact assessment models for the European energy sector.

4.2.1 Methodology – constructing the run-of-river hydro power generation profiles

Data used

The chapter describes the two main datasets used and preprocessing steps:

- dataset containing daily sub basins' river discharge at the European scale prepared by the E-HYPE model (Europe Hydrological Predictions for the Environment),
- dataset containing hourly time series on power generation from run-of-river (ROR) hydro, available from ENTSO-E Transparency Platform.

River discharge. Sub basins' river discharge data was provided by SMHI European hydrological model E-HYPE (Europe Hydrological Predictions for the Environment). The dataset contains information on 35408 sub basins and their geometry. For the reference period, the dataset has a daily resolution from Jan 1st, 2015 to Jan 31st, 2023. Daily river discharge projections are available from 2025 to 2100 for 9 EURO-CORDEX model ensembles under RCP2.6, RCP4.5 and RCP8.5.

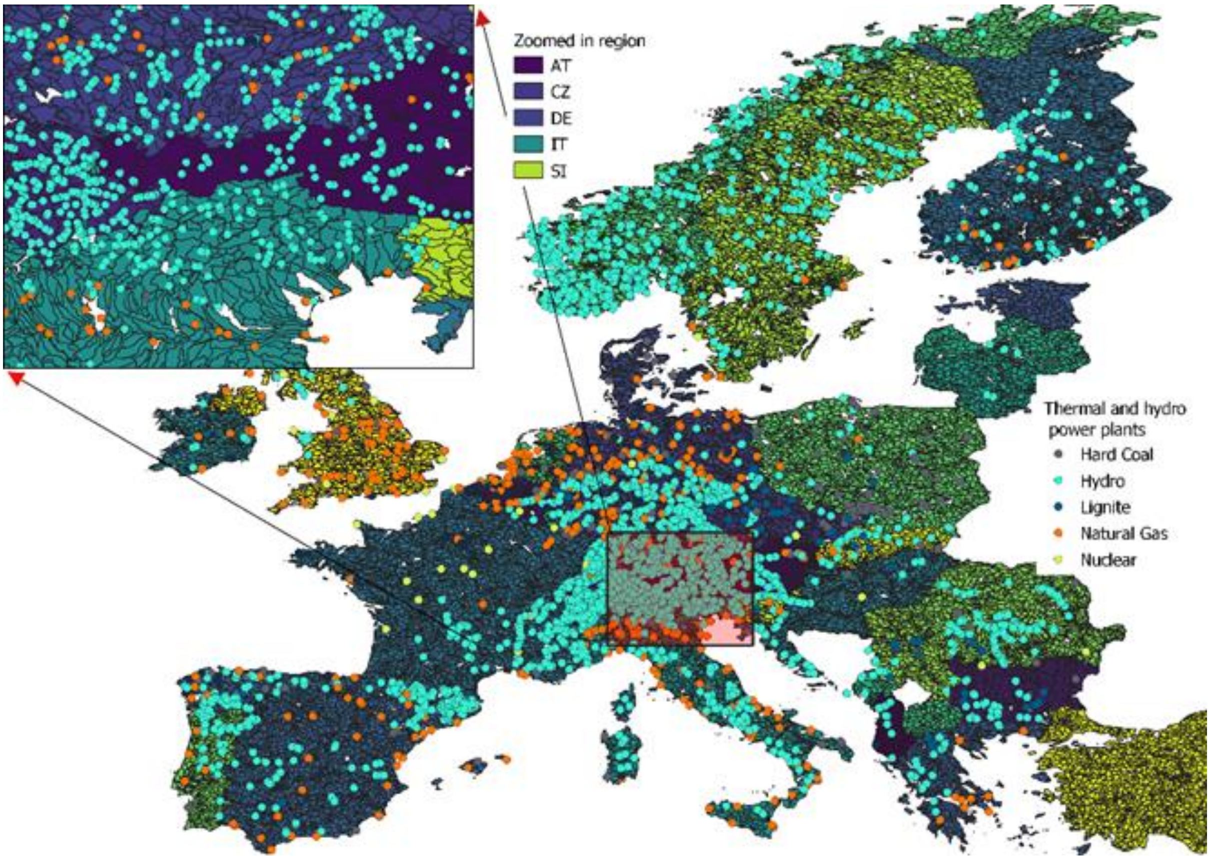


Figure 3-13 E-HYPE model subbasins (catchments) and hydropower generation plants.

Source: catchments: SMHI shape file; power plants' locations: hydro: JRC Hydro-power plants database, thermal power plants: JRC Open Power Plants Database (JRC-PPDB-OPEN) based on v. 1.0, PLATTS (2020).

Each subbasin in the E-HYPE dataset was matched to a given European country. As the E-HYPE data contains sub basins outside the boundaries of the Eurostat data on hydro power generation (i.e. Russia, Turkey, etc.), the number of sub basins reduced from 35408 to 24597.

ROR power generation. Data on power generation from run-of-river (ROR) hydro power plants, available from ENTSO-E Transparency Platform, contains hourly data on ROR power generation in MWh, aggregated by each country for a period from 2015 to 2023. This dataset is resampled to daily frequency to match the daily resolution of the E-HYPE river discharge dataset.

The time series of the ROR power generation dataset was shortened to eliminate an impact from structural changes identified in the data (such as changes in reporting format). These changes were visible in the first few years of ENTSO-E data collection compared to later years. Instead of 1st Jan 2015, the dataset starts at 1st Jan 2017. Consequently, the training dataset for machine learning methods will be smaller.

Data pre-processing. Several steps of data pre-processing were carried out to prepare the data for machine learning methods. For the reference period, from 1st Jan 2017 to 31st Jan 2023, both E-HYPE and ENTSO-E data needed to be aligned before being used for training the machine learning regression models. Both E-HYPE and ENTSO-E datasets were processed, to identify missing entries. If found, the observation was removed from both the river discharge and ROR generation datasets. The data on sub basins and generation is split into country-specific samples.

Methodology: machine learning methods

Two machine learning methods were explored to find the best fit: XGBoost and Artificial Neural Networks (ANN). XGBoost (also known as Extreme Gradient Boosting) is an ensemble learning technique and more specifically a version of Gradient Boosting models that add sequential models to correct the performance of the previous model. ANN, on the other hand, uses artificial neurons to find the relationship between the input and output data. It can contain multiple layers of neurons and uses weights to calibrate the strength of each neuron. Backpropagation is used to optimize the weights over multiple iterations.

A supervised machine learning regression model can be used to find the relationship between the multivariate features dataset (in our example this is river discharge for each subbasin) and the target variable (ROR hydro generation). After preprocessing the data, the discharge of each subbasin in each country becomes part of the country's model, while the aggregated ROR generation in the same country is used as the target variable. The dataset was split into a training and testing sample, where the testing sample accounted for the last 20% of days in the dataset. The training sample was then shuffled to reduce autocorrelation and consequently the chance of overfitting.

The two models, ANN and XGBoost, were fitted and their performance was compared for all countries at European scale. To obtain the optimal parameters for the XGBoost model, a grid search was undertaken[6]. Austria was chosen as the country to find the optimal hyperparameters, due to its medium size with a significant amount of run-of-river and poundage hydro generation capacity (5.25 GW in 2020, see Table 3-2). Comparing default parameters with hyperparameters optimized for Austria, it was observed that default parameters performed better in general across all countries than hyperparameters optimized for Austria. Therefore, all countries were given the default XGBoost parameters.

At the country level, the features importance method in XGBoost was used to determine the sub basins with the best predictive power for hydro power generation. The most important sub basins were plotted on a map using GeoPandas together with the locations of ROR hydro generators available from the JRC Hydro power plants database.

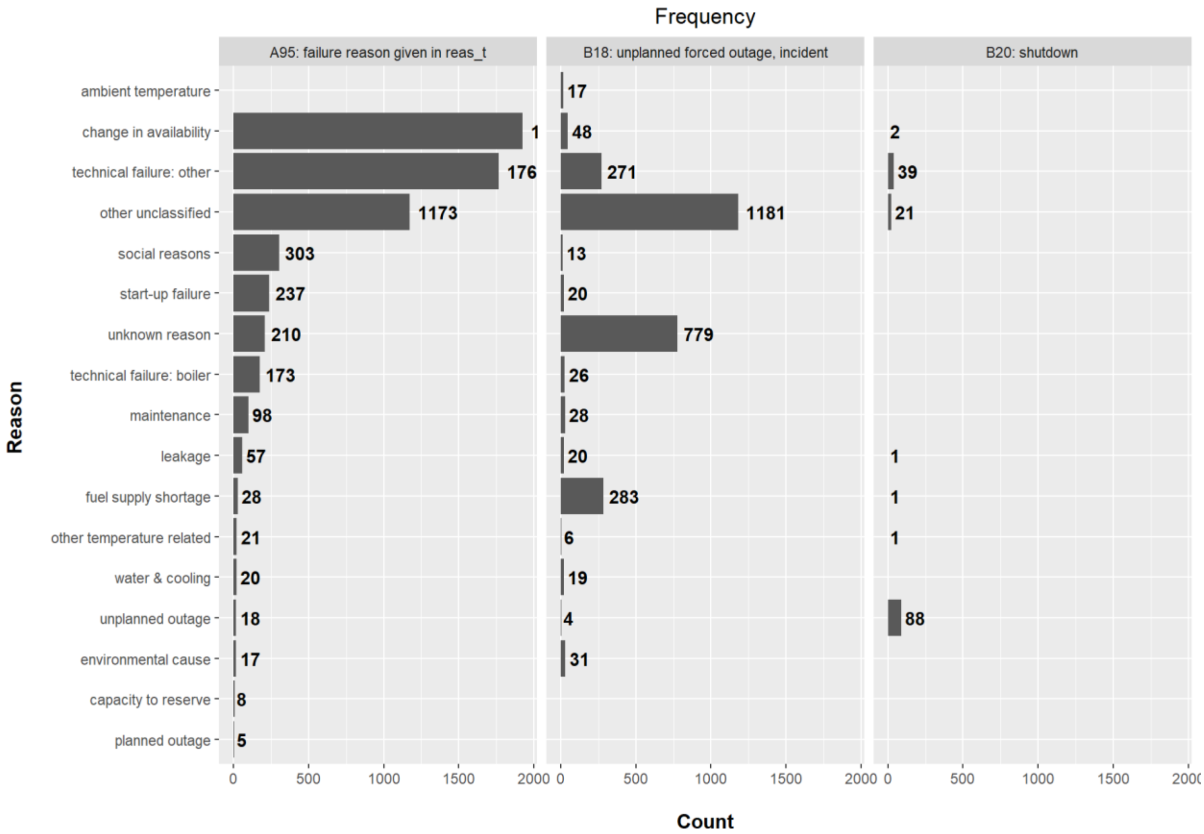
For the ANN model, a sequential model was used with an input layer of nodes equal to the number of sub basins in a country, and a rectified linear unit (ReLU) activation function was applied. A single output node was added with a linear activation function. The Adam[7] optimizer was used to compile the model using the mean squared error to minimize the loss function.

Limitations of the method

There may be other external factors affecting the results that cannot be explained by changes in the river discharge. For example, generator outages can be caused by exogenous impacts not related to weather conditions. Planned outages for maintenance, technical failures, start-up failures, etc., are completely exogenous and will impact the generation levels of the plants, Figure 3-14. Non-ROR-dominant countries with few ROR hydro generators, such as the Czech Republic, will likely be more affected.

[6] The *RandomizedSearchCV* module from scikit-learn (Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011) was used searching over the following hyperparameters in the XGBoost model: learning rate, number of estimators, max depth and alpha. The learning rate determines the step size when searching for the minimum point of the loss function. The number of estimators is the number of boosting rounds in XGBoost. The max depth is the maximum tree depth of the decision tree. The alpha hyperparameter determines the L1 regularization term sensitivity, where a higher number makes the model more conservative. The grid search was done using a 5-fold cross-validation sample. Based on this search, the optimal hyperparameters for Austria were learning rate = 0.1, max depth = 2, alpha = 5, and number of estimators = 300.

[7] The Adam optimizer stands for Adaptive Moment Estimator.



Note: Unplanned outages of the production units available from ENTSO-E Transparency platform from 2015-12-31 to 2022-03-31 for the European bidding zone. A95, B18, B20 are code reasons of the unplanned outages as given by the operator to ENTSO-E, the full list of codes is available at ENTSO-E Code List, Apr 28, 2023, Version 88.

Figure 3-14 Frequency of unplanned outages in the European bidding zone from 2015 to 2022.

Additionally, changes in the ROR generation capacity in a country can impact the training of the model and future predictions. Changes to ROR capacity that occur within the ML model training period, can bias the results. And changes in future capacity can bias the future predictions, if for example, the Czech Republic is to double ROR capacity from between 2020 and 2040. To overcome this problem, a version of the ML model has been trained on the ROR generation data normalized by the installed ROR capacity in the historic period. For each country, future ROR generation patterns can be derived in relation to the predicted normalized ROR generation.

Another limitation of the results is that the future predictions (2025-2100) are over a much longer time horizon than the historical data on ROR generation available (2017-2023). A longer time horizon of training data could make the ML model more robust (De Felice, 2020). As part of the ENTSO-E generation dataset, pondage generation is also included, so the ROR pattern for training can also include capacities that store water to maximize revenues. There is also a concern regarding the data quality of the ENTSO-E Transparency Platform, where there are missing values in some generation datasets (Hirth, Muhlenpfordt, & Bulkeley, 2018).

4.2.2. Results – constructing hydro power generation profiles

This section presents the results and performance of the models. Both the results from XGBoost and Neural Networks model are presented for comparison. The mean absolute error (MAE) and normalized root mean squared error (NRMSE) are used as metrics to evaluate and compare the two models. The NRMSE is normalized by dividing by the difference between the maximum and minimum value observed in the test dataset.

Table 3-2 below shows the metric for each country. When using the NRMSE as the metric, ANN performs better on 56% (8 out of 18) of countries, while XGBoost performs better for the remaining 44% countries (8 out of 18).

Table 3-2 Metrics on test data of different models.

Country	MAE (XGBoost)	NRMSE (XGBoost)	MAE (Neural networks)	NRMSE (Neural networks)	NRMSE (XGBoost, weekly)	NRMSE (Neural Networks , weekly)	Preferred model (NRMSE, daily)
Austria	353.96	12.03%	392.20	11.96%	14.08%	17.17%	ANN
Belgium	11.10	31.49%	10.43	29.50%	32.19%	24.45%	ANN
Croatia	44.85	29.72%	40.91	26.50%	38.43%	37.69%	ANN
Czech Republic	27.64	27.67%	28.87	29.52%	30.16%	32.34%	XGBoost
Finland	243.02	18.58%	302.00	22.21%	15.85%	34.04%	XGBoost
France	607.44	15.05%	487.67	12.24%	16.79%	14.37%	ANN
Germany	291.14	30.59%	151.43	16.99%	36.44%	30.68%	ANN
Hungary	2.19	18.53%	2.33	18.57%	22.86%	21.75%	XGBoost
Ireland	22.40	17.20%	25.47	18.64%	-	-	XGBoost
Italy	612.68	21.86%	422.90	15.25%	26.28%	31.98%	ANN
Latvia	94.15	12.38%	118.88	15.08%	11.54%	13.18%	XGBoost
Lithuania	17.30	29.90%	14.58	24.03%	26.01%	21.61%	ANN
Poland	15.74	8.53%	39.66	19.35%	9.89%	34.88%	XGBoost
Portugal	251.03	14.24%	260.96	16.38%	12.45%	45.67%	XGBoost
Romania	164.81	13.67%	142.79	14.71%	14.21%	49.52%	XGBoost
Slovakia	110.62	24.51%	103.94	22.79%	31.20%	30.38%	ANN
Slovenia	196.53	24.17%	169.82	19.76%	26.93%	61.73%	ANN
Spain	353.96	12.03%	392.20	11.96%	26.28%	31.98%	ANN

Based on the results, we define countries into groups of ROR-dominant and non-ROR dominant countries for further analysis. We define a ROR-dominant country where ROR represents more than 5% of total generation or more than 0.5 GW of ROR capacity in the reference year 2020. The differences between the models for ROR-dominant and non-ROR-dominant countries are discussed below. The performance of the models for each country is presented in the Annex 1.

Table 3-3 Run-of-river generation share and capacity in each country (2020)

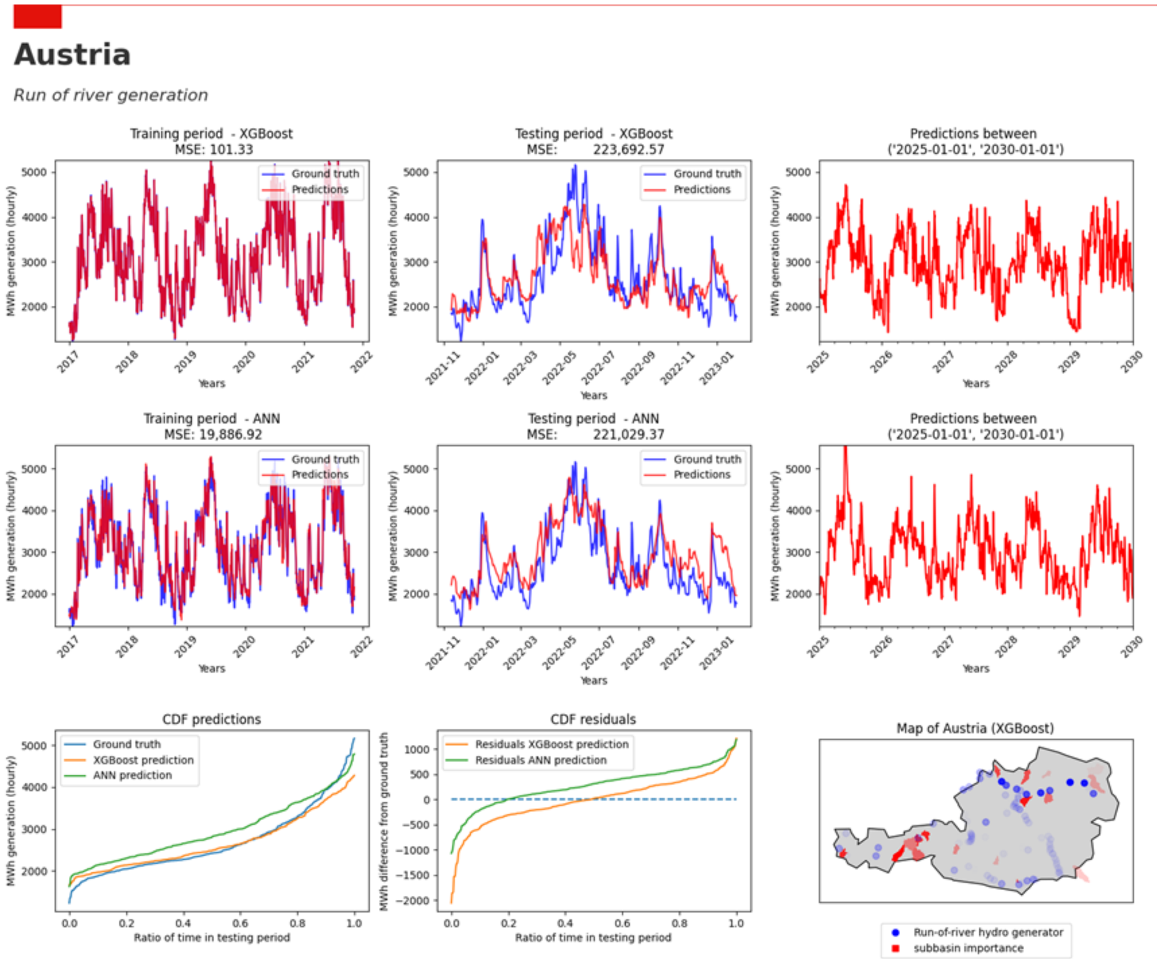
Country	ROR % of overall generation (2020)	ROR capacity (GW, 2020)	Grouping
Austria	27.6%	5.25	ROR-dominant

Belgium	0.4%	0.09	Non-ROR-dominant
Bulgaria	0.4%	0.31	Non-ROR-dominant
Croatia	20.3%	0.93	ROR-dominant
Czech Republic	1.8%	0.29	Non-ROR-dominant
Finland	11.4%	1.85	ROR-dominant
France	1.4%	1.71	ROR-dominant
Germany	2.4%	4.01	ROR-dominant
Hungary	0.7%	0.06	Non-ROR-dominant
Ireland	0.0%	0.00	Non-ROR-dominant
Italy	7.0%	3.12	ROR-dominant
Latvia	48.2%	1.59	ROR-dominant
Lithuania	13.0%	0.12	Non-ROR-dominant
Poland	0.9%	0.48	Non-ROR-dominant
Portugal	13.9%	3.05	ROR-dominant
Romania	15.5%	3.62	ROR-dominant
Slovakia	2.3%	0.15	Non-ROR-dominant
Slovenia	3.9%	0.21	Non-ROR-dominant
Spain	6.4%	4.51	ROR-dominant

Focus on ROR-dominant countries

The run-of-river hydro-dominant countries were identified: Austria, Croatia, Finland, France, Germany, Italy, Latvia, Portugal, Romania, and Spain.

In the example below, we demonstrate how the method is implemented for Austria, which is a ROR-dominated country. Reports on other countries in the group are available in the Annex 1. Figure 3-15 below presents the XGBoost and ANN results for Austria, using daily data for the period from 2025 to 2030. The first row shows the training, test and future prediction using the XGBoost model. The predictions for future ROR generation in each country was prepared using the dataset as an input to the ML model. The second row presents the same plots using the ANN model. The third row shows the cumulative distribution function (CDF) of the ground truth and the two predictions (XGBoost and ANN), the cumulative distribution function of the residuals, and a map of Austria with its ROR generators in blue and the most relevant subbasins (red) in the XGBoost model. The table in the last row of Figure 3-15 demonstrates that ANN has lower RMSE and NRMSE, but higher MAE, than XGBoost. The cumulative distribution functions (CDF) show that the ANN model is upwards biassed on the test data, where more than 80% of the predictions are above the ground truth. The XGBoost model predicts roughly 50% of predictions above ground truth and is thus more balanced in its prediction in relation to the ground truth. The map on the Figure 3-15 below, shows that the XGBoost model picks up a diverse set of sub basins located in the west, north and east of the country to explain the country-aggregated ROR target variable. These sub-basins are located in the mountainous regions of Austria.



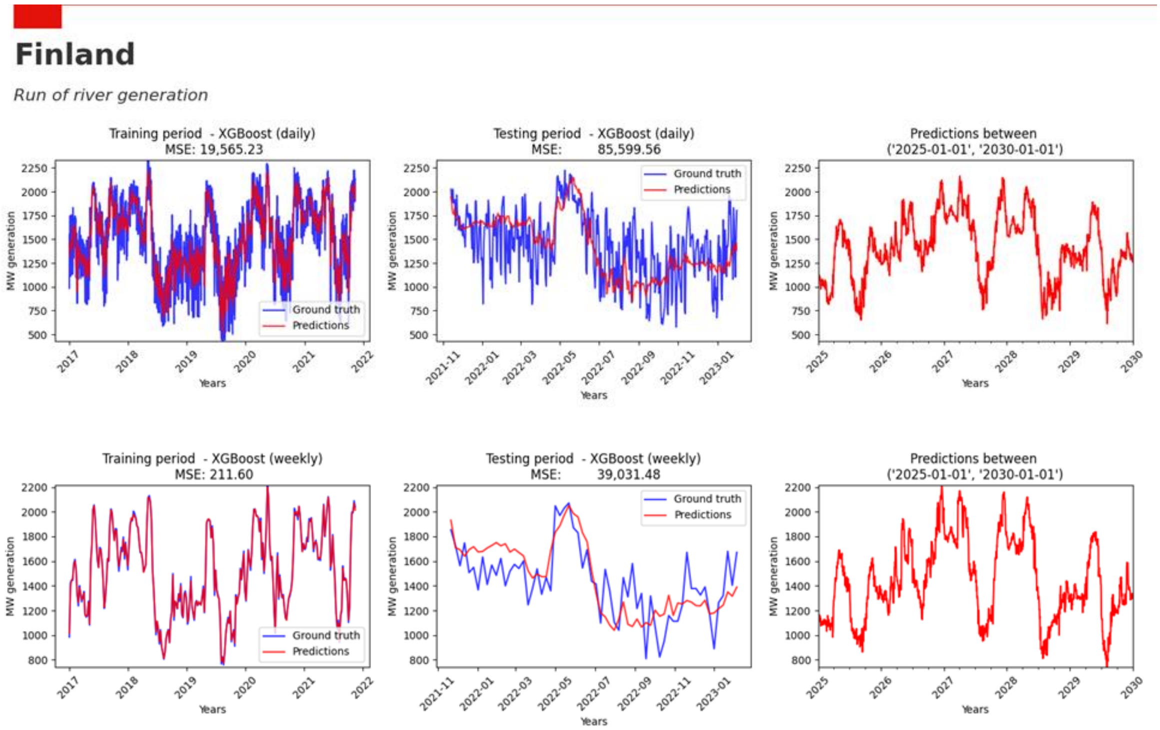
Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	7.59706	353.96236	104.85891	392.19504
MSE	101.33155	223692.57067	19886.92346	221029.37423
RMSE	10.06636	472.96149	141.021	470.13761
NRMSE	0.24909%	12.02741%	3.48952%	11.9556%

Note: The first row shows the training, test and future prediction using the XGBoost model and the second row - for the ANN model. The third row shows the CDF of the residuals. The last plot in the third row shows the map of Austria, and locations of ROR generators in blue dots (hue intensity is scaled by the installed ROR generation capacity) and the most relevant sub basins in red (hue intensity is scaled by the importance feature in the ML model) identified by the XGBoost model. A table of relevant metrics for training and test data is shown in the fourth row.

Figure 3-15 Results from XGBoost and Neural Networks for Austria using daily data.

Another ROR-dominant country is Finland. Finland has large variations in its generation at a daily resolution, which is explained by its large lakes used to regulate the rivers (De Felice, 2020). It can then optimize the river flow based on the prevailing prices in the spot market.

Figure 3-16 shows the comparison of the XGBoost model using daily and weekly data. The MSE is significantly lower in the weekly averaged data compared to daily data, and from a visual inspection it is clear that the model cannot explain the daily variation using exogenous river flows.



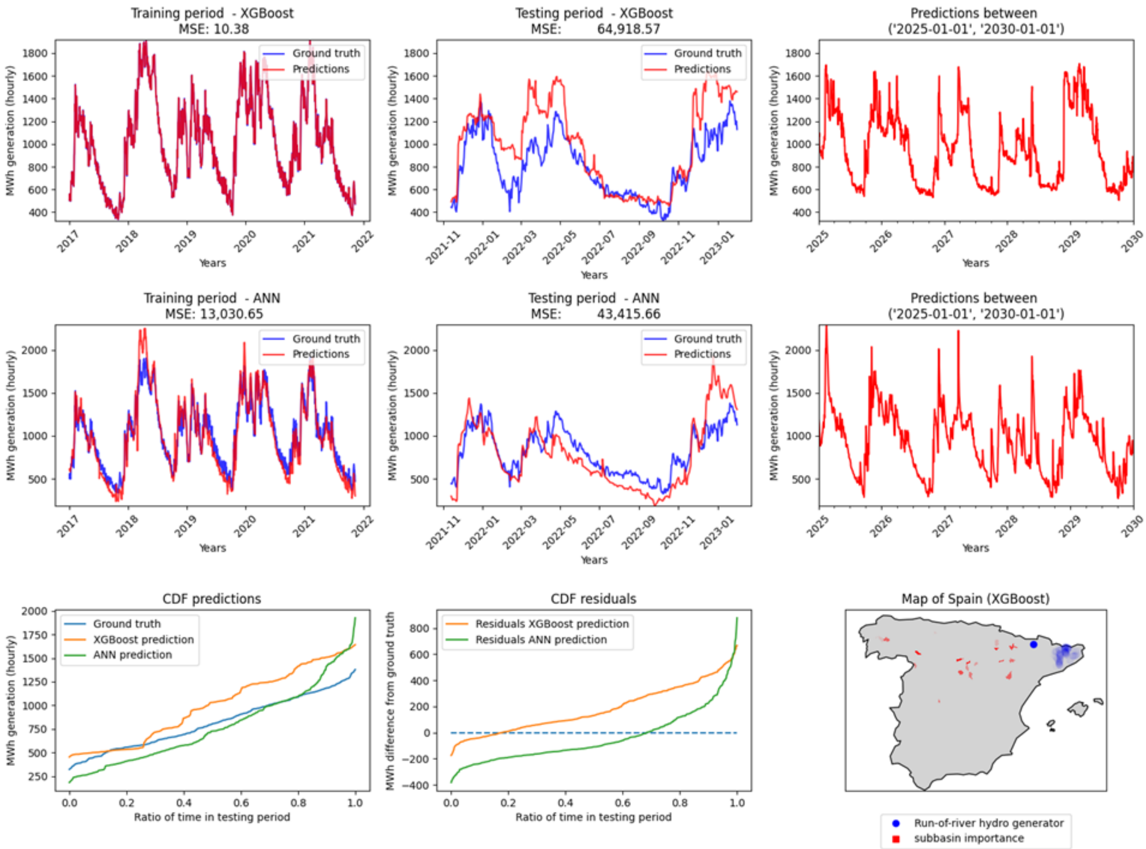
Note: The first row shows the training, test and future prediction using the XGBoost model and the second row - for the ANN model.

Figure 3-16 Daily and weekly simulation of ROR generation for Finland.

Spain, also a ROR-dominant country, shows a different performance than Austria comparing ANN and XGBoost model results in the testing period. Figure 4-15, table in the third row, demonstrates that ANN has lower RMSE, NRMSE and MAE than XGBoost. The cumulative distribution functions (CDF) show that the XGBoost model is upwards biassed on the test data, where more than 80% of the predictions are above the ground truth. Roughly 70% of the ANN model predictions are above the ground truth. The map on the Figure 3-15 below, shows the location of the set of sub basins located in the eastward to the country ROR capacities, revealing the likely bias of the selected river catchments to the ones located within the national borders, in the valley of the river Ebro and adjacent regions of Pyrenees.

Spain

Run of river generation



Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	2.32565	196.52812	87.92108	169.81703
MSE	10.38321	64918.56613	13030.65296	43415.66233
RMSE	3.2223	254.79122	114.15189	208.36425
NRMSE	0.20357%	24.16801%	7.21149%	19.76422%

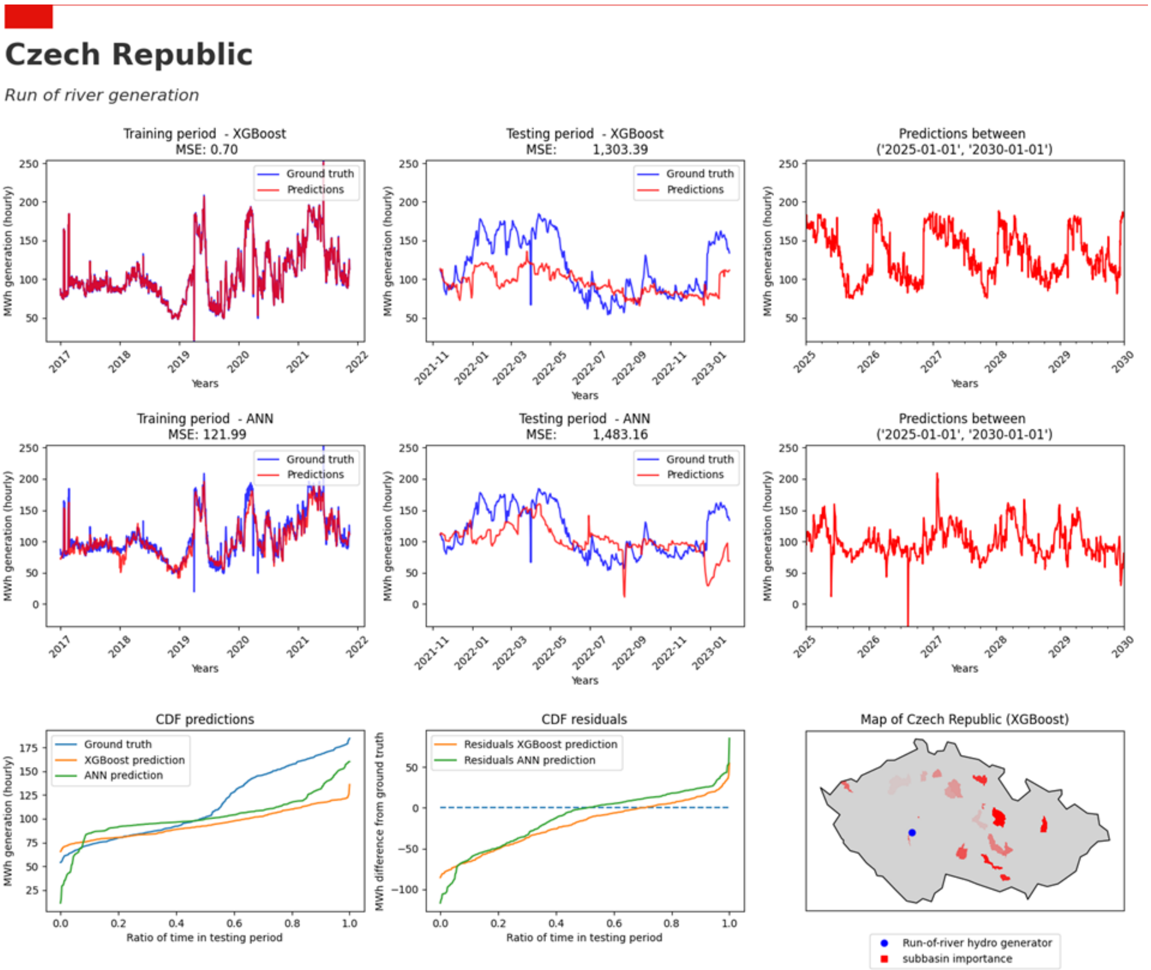
Note: The first row shows the training, test and future prediction using the XGBoost model and the second row - for the ANN model. The third row shows the CDF of the residuals. The last plot in the third row shows the map of Austria, and locations of ROR generators in blue dots (hue intensity is scaled by the installed ROR generation capacity) and the most relevant sub basins in red (hue intensity is scaled by the importance feature in the ML model) identified by the XGBoost model. A table of relevant metrics for training and test data is shown in the fourth row.

Figure 3-17 Results from XGBoost and Neural Networks for Spain using daily data.

Focus on non-ROR-dominant countries

Group of the non-ROR-dominant countries include Belgium, Bulgaria, Czech Republic, Hungary, Ireland, Lithuania, Poland, Romania, Slovakia, and Slovenia.

The Czech Republic serves as an example of a non-ROR dominant. country. Figure 3-18 shows the models' performance. The model has a good fit on the training data, but it does not perform particularly well on the test data. There is only one significant ROR hydro generator, see the map of the Czech Republic on Figure 3-18 third row, while the XGBoost model picks up subbasins from around the country. It is possible that the model is overfitting. This model has identified a pattern from the surrounding sub basins that can predict the one ROR hydro generator, but it has not identified the single catchment that this generator depends upon, and therefore the results are not generalizable out of sample (i.e. in the test period). The models for non-ROR-dominant countries are more difficult to fit and are likely to overfit in the training period. The model attempts to find a combination of sub basins to explain the few run-of-river hydro generators in the training sample with a low level of accuracy.



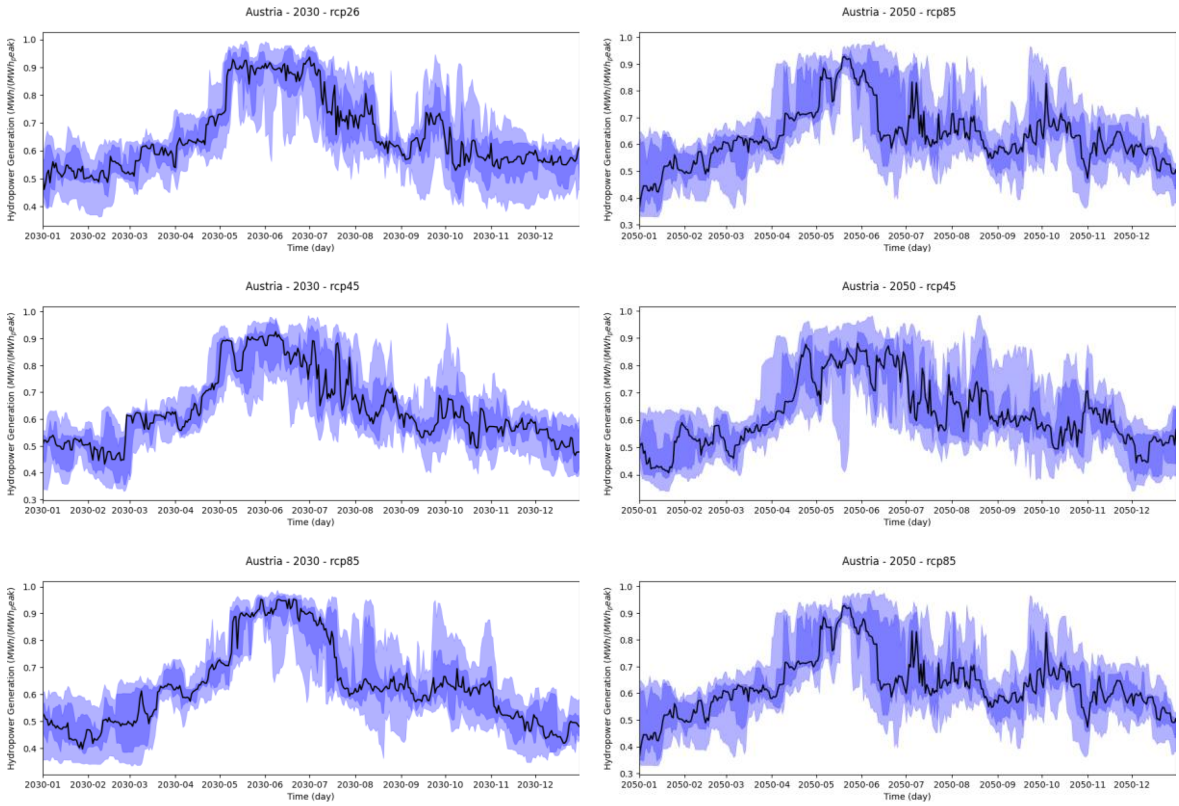
Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.61694	27.63753	7.42213	28.87376
MSE	0.69564	1303.3946	121.98932	1483.15602
RMSE	0.83405	36.10256	11.04488	38.51176
NRMSE	0.356%	27.67496%	4.71433%	29.52177%

Note: The first row shows the training, test and future prediction using the XGBoost model and the second row - for the ANN model. The third row shows the CDF of the residuals. The last plot in the third row shows the map of Austria, and locations of ROR generators in blue dots (hue intensity is scaled by the installed ROR generation capacity) and the most relevant sub basins in red (hue intensity is scaled by the importance feature in the ML model) identified by the XGBoost model. A table of relevant metrics for training and test data is shown in the fourth row.

Figure 3-18 Daily simulation of ROR generation for Czech Republic.

Focus on future climate - RCP2.6, RCP4.5 and RCP8.5

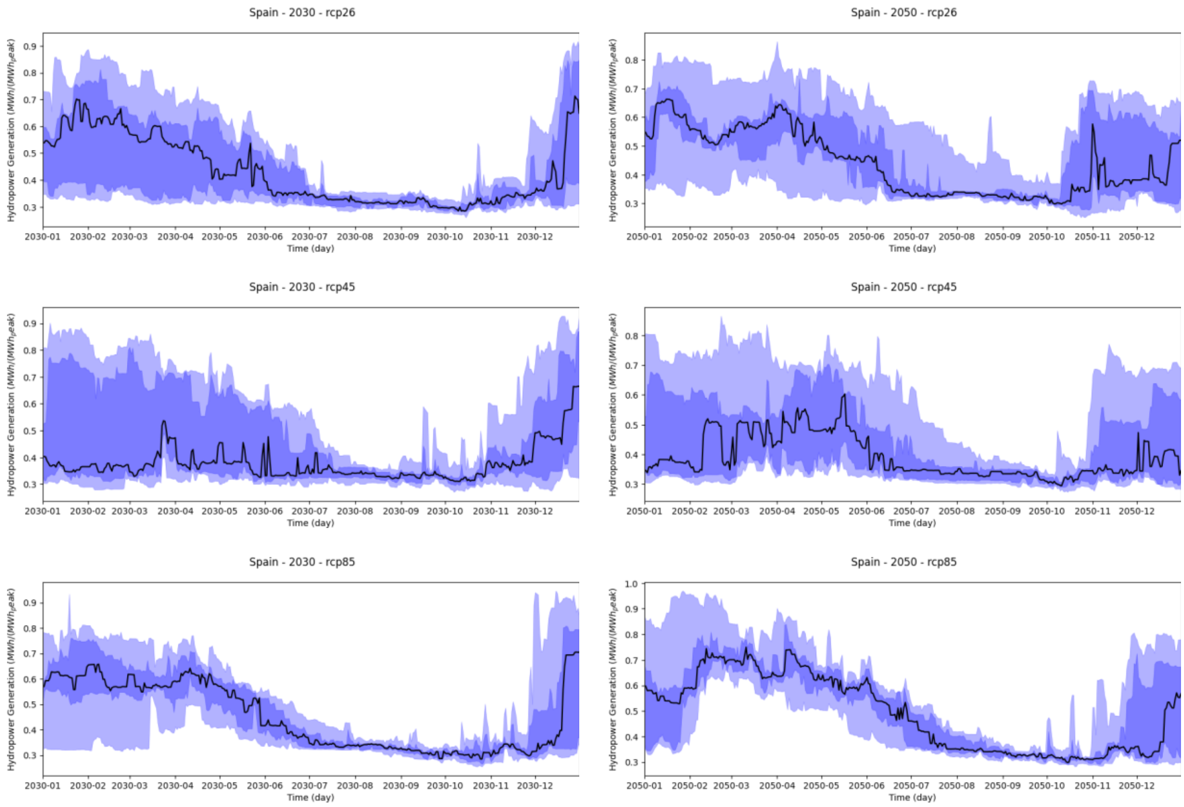
The future projections of ROR generation patterns are available for the period from 2025 to 2100 for RCP2.6, RCP4.5 and RCP8.5. The results were simulated for the 9 EURO-CORDEX model ensembles. Below, Figure 3-19 and Figure 3-20 show the future projections of ROR generation patterns for Austria and Spain.



Note: Black line shows the median ROR generation projected by ML method, based on the ensemble of 9 EURO-CORDEX models' simulation of river discharge. Light blue area represents the 5th to 95th percentile range, and the 25th to 75th percentile is represented by the dark blue shaded region.

Figure *3-19 Normalized ROR generation patterns for 2030 and 2050 years for Austria under RCP2.6, RCP4.5 and RCP8.5.*

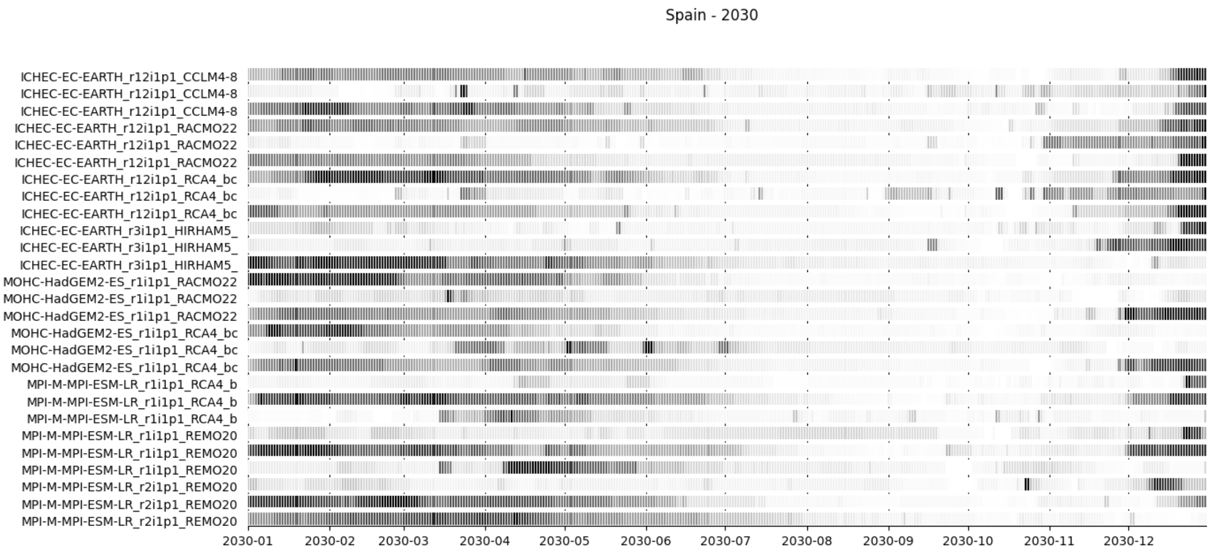
The uncertainty in the ROR generation patterns can be described by analyzing the EURO-CORDEX model ensembles for each of the three RCPs. For Austria (Figure 3-19), model's ensembles indicate rather high agreement, with narrow ranges for the 5th to the 95th percentile range, and the 25th to the 75th percentile range respectively. Whereas for Spain (Figure 3-20), especially in the RCP2.6 and RCP4.5 for 2030, changes in the ROR generation patterns indicate lower agreement between the models and higher uncertainty in realization of the climate impacts on hydro power.



Note: Black line shows the median ROR generation projected by ML method, based on the ensemble of 9 EURO-CORDEX models' simulation of river discharge. Light blue area represents the 5th to 95th percentile range, and the 25th to 75th percentile is represented by the dark blue shaded region.

Figure 3-20 Normalized ROR generation patterns for 2030 and 2050 years for Spain under RCP2.6, RCP4.5 and RCP8.5.

In the next steps, the ML-enhanced dataset for each country's ROR generation profiles will be used in the PRIMES-IEM model. The prepared energy scenarios include different RCP scenarios: RCP2.6, RCP4.5 and RCP8.5. Additionally, to explore the uncertainty and realization of extreme ROR generation patterns, it is essential to include scenarios with the uncertainty ranges of the models' ensemble: for the 5th to the 95th percentile range. To maintain consistency of the ROR generation patterns within a year, on a daily basis, the selected extreme scenarios are based on the river discharge future projections of the models in the ensemble (Figure 3-21).



Note: Shades of black color indicate the value for each daily normalized ROR generation, with white color equal to lower end of the scale (0) and black to higher end of the scale (1).

Figure 3-21 Median daily ROR generation in Spain for 2030, for the EURO-CORDEX model ensemble, for RCP2.6.

Conclusions and further work

Further improvements can be done to validate the findings and increase the robustness and confidence in the ML model trained. Understanding the biases of the model can provide more insight into its performance. The use of black box models like ANN also suffers from interpretability. More research into the interpretation of AI methods can provide further insight into the decisions made by the ANN model. This is important when relying on complicated models for policy making.

Further work can also be done to optimize the parameterization of the model for each country. This is done using the Adam optimizer for Neural Networks, but the XGBoost method relies on the default parameters. It is possible that the XGBoost fit could be improved by undertaking a grid search for each country and using the country-specific optimized hyperparameters. Moreover, more lagging features will be explored as it is expected that they can increase model accuracy.

More exploration into cross-border effects, especially for countries that share river basins with other neighboring countries, may provide additional insights. For example, including sub basins in countries near the border of a country could be explored to see if those sub basins have additional predictive power. This further work can contribute to assessment of the spatial coordination of the hydro power production (Wörman et al., 2020).

The following work steps within the CLINT project include the elaboration of the PRIMES-IEM scenarios that consider climate-driven changes in hydro power generation and show the role of hydro power in the decarbonized energy sector.

4 INVESTIGATION ON AI-ENHANCED CLIMATE SERVICES FOR EXTREME IMPACTS - FOOD SECTOR

4.1 AI-BASED CROP GROWTH EFFORT - BACKGROUND AND OBJECTIVES

The main idea behind the crop model emulator development is to enhance the skill of crop yield prediction when it comes to integration of seasonal climate forecasts. In contrast to the mechanistic crop model, which generally relies on six input meteorological variables (daily minimum and maximum temperature, precipitation, wind speed, global solar radiation and relative humidity), the crop model emulator will rely on three input variables: daily minimum and maximum temperatures and daily total precipitation. Minimum and maximum temperatures are chosen here due to their relevance for crop growth, as they better capture the impact of cold and heat stress, respectively, than daily mean temperature. With a lower number of input variables, we aim to reduce the forecast uncertainty arising from integration of seasonal forecasts of multiple meteorological variables, originally required for the mechanistic model. Furthermore, running the crop growth model is computationally expensive, making the model infeasible for producing large ensembles for which the emulator can be used as well.

In this stage, we are aiming in training and calibrating a crop growth emulator using the ERA5 reanalysis dataset from ECMWF. ERA5 serves as the observational dataset that is utilized in terms of predictor variables in order to train and test the AI surrogate model.

The initial step of the overall process is the actual crop growth simulation for the target years for the chosen crop, grain maize. The input data for the ECroPS model are derived from the Joint Research Center of the European Commission (EC-JRC).

In order to perform the simulations, we feed the ECroPS model with the following gridded input data:

Weather The daily ERA5 variables used for the simulations are maximum temperature, minimum temperature, shortwave downwelling radiation at the surface, precipitation, relative humidity, and wind calculated by its vector components. ECroPS requires three types of water balance variables: potential evaporation from a free water surface, potential evaporation from a moist bare soil surface, both of which are calculated with the modified Penman approach, and the canopy evapotranspiration which is calculated with Penman-Monteith approach.

Soil The input data are gridded static data derived from the databases of JRC. According to the database details, the European soil is divided into soil mapping units (SMU) representing differentiated zones that reflect the soil types' homogeneity. The original soil database SMUs consist of one or more soil types, represented with soil typological units (STUs), with STUs occupying measured percentages per SMU. Overall, the provided soil map used for crop growth determine aspects such as the rooting depth, the available water capacity (AWC) and the infiltration capacity, variables that are associated to the STU and are therein introduced as soil moisture characteristics. Without losing the generalizability of our methodology, we avoid introducing the complexity of accounting for the possibly multiple STUs associated to every grid cell of the European domain, and thus requiring as many simulations as the STUs per grid cell. Instead, we perform the simulations once per grid cell regarding the soil type, and assume the most dominant one as the one to be associated with each corresponding cell.

Crop-specific data and agro-management traits per cell Sowing date and number of degree-days for the emergence-anthesis period and the number of degree-days for the anthesis-maturity period, vernalization parameters, dry matter partitioning parameters and many others, required for the simulation runs for grain maize.

We consider the crop and soil parameters static while ERA5 variables are the time-varying components that force the ECroPS model for 31 years, from 1993 to 2023. Each cell output is an independent simulation, forced from sowing date to the end of each year, producing output variables such as the Total Weight of Storage Organs (TWSO) on a daily basis as the crop grows. We focus our interest on the crop growth development stages between flowering and maturity (Development Stage 1 (DVS1) to Development stage 2 (DVS2), in terms of the ECroPS characterization of crop development stages).

The objective of the reported advances is the surrogate modeling using an AI-based model of the crop growth between the development stages DVS1 and DVS2.

In this sense, the goal is to reproduce the underlying processes that dictate crop growth in terms of weather inputs. The model is forced with ERA5, which also determines the spatial resolution of the simulation, which is accordingly, the climate model's native resolution of 0.25 degrees.

4.2. METHODOLOGY – AI PIPELINE

4.2.1. Preprocessing of the data

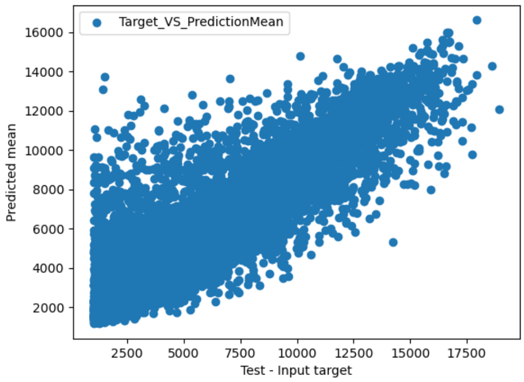
The database consists of the sample points deriving from each independent run of the ECroPS model for each grid cell and each year, for the defined European domain, accounting for the cells for which we have crop parameters, more specifically an actual sowing date.

The database contains N samples/ECroPS simulations structured as a datacube of crop growth timelines (K days) for the three chosen climate variables: minimum and maximum temperature and total precipitation.

4.2.2. Benchmarking and next steps

We performed benchmarking model runs with the full database, unfolded and flattened in space and time into a big tabular dataset. As a result, the database used for the benchmarking process is of the shape (N, K*3). This means that the feature space now consists of a few hundred features which are structured as {Day0Tmax, Day0Tmin, Day0Pr, …, DayKTmax, DayKTmin, DayKPr}.

The tests are performed using ensembles of Random Forests and ensembles (15 members) of XGBoost runs, utilizing in the first run greedy grid searches for the hyperparameters, choosing the best performing ones in order to optimize the output predictions.



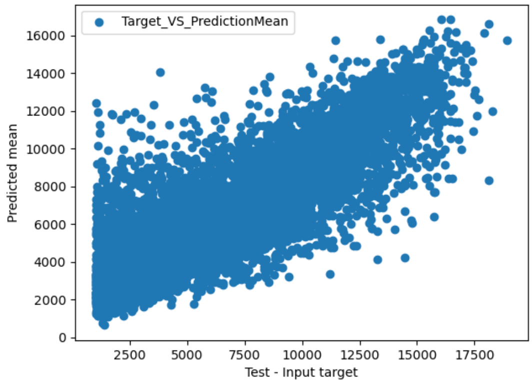


Figure *4-1* Random Forests (left) and XGBoost (right) performance. Target values VS the average of the ensemble outputs.

Since the feature space is high dimensional, thus imposing problems in terms of noise and redundancy, we perform a series of feature selection steps to 1) minimize noise, 2) condense the feature space, 3) provide quicker and more robust convergence for the AI pipeline and 4) provide a more explainable side to the problem and the solution.

Our initial trials utilized the Recursive Feature Elimination with Cross-Validation (RFECV) algorithm (Guyon et al., 2002).

We tested the RFECV with Extreme Learning Machines (ELMs) (Huang et al., 2006), using 5-fold cross-validation, a step of one feature removal per iteration and negative MSE as the metric of the model performance. RFECV is an iterative feature selection technique that recursively removes features from the dataset while evaluating the model's performance using cross-validation (CV). It starts by training the model on the full feature set for computing a performance metric (e.g., accuracy, mean squared error) using CV. Then, it ranks the features based on their importance (e.g., coefficients in linear models, feature importances in tree-based models) and eliminates the least important feature. This process is repeated iteratively until the desired number of features is reached or until the performance metric stops improving. In our case, RFECV did not converge to a solution most probably due to some common pitfalls of the algorithm:

1) very high dimensionality: In datasets with a high number of features, RFECV may encounter difficulties in finding an optimal subset of features due to the large search space.

2) noisy data: If the dataset contains noisy or irrelevant features that do not contribute useful information to the model, RFECV may struggle to converge to an optimal subset of features. Noise in the data can obscure the true signal and make it challenging for RFECV to distinguish between informative and non-informative features.

3) lacks diversity in its search: RFECV may become trapped in a local minimum or fail to find a globally optimal solution because it doesn't adequately explore all possible solutions. Instead, it may focus too much on a limited subset of the search space, potentially missing better solutions that exist in other regions.

In continuation, we tried an optimization pipeline for feature selection and tried to optimize an objective function using the minimize functions from scikit-optimize[8]. In the context of feature selection, this function typically evaluates the performance of a model using a specific subset of features. The minimize function searches for the subset of features that minimizes (or maximizes) the performance metric (here, negative MSE) defined by the objective function. The result provides information about the optimized subset of features and the corresponding performance metric. We tested the 'minimize' optimization approach with ELMs using a 5-fold cross-validation scheme, a maximum evaluation value of 10000, and two methods, namely the Nelder-Mead and the Powell ones. Since for both derivative-free methods the algorithms did not converge to a solution that satisfies the convergence criteria and did not find a minimum (or maximum) within the given iterations, we moved to more complex feature selection operations.

The final feature selection actions are performed using the Coral Reef Optimization with Substrate Layers (CRO-SL) algorithm (Pérez-Aracil et al., 2023). The CRO-SL algorithm is a nature-inspired optimization algorithm developed based on the ecological principles of coral reefs and their substrate layers. Coral reefs are complex and diverse ecosystems with various layers, each supporting different species and playing a crucial role in the overall health of the reef. Inspired by this natural system, the CRO-SL algorithm aims to efficiently solve optimization problems by mimicking the dynamics of coral reefs and substrate layers. The CRO-SL algorithm initializes a population of potential solutions (corals) within the search space. Each coral represents a candidate solution to the optimization problem. The Coral Reef is formed by the interaction between the corals and their environment, mimicking the competition and cooperation observed in coral reef ecosystems. In such coral reef ecosystems, substrate layers provide a habitat for various organisms and contribute to the overall ecosystem's stability. Similarly, in the CRO-SL algorithm, substrate layers are created to represent different levels of solution quality. These layers are dynamically adjusted based on the performance of the corals. Through their reproduction and mutation operations, corals undergo genetic variations to explore new areas of the search space. This process helps in maintaining diversity within the population and prevents premature convergence to suboptimal solutions. Through evaluation and selection, the fitness of each coral is assessed based on the objective function of the optimization problem. Corals with higher fitness values are more likely to survive and reproduce, influencing the composition of the population in subsequent generations. Finally, environmental pressure and adaptation are performed within the selection to encourage adaptation, driving the search towards better solutions. The algorithm terminates when a termination criterion is met, such as reaching a maximum number of evaluations or achieving a satisfactory solution after effectively exploring complex solution spaces and finding high-quality solutions to optimization problems.

The optimization problem we tested is based on the backbone of ELMs runs, evaluated 10000 times using a 5-fold CV scheme, reaching on average 120 generations. This process is run 5 times and the final features selected are the ones that are chosen at least 4 times from all 5 runs, thus providing a more statistically robust feature subset.

[8] https://github.com/scikit-optimize/scikit-optimize/tree/master

The selected feature space is further reduced using Principal Components Analysis (PCA) in order to explore deeper the noise components in the dataset and evaluate the usefulness of tailored AI pipelines. The higher the noise, the more suitable a neural network-based solution may be and the higher the value of probabilistic approaches.

The PCA is conducted on the CRO-SL reduced feature space. From the principal components (PCs) we select the top-ranking ones that are responsible for explaining the 90% of the variability. Finally, from each of those selected PCs, we single out for our further experiments the top 5 features contributing to that PC, thus resulting with a minimal subset of unique features from the original feature space (Figure 4-2).

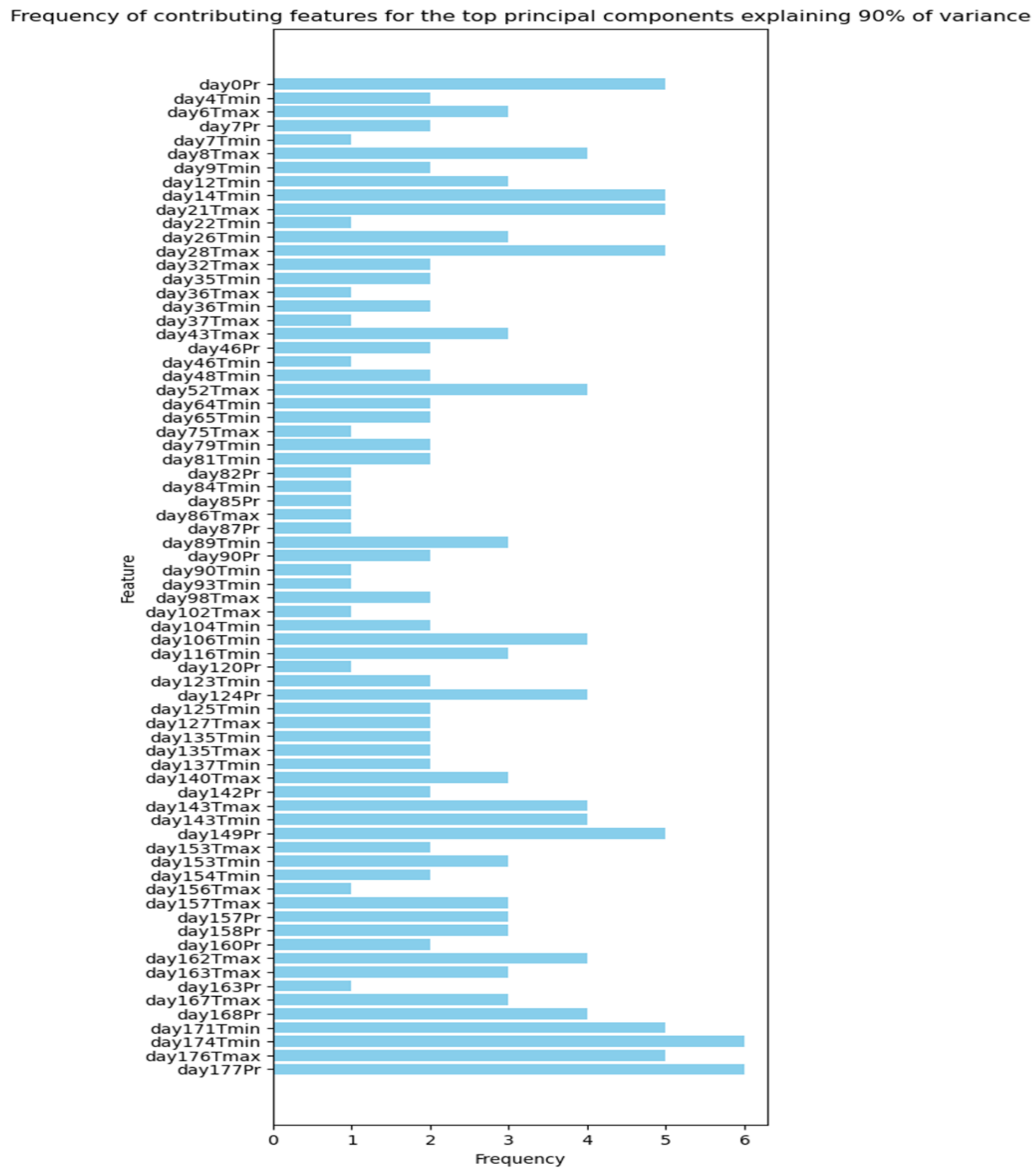


Figure 4-2 Frequency of the contributing features from all the PCs that are responsible for the 90% of the explained variance of the features subset from the CRO-SL process.

Since all the intuitive tests and data characteristics point to the need for irregular representations that cannot be learnt with ML algorithms, we naturally turn our focus to Neural Networks (NNs). After being introduced in the 1980s, NNs have become today one of the widest used AI toolsets due to their ability to learn highly non-linear and complex relationships from extensive datasets. The successful application of NNs was shown in a wide variety of different scientific fields with important achievements also in climate science (e.g. Schultz et al., 2021; Camps-Valls et al., 2021).

Finally, the new skeleton database is the input to a feedforward neural network (FNN) with 5 hidden layers, dropout regularization and batch normalization. The predictions of the FNN do not differ greatly in comparison to the RF and XGB ones (Figure 4-3), implying that a more complex and different network architecture is required to capture the processes.

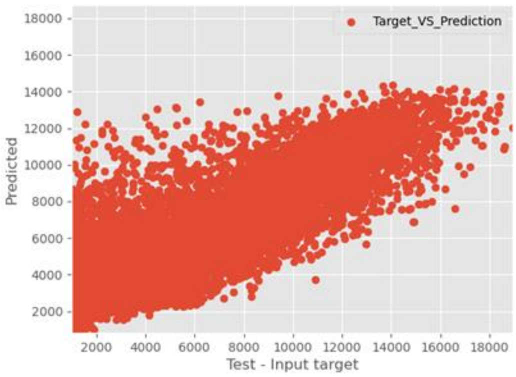


Figure 4-3 FNN scatter plot of test dataset inputs VS predicted outputs.

However, since we are still exploring the feature space in terms of tabular, flattened data, thus eliminating the contribution and the underlying processes that are related to the time domain and its value in crop growth emulation, it is of high interest to determine other AI pipelines that rely on recurrent neural network (RNN) architectures consisting of Long-Short-Term-Memory (LSTM) components as learning layers.

Our initial tests with RNNs are based on operations that allow the network to predict full timelines per sample (or cell). In this sense, the most recent tests require a database of timeline samples that are fed altogether to various shallow RNNs. Tests with up to 4 hidden layers with various neuron numbers per layer have failed in convergence, exhibiting gradient explosions or gradient vanishing phenomena. The database is constructed by the 3 original features, (min T_{min} and max T_{max} temperature, precipitation Pr) along with a feature engineering process that adds temporally lagging features for 5 days, thus effectively adding 15 new features to the feature space. These converging failures of these RNNs have pointed towards our future work involving RNNs. The designed pipeline evolves around a different strategy regarding the database construction and the RNN prediction process. The single-temporal prediction RNNs currently designed are going to predict single timesteps, instead of full timelines.

Concurrently, we are tailoring our efforts towards a very promising direction for constructing the crop growth emulator that may not only provide more consistent AI modeling but also quantify inherent uncertainties. While a NN can theoretically learn any nonlinear function and hence be considered a universal approximator (Efron & Hastie, 2016), in practice, prediction accuracy will always be dependent on the data set at hand, resulting in uncertainty. This is especially important for NNs as they are by their nature prone to overfitting thus questioning their generalization capabilities (Szegedy et al., 2013), which can result in overconfident forecasts making them potentially unsuitable for high-risk domains (Goan et al., 2020). Because of their high complexity, these models have been dubbed "black boxes" (McGovern et al., 2019), which have been found to learn spurious associations (Lazer et al., 2019) that do not necessarily reflect the physical relationships of the Earth system (Kashinath et al., 2021). As a result, determining the uncertainty of the predictions is critical for evaluating NNs. In statistics, this problem is frequently solved by including a stochastic noise component in the system to reflect the uncertainty. This approach has also been carried over to NNs either by using a stochastic activation function or by stochastic weights (Jospin et al., 2007), which motivated the class of Stochastic Neural Networks (SNNs).

A class of particularly popular SNNs has been introduced with Bayesian Neural Networks (BNNs), which is a stochastic neural network trained using Bayesian inference (Jospin et al. 2020). The Bayesian theory treats model parameters as sources of uncertainty (i.e., as random variables) and determines their distribution by specifying prior distributions based on experts' knowledge, which are then updated with the information in the observed data using Bayes theorem (van de Schoot et al., 2021). In BNNs, the Bayesian perspective is used to address some potential shortcomings of NNs by merging techniques of the NNs and Bayesian theory: this is essentially being accomplished by imposing a distribution across the network parameters, which is determined using methods from Bayesian theory. The BNN is then able to produce probabilistic predictions, which allows one to deduce the nature and distribution of these parameters (Mullachery et al. 2018), therefore giving insights into the network structure. Furthermore, the learned distribution of these parameters can be used to create an ensemble of multiple neural networks or potentially high-performing models and taking the average of these models (i.e. the ensemble average) can yield significantly improved accuracy in comparison to other modern deep learning models (see for more details, Wilson & Izmailov, 2020). These methods have also been successfully used in climate science to study, for instance, the hydrological impact of climate change (Khan & Coulibaly, 2010), to assist calibration of global climate models (Hauser et al., 2012), to fuse output from climate ensembles with observations (Amos et al., 2021) and for ensemble techniques of geophysical models (Sengupta et al., 2020).

Our initial tests with BNNs are limited to shallow architectures with dense variational layers, assuming multivariate gaussian prior distributions for the model parameters and gaussian posterior distributions for the learnt means, variances and covariances. However, we are still on a very first phase of experiments and have not yet managed to successfully train and predict with our naïve networks.

4.2.3. Conclusion and further work

In terms of initial surrogate modeling efforts, we can claim that RNNs show the most structurally consistent characteristics in terms of crop growth emulation due to their time-dependent nature. The current effort centers around testing simple RNN architectures with a different sample database preprocessing and shaping strategy. BNNs are also within the scope of our tests, and while the initial tests do not show promising results, we will pursue that direction as well.

5 SUMMARY AND MOVING FORWARD

In the rapidly evolving landscape of climate research, the integration of AI/ML in climate services is redefining how we comprehend and respond to climatic changes across various sectors. These services crucially depend on the availability and quality of data as used in the case studies and methods application in the current deliverable that is based on a wide spectrum of environmental parameters such as temperatures, river discharge and precipitation. There is a pressing need for access to robust national and international databases that not only provide historical averages and detailed risk assessments but also offer long-term projections and scenarios (Buontempo et al. 2018). This foundational data is essential for developing accurate and actionable climate services that can effectively inform policy and decision-making processes.

Furthermore, the interpretation of the highly heterogeneous data provided by AI/ML CS plays a critical role in climate-informed decision-making and the development of climate-smart policies and planning. AI/ML CS not only deliver detailed impact indicators and summary reports on climate change impacts but also enhance these insights with spatial context illustrations and robustness projections for each indicator. Additionally, AI/ML CS come equipped with user guidance materials designed to make complex climate information accessible to decision-makers who may not have specialized knowledge in climate science.

The methods and applications outlined in the deliverable represent preliminary findings that move us toward a deeper comprehension of the specific needs for CS and the extent of data required across the water, energy, and food sectors. Each sector has endeavored to identify the most relevant climate parameters necessary for effectively integrating climate data into their respective impact models. Additionally, these sectors have focused on selecting appropriate AI/ML methods to manage the complexity of the available data and its integration to the modeling framework. Below, for each sector we summarize the work performed and outline next steps.

6.1. AI-ENHANCEMENT OF CLIMATE SERVICES FOR THE WATER SECTOR

6.1.1. Conclusions from the preliminary results

In this deliverable, we investigated AI-enhancement of climate services for the water sector in two aspects: (1) hybrid hydrological modeling to enhance performance at local scale, and (2) attribution of runoff changes to climatic drivers in present and future conditions.

Regarding the hybrid hydrological modeling, we applied four post-processing methods to runoff simulations generated by the E-HYPE model and evaluated their performance using three different metrics while also analyzing the spatial distribution of the skills. Moreover, the investigation extended to identifying the primary factors driving performance enhancements gained in each post-processing method. The key findings are:

- The analysis reveals a notable improvement by post-processing the raw simulations, in terms of both total volume and high and low extremes. This is evidenced by a decrease in SMAE and an increase in NSE and logNSE, which suggests that post-processed models provide a more accurate representation of hydrological dynamics than the raw simulations.
- Across the different post-processing techniques, a similar spatial pattern of skill improvements is observed, showing higher skills in stations located in central Europe. This pattern is more enhanced in the context of extreme events, which also indicates the added value from post processing methods on high and low streamflow extremes.
- Key drivers were identified for influencing the model performance after post-processing and these are: mean precipitation, mean temperature, basin hydrological regimes and elevation. Each driver ranked differently across the various metrics, indicating their different impacts on model performance. Notably, the recurrent identification of hydrological clusters as a significant factor for both volume and extremes emphasizes its importance in refining model accuracy, in terms of volume and extremes.

Regarding attributing runoff changes to climatic drivers across Europe under present and future conditions, we applied the Budyko framework to provide a quantitative understanding of how changes in climatic factors affect alterations in runoff during the early, mid and late century under low, medium and high emission scenarios. The key findings include:

- Under future conditions precipitation showed a contrasting spatial behavior with increasing precipitation, specifically in north-eastern Europe and Scandinavia and a substantial decrease in southern Europe under the high emission scenario. Evaporation increased under a high emission scenario for the whole of Europe.
- On average, the sensitivity of runoff to precipitation changes was more substantial than to evaporation across Europe. Northern and central Europe showed a higher sensitivity to precipitation than evaporation changes, while in other regions, the sensitivity to both climatic factors was comparable.
- Precipitation was the main factor of increased runoff in Scandinavia and north-eastern Europe, especially under medium and high emission scenarios, while it contributed to decrease in runoff in southern Europe.
- Evaporation changes were the main factor of runoff alterations in central Europe from the early to late century and with increasing the severity of the emission scenario.

6.1.2. Moving forward

The current results from post-processing hydrological predictions, where skills show spatial compensation among different post-processing methods, requires a further investigation on ensemble techniques, e.g. probabilistic multi-model-ensemble approaches, to benefit from multiple model outputs. Therefore, the next step would be to explore possibilities of such techniques, namely copula-based Bayesian Model Averaging, to combine post-processing results from individual models and characterize the uncertainty induced. This would allow more reliable predictions by weighing and combining their individual results according to the bias against the observations. In addition, we plan to apply this method in a seasonal forecasting context and consequently generate a new AI-enhanced hydrological forecasting service.

With regard to advancing the attribution of runoff extremes to climate change, the next steps will focus on linking the attribution to the local hydrological regime using machine learning, as the results from the post-processing showed potential towards this direction. This investigation can help understand whether the contribution of climatic factors to runoff extreme events is similar in catchments of similar streamflow regimes; for instance, catchments responding fastly or slowly to precipitation signals. The clusters applied in post-processing will be used to identify similarities and differences in the relative contribution of precipitation and evaporation to runoff extremes under different periods and emission scenarios across Europe.

6.2. AI-ENHANCEMENT OF CLIMATE SERVICES FOR THE ENERGY SECTOR

6.2.1. Conclusions from the preliminary results

In this deliverable, we give the preliminary results for the enhancement of the PRIMES-IEM model and linking of weather and climate data. The methods and applications focused on (1) adjustment of the power demand to changes in the air temperatures and (2) adjustment of the hydropower generation profiles to changes of the river discharge. The introduced model enhancements focus on improving the preprocessing climate datasets, and mainly focusing on river discharge and air temperature data for the preliminary analysis. The results for the following applications were demonstrated:

- The analysis of the relationship between temperature and heating or cooling demand for electricity for the historical periods characterize the country's specific power needs during seasonal temperature changes. For the future projections under the three RCP scenarios (RCP2.6, RCP4.5 and RCP8.5) daily temperature changes provided by impact models, result in more accurate representation of power demand dynamics under different climatic conditions.
- Changes in river discharge under climatic drivers in present and under future conditions (RCP2.6, RCP4.5 and RCP8.5) shows great variability in the availability in hydropower generation profiles at national level. Using supervised machine learning regression models, we are able to find the relationship between river basins and the ROR annual generation profiles and

emphasize the importance of taking spatial dimensions into account when analyzing and studying energy sector climatic drivers.

Introducing enhancements to the dedicated impact model for the energy sector, enhances our ability to deliver information - based decision making.

6.2.2. Moving forward

This study demonstrates the preparation of climate-specific data for use in the European scale PRIMES-IEM model. In spite of specific implementations and methods for each model, the general processing steps and methods must be able to translate the variability of future climate conditions. The ML-enhanced techniques can enable a more accurate linkage of climate datasets with energy modeling and contribute to a better understanding of climate and energy dynamics. In this way, the identification of climate change risks and solutions in the energy sector can be greatly improved.

Further steps may include better representation of climate induced operational risks to thermal power generation capacities. Ultimately, the policy scenarios with climate-aware input data can demonstrate the projections of the power generation and power demand taking into account climate variability under current and planned policies at the European scale. In the Deliverable 6.4 some climate-informed policy scenarios will be generated demonstrating the impact of extreme events in the energy sector.

Box 1: *Efficiency losses and cooling needs*: heats and droughts cause decreases in efficiency of thermal power plants both due to high ambient temperature (Burillo et al. 2019) and due to cooling system failures (van Vliet et al. 2016). Heats and droughts affect the cooling of thermal power plants: nuclear generation limited by regulations for border temperatures of water intake for cooling (van Vliet et al. 2016); thermal power plants with water cooling systems – natural gas and hard coal (Byers et al. 2016).

6.3. AI-ENHANCEMENT OF CLIMATE SERVICES FOR THE FOOD SECTOR

Summarizing, this study aims to deliver an AI-enhanced climate service for food security by:

- developing a crop model emulator for Europe and force it using calibrated and downscaled seasonal climate forecasts,
- understanding the impact of concurrent climate extremes on crop yields and assess predictability of crop yields when concurrent climate extremes occur,
- assess the impact of climate change on crop yields in Europe and evaluate several adaptation options for optimizing crop productivity.

6.3.1. Conclusions from the preliminary results

The preliminary results towards the construction of the AI-based emulator show that in terms of a tabular database generation, there is an extensive need to assess the inherent noise. Finalizing the optimization-for-feature-selection pipeline with CRO-SL showed promising results in terms of convergence, but it is evident that in terms of AI modeling, ECroPS shows heavy presence of noise which makes the detection of patterns quite challenging.

This pushes the effort towards RNNs with clearer focus due to the trials with tabular datasets, using an architecture that encompasses a consequent mode of operations.

In this regard, we are currently building a pipeline with a database that inherits not only the raw data but also their temporal dimension.

Additionally, we are working towards a probabilistic AI framework, using BNNs, which is still in a nascent stage.

6.3.2. Moving forward

The workflow for development of AI-enhanced climate service is presented in Figure 5-1. Once the AI-enhanced crop yield emulator is developed, seasonal climate predictions will be used to run the emulator to estimate seasonal predictions of crop yields. Seasonal predictions will be applied to the crop model emulator during different times of the growing season. Given the seasonal forecast initialization time, the estimation of crop yield by emulator depends on the integration of observed and forecast precipitation and temperature data. Observed values are used until the time of forecast initialization, and are thereafter merged with seasonal forecasts to cover the entire period for each indicator. We can justify this merging of reanalysis and seasonal forecasts data as the latter are bias-adjusted. For example, if we consider winter wheat and the November initialization of seasonal forecast, the majority of indicators are largely based on seasonal forecast values since the timing largely coincides with either sowing or the initial growth phases of winter wheat across Europe. While, higher share of observed data is integrated into emulators' estimation in subsequent forecast initializations.

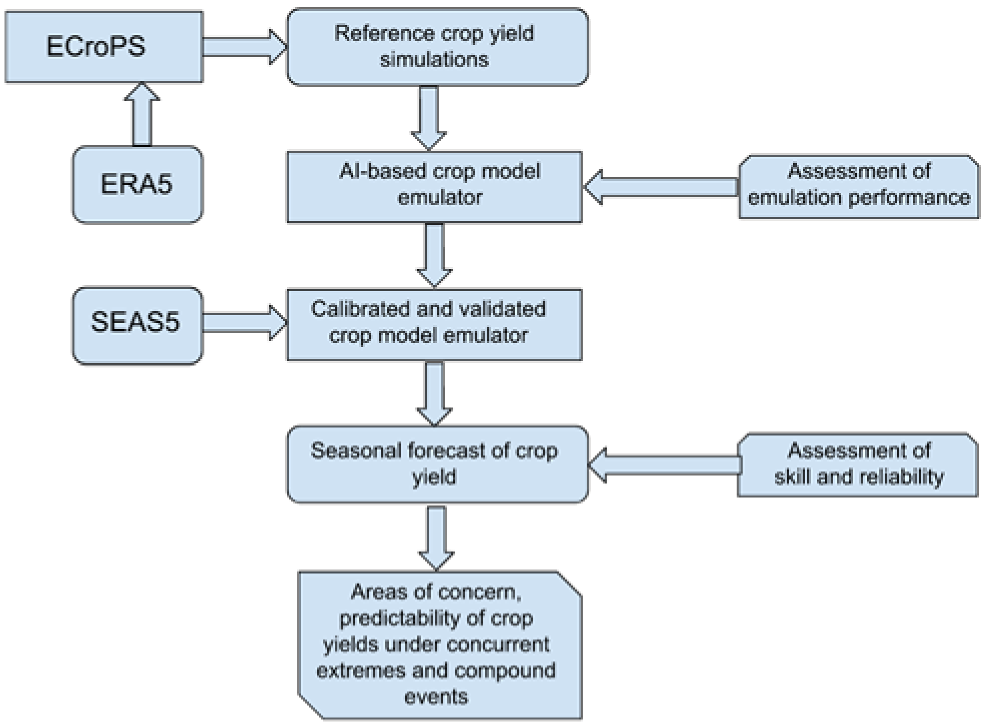


Figure 5-1 Workflow for development of AI-enhanced climate service integrating seasonal climate forecasts.

Downscaled and calibrated seasonal hindcasts of daily minimum and maximum temperatures and precipitation from the ECMWF SEAS5 system (Johnson et al., 2019) were provided for the purpose of this task by TCDF[9]. The reference climate data for the statistical post-processing of the seasonal forecast hindcast comes from ERA5 reanalysis, which is used also for running the ECROPS mechanistic crop model.

The principal results of this simulation framework will provide crop yield forecasts at the end of growing season for three main crops grown in Europe. The simulations will be summarized in terms of area of concern maps exposing the regions where grain yield drop might occur due to evolving and/or predicted extreme climate events, focusing primarily on heat waves and extreme droughts. This will provide useful insights on two aspects: (i) better understanding of crop yield variation when concurrent extremes and compound events occur, and (ii) the predictability of crop yields on European scale during these events.

Contrarily to seasonal forecasts, climate projections consider time scales until the end of the 21st Century. The climate projections are based on downscaled and bias adjusted CMIP6 simulations. The ERA5 reanalysis will be used as a reference dataset to perform downscaling to 0.1degrees and bias adjustment. To simulate the impact of climate change on crop yields, the AI-based crop model emulator needs to be complemented by mechanistic crop model simulations, especially when it comes to assessment of adaptation options on crop productivity. Therefore, a twofold process is planned: (i) ECroPS model will be used to simulate the impact of climate change on crop yields in the future and (ii) the AI based crop model emulator will be used to study the impact of climate change on crop yields under assumption that no agricultural practices would change in the future. This will give us useful insights especially on the impact of concurrent extremes and compound events in the future on crop productivity in the case no adaptation would take place.

[9] Loukos H., Pechlivanidis, I., Xoplaki, E. Ceglar, A., Ficchi, A., Cavicchia, L., Alvarez-Castro, C. 2022. Data provision for EU climate services. Milestone 3 of the CLINT project

REFERENCES

AlDahoul, N., Momo, M.A., Chong, K.L., Ahmed, A.N., Huang, Y.F., Sherif, M., El-Shafie, A., 2023. Streamflow classification by employing various machine learning models for peninsular Malaysia. Sci. Rep. 13, 14574. https://doi.org/10.1038/s41598-023-41735-9

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration, Guidelines for computing crop water requirements. In FAO Irrigation and drainage paper 56. Rome

Amos, M., Sengupta, U., Young, P. and Hosking, J.S., 2021. A continuous vertically resolved ozone dataset from the fusion of chemistry climate models with observations using a Bayesian neural network.

Barendrecht, M.H., Viglione, A., Kreibich, H., Merz, B., Vorogushyn, S., Blöschl, G., 2019. The Value of Empirical Data for Estimating the Parameters of a Sociohydrological Flood Risk Model. Water Resour. Res. 55, 1312–1336. https://doi.org/10.1029/2018WR024128

Berg, P., Almén, F., & Bozhinova, D. (2021). HydroGFD3.0 (Hydrological Global Forcing Data): A 25 km global precipitation and temperature data set updated in near-real time. Earth System Science Data, 13(4), 1531–1545. https://doi.org/10.5194/essd-13-1531-2021

Berg, P., Photiadou, C., Simonsson, L., Sjokvist, E., Thuresson, J., and Mook, R., (2021): Temperature and precipitation climate impact indicators from 1970 to 2100 derived from European climate projections. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.9eed87d5

Bézenac, E. de, Pajot, A., Gallinari, P., 2019. Deep learning for physical processes: incorporating prior scientific knowledge*. J. Stat. Mech. Theory Exp. 2019, 124009. https://doi.org/10.1088/1742-5468/ab3195

Bloomfield, H.C., Brayshaw, D.J., Deakin, M. and Greenwood, D., 2022. Hourly historical and near-future weather and climate variables for energy system modelling. Earth System Science Data Discussions, 2022, pp.1-26.

Bloomfield, H.C., Brayshaw, D.J., Gonzalez, P.L. and Charlton-Perez, A., 2021. Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables. Meteorological Applications, 28(4), p.e2018.

Botzen, W.J.W., Aerts, J.C.J.H., van den Bergh, J.C.J.M., 2009. Dependence of flood risk perceptions on socioeconomic and objective risk factors. Water Resour. Res. 45. https://doi.org/10.1029/2009WR007743

Brás, T.A., Seixas, J., Carvalhais, N. and Jägermeyr, J., 2021. Severity of drought and heatwave crop losses tripled over the last five decades in Europe. Environmental Research Letters, 16(6), p.065012.

Byers, E.A., Hall, J.W., Amezaga, J.M., O'donnell, G.M. and Leathard, A., 2016. Water and climate risks to power generation with carbon capture and storage. Environmental Research Letters, 11(2), p.024011.

Camps-Valls, G., Tuia, D., Zhu, X.X. and Reichstein, M. eds., 2021. Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences. John Wiley & Sons.

CJägermeyr, J., Müller, C., Ruane, A.C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J.A. and Fuchs, K., 2021. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. Nature Food, 2(11), pp.873-885.

Collins, M., Sutherland, M., Bouwer, L., et al. 2019. Extremes, Abrupt Changes and Managing Risks. In IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 589–655. https://doi.org/10.1017/9781009157964.008

De Felice, M., 2020. Hydropower information for power system modelling: the JRC-EFAS-Hydropower dataset.

de Wit, A., and Coauthors, 2019: 25 years of the WOFOST cropping systems model. Agricultural Systems, 168, 154–167, https://doi.org/10.1016/j.agsy.2018.06.018.

DG Energy 2018. Q3 Quarterly report on European electricity markets, DG Energy, Volume 11, issue 3 [Available online: https://drive.google.com/file/d/1XMDureROYzbAj9pY7cK7Ra X91-VZ9-Lz/view?usp=sharing]

DG Energy 2020. Q1 Quarterly report on European electricity markets, DG Energy, Volume 13, issue 1. [Available online: https://drive.google.com/file/d/1yH6zbNsHsz7MFTNE nHJxNPF0I_yMZlCM/view?usp=sharing]

Di Baldassarre, G., Sivapalan, M., Rusca, M., Cudennec, C., Garcia, M., Kreibich, H., Konar, M., Mondino, E., Mård, J., Pande, S., Sanderson, M.R., Tian, F., Viglione, A., Wei, J., Wei, Y., Yu, D.J., Srinivasan, V., Blöschl, G., 2019. Sociohydrology: Scientific Challenges in Addressing the Sustainable Development Goals. Water Resour. Res. 55, 6327–6355. https://doi.org/10.1029/2018WR023901

Donnelly, C., Greuell, W., Andersson, J., Gerten, D., Pisacane, G., Roudier, P., Ludwig, F., 2017. Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level. Climatic Change 143, 13–26. https://doi.org/10.1007/s10584-017-1971-7

Du, Y., Clemenzi, I., Pechlivanidis, I.G., 2023. Hydrological regimes explain the seasonal predictability of streamflow extremes. Environ. Res. Lett. 18, 094060. https://doi.org/10.1088/1748-9326/acf678

Dubus, L., Saint-Drenan, Y.M., Troccoli, A., De Felice, M., Moreau, Y., Ho-Tran, L., Goodess, C., Amaro e Silva, R. and Sanger, L., 2023. C3S Energy: A climate service for the provision of power supply and demand indicators for Europe based on the ERA5 reanalysis and ENTSO-E data. Meteorological Applications, 30(5), p.e 2145.

E3-Modelling 2018. PRIMES model. Detailed model description. Available at: https://e3modelling.com/modelling-tools/primes/.

EDO. 2018. Drought in Central-Northern Europe, September 2018. In EDO Combined Drought Indicator (CDI), edited by European Drought Observatory (EDO): European Commission Joint Research Centre.

Efron, B. and Hastie, T., 2022. Computer age statistical inference.

Enayati, M., Bozorg-Haddad, O., Bazrafshan, J., Hejabi, S., Chu, X., 2020. Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. J. Water Clim. Change 12, 401–419. https://doi.org/10.2166/wcc.2020.261

European Commission 2011. European Commission and the Council, Commission staff working document. Final report of the Advisory Group on the Energy Roadmap 2050 Summary record of the PRIMES Peer review Meeting Results of the public consultation on the Energy Roadmap 2050. Accompanying the document CCommunication from the Commission to the European Parliament, the Council, he European Economic and Social Committee and the Committee of the Regions. Energy Roadmap 2050. SEC(2011) 1569.

European Commission 2016. European Commission, Directorate-General for Climate Action, Directorate-General for Energy, Directorate-General for Mobility and Transport, Zampara, M., Obersteiner, M., Evangelopoulou, S. et al. EU reference scenario 2016 : energy, transport and GHG emissions : trends to 2050. Publications Office. https://data.europa.eu/doi/10.2833/001137

European Commission 2021. European Commission, Directorate-General for Climate Action, Directorate-General for Energy, Directorate-General for Mobility and Transport, De Vita, A., Capros, P., Paroussos, L. et al. EU reference scenario 2020: energy, transport and GHG emissions : trends to 2050. Publications Office. https://data.europa.eu/doi/10.2833/35750

European Commission 2022. Commission staff working document Implementing the RepowerEU action plan: Investment needs, hydrogen accelerator and achieving the bio-methane targets. Accompanying the document "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions REPowerEU Plan" SWD/2022/230 final

European Commission 2024. Impact Assessment Report Part 1 Accompanying the document "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Securing our future Europe's 2040 climate target and path to climate neutrality by 2050, building a sustainable, just and prosperous society". SWD/2024/63 final

Felice, M.D. and Kavvadias, K., 2022. ERA-NUTS: Meteorological time-series based on C3S ERA5 for European regions (1980–2021).

Fernandez Blanco Carramolino, R., Kavvadias, K. and Hidalgo Gonzalez, I., Water-related modelling in electric power systems: WATERFLEX Exploratory Research Project: version 1, EUR 29039 EN, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-77631-1, doi:10.2760/482511, JRC109941.

Friedman, J.H., 1991. Multivariate adaptive regression splines. The annals of statistics, 19(1), pp.1-67.

Geer, A.J., 2021. Learning earth system models from observations: machine learning or data assimilation? Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 379, 20200089. https://doi.org/10.1098/rsta.2020.0089

Gernaat, D.E., de Boer, H.S., Daioglou, V., Yalew, S.G., Müller, C. and van Vuuren, D.P., 2021. Climate change impacts on renewable energy supply. Nature Climate Change, 11(2), pp.119-125.

Goan, E. and Fookes, C., 2020. Bayesian neural networks: An introduction and survey. Case Studies in Applied Bayesian Data Science: CIRM Jean-Morlet Chair, Fall 2018, pp.45-87.

Gu, L., Yin, J., Slater, L.J., Chen, J., Do, H.X., Wang, H.-M., Chen, L., Jiang, Z., Zhao, T., 2023. Intensification of Global Hydrological Droughts Under Anthropogenic Climate Warming. Water Resources Research 59, e2022WR032997. https://doi.org/10.1029/2022WR032997

Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. Hydrol. Earth Syst. Sci. 16, 3383–3390. https://doi.org/10.5194/hess-16-3383-2012

Gudmundsson, L., Leonard, M., Do, H.X., Westra, S., Seneviratne, S.I., 2019. Observed Trends in Global Indicators of Mean and Extreme Streamflow. Geophysical Research Letters 46, 756–766. https://doi.org/10.1029/2018GL079725

Guse, B., Fatichi, S., Gharari, S., Melsen, L.A., 2021. Advancing Process Representation in Hydrological Models: Integrating New Concepts, Knowledge, and Data. Water Resour. Res. 57, e2021WR030661. https://doi.org/10.1029/2021WR030661

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine learning, 46, pp.389-422.

Hannaford, J., Buys, G., Stahl, K., Tallaksen, L.M., 2013. The influence of decadal-scale variability on trends in long European streamflow records. Hydrology and Earth System Sciences 17, 2717–2733. https://doi.org/10.5194/hess-17-2717-2013

Haupt, S.E., Chapman, W., Adams, S.V., Kirkwood, C., Hosking, J.S., Robinson, N.H., Lerch, S. and Subramanian, A.C., 2021. Towards implementing artificial intelligence post-processing in weather and climate: proposed actions from the Oxford 2019 workshop. Philosophical Transactions of the Royal Society A, 379(2194), p.20200091.

Hauser, T., Keats, A. and Tarasov, L., 2012. Artificial neural network assisted Bayesian calibration of climate models. Climate dynamics, 39, pp.137-154.

Hauswirth, S.M., Bierkens, M.F.P., Beijk, V., Wanders, N., 2021. The potential of data driven approaches for quantifying hydrological extremes. Adv. Water Resour. 155, 104017. https://doi.org/10.1016/j.advwatres.2021.104017

Hirth, L., Mühlenpfordt, J. and Bulkeley, M., 2018. The ENTSO-E Transparency Platform–A review of Europe’s most ambitious electricity data platform. Applied energy, 225, pp.1054-1067.

Hiruta, Y., Gao, L. and Ashina, S., 2022. A novel method for acquiring rigorous temperature response functions for electricity demand at a regional scale. Science of The Total Environment, 819, p.152893.

Ho, L.T., Dubus, L., De Felice, M. and Troccoli, A., 2020. Reconstruction of multidecadal country-aggregated hydro power generation in Europe based on a random forest model. Energies, 13(7), p.1786.

Huang, G.B., Zhu, Q.Y. and Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), pp.489-501.

Hundecha, Y., Arheimer, B., Berg, P. et al. Effect of model calibration strategy on climate projections of hydrological indicators at a continental scale. Climatic Change 163, 1287–1306 (2020). https://doi.org/10.1007/s10584-020-02874-4

Hundecha, Y., Arheimer, B., Donnelly, C., & Pechlivanidis, I. G. (2016). A regional parameter estimation scheme for a pan-European multi-basin model. Journal of Hydrology: Regional Studies, 6, 90–111. https://doi.org/10.1016/j.ejrh.2016.04.002

Hundecha, Y., Arheimer, B., Donnelly, C., Pechlivanidis, I., 2016. A regional parameter estimation scheme for a pan-European multi-basin model. J. Hydrol. Reg. Stud. 6, 90–111. https://doi.org/10.1016/j.ejrh.2016.04.002

Hundecha, Y., Merz, B., 2012. Exploring the relationship between changes in climate and floods using a model-based analysis. Water Resources Research 48. https://doi.org/10.1029/2011WR010527

IPCC, 2018. Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 3-24, doi:10.1017/9781009157940.001.

IPCC, 2022. Climate Change 2022: Impacts, Adaptation and Vulnerability. Summary for Policymakers.

Jiang, Z., Li, W., Xu, J., Li, L., 2015. Extreme Precipitation Indices over China in CMIP5 Models. Part I: Model Evaluation. J. Clim. 28, 8603–8619. https://doi.org/10.1175/JCLI-D-15-0099.1

Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G. and Keeley, S.P., 2019. SEAS5: the new ECMWF seasonal forecast system. Geoscientific Model Development, 12(3), pp.1087-1117.

Jospin, V.L., Buntine, W., Boussaid, F., Laga, H. and Bennamoun, M., 2020. Hands-on Bayesian Neural Networks--a Tutorial for Deep Learning Users. arXiv e-prints, pp.arXiv-2007.

Kashinath, K., and Coauthors, 2021: Physics-informed machine learning: case studies for weather and climate modelling. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 379, 20200093, doi: 10.1098/rsta.2020.0093

Khan, M.S. and Coulibaly, P., 2010. Assessing hydrologic impact of climate change with uncertainty estimates: Bayesian neural network approach. Journal of Hydrometeorology, 11(2), pp.482-495.

Kornhuber, K., Osprey, S., Coumou, D., Petri, S., Petoukhov, V., Rahmstorf, S. and Gray, L., 2019. Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. Environmental Research Letters, 14(5), p.054002.

Kornhuber, K., Osprey, S., Coumou, D., Petri, S., Petoukhov, V., Rahmstorf, S. and Gray, L., 2019. Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. Environmental Research Letters, 14(5), p.054002.

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. Hydrol. Earth Syst. Sci. 22, 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Lazer, D., Kennedy, R., King, G. and Vespignani, A., 2014. The parable of Google Flu: traps in big data analysis. science, 343(6176), pp.1203-1205.

Lekkas, P.T., 2013. Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications [WWW Document]. URL https://journal.gnest.org/publication/778 (accessed 2.27.24).

Lesk, C., Anderson, W., Rigden, A., Coast, O., Jägermeyr, J., McDermid, S., Davis, K.F. and Konar, M., 2022. Compound heat and moisture extreme impacts on global crop yields under climate change. Nature Reviews Earth & Environment, 3(12), pp.872-889.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., & Arheimer, B. (2010). Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrology Research, 41(3–4), 295–319. https://doi.org/10.2166/nh.2010.007

Liu, J., You, Y., Zhang, Q., Gu, X., 2021. Attribution of streamflow changes across the globe based on the Budyko framework. Science of The Total Environment 794, 148662. https://doi.org/10.1016/j.scitotenv.2021.148662

Liu, J., Zhang, Q., Zhang, Y., Chen, X., Li, J., Aryal, S.K., 2017. Deducing Climatic Elasticity to Assess Projected Climate Change Impacts on Streamflow Change across China. Journal of Geophysical Research: Atmospheres 122, 10,228-10,245. https://doi.org/10.1002/2017JD026701

Liu, S., Wang, J., Wang, H., Wu, Y., 2022. Post-processing of hydrological model simulations using the convolutional neural network and support vector regression. Hydrol. Res. 53, 605–621. https://doi.org/10.2166/nh.2022.004

Madsen, H., Thyregod, P., 2010. Introduction to General and Generalized Linear Models. CRC Press.

McGovern, A., Lagerquist, R., Gagne, D.J., Jergensen, G.E., Elmore, K.L., Homeyer, C.R. and Smith, T., 2019. Making the black box more transparent: Understanding the physical implications of machine learning. Bulletin of the American Meteorological Society, 100(11), pp.2175-2199.

Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Resour. 28, 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Mullachery, V., A. Khera, and A. Husain, 2018: Bayesian Neural Networks. https://doi.org/10.48550/arXiv.1801.07710.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Journal of Hydrology, 10, 282–290.

Orlov, A., Sillmann, J. and Vigo, I., 2020. Better seasonal forecasts for the renewable energy industry. Nature Energy, 5(2), pp.108-110.

Oudin, L., Michel, C., & Anctil, F. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? Journal of Hydrology, 303(1–4), 275–289. https://doi.org/10.1016/j.jhydrol.2004.08.025

Papacharalampous, G., Tyralis, H., 2022. A review of machine learning concepts and methods for addressing challenges in probabilistic hydrological post-processing and forecasting. Front. Water 4.

Papacharalampous, G., Tyralis, H., Pechlivanidis, I.G., Grimaldi, S., Volpi, E., 2022. Massive feature extraction for explaining and foretelling hydroclimatic time series forecastability at the global scale. Geosci. Front. 13, 101349. https://doi.org/10.1016/j.gsf.2022.101349

Pechlivanidis I.G., Arheimer, B., 2015, 'Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case', Hydrol. Earth Syst. Sci., 19, 4559-4579, doi: 10.5194/hess-19-4559-2015

Pechlivanidis I.G., Jackson B., McIntyre N., Wheater H.S., 2011, 'Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications', Global NEST Journal, 13 (3): 193-214

Pechlivanidis, I.G., Crochemore, L., Rosberg, J., Bosshard, T., 2020. What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts? Water Resour. Res. 56, e2019WR026987. https://doi.org/10.1029/2019WR026987

Pérez-Aracil, J., Camacho-Gómez, C., Lorente-Ramos, E., Marina, C.M., Cornejo-Bueno, L.M. and Salcedo-Sanz, S., 2023. New probabilistic, dynamic multi-method ensembles for optimization based on the CRO-SL. Mathematics, 11(7), p.1666.

Platts. 2018. "European Power Daily." S&P Global 20 (151).

Platts. 2018. "European Power Daily." S&P Global 20 (151).

Roderick, M.L., Farquhar, G.D., 2011. A simple framework for relating variations in runoff to variations in climatic conditions and catchment properties. Water Resour. Res. 47. https://doi.org/10.1029/2010WR009826

Rottler, E., Francke, T., Bürger, G., Bronstert, A., 2020. Long-term changes in central European river discharge for 1869–2016: impact of changing snow covers, reservoir constructions and an intensified hydrological cycle. Hydrology and Earth System Sciences 24, 1721–1740. https://doi.org/10.5194/hess-24-1721-2020

Schneider, C., Laizé, C.L.R., Acreman, M.C., Flörke, M., 2013. How will climate change modify river flow regimes in Europe? Hydrology and Earth System Sciences 17, 325–339. https://doi.org/10.5194/hess-17-325-2013

Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A. and Stadtler, S., 2021. Can deep learning beat numerical weather prediction? Philosophical Transactions of the Royal Society A, 379(2194), p.20200097.

Sengupta, U., Amos, M., Hosking, S., Rasmussen, C.E., Juniper, M. and Young, P., 2020. Ensembling geophysical models with Bayesian neural networks. Advances in Neural Information Processing Systems, 33, pp.1205-1217.

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. Hydrol Earth Syst Sci 27, 1865–1889. https://doi.org/10.5194/hess-27-1865-2023

Staffell, I., Pfenninger, S. and Johnson, N., 2023. A global model of hourly space heating and cooling demand at multiple spatial scales. Nature Energy, 8(12), pp.1328-1344.

Stahl, K., Tallaksen, L.M., Hannaford, J., van Lanen, H. a. J., 2012. Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble. Hydrology and Earth System Sciences 16, 2035–2047. https://doi.org/10.5194/hess-16-2035-2012

Szegedy, C., W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow and R. Fergus, 21.12.2013: Intriguing properties of neural networks.

Tang, X., Fu, G., Zhang, S., Gao, C., Wang, G., Bao, Z., Liu, Y., Liu, C., Jin, J., 2022. Attribution of climate change and human activities to streamflow variations with a posterior distribution of hydrological simulations. Hydrology and Earth System Sciences 26, 5315–5339. https://doi.org/10.5194/hess-26-5315-2022

Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rakovec, O., Samaniego, L., Sheffield, J., Wood, E.F., Zink, M., 2018. Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming. Environ. Res. Lett. 13, 014003. https://doi.org/10.1088/1748-9326/aa9e35

Toreti, A., Maiorano, A., De Sanctis, G., Webber, H., Ruane, A.C., Fumagalli, D., Ceglar, A., Niemeyer, S. and Zampieri, M., 2019. Using reanalysis in crop monitoring and forecasting systems. Agricultural systems, 168, pp.144-153.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M.G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J. and Yau, C., 2021. Bayesian statistics and modelling. Nature Reviews Methods Primers, 1(1), p.1.

Van Ruijven, B.J., De Cian, E. and Sue Wing, I., 2019. Amplification of future energy demand growth due to climate change. Nature communications, 10(1), p.2762.

Van Vliet, M.T., Sheffield, J., Wiberg, D. and Wood, E.F., 2016. Impacts of recent drought and warm years on water resources and electricity supply worldwide. Environmental Research Letters, 11(12), p.124021.

Van Vliet, M.T., Wiberg, D., Leduc, S. and Riahi, K., 2016. Power-generation system vulnerability and adaptation to changes in climate and water resources. Nature Climate Change, 6(4), pp.375-380.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M., & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resources Research, 50(9), 7505–7514. https://doi.org/10.1002/2014WR015638

Wenz, L., Levermann, A. and Auffhammer, M., 2017. North–south polarization of European electricity consumption under future warming. Proceedings of the National Academy of Sciences, 114(38), pp.E7910-E7918.

Wilson, A.G. and Izmailov, P., 2020. Bayesian deep learning and a probabilistic perspective of generalization. Advances in neural information processing systems, 33, pp.4697-4708.

Wörman, A., Uvo, C.B., Brandimarte, L., Busse, S., Crochemore, L., Lopez, M.G., Hao, S., Pechlivanidis, I. and Riml, J., 2020. Virtual energy storage gain resulting from the spatio-temporal coordination of hydropower over Europe. Applied energy, 272, p.115249.

Xu, L., Feng, K., Lin, N. et al. Resilience of renewable power systems under climate risks. Nat Rev Electr Eng 1, 53–66 (2024). https://doi.org/10.1038/s44287-023-00003-8

Xu, L., Feng, K., Lin, N. *et al.* Resilience of renewable power systems under climate risks. *Nat Rev Electr Eng* 1, 53–66 (2024). https://doi.org/10.1038/s44287-023-00003-8

Xu, L., Feng, K., Lin, N., Perera, A.T.D., Poor, H.V., Xie, L., Ji, C., Sun, X.A., Guo, Q. and O'Malley, M., 2024. Resilience of renewable power systems under climate risks. Nature Reviews Electrical Engineering, 1(1), pp.53-66.

Xu, T., Liang, F., 2021. Machine learning for hydrologic sciences: An introductory overview. WIREs Water 8, e1533. https://doi.org/10.1002/wat2.1533

Yalew, S.G., van Vliet, M.T., Gernaat, D.E., Ludwig, F., Miara, A., Park, C., Byers, E., De Cian, E., Piontek, F., Iyer, G. and Mouratiadou, I., 2020. Impacts of climate change on energy systems in global and regional scenarios. Nature Energy, 5(10), pp.794-802.

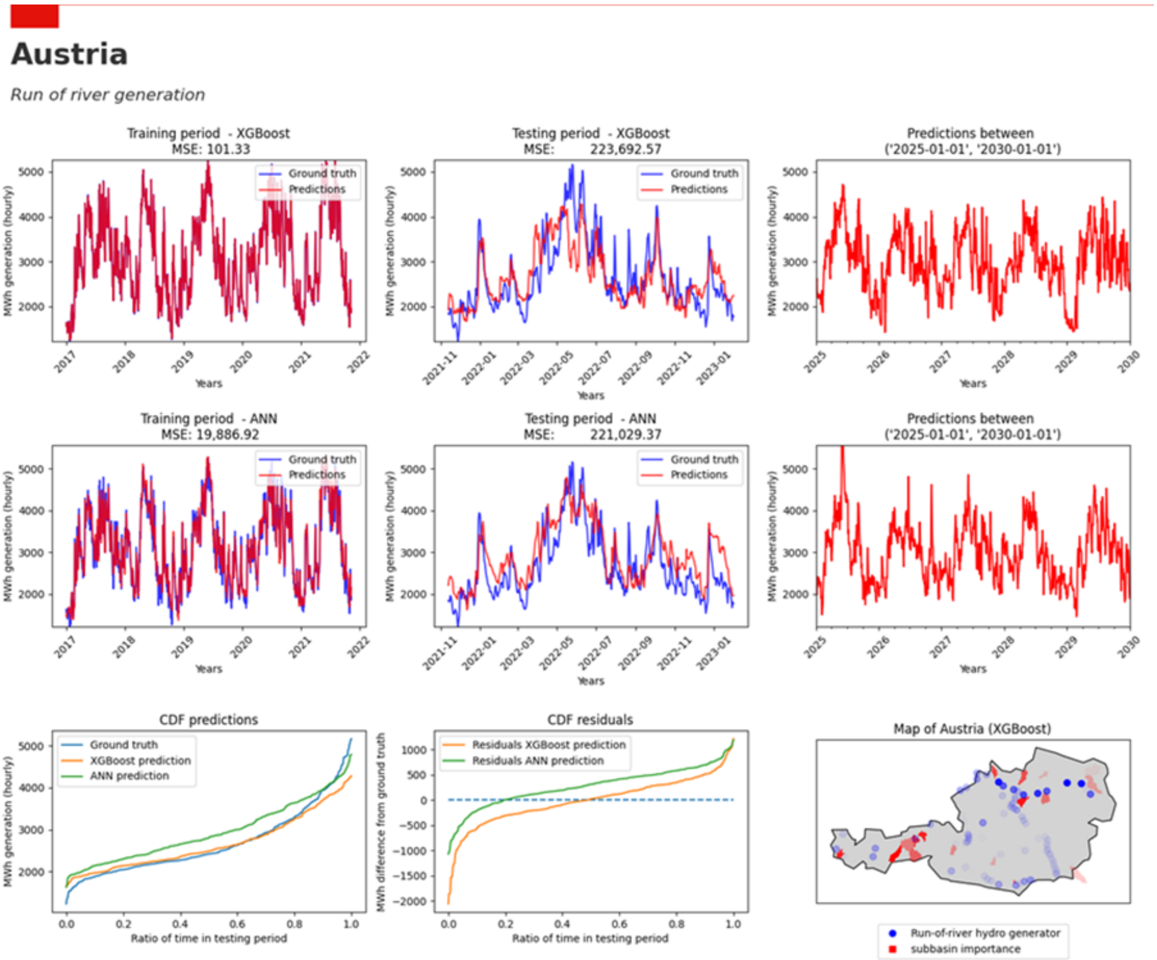
Yang, D., Yang, Y., Xia, J., 2021. Hydrological cycle and water resources in a changing world: A review. Geogr. Sustain. 2, 115–122. https://doi.org/10.1016/j.geosus.2021.05.003

Yeste, P., Rosa-Cánovas, J.J., Romero-Jiménez, E., García-Valdecasas Ojeda, M., Gámiz-Fortis, S.R., Castro-Díez, Y., Esteban-Parra, M.J., 2021. Projected hydrologic changes over the north of the Iberian Peninsula using a Euro-CORDEX multi-model ensemble. Science of The Total Environment 777, 146126. https://doi.org/10.1016/j.scitotenv.2021.146126

Zscheischler, J. and Fischer, E.M., 2020. The record-breaking compound hot and dry 2018 growing season in Germany. Weather and Climate Extremes, 29, p.100270.

ANNEX

1. MODEL FIT RESULTS PER COUNTRY FOR SECTION 4.2

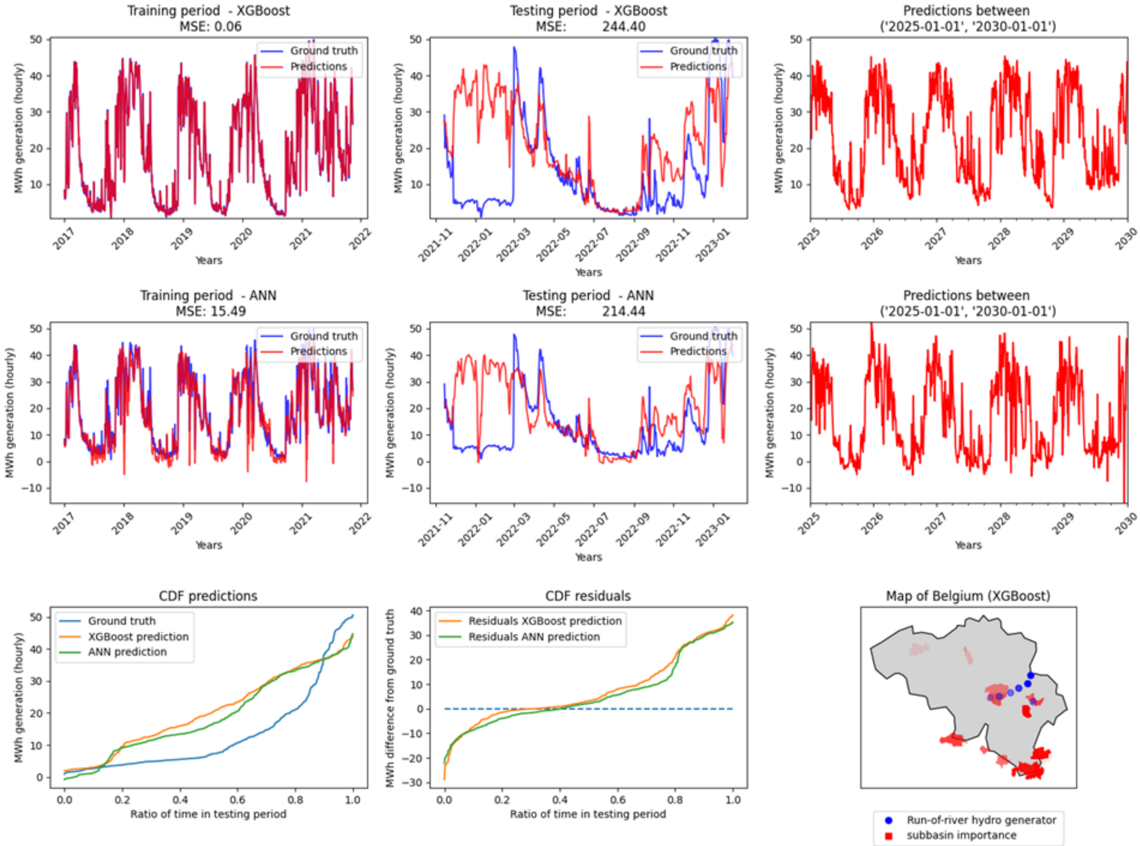


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	7.59706	353.96236	104.85891	392.19504
MSE	101.33155	223692.57067	19886.92346	221029.37423
RMSE	10.06636	472.96149	141.021	470.13761
NRMSE	0.24909%	12.02741%	3.48952%	11.9556%

Annex Figure *1* Results from XGBoost and Neural Networks for Austria using daily data for Austria

Belgium

Run of river generation

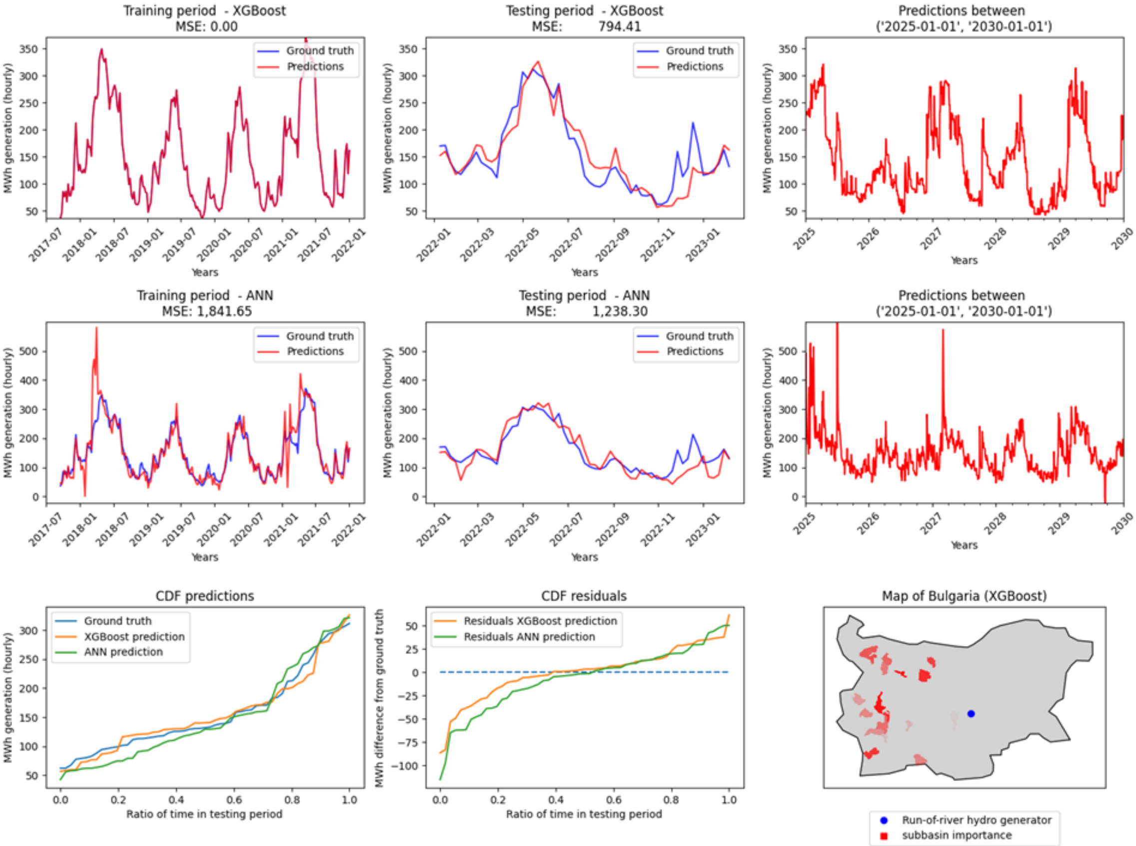


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.18336	11.1016	2.63121	10.42783
MSE	0.06207	244.4008	15.49274	214.43615
RMSE	0.24914	15.63332	3.93608	14.64364
NRMSE	0.50567%	31.48996%	7.98893%	29.49646%

Annex Figure *2* Results from XGBoost and Neural Networks for Austria using daily data for Belgium

Bulgaria

Run of river generation

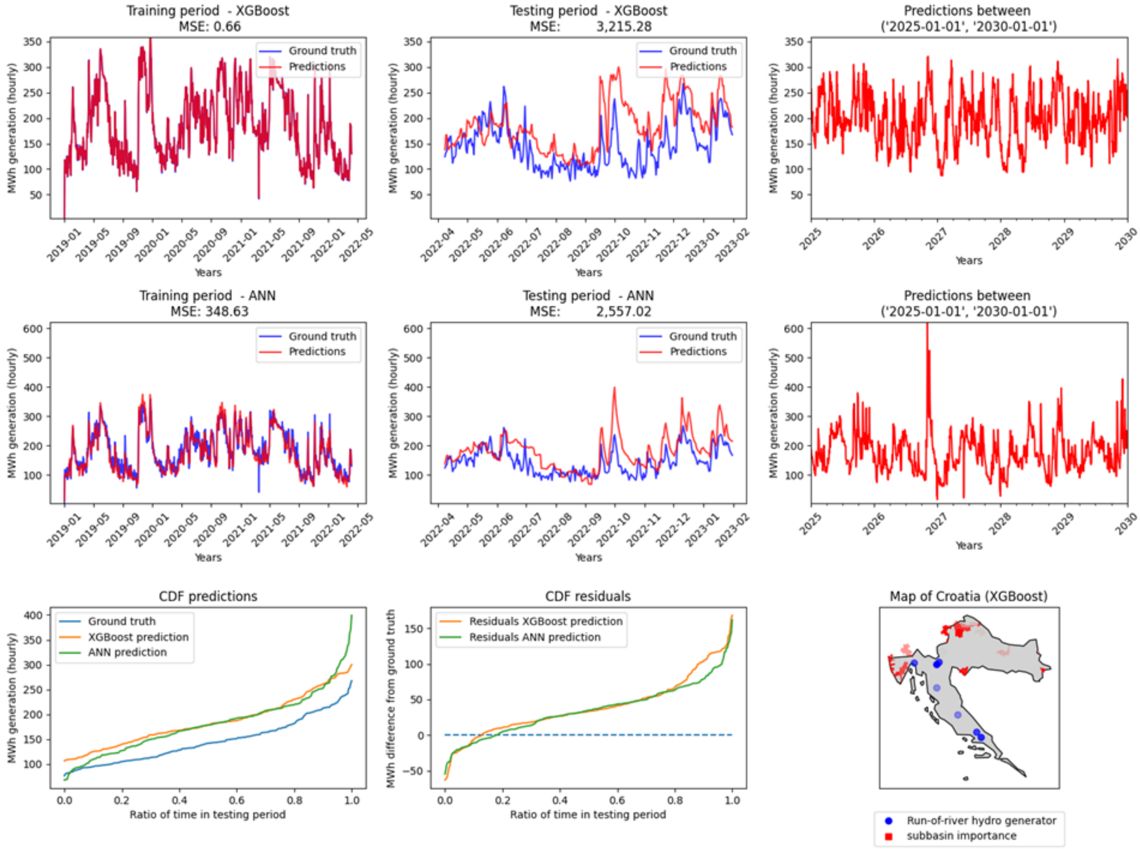


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.00077	20.43107	22.78674	25.80664
MSE	0.0	794.4147	1841.6531	1238.30347
RMSE	0.0	28.18536	42.91449	35.18954
NRMSE	0.0%	11.30017%	12.83493%	14.10831%

Annex Figure *3* Results from XGBoost and Neural Networks for Austria using daily data for Bulgaria

Croatia

Run of river generation

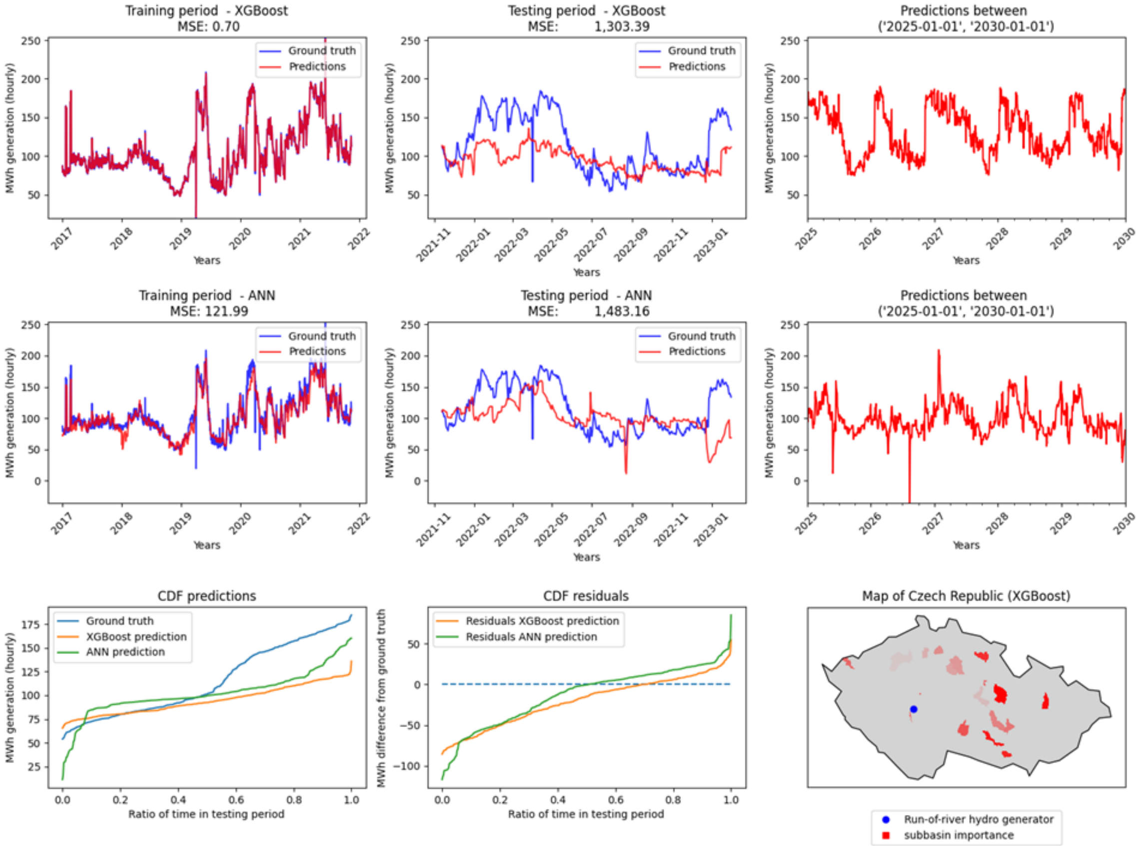


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.60115	44.84912	14.1747	40.90782
MSE	0.66138	3215.28423	348.63426	2557.02219
RMSE	0.81325	56.70348	18.67175	50.56701
NRMSE	0.22922%	29.71621%	5.26287%	26.50031%

Annex Figure *4* Results from XGBoost and Neural Networks for Austria using daily data for Croatia

Czech Republic

Run of river generation

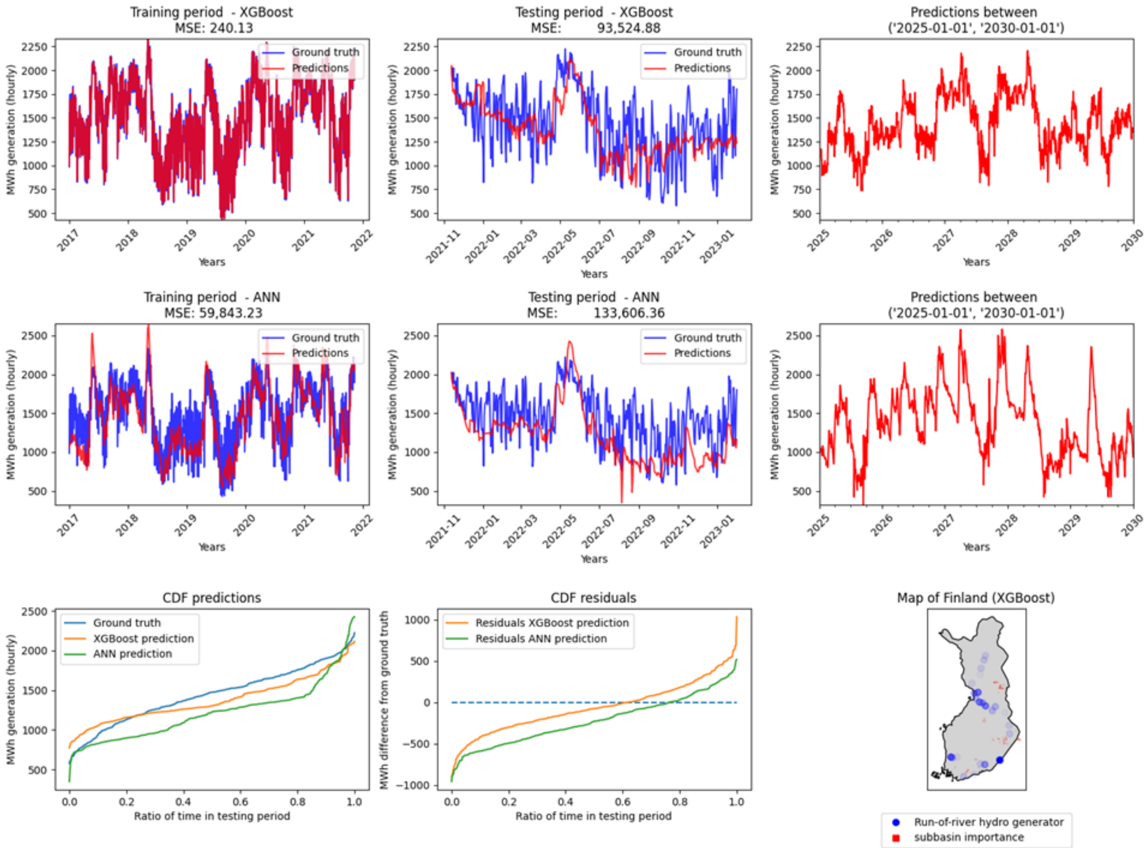


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.61694	27.63753	7.42213	28.87376
MSE	0.69564	1303.3946	121.98932	1483.15602
RMSE	0.83405	36.10256	11.04488	38.51176
NRMSE	0.356%	27.67496%	4.71433%	29.52177%

Annex Figure 5 Results from XGBoost and Neural Networks for Austria using daily data for Check Republic

Finland

Run of river generation

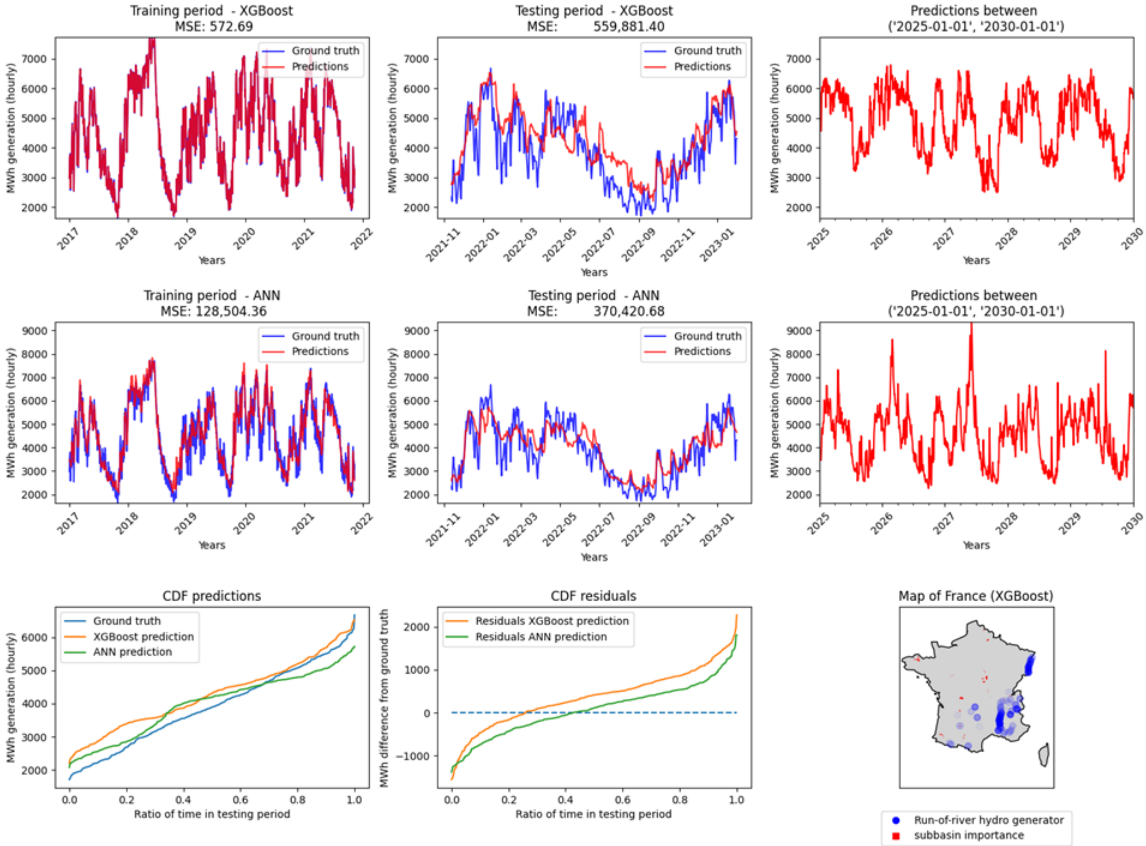


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	11.44436	243.02183	199.4468	302.00465
MSE	240.13379	93524.88466	59843.22541	133606.35969
RMSE	15.49625	305.81839	244.62875	365.52204
NRMSE	0.81491%	18.58085%	12.86451%	22.20831%

Annex Figure *6* Results from XGBoost and Neural Networks for Austria using daily data for Finland

France

Run of river generation

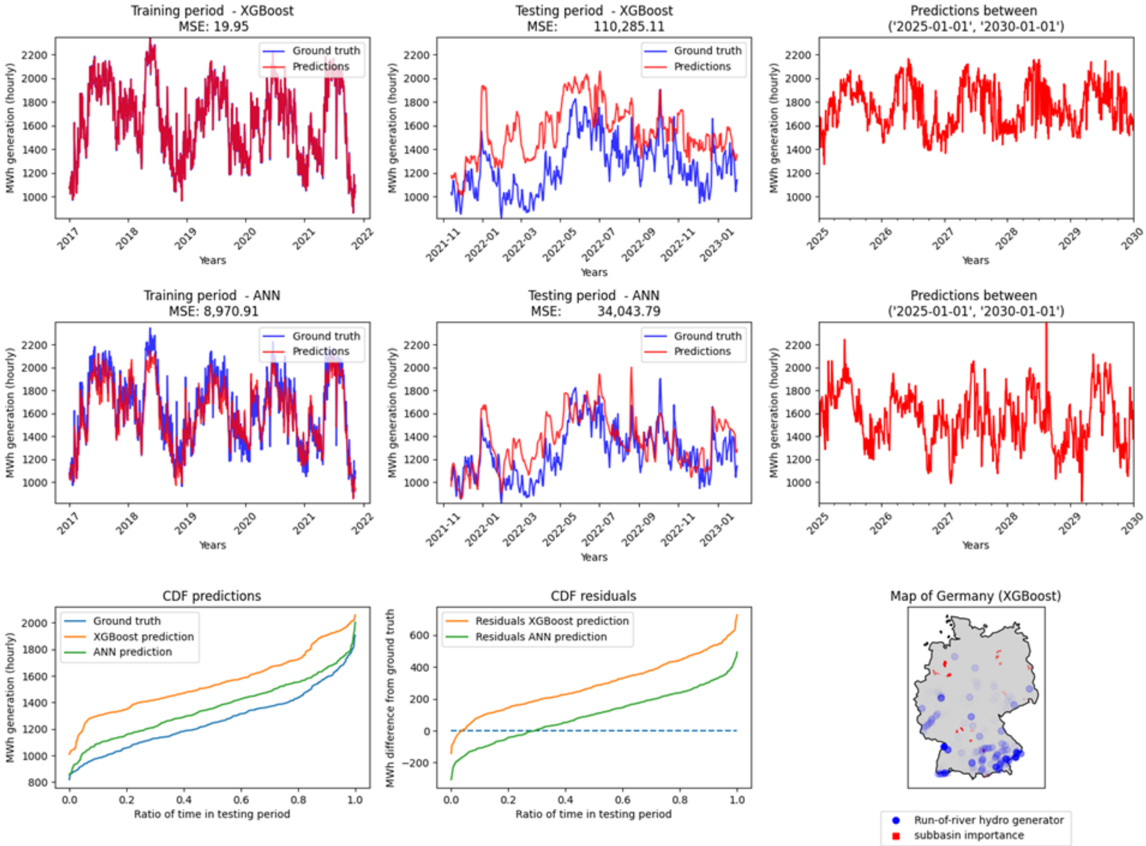


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	17.50612	607.43704	280.51315	487.66707
MSE	572.68767	559881.40232	128504.36352	370420.68429
RMSE	23.93089	748.25223	358.47505	608.62196
NRMSE	0.39296%	15.05008%	5.88645%	12.2416%

Annex Figure 7 Results from XGBoost and Neural Networks for Austria using daily data for France

Germany

Run of river generation

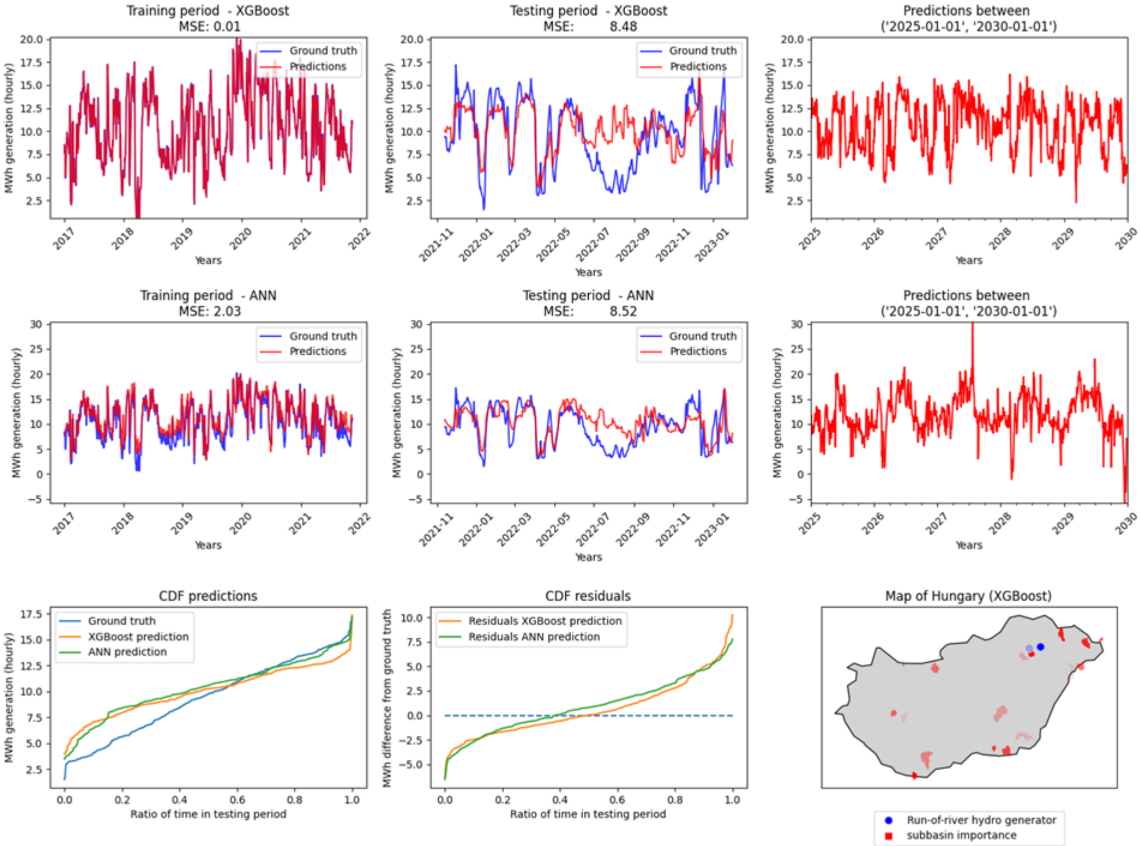


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	3.29497	291.13519	76.44848	151.43123
MSE	19.94948	110285.10619	8970.90779	34043.79407
RMSE	4.46648	332.09201	94.71488	184.5096
NRMSE	0.30096%	30.58527%	6.38198%	16.99311%

Annex Figure *8* Results from XGBoost and Neural Networks for Austria using daily data for Germany

Hungary

Run of river generation

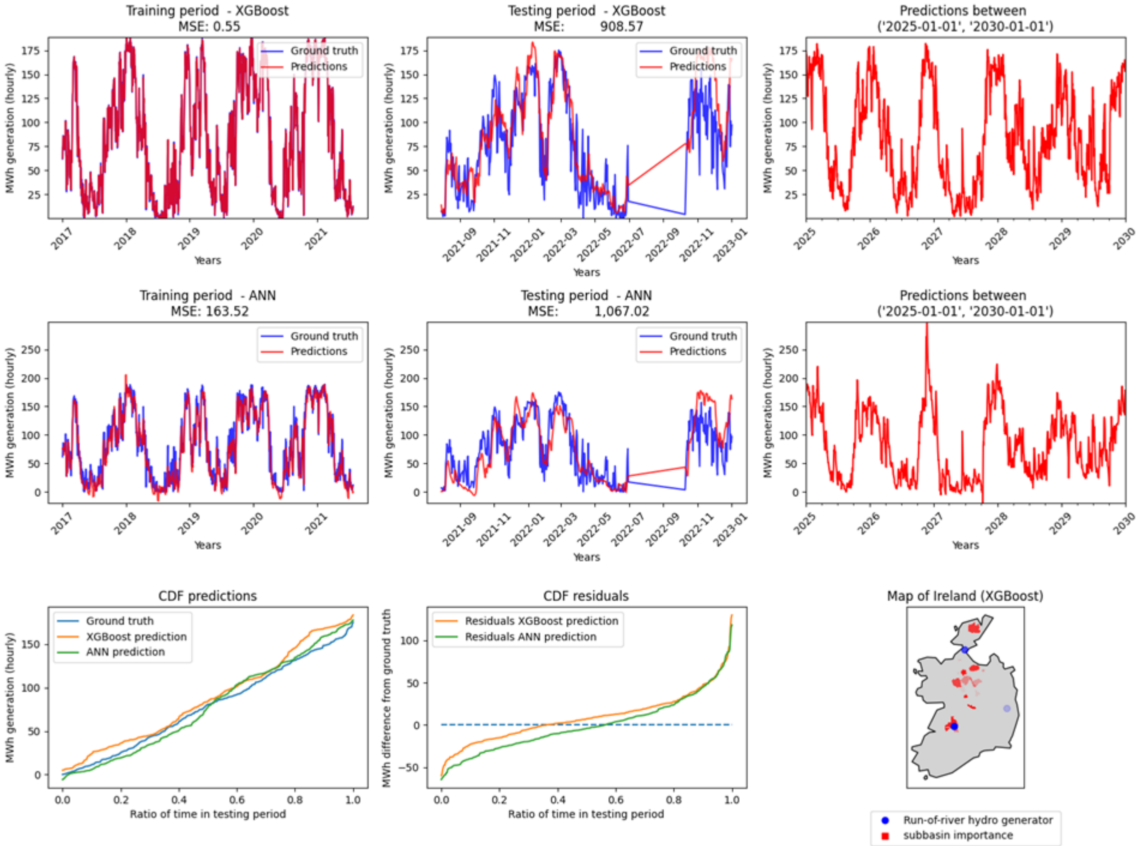


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.05237	2.18614	1.1175	2.32734
MSE	0.00511	8.47612	2.02748	8.52002
RMSE	0.07148	2.91138	1.4239	2.91891
NRMSE	0.36413%	18.52612%	7.2535%	18.57404%

Annex Figure *9* Results from XGBoost and Neural Networks for Austria using daily data for Hungary

Ireland

Run of river generation

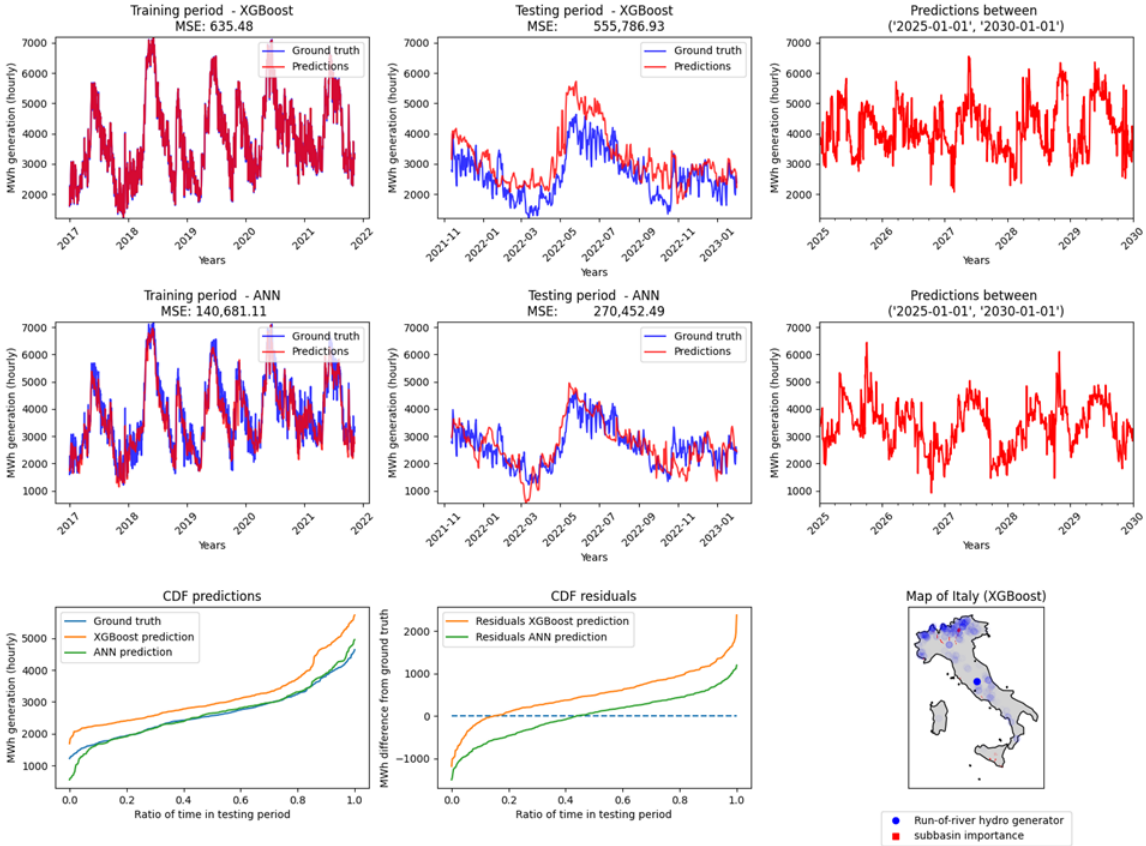


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.53925	22.39798	9.77697	25.46512
MSE	0.54696	908.57137	163.51626	1067.0157
RMSE	0.73957	30.14252	12.78735	32.66521
NRMSE	0.39223%	17.20433%	6.78184%	18.64419%

Annex Figure *10* Results from XGBoost and Neural Networks for Austria using daily data for Ireland

Italy

Run of river generation

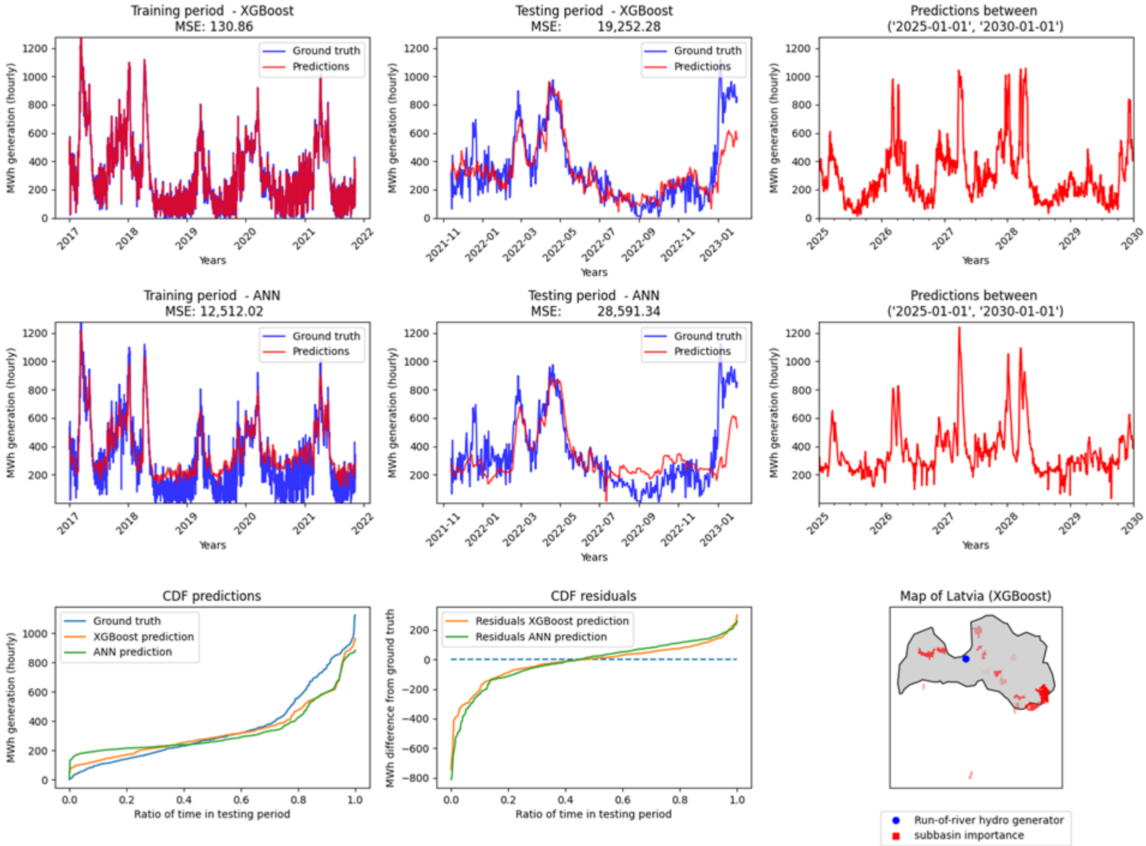


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	18.85388	612.67962	297.59762	422.90467
MSE	635.4812	555786.92681	140681.1117	270452.49448
RMSE	25.20875	745.51118	375.07481	520.05047
NRMSE	0.42164%	21.85636%	6.27351%	15.24646%

Annex Figure *11* Results from XGBoost and Neural Networks for Austria using daily data for Italy

Latvia

Run of river generation

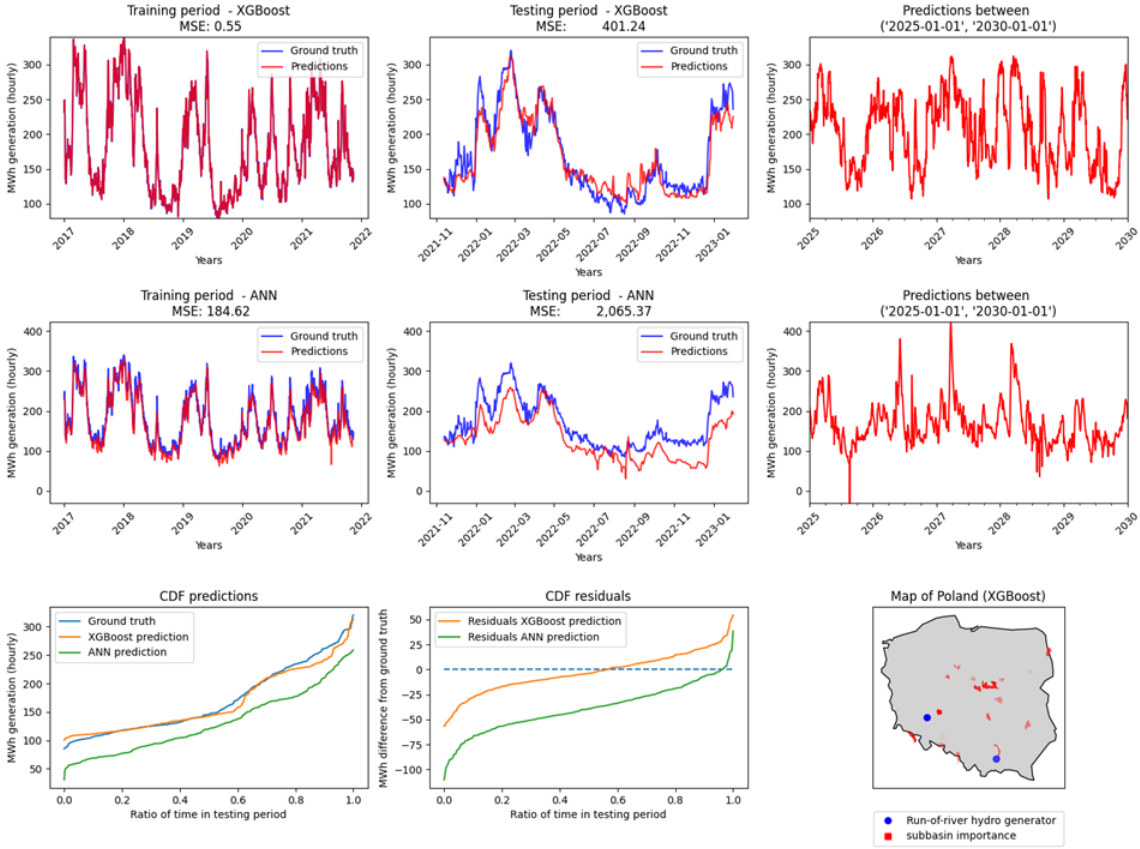


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	8.4696	94.14698	88.54665	118.87747
MSE	130.86006	19252.28346	12512.02247	28591.33688
RMSE	11.43941	138.7526	111.85715	169.08973
NRMSE	0.89554%	12.37527%	8.7568%	15.08103%

Annex Figure *12* Results from XGBoost and Neural Networks for Austria using daily data for Latvia

Poland

Run of river generation

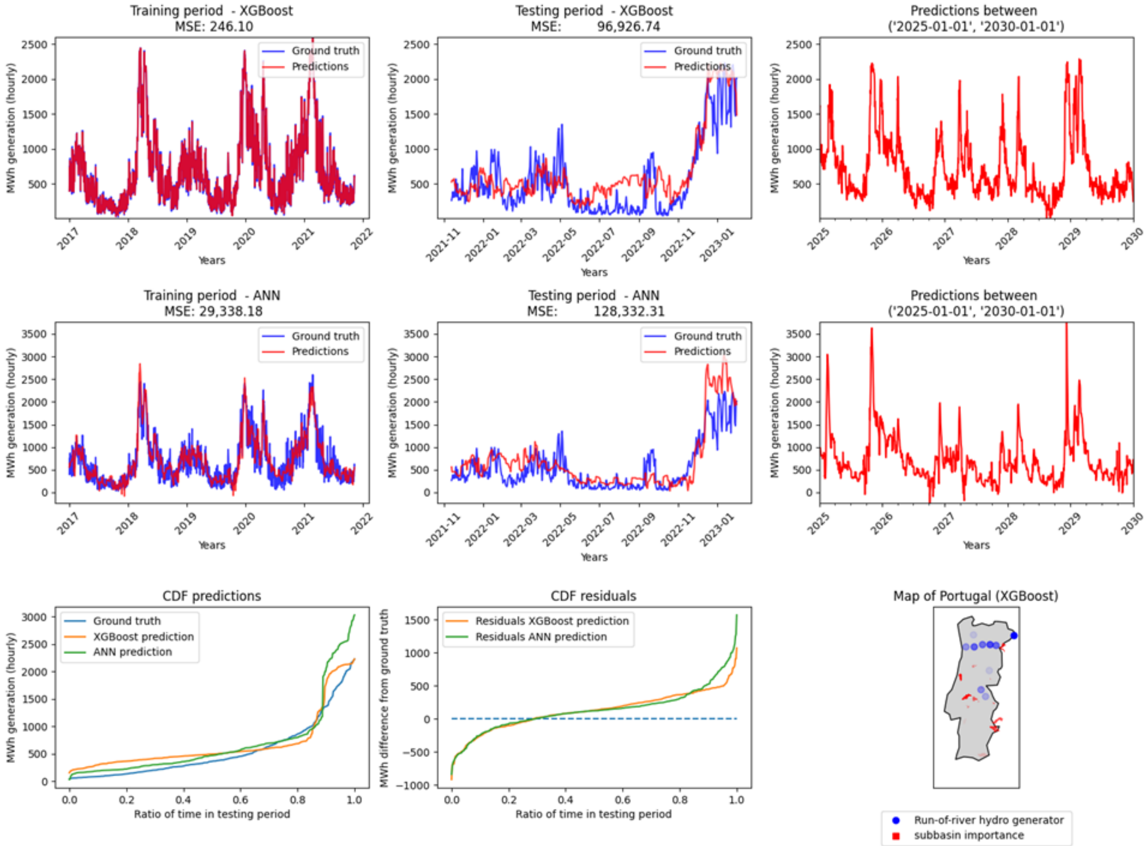


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	0.55162	15.73587	11.05881	39.6641
MSE	0.55237	401.23889	184.62282	2065.37378
RMSE	0.74322	20.03095	13.5876	45.44638
NRMSE	0.28538%	8.52706%	5.2173%	19.34626%

Annex Figure *13* Results from XGBoost and Neural Networks for Austria using daily data for Poland

Portugal

Run of river generation

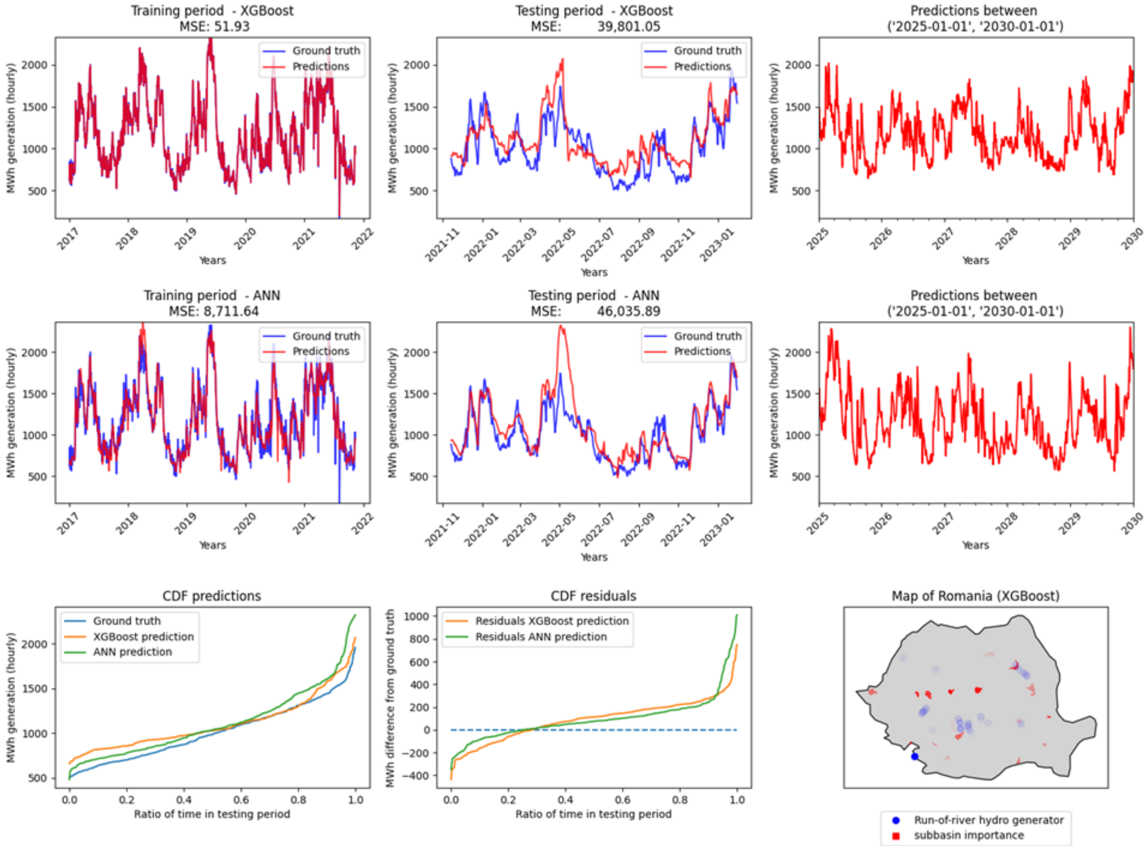


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	11.541	251.03196	130.60646	260.95621
MSE	246.09973	96926.73746	29338.17604	128332.30502
RMSE	15.68757	311.33059	171.2839	358.23499
NRMSE	0.61054%	14.23649%	6.66614%	16.38132%

Annex Figure *14* Results from XGBoost and Neural Networks for Austria using daily data for Portugal

Romania

Run of river generation

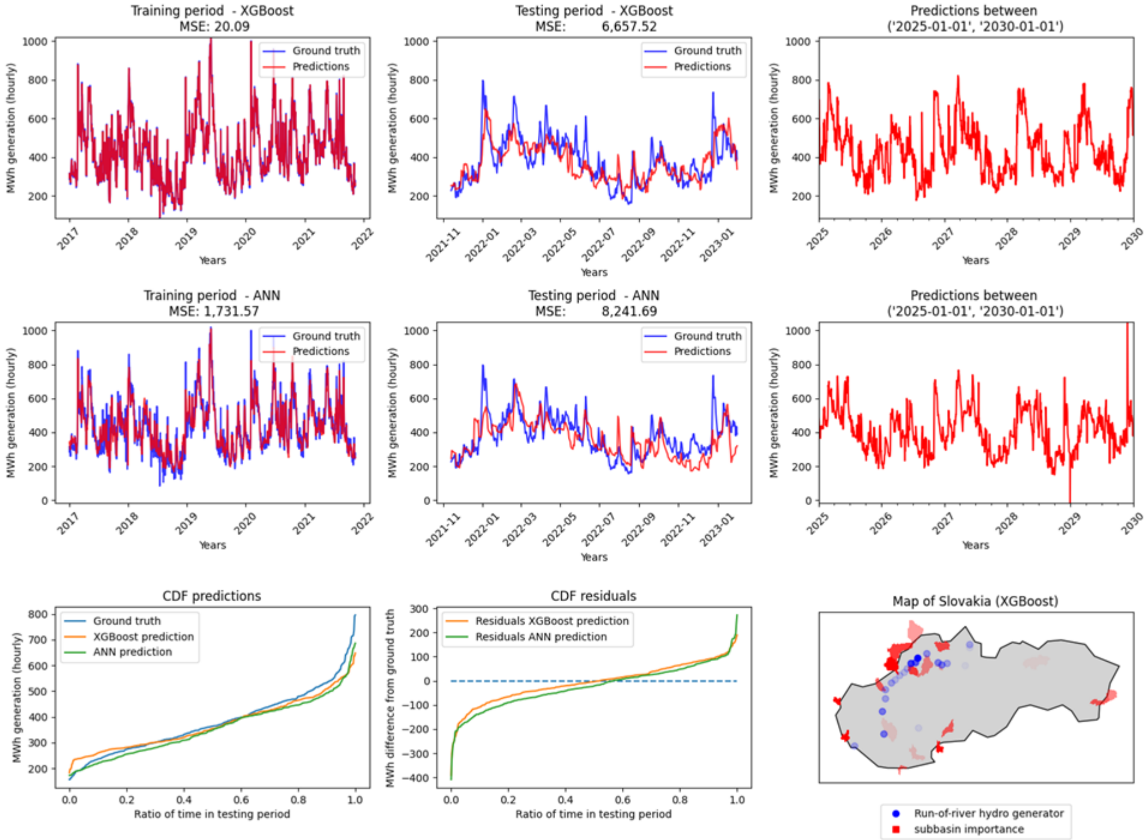


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	5.20856	164.80664	66.69318	142.78847
MSE	51.93431	39801.05	8711.63982	46035.89039
RMSE	7.20655	199.50201	93.33617	214.55976
NRMSE	0.3348%	13.6739%	4.33623%	14.70596%

Annex Figure *15* Results from XGBoost and Neural Networks for Austria using daily data for Romania

Slovakia

Run of river generation

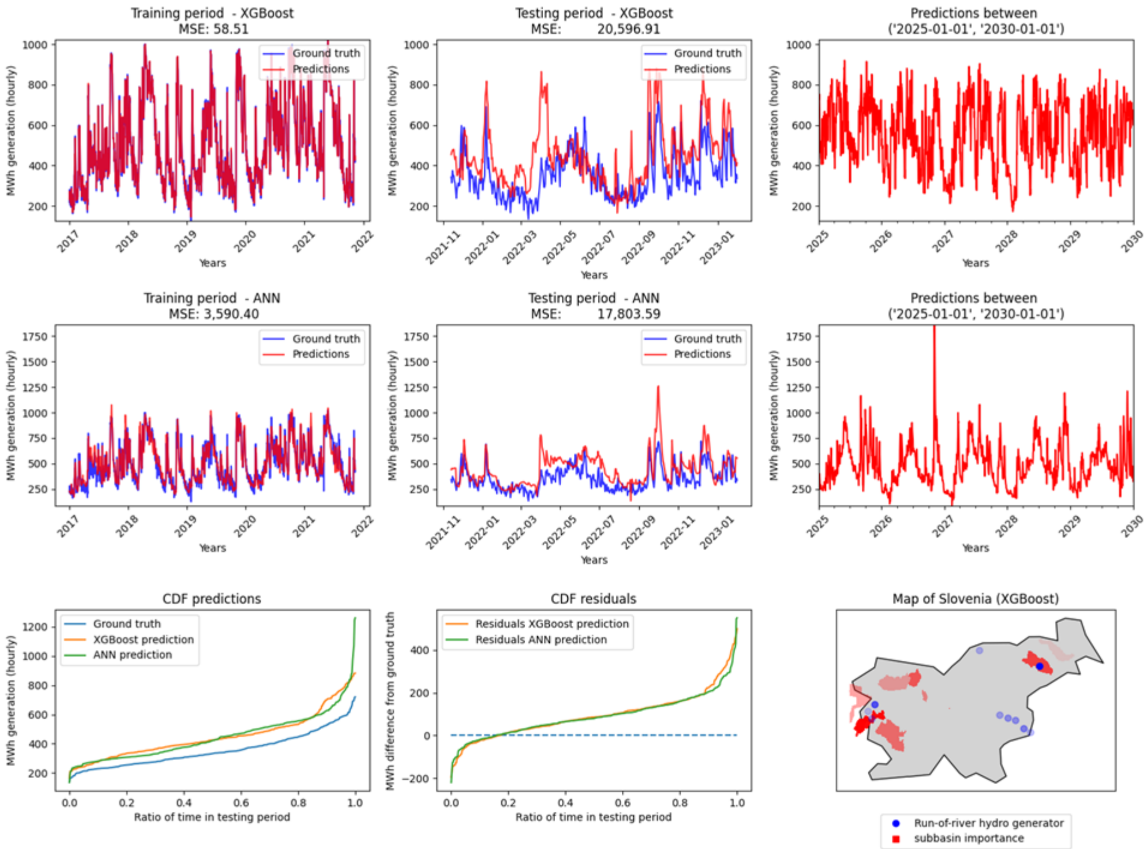


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	3.34109	62.2003	31.50068	70.48756
MSE	20.08934	6657.51565	1731.56592	8241.69102
RMSE	4.48211	81.5936	41.61209	90.78376
NRMSE	0.47906%	12.7583%	4.4476%	14.19531%

Annex Figure *16* Results from XGBoost and Neural Networks for Austria using daily data for Slovakia

Slovenia

Run of river generation

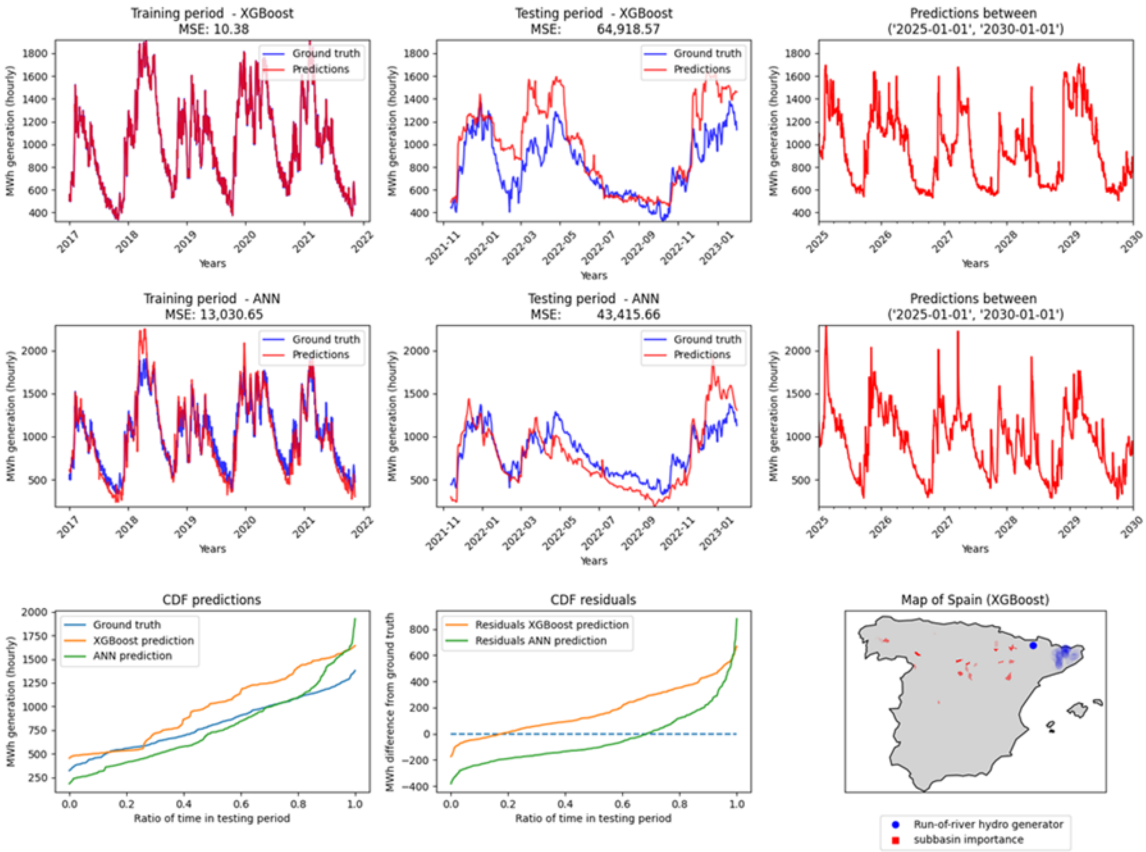


Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	5.6478	110.61885	43.70026	103.94187
MSE	58.50928	20596.91108	3590.40408	17803.58946
RMSE	7.64914	143.51624	59.91998	133.43009
NRMSE	0.85486%	24.50834%	6.69662%	22.78592%

Annex Figure *17* Results from XGBoost and Neural Networks for Austria using daily data for Slovenia

Spain

Run of river generation



Metric	Training: XGBoost	Testing: XGBoost	Training: ANN	Testing: ANN
MAE	2.32565	196.52812	87.92108	169.81703
MSE	10.38321	64918.56613	13030.65296	43415.66233
RMSE	3.2223	254.79122	114.15189	208.36425
NRMSE	0.20357%	24.16801%	7.21149%	19.76422%

Annex Figure *18* Results from XGBoost and Neural Networks for Austria using daily data for Spain