CLINT

CLIMATE INTELLIGENCE

AI-ENHANCED ATTRIBUTION AND PROJECTIONS OF EXTREME EVENTS

June, 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:	WP5
Deliverable #:	D5.2
Deliverable Type:	Report
Contractual Date of Delivery:	30 June 2024
Actual Date of Delivery:	30 June 2024
Title of Document:	AI-enhanced attribution and projections of extreme events
Responsible partner:	Consejo Superior de Investigaciones Científicas (CSIC)
Author(s):	David Barriopedro, Étienne Plésiat, Antonello Squintu, Niklas Luther, Pablo G. Zaninelli, Elena Xoplaki, Eugenio Lorente Ramos, Jorge Pérez Aracil
Content of this report:	The deliverable reports on the applications of ML algorithms to EE attribution, the detection of observed trends and the construction of future storylines
Availability:	This report is public

Document revisions		
Author	*Revision content*	*Date*
David Barriopedro, Étienne Plésiat, Antonello Squintu, Niklas Luther, Pablo G. Zaninelli, Elena Xoplaki, Eugenio Lorente Ramos, Jorge Pérez Aracil	First draft of the document	03.06.2024
Lucia De Stefano	Internal review	17.06.2024
Schalk Jan van Andel	Internal review	24.06.2024
David Barriopedro, Étienne Plésiat, Antonello Squintu, Niklas Luther, Pablo G. Zaninelli	Second draft of the document	28.06.2024
Guido Ascenso, Matteo Giuliani, Andrea Castelletti	Final check and submission	30.06.2024

Table of Content

LIST OF ACRONYMS

ADW: Angular Distance Weighting
AE: AutoEncoder
AI: Artificial Intelligence
ALE: Accumulated Local Effect
AM: Analogue Method
BVHMD: BiVariate Heat Magnitude Day
CAM: Climate Anomaly Method
CDF: Conditional Distribution Function
CEEI: Concurrent Extreme Event Index
CMIP: Coupled Model Intercomparison Project
CRO: Coral Reef Optimisation
CWS: Common Warming State
D: Deliverable
DL: Deep Learning
DR: Drought
EE: Extreme Event
ETCCDI: Expert Team on Climate Change Detection and Indices
GEV: Generalised Extreme Value
GCM: Global Climate Model
GSAT: Global mean near-Surface Temperature
GWL: Global Warming Level
HH: High-High quadrant
HL: High-Low quadrant
HP: Heavy Precipitation
HW: Heat Wave
HWMI: Heat Wave Magnitude Index
JJA: June-July-August
KRGCCA: Kernel Regularized Canonical Correlation Analysis
LH: Low-High quadrant
LL: Low-Low quadrant
MAE: Mean Absolute Error
ML: Machine Learning
MS: Milestone
MSE: Mean Squared Error
NDQ90: Number of days above the 90th percentile
NN: Neural Network
NPSPEI: Non-Parametric SPEI
PCA: Principal Component Analysis
QUINN: Quantile Regression Using I-Spline Neural Network
RMSE: Root Mean Squared Error
SLP: Sea Level Pressure
SSP: Shared Socioeconomic Pathway
SST: Sea Surface Temperature
SPEI: Standardised Precipitation Evapotranspiration Index

TN: Minimum temperature
TX: Maximum temperature
TX90p: 90th percentile of TX
VAE: Variational AutoEncoder
WB: Water Balance
WP: Work Package
Z500: 500 hPa Geopotential Height

LIST OF FIGURES

VAE. TX values are expressed as percentiles of the empirical cumulative distribution function of the train dataset. VAE anomalies are detected using different thresholds (columns): maximum, 95th percentile, 75th percentile and 50th percentile of MSE. Bottom: Index of plausibility based on correlations of the spatial patterns for the 100 days with the highest and lowest anomaly, as defined by (right) MSE and (left) Kullback-Leibler divergence error.

Figure 14 Daily fields for HWs simulated in the historical period of HadGEM3-GC31-LL GCM affecting (top) the Iberian Peninsula, (middle) central Europe and (bottom) Scandinavia. Columns denote the temperature departure (°C), the VAE anomaly (only grid points where the MSE values are greater than the median MSE of the training dataset are shown) and the percentile of the TX value computed with a GEV fitted to the training dataset.

Figure 15 Regional means of warm days (left) and cool nights (right) for the original HadEX3 dataset, the original HadEX-CAM dataset and its AI reconstruction over the full grid (Europe) and three European regions defined by the IPCC report (IPCC 2021): NEU, WCE, MED. Units are percentages of warm days and cold nights in a month.

Figure 16 Linear trends (in days decade^{-1}) of warm days for the period 1901-2018. Left panel: original HadEX3 dataset (considering only grid boxes with at least 66% of valid data across the whole time period). Right panel: AI reconstruction.

Figure 17 Same as Figure 16 for cool nights.

Figure 18 Linear trends (in days/decade) of warm days for the period 1980-2018. Left panel: original HadEX3 dataset (considering only grid boxes with at least 66% of valid data across the whole time period). Central panel: AI reconstruction. Right panel: ERA5 dataset.

Figure 19 Same as Figure 18 for cool nights.

Figure 20 Administrative regions of Piedmont (yellow), Lombardy (blue), Veneto (red) and Emilia-Romagna (green) and the corresponding grid-points of ERA5.

Figure 21 Examples of geographical clusters obtained for some of the variables: Z500, total precipitation, SLP and SST.

Figure 22 Selection rate related to the ERA5 clusters (columns, only spatial average statistics are presented) for different lags (rows). Purple colour denotes selection rates below 0.67, green stands for values above the threshold, indicating the selected benchmark drivers.

Figure 23 Selection rates for the same drivers as in Figure 22 (rows) and the three statistics (squares in each cell). Columns indicate the analysed simulations. Green is used for ERA5 benchmark drivers and blue for the selection rates of the simulations. Orange vertical bands indicate the on-site benchmark drivers. Purple vertical bands highlight the two clusters used for the construction of storylines.

Figure 24 Number of simulations validating each pair of ERA5 benchmark drivers.

LIST OF TABLES

EXECUTIVE SUMMARY

This report describes Machine Learning (ML) developments for the study of climate change impacts on extreme events (EEs) in order to support the attribution of single EEs, the detection of observed trends, and the quantification of future changes in EEs and concurrent EEs. The AI-added value is assessed with respect to existing datasets and/or methods, and illustrated with examples covering the ample spectrum of spatio-temporal scales and types of EEs considered in CLINT, from heat waves (HWs) at hotspot area to concurrent hot-dry EEs at continental scale.

First, two AI-based algorithms for attribution of EEs contributing to Deliverable 2.3 are presented. The first one combines Autoencoders and the classical Analogue Method (AE-AM) to quantify climate change influences on EEs by yielding a probabilistic reconstruction of meteorological fields associated with EEs for two periods with different levels of anthropogenic forcing. By encoding relevant information of predictor fields, the AE-AM outperforms the classical AM in reconstructing major historical European HWs, and detects climate change signals in a selection of historical EEs. The second AI-based model for attribution uses a Variational Autoencoder (VAE) for a probabilistic detection of climate change signals in the spatio-temporal evolving patterns of EEs. The learning of the VAE from a simulated natural world is employed to reconstruct a naturalised version of EEs that occur in the present climate. As illustrated in a selection of simulated HWs, the VAE arises as a powerful tool for near real-time attribution.

Secondly, the report describes the results of an AI-based algorithm to fill gaps in observational datasets (Deliverable 2.2). The AI model, a U-Net with partial convolutional layers, is trained with complete datasets (e.g. historical runs of Global Climate Models, GCMs). When applied to the monthly extreme temperature indices of HadEX (1901-2018), the algorithm produces consistent results and outperforms traditional infilling algorithms, being very effective in periods and regions with a large amount of missing information. The completed dataset provides a long record of observational series with continuous spatial coverage at a horizontal resolution similar to that of historical reanalyses for an improved assessment of regional trends in EEs. The inspection of the infilled product reveals more complex historical patterns of extreme temperature indices than those reported before.

Finally, an AI model contributing to Deliverable 2.3 is developed for the construction of storylines (i.e. feasible climate change responses in EEs from the combined effect of changes in relevant drivers). The identification of drivers follows two complementary strategies: spatial clustering for dimensionality reduction and evolutionary algorithms for an optimised feature extraction. The novel use of Common Warming States for the computation of changes in EEs and their drivers further circumvents model bias issues. As an illustration, the AI model is applied to summer HWs in the Po Valley. The large spread of future projections is reduced to four storylines, which entail substantial differences depending on the future evolution of southwestern European precipitation and North Atlantic pressure. Storylines are constructed for summer hot-dry concurrent EEs in central Europe, which require specific AI-based developments to account for the multivariate nature of concurrent EEs, state-dependent feature importance and non-linear relationships between the drivers. For most storylines hot and dry conditions become more severe, but their future evolution is largely mediated by changes in summer atmospheric circulation and secondarily by a spring dipole in sea surface temperatures.

Our results highlight the added value of AI for enhancing attribution, trend detection and the identification of drivers of EEs, underlining at the same time the need for additional work on process understanding. In this sense, future steps should build upon hybrid approaches / explainable AI models and move towards the impacts of EEs. Efforts are underway to expand CLINT experience to other pilot areas and EEs, including the attribution of EEs in the real world, the detection of trends for challenging extreme indices (e.g. precipitation), or the identification of drivers and their dependencies for future EEs and concurrent EEs in other European regions.

1. INTRODUCTION

Work Package (WP) 5 (Attribution Analysis and Future Projections of Extreme Events) aims to explore Machine Learning (ML) approaches for the study of Extreme Events (EE), in particular their links with anthropogenic forcing and future changes. This goal is aligned with Objective 2 in CLINT, pursuing the development of Artificial Intelligence (AI)-enhanced Climate Science to support the attribution of single EEs and their observed trends to man-made climate change, and the quantification of future changes in EEs. WP5 is structured in three Tasks, targeting the attribution of individual EEs (Task 5.1), observed trends (Task 5.2) and future changes (Task 5.3) in EEs and concurrent EEs, respectively (see Table 1). This report is the final deliverable of WP5, describing the results from the work carried out in Tasks 5.1, 5.2 and 5.3.

AI/ML techniques have recently been applied to identify anthropogenic signals in the global spatial patterns of annual mean temperature and precipitation (e.g., Barnes et al. 2019), to evaluate the relative contribution of natural and anthropogenic forcings to global warming (e.g. Pasini et al. 2017), or to map the evidence of anthropogenic influences in mean temperature and precipitation (e.g. Callaghan et al. 2021). However, to the best of our knowledge, AI/ML algorithms have not been explicitly applied to quantify climate change influences on individual EEs, revisit the observed trends in EEs or construct feasible (driver-based) future scenarios of changes in EEs and concurrent EEs. Therefore, to assess the potential added value of AI on these issues, a three-step framework has been designed in WP5 (Figure 1):

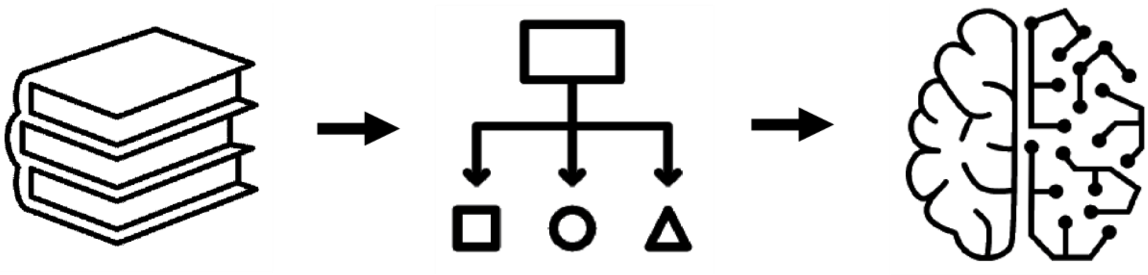

Figure 1 Three-step CLINT framework: review, benchmarking and AI-based developments.

1. Review of existing datasets or approaches (what has been done?): This step aims to provide a critical review in order to assess current capabilities, challenges and ways of improvement for attribution, trend detection and future changes in EEs and concurrent EEs (Table 1). The review relied on papers published in peer-reviewed journals, including review and perspective articles, as well as public reports, books and open-source collections of case studies.

2. Benchmarking (what can be done?): This step identifies 'windows of opportunity' for a ML-based improvement in the attribution of EEs, and the quantification of their observed trends and future projections. The specific aspects that are subject to AI-enhancement must be: i) feasible to implement, thus requiring coordination with WP2 and WP8 for the design of AI models; 2) interpretable to provide process-based understanding of EEs (in support of WP3 and WP4); 3) informative for the impacted sectors (WP6 and WP7). Some of the applications developed by CLINT are novel, whereas others represent improvements with respect to existing ones, in both cases

providing an AI-added value. The specific challenge varies with the Task considered (Table 1), ranging from methods (Task 5.1), datasets (Task 5.2) and understanding (Task 5.3).

3. AI-based applications (what is done in CLINT?): This step aims to explore, develop and implement the most suitable AI-based models to address the specific issues in EEs and concurrent EEs that have been identified in Step 2. It implied the adaptation of existing algorithms to the specific problem at hand, and in some cases, the development of new ones. In general, for the same problem multiple approaches, algorithms and/or architectures have been considered and tested. However, we only discuss herein the definitive ones at this stage (Table 1). AI-based models were designed to be portable and transferable so that, in principle, they can be applied to different regions, seasons and/or EEs. However, not all of these combinations have been tested, and one should not expect that the algorithms serve for all purposes. In these cases, hyper-parameter tuning or transfer learning can be applied. The chosen AI-based models vary in complexity, ranging from AutoEncoders (AEs) and Variational AutoEncoders (VAEs, Task 5.1), to U-Net Neural Networks (NNs) with convolutional layers (Task 5.2), and a combination of deep-learning (DL) algorithms for an optimised feature extraction (e.g. Coral Reef Optimisation, CRO) with ML techniques for dimensionality reduction (Task 5.3). Details of the ML methods and algorithms can be found in the Deliverable (D) D2.2 (ML algorithms for EE forecast and reconstruction) and D2.3 (ML algorithms for climate science).

Table 1 List of tasks addressed in WP5, along with their objectives, benchmarks, ML approaches, expected results and the type of EEs described in this report.

Task	Objective	Benchmark	ML approach	Expected result	EE considered
T5.1	Climate change signals in individual EEs	Reliance on 'framing' choices (spatio-temporal scales, event definition)	Hybrid model: autoencoder - analogue method Variational Autoencoder	Novel approaches for the attribution of the spatio-temporal evolution of individual EEs	Summer heat waves and heavy precipitation at subcontinental (regional) scales
T5.2	Trends in the spatial patterns of extreme indices	HadEX3 database of extreme indices	U-Net Neural Network with partial convolutions	Infilled datasets for a better detection of historical trends in EEs	Monthly frequency of hot and cold days over Europe all year-round
T5.3	Future changes of EEs based on robust responses in their drivers	Identification of climate change drivers of EEs	Dimensionality reduction & optimised feature extraction Kernel Regularised Generalised Canonical Correlation Analysis	Unreported remote drivers and best / worst-case scenarios of future changes in EEs	Summer heat waves in the Po valley Summer hot-dry days in central Europe

This report will focus on the application of ML algorithms for the attribution, detection of historical trends and future changes of EEs (Step 3). Steps 1 and 2 have already been addressed, and can be found in previous reports, including D5.1 (EE attribution), and Milestone (MS) MS25 (Historical EE selected for attribution studies), MS26 (Trends in EEs detected) and MS27 (Storylines of projected changes in each type of EE constructed). Except for the datasets (Section 2), the rest of the report, including the Methodology (Section 3) and Results (Section 4) will be divided following the three tasks of WP5, as they employ different ML algorithms and have their own challenges.

Not all types of EEs, regions/hotspots, or seasons that may be of interest have been described. Instead, a prioritised selection has been made, trying to cover the ample spectrum of spatio-temporal scales and types of EEs considered in CLINT. They range from heat waves (HWs) at hotspot level (e.g. Po Valley) to concurrent hot-dry EEs at large continental scale, which can be informative for WP7 and WP6, respectively. Preference is given to HWs and droughts (DRs), since they represent one of the main problems derived from climate change, due to the observed and projected increase in their frequency, duration and/or severity (Seneviratne et al. 2021). Moreover, these EEs can occur in isolation or in combination (i.e. concurrent EEs), magnifying their effects, particularly in transitional climates characterised by strong land-atmosphere feedbacks such as Europe. Indeed, combined heat and dryness is the most studied type of concurrent EE because of its demonstrated impacts on socio-economic sectors (see e.g. Hao et al. 2022 for a review). In the food sector, for example, combined HWs and DRs can reduce cereal yields by 9-10% at national level, and explain 40% of the yield interannual variability (Lesk et al. 2016; Zampieri et al. 2017).

Based on the successive reports of the Intergovernmental Panel on Climate Change (IPCC), there is a widespread agreement that hot-dry days will increase in both intensity and frequency (IPCC 2021). Recent studies for central Europe have indicated, for instance, that summers like the record-breaking one in 2018 could become the norm in the middle century (Toreti et al. 2019). The increasing trend in hot-dry days is dominated by the intensification of HWs (Wu et al. 2021), which can occur regardless of DR occurrence. HWs cause devastating impacts on human health, ecosystems, agriculture and economy.

Therefore, as an illustration of the ML algorithms developed in WP5, this deliverable will pay special attention to HWs (or temperature extremes in general) as isolated EEs, and hot-dry EEs, as concurrent EEs, although Heavy precipitation (HP) is also considered. The analysis of HWs will focus on continental scales relevant to WP6, as well as regional vulnerable hotspots such as the Po Valley, which is a sensitive area due to the large population and concentration of agricultural and industrial activities. The analysis of concurrent EEs will consider hot-dry days in central Europe (including Rhine Delta and Lake Como basin hotspots), where summer HWs tend to be strongly coupled to DRs, and can significantly harm agricultural crops and hydrological resources.

1.1 Attribution of EEs

EE attribution quantifies the influence of climate change on the probability of occurrence, magnitude or driving factors of a particular EE (NAS 2016; Shepherd 2016; Stott et al. 2016; Otto 2017). Climate change influences can be inferred by comparing the distributions of a class of EE (similar to the observed one) in two climates, one with and one without human influences, also referred to as factual world (the actual world with anthropogenic forcings) and counterfactual world (the world that would have been without climate change). D5.1 reviewed existing knowledge, data and approaches for attribution of EEs to climate change and provided a database collection of about one hundred attribution case studies of different types of EEs, setting up the roadmap for avenues on the use of ML techniques in attribution.

Therein, the analogue method (AM) was presented as a powerful method for fast attribution of EEs, which can be applied to observations and Global Climate Model (GCM) simulations (e.g., Cattiaux et al. 2010; Yiou et al. 2017; Jézéquel et al. 2018; Faranda et al 2022). Besides attribution, the AM is a classical technique for field reconstruction (Zorita and Von Storch 1999). In essence, this is a k-nearest neighbour method, which relies on the fact that two similar atmospheric states (predictors) cause similar surface conditions (target). Specifically, two days are considered analogues when their atmospheric states (or any other conditional factor of the target field to reconstruct) resemble each other in terms of a similarity criterion. These analogues are considered random 'replicates', which allow reconstructing the expected distribution of the target variable.

In EE attribution, the AM searches for historical states of the atmospheric circulation that resemble the one observed during the EE. Changes in the EE are inferred by comparing the distributions of the target reconstructed from analogues of two periods with different levels of anthropogenic forcings. That way, the AM reconstructs how the EE would have been in recent and past periods given the occurrence of the same conditional factor. As the dynamics are the same in both periods, the difference emphasises the influence of thermodynamic changes on the EE, which are easier to detect and attribute to climate change (Shepherd 2016).

Due to the efficiency and low computational cost of AI models, they have the potential of providing fast (near real-time) attribution, circumventing some of the current limitations in attribution (e.g. the need of time-consuming GCM simulations), and bringing novel developments (e.g. the detection of climate change signals in the spatio-temporal pattern of the EE). In particular, an explicit attribution on the evolving patterns of the EE would allow extending the classical attribution question to other attributes of the EE (e.g., duration, spatial extent, trajectory, etc.), and minimise the sensitivity of the results to the EE definition (see D5.1 for more details).

This deliverable describes two novel AI-based approaches for attribution. The first one aims to improve the classical AM through the use of ML algorithms for a more efficient search of analogues than in the classical AM. This method will be tested in two high-impact EEs. A pure AI-based approach (under development) will also be introduced, since it has the potential of providing near-real-time attribution of the changes in the spatio-temporal patterns of individual EEs.

1.2 Trends in EEs

Task 5.2 focuses on ML methods developed in Task 2.5 to detect observed trends in EEs and related extreme indices. For this purpose, different observational datasets widely employed by the climate community have been investigated. These datasets contain diverse climate variables that are derived from weather stations and then interpolated onto a globally uniform spatial grid. Due to inherent limitations of observational measurements (see D2.2), these datasets exhibit missing values that vary in space and time. Data scarcity is especially prominent before the mid-20th century and poses major challenges in the analysis of trends in EEs, which is essential for elaborating effective climate risk assessments and policies. To circumvent this problem, it is common to infill the missing data using statistical methods such as the Angular Distance Weighting (ADW, Shepard

1968), the thin plate spline interpolation (Hutchinson 1995) or Kriging (Oliver and Webster 1990). Nevertheless, these methods suffer from well-known limitations that reduce their effectiveness when applied to climate data with a large amount of missing values.

The emergence of DL-based inpainting techniques in the past years offered the possibility to overcome these limitations through the application of transfer learning with GCM or reanalysis data (Shibata et al. 2018; Dong et al. 2019; Geiss and Hardin 2021; Kadow et al. 2020; Hu et al. 2023; Yao et al. 2023; Olonscheck et al. 2023). For instance, the global reconstruction of the HadCRUT4 monthly temperature dataset (Morice et al. 2021) by Kadow et al. (2020) highlights the remarkable performance of NNs compared to Kriging and Principal Component Analysis (PCA). This DL methodology, based on the U-Net architecture (Ronneberger et al. 2015) and partial convolutional layers (Liu et al. 2018), has been adapted in Task 2.5, and its Python implementation is available at https://github.com/FREVA-CLINT/climatereconstructionAI. The method, described in D2.2, has been applied successfully to the reconstruction of observational datasets of different monthly mean variables (e.g. temperature and precipitation). This deliverable focusses on its application to extreme indices, as is the scope of CLINT, describing the characteristics of the observational dataset, the details of the training and evaluation of the AI models used for the reconstruction and analysis of the results.

1.3 Future changes in EEs and concurrent EEs

Task 5.3 aims to investigate future projections of EEs and their interaction with changes that contextually occur in the drivers of the EEs themselves. EEs and concurrent EEs are linked to large-scale climate variables, which are in turn often influenced by climate change (Horton et al. 2015). Understanding the relationship between large-scale drivers and EEs can help to improve predictions and the understanding of future evolution of EEs. The focal point of this Task is the identification of drivers of EEs and the quantification of changes in the drivers and their effects on EEs. This analysis has been performed using storylines (Shepherd et al. 2018; Zappa 2019), which represent an expected outcome based on one physically self-consistent combination of climate change responses in certain drivers of regional climate (Zappa and Shepherd 2017). This technique aims at dealing with GCM uncertainty by selecting plausible future climate configurations and inspecting the different evolutions of EE that occur in them. These storylines span the range of uncertainty within the multi-model ensemble of future projections through several combinations of changes in a manageable number of drivers. The drivers of an EE can be simplified in the form of time series describing a large-scale internal phenomenon (such as the North Atlantic Oscillation or El Niño-Southern Oscillation) or a climate variable averaged over a specific area.

Storylines help provide further insight into the driving factors of EEs and their regional implications. Several studies have employed this approach to discretise the uncertainty in different regional indicators (Mindlin et al. 2020; Monerie et al. 2023). Storylines of dynamical variables such as precipitation and wind are typically generated from the combined response of remote drivers and global teleconnections of the atmospheric circulation (Woollings 2010). Differently, thermodynamically-driven changes or EEs are affected by regional drivers (e.g. land-atmosphere coupling) and remote factors influencing regional conditions (e.g. sea surface temperatures, SSTs) (Garrido-Pérez et al. 2024 represents a recent application within CLINT).

Storylines rely on the identification of robust drivers of EEs. AI algorithms can help in this task through (see Salcedo-Sanz et al. 2024): i) dimensionality reduction techniques for the construction of candidate drivers; ii) optimisation algorithms for feature selection in high-combinatorial problems in order to find the best set of drivers; iii) advanced regression procedures to account for non-linear interactions between the drivers. This combination of AI-based tools allows to browse large datasets of candidate drivers and identify the most relevant ones for each type of EE.

This deliverable illustrates applications on the AI-based identification of relevant drivers and the construction of future storylines for two types of EEs: HWs in a CLINT hotspot (Po Valley), and concurrent (hot and dry) EEs at European scales, which are useful for WP7 and WP6, respectively. The methods used to define these EEs and their drivers have been applied to reanalysis data, which are used as a benchmark for the evaluation of GCMs, therefore identifying the GCMs that reproduce the relationships between EE and drivers found on reanalysis. The responses of these drivers to climate change are then combined in a meaningful way to construct storylines of future changes in EEs and concurrent EEs.

2. DATASETS

A summary of the datasets employed in this report is presented in Table 2.

Table 2 List of observational, reanalysis and GCM datasets used in this report. Name = short name of the dataset, Source = original producer, handled by = partner doing the download/extraction of the dataset over the region of interest, processed by = partner applying "model output statistics" if any (downscaling, calibration etc.), stored by = partner distributing the data for the consortium partners (e.g. centralised CLINT repository at DKRZ), used by = partner using the data (case study).

Name	Source	Handled by	Processed by	Stored by	Used by
ERA5	C3S-CDS	CMCC	N/A	CMCC	CMCC, JLU, CSIC, UAH
HadEx	MetOffice	DKRZ	DKRZ	DKRZ	DKRZ
MPI-ESM-MR	Hereon	Hereon	Hereon	DKRZ	CMCC, Hereon
GCMs	CMIP6	DKRZ	N/A	DKRZ	CMCC, JLU, DKRZ, CSIC

2.1 Observational products

Observational products of monthly extreme indices with missing values are infilled in Task 5.2 using an AI-based model. They are obtained from the HadEX3 dataset (Dunn et al. 2020), which provides 29 land-based extreme indices derived from daily precipitation and temperature records of 30,000 weather stations since 1901. These monthly indices follow the definition of the Expert Team on Climate Change Detection and Indices (ETCCDI). They are computed at the weather stations and have undergone quality checks before interpolation onto the regular grid with the ADW method.

Extreme indices of the HadEX3 data, available at https://www.metoffice.gov.uk/hadobs/hadex3 (last access 12/10/2022), are not suitable for AI-based infilling, since they have already been interpolated with the ADW approach. Therefore, an intermediate product of the dataset has been created using the Climate Anomaly Method (CAM, Jones 1994), which avoids interpolation. This method computes monthly indices from in-situ data by using the Climpact2 software before its conversion to a gridded form with CAM. The resulting dataset, named HadEX-CAM, focuses on hot and cold temperature extremes across the European continent, which are defined in Section 3.1.

Quality controls, including the identification and exclusion of anomalous data, are conducted to ensure the integrity of the dataset, which covers the 1901-2018 period. As shown in Figure 2, the resulting extreme indices present a large amount of missing values, providing a challenging benchmark for our AI method.

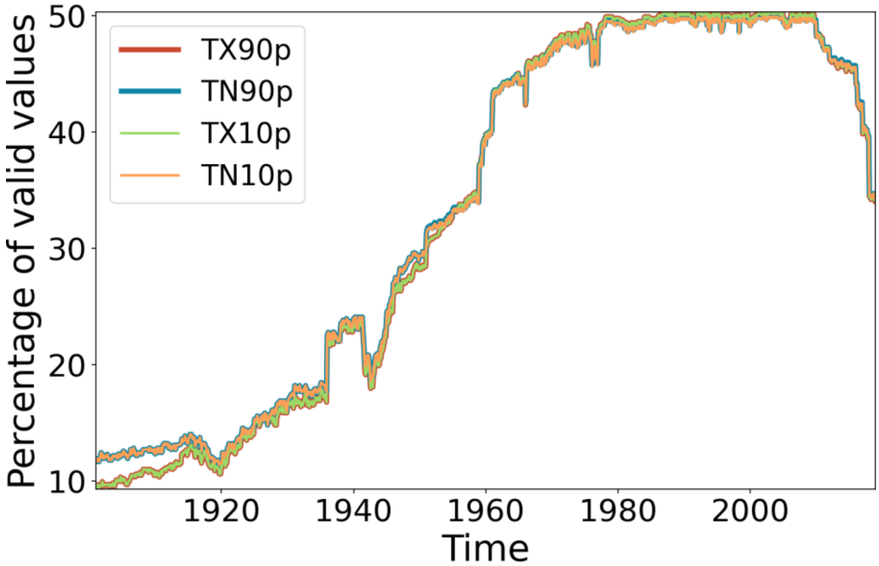

Figure 2 Time evolution of the percentage of valid values in the original HadEX-CAM dataset over Europe from January 1901 to December 2018. Colour lines represent different extreme indices. Note that the time series tend to overlap. For better visualisation, they are plotted with colour lines of different thickness. 100% of valid values corresponds to the total number of values per month covering the land, i.e. 763 values.

2.2 Reanalyses

In all tasks of WP5, we use data from the ERA5 reanalysis data (Hersbach et al. 2020) over the 1940-2022 period. This reanalysis is employed as a ground truth and for training some AI-based models in real world (benchmarking). In some applications, it is also used for evaluation of the results retrieved from GCMs. The variables, spatial resolution and domain and temporal frequency of the data vary depending on the application (details are provided in the corresponding section).

2.3 Global Climate Models

Simulated data come from GCMs participating in the Climate Model Intercomparison Project phase 6 (CMIP6). CMIP6 data were obtained from the Earth System Grid Federation node of DKRZ, accessible on the Levante platform (Table 3).

Table 3 Details of the CMIP6 data used in WP5. (SSP = all SSP scenarios; mmb01 = member number one).

Model	Number of members	Available simulations	Nominal resolution
AWI-CM-1-1-MR	5	Historical, SSP (mmb01), SSP370 (others)	100 km
BCC-CSM2-MR	1	Historical, SSP	100 km
CAMS-CSM1-0	1	Historical, SSP	100 km
CMCC-ESM2	1	Historical, SSP126, SSP245, SSP585	100 km
CNRM-CM6-1	3	Historical, SSP (mmb01), SSP370 (other)	250 km
CNRM-CM6-1-HR	1	Historical	50 km
EC-Earth3	21	Historical, various SSP for each mmb	100 km
EC-Earth3-CC	1	Historical	100 km
EC-Earth3-Veg	1	Historical	100 km
EC-Earth3-AerChem	1	Historical	100 km
GFDL-CM4	1	Historical, SSP245, SSP585	100 km
HadGEM3-GC31-MM	1	Historical	100 km
HadGEM3-GC31-LL	10	Historical, Natural, SSP245	250 km
INM-CM4-8	1	Historical, all	100 km
MPI-ESM1-2-HR	10	Historical, SSP (mmb01), SSP370(others)	100 km
MRI-ESM2-0	5	Historical, SSP (mmb01), SSP370(others)	100 km
NorESM2-MM	1	Historical, all	100 km

Task 5.1 also employs simulated data from the HadGEM3-GC31-LL GCM. In this case, historical and natural simulations (both covering 1850-2014) are considered, which are commonly employed in the CMIP6-component Detection & Attribution Model Intercomparison Project to facilitate an improved estimation of the climate response to anthropogenic forcings over the historical period (Gillett et al. 2016). Natural forcing simulations are equivalent to historical runs in that they impose the observed evolution of natural forcings (solar, volcanic, etc.), but anthropogenic forcings are fixed at preindustrial levels. Therefore, natural experiments allow inferring how the historical period could have been without increasing levels of anthropogenic forcings (but with the same natural forcings as in observations). The HadGEM3-GC31-LL model has been selected because it provides a high number of realisations. Furthermore, it performs well in the simulation of many EEs over

Europe, and its atmospheric-only version, HadGEM3-A, is the core of the near-real time attribution system of the Hadley Centre employed in many attribution studies (e.g., Ciavarella et al. 2018).

Historical simulations (1850-2014) of eight GCMs participating in CMIP6 have been used as an independent and complete dataset to train the AI-based infilling models in Task 5.2. In this case, GCMs are selected by imposing a minimum horizontal resolution that must be higher or equal to that of the target (HadEX-CAM) dataset (see Table 3).

GCM data are also used for the calculation of EEs, concurrent EEs and drivers in Task 5.3. They include historical experiments and Shared Socio-economic Pathway (SSP) scenarios (2015-2100) of CMIP6 GCMs. Future climate projections include the ScenarioMIP pathways: SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5. The computation of EEs and drivers in GCMs follows the same approach as in ERA5. Only members and experiments with complete series are selected. In some cases, the low resolution of the GCM provided an insufficient number of grid points for the calculation of EEs. For this reason, only models with horizontal resolution finer than ~150 km are considered (Table 3).

3. METHODS

3.1 Extreme indices and concurrent events

Temperature extreme indices: For the analysis of trends (Task 5.2), we use the HadEX-CAM dataset (Section 2.1), which includes monthly temperature extreme indices, based on daily maximum (TX) and daily minimum (TN) temperature:

· Warm days: the percentage of days in the month when the TX is higher than the 90th percentile
· Warm nights: the percentage of days in the month when the TN is higher than the 90th percentile
· Cool days: the percentage of days in the month when the TX is lower than the 10th percentile
· Cool nights: the percentage of days in the month when the TN is lower than the 10th percentile

The daily percentiles are calculated over a 1981-2010 base period using a 5-day window centred on each calendar day, following the ETCCDI definition.

HWs: HW indices were calculated for ERA5 and each combination of model, member and scenario (from now on generally referred to as simulation) by using the scripts developed in D2.2. HW definition relies on the daily series of TX exceedances above a seasonally-varying threshold, defined as the daily 90th percentile of TX over a baseline period. From these series, various yearly indices are obtained, such as the 90th percentile of TX (TX90p), the number of days above the 90th percentile (NDQ90), or the Heat Wave Magnitude Index (HWMI; Russo et al. 2015). The series of TX exceedances determines HW occurrence, which is used as a target for the AI-based identification of drivers (see Section 3.5), while the series of yearly indices are employed for the evaluation of GCMs in the future climate and the construction of storylines (Section 4.3). The HWMI is used to identify the top European HWs of 1950-2010, which are employed to test the performance of the hybrid model for attribution (Section 3.3). For the computation of percentiles in ERA5, the baseline is the 1981-2010 period, which has a global mean near-surface temperature (GSAT) of 14.2 °C. For each GCM simulation, the 30-yr period with the same GSAT as ERA5 is chosen as baseline (see Section 3.2).

Bivariate heat magnitude day (BVHMD): To characterise HWs in the context of concurrent EEs, we extend the Heat Magnitude Day of Zampieri et al. (2017) by including both TX and TN. This allows us to describe temperature-related impacts more comprehensively, since some sectors are also affected by TN (Perkins and Alexander 2013). These methods and the derived BVHMD will be described in D3.3 (AI-enhanced Extreme Events detection).

Non-parametric SPEI (NPSPEI): For DR detection, we use our recently proposed non-parametric version of the Standardised Precipitation and Evapotranspiration Index (SPEI). This approach uses a non-parametric local likelihood-based method (Loader 1996; Geenens et al. 2017) to estimate the distribution function instead of a parametric distribution like the classical SPEI (Vicente-Serrano et al. 2010). More details can be found in D3.1 (EE Detection). The index is based on water balance (WB), defined as the difference of total precipitation and potential evapotranspiration. For the latter, we use the Penman-Monteith approach (Allen et al. 1998). This physically-based approach is recommended by the IPCC for studying DRs, as temperature-based approximations (e.g., Hargreaves, Thornthwaite) tend to overestimate trends and the magnitude of DRs, especially in the context of global warming. In the report, we will use the NPSPEI-1 for DR characterization, here called NPSPEI for ease of notation, which takes into account the monthly accumulated water unbalance.

Figure 3 shows the co-variability of BVHMD and NPSPEI for the summers (June-to-August) of 1981-2010 in ERA5, demonstrating the capability of these indices to capture the HW-DR dependence.

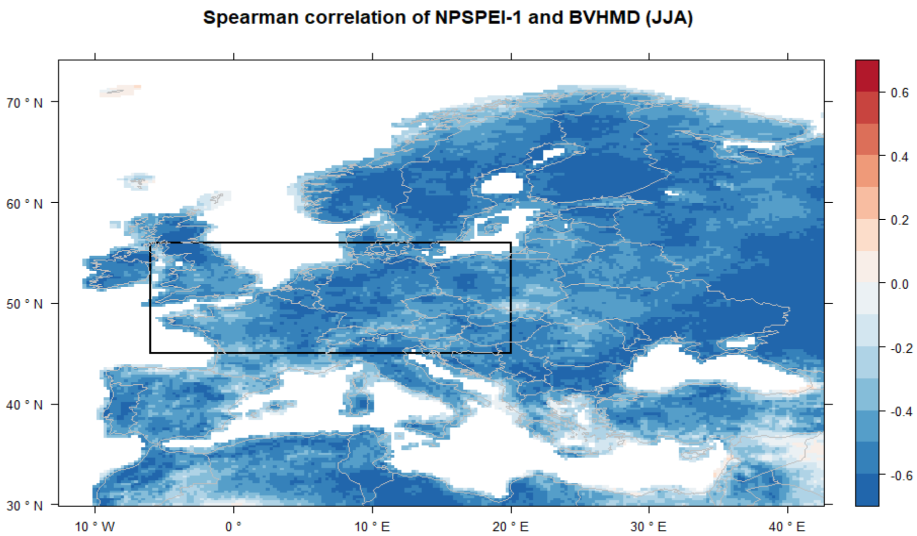

Figure 3 Spearman correlation of monthly NPSPEI-1 and bivariate HMD for June, July and August (JJA) 1981-2010. The black square identifies the central European region [6°W-20°E, 45-56°N] employed for the assessment of hot-dry concurrent EEs.

Concurrent extreme event index: The Concurrent Extreme Event Index (CEEI) is used for the analysis of concurrent EEs. It is a one-dimensional unitless time series constructed from the above mentioned HW- and DR-related indices by using Kernel regularised generalised canonical correlation analysis (KRGCCA; Tenenhaus et al. 2015, 2017). This method identifies the dominant patterns of BVHMD and NPSPEI, as well as those of selected drivers (herein SST and 500 hPa geopotential height, Z500), while concurrently maximising the correlation among these components. By examining the interactions between BVHMD and NPSPEI, the CEEI reflects the dominant patterns of the combined space of HW and DR indices. That way, the CEEI characterises the connection of HWs and DRs: positive values indicate warm and dry conditions, with increasing values denoting increasing severity. The CEEI will also be indirectly linked to the drivers of these two EEs through correlation properties. See D2.2 for technical details.

3.2 Common Warming States

The evaluation of CMIP6 GCMs (with respect to ERA5) is performed by considering the 30-yr periods in which each GCM simulates current climate conditions. These reference periods (and the corresponding future climate periods) are termed Common Warming States (CWSs). They require the computation of GSAT and provide an alternative and complementary approach for model evaluation and projections to traditional methods based on time-fixed periods or Global Warming Levels (GWLs). The simulated period set as reference of current conditions is directly compared to the benchmark (ERA5) and used for GCM validation. The definition of future climate periods builds upon that of current conditions.

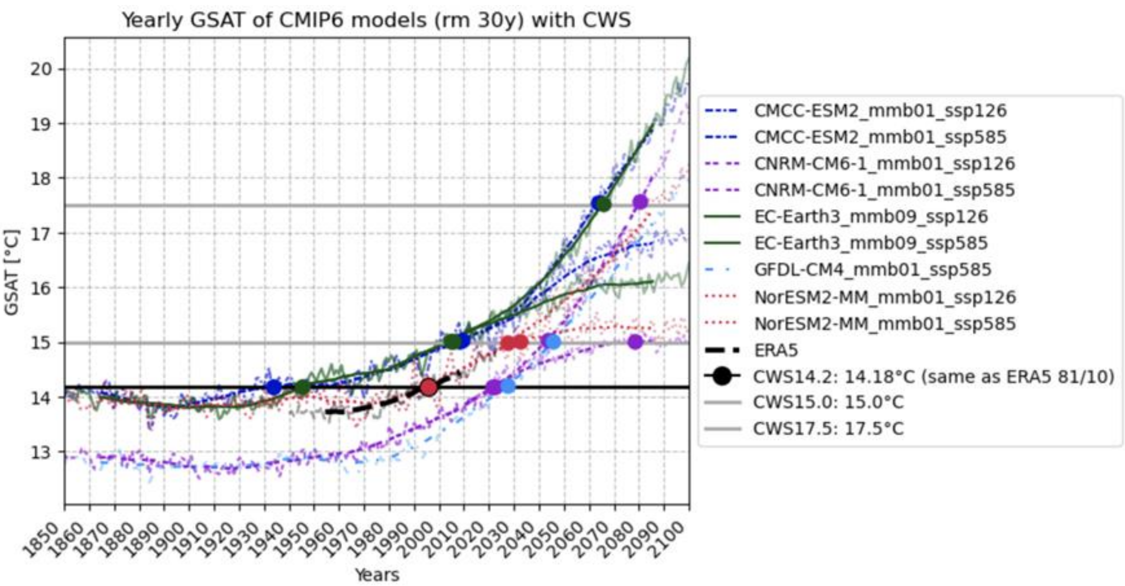

Figure 4 Yearly GSAT (with 30-year running mean) of a sample of GCMs and ERA5. Intersections with horizontal lines indicate the central year of the corresponding CWSs.

CWSs identify periods in which the GCMs have the same (predefined) absolute value of GSAT. This approach aims at assessing the effects of global conditions on EEs (regardless of the pre-industrial values simulated by the GCM). In particular, CWSs allow us to inspect how EE occurrence and intensity change when the GSAT reaches a specific threshold. As current climate, we consider the 1981-2010 period, which has a GSAT of 14.2 °C in ERA5. Then, for each GCM simulation we identify the 30-yr period when this GSAT value is simulated, which is referred to as CWS14.2. Note that simulations can have their CWS14.2 ending after 2014. In that case the CWS14.2 is determined for each scenario separately. Similarly, for future projections a threshold of 15 °C (CWS15) is established. The process is illustrated in Figure 4, which shows the 30-yr centred periods in which different GCMs cross the aforementioned periods representative of current (CWS14.2) and future

(CWS15) conditions. This novel approach is chosen as an alternative framework to the well-known GWLs, which are defined as periods with the same GSAT increase over the pre-industrial average.

3.3 ML algorithms for attribution

We consider two AI-based algorithms for attribution: a hybrid approach, AE-AM (in collaboration with WP4), combining AE with the classical AM, and a pure AI-based method (VAE). As stated in the introduction, the AM reconstructs the conditions expected during the EE (target) by considering analogues, i.e. days with similar conditional factors (predictors) to the one observed at the time of the EE (bottom panel in Figure 5). For each calendar day d of the EE, we identify the best K = 20 analogues of the ERA5 record (1940-2022), excluding the year of the EE. The similarity metric is the root mean squared error (RMSE) of the atmospheric circulation over a predefined spatial domain. As a predictor, different atmospheric circulation variables are tested, including sea level pressure (SLP) and Z500, without reporting substantial differences. The distribution of the targeted field is reconstructed by randomly picking one of the K best analogues for each day of the EE, and repeating this process N = 1,000 times. The associated error is measured as the departure between the reconstructed and observed values.

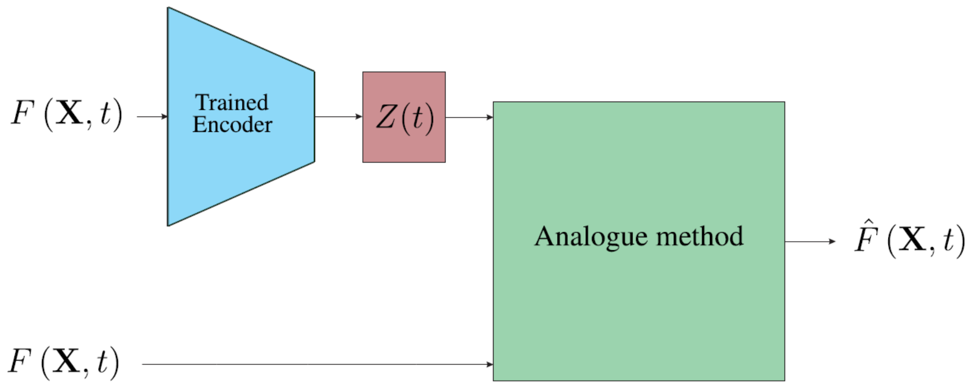

Figure 5 AE-AM model (top) and classical AM (bottom). The EE under study at time t is defined by the variable F over the domain X. In the AE-AM approach F(X,t) feeds the trained encoder to obtain the latent space Z(t). The AM is then applied to obtain the reconstruction of F(X,t) by using the closest neighbours of F(X,t) (AM) and of Z(t) (AE-AM). Adapted from Pérez-Aracil et al. (2024)

Recently, ML has been responsible for significant advances in AM-based modelling (Salcedo-Sanz et al. 2024). As a joint effort of WP2, WP4 and WP5, CLINT has designed a novel ML-based hybrid approach for the attribution of EEs, combining AEs and the AM. An AE is an unsupervised method that comprises two deep NNs, one to encode information (encoder) and the other to decode it (decoder). The intermediate (encoded) representation is called latent space, which provides a meaningful and compact representation of the input data, from which the original data can be reconstructed with the decoder. The hypothesis of the AE-AM is that the AM can find better analogues by working on this optimised space of reduced dimensionality than on spatially resolved fields of the predictor. The method has been developed by WP2 parners for WP4 and WP5. For details on the architecture of the AE-AM model, see D2.3 and Pérez-Aracil et al. (2024). The AE is trained in ERA5 data (excluding the observed EE), using the difference between the input and the decoded field as loss function. Once it has been trained, the encoder is employed for the reconstruction of the EE of interest: for each day of the EE, the encoder transforms the predictor field to its latent space, and the AM searches for the K closest analogues in this reduced space (top panel in Figure 5). Note that the decoder is not used for reconstruction. Therefore, AE-AM and AM only differ in that they search analogues of the filtered and raw predictor fields, respectively, which makes the AE-AM highly interpretable.

As the AM-based attribution relies on field reconstruction, the added value of the AE-AM method will be demonstrated by comparing the AM (benchmark) and AE-AM reconstructions of the TX fields observed during eight major European HWs (Pérez-Aracil et al. 2024). Table 4 lists the spatial domains used for the predictor (SLP) and target (TX) fields of each HW event, which vary from case to case. In a second stage, the AE-AM is applied for the attribution of a reduced number of selected EEs, including a HW and a heavy precipitation (HP) event. The attribution is performed by comparing the variables of these EEs (TX for HWs and precipitation for HP), as reconstructed by analogues from two periods with different levels of anthropogenic forcing: 1940-1980 (old or counterfactual world) and 1981-2022 (new or factual world). The difference between the reconstructions of the factual and counterfactual worlds is attributed to recent climate change.

Table 4 Summary information of the HWs considered for the assessment of AE-AM vs. AM approaches.

HW event	Dates	Mean TX (°C)	SLP domain	TX domain
France 2003	01–19 Aug	28.4	[32–70]°N, [28°W–30°E]	[42–50]°N, [6°W–8°E]
Spain 1995	16–24 Jul	30.5	[32–70]°N, [28°W–30°E]	[34–42]°N, [10°W–4°E]
Greece 1987	18–27 Jul	30.2	[28–66]°N, [8°W–50°E]	[34–44]°N, [18–32]°E
Germany 2006	09–31 Jul	26.2	[24–72]°N, [28°W–30°E]	[44–54]°N, [4°W–16°E]
Poland 1994	21 Jul–11 Aug	28.8	[32–70]°N, [18°W–40°E]	[48–56]°N, [14–26]°E
Balkans 2007	15–28 Aug	29.2	[32–70]°N, [8°W–50°E]	[40–52]°N, [18–42]°E
Russia 2010	16 Jul – 19 Aug	32.6	[32–70]°N, [22–80] °E	[38–60]°N, [40–60]°E
Russia 1954	01–12 Jul	28.8	[32–70]°N, [8°W–50°E]	[44–60]°N, [28–48]°E

The second approach for attribution uses a VAE, which represents a probabilistic variant of the AE (Klampanos et al. 2018; Kingma and Welling 2019). In a VAE the representation of the latent space maps each input state into a probability distribution, from which a reconstruction ensemble is retrieved (instead of just one single decoded field, as in AEs). Describing the parameters of the latent space with probability distributions is highly desirable to account for uncertainty in reconstructions. VAEs (and AEs) can be used to detect anomalies, i.e. values in the input field that exceed a given threshold, usually inferred from the maximum reconstruction error during training (e.g. Camps-Valls et al. 2021). This rationale is employed in CLINT to detect anthropogenic signals in EEs (Figure 6).

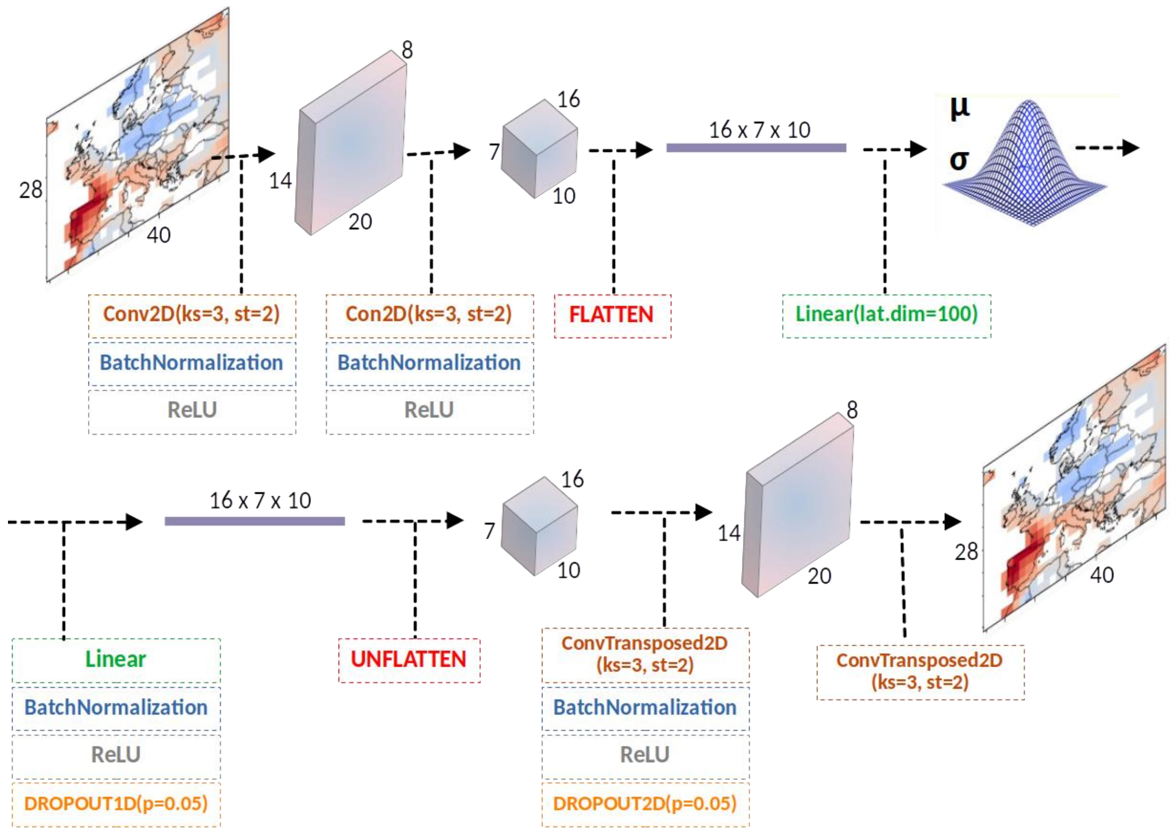

Figure 6 Schematic representation of the VAE employed for attribution in CLINT.

To do this, the VAE is trained with daily fields (herein TX) from an ensemble of natural simulations (i.e. without anthropogenic forcings). After training, it is used to reconstruct and detect anomalies in daily fields from historical experiments of the same GCM (i.e. with anthropogenic forcings). For a given input field of the historical run (e.g. the TX field of a HW), the VAE will reconstruct the features of the natural world learnt during the training, thus delivering an ensemble of naturalised reconstructions of the input field (e.g. multiple realisations of how the EE could have been without climate change). If the threshold for the anomaly detection is chosen as the maximum reconstruction error during training, the detected anomalies will represent virtually impossible outcomes for a natural world. The architecture of the VAE is described in D2.3. For testing this AI-based attribution approach, we have used an ensemble of historical and natural simulations of the HadGEM3-GC31-LL GCM (Section 2.3). The VAE is trained with two input channels: daily European maps of TX for the summer seasons of the natural simulations (1850-2014), and the corresponding calendar day to account for intra-seasonal variations. After training, the VAE is applied to detect climate change signals in HWs simulated by the GCM in the historical simulation.

3.4 ML algorithms for reconstruction

In recent years, DL techniques have emerged as efficient approaches to tackle diverse problems in climate science. Examples of these DL models are densely connected networks, convolutional NNs (convnets) and recurrent NNs. The DL-based method employed for the reconstruction of the HadEX-CAM observational dataset (Section 2.1) is detailed in D2.2. Herein, we summarise the main characteristics of the method, as well as the features relevant to the reconstruction of the specific variables under consideration in Task 5.2.

As depicted in Figure 7, the algorithm employs a U-Net that is made of partial convolutional layers. As shown by Liu et al. (2018), this type of convolutions is more effective in the infilling of large and irregular regions of missing values compared to standard convolutions. This characteristic is particularly beneficial for the reconstruction of observed climate variables, since the data can be very scarce in earlier periods, as in the HadEX-CAM dataset.

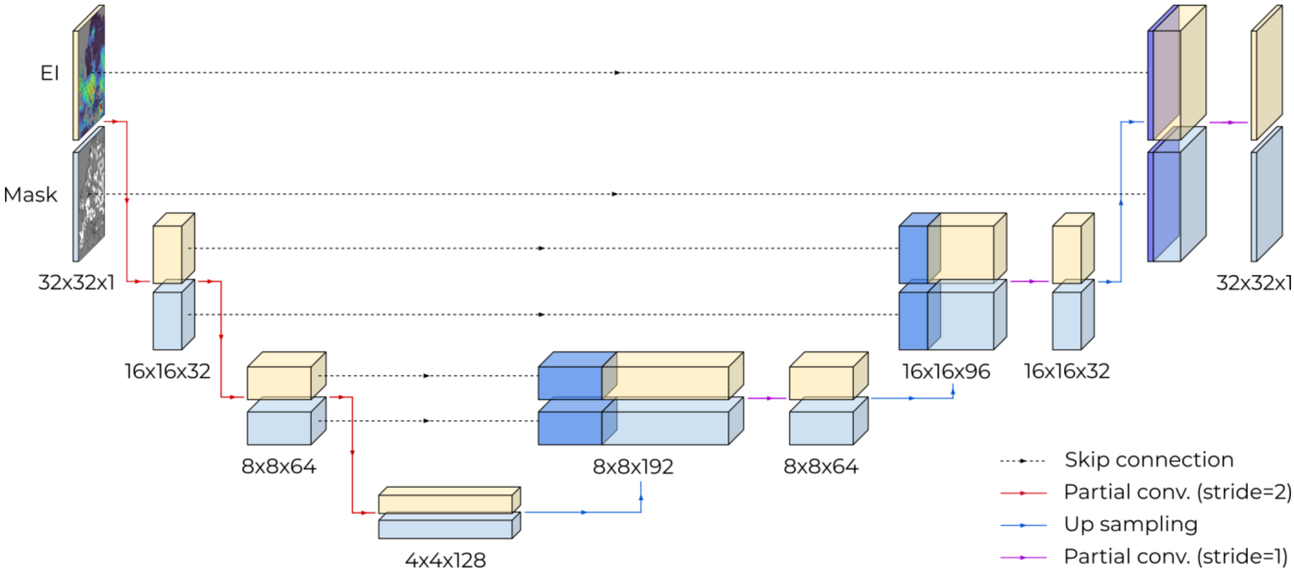

Figure 7 Example of U-Net architecture used to reconstruct the HadEX-CAM dataset.

Multiple loss functions can be employed in combination with this architecture of NNs (the specific choice is expected to depend on the dataset). Two types of loss functions have been implemented: a standard Mean Absolute Error (MAE) loss that is more relevant for pixel-level accuracy; and a sophisticated loss function (Liu et al. 2018) that favours physical realism. The latter, hereafter referred to as inpainting loss function, is a combination of five terms whose relative contribution to the loss function is determined empirically. For a full description of the terms, see D2.2.

The hyperparameters of the infilling models, such as the number of encoding/decoding layers or the learning rate, are also expected to be specific to each dataset and hence they have been determined by performing a hyperparameter search. Additionally, some optional features have been implemented, such as the data normalisation, the circular padding for global data and the binding of the predictions to a limited range of values. They permit to improve the accuracy of the U-Net and can be used when they are relevant to the characteristics of the dataset.

Given the limited length and the incomplete nature of the observational datasets, it is required to employ a transfer learning methodology by training the infilling AI models in an independent complete dataset, herein taken from historical simulations of CMIP6 GCMs (Section 2.3). The GCM data is pre-processed to match the definition and characteristics of the observational HadEX-CAM variables by calculating the extreme temperature indices of Section 3.1 with Climate Data Operator (Schulzweida 2022). The data is finally regridded to match the spatial resolution of the target dataset and split into a training, a validation and a test set (Table 5).

Table 5 Total number of samples and number of samples per month for the training, validation and test sets.

Dataset	Training set		Validation set		Test set	
	Total	/month	Total	/month	Total	/month
HadEX-CAM	50616	37	9576	7	1368	1

The NN (Figure 7) employs two types of inputs: the gridded variable and the masks of missing values derived from the dataset to be infilled. During the training, the masks are selected randomly for each sample and used to create artificial missing values in the training dataset of complete GCMs. The inputs are propagated through the NN following the mask-update procedure described in D2.2, and the loss function is calculated for the missing values only by comparing the predictions of the infilled dataset with the original one. The training is performed iteratively and stopped when optimal training and validation loss values have been reached. The validation set is used during the training step to calculate the learning curves and prevent overfitting of the models.

The test set is used to evaluate the accuracy of the trained models by using unseen GCM data. For this purpose, artificial missing data is created in the test set of CMIP6 GCMs for each month by using the corresponding mask of missing values extracted from the observational dataset. Evaluation metrics, such as RMSE or Spearman rank correlation coefficient are calculated to assess the performance of the infilled predictions with respect to the unmasked test data. Once the infilling model has been trained, validated and tested, it is applied to the observational incomplete dataset in order to reconstruct the spatial fields of the targeted observables for each month of the considered period (1901-2018).

3.5 ML algorithms for storylines

Storylines describe a set of plausible evolutions of a variable or event of interest in a changing climate (Zappa and Shepherd 2017). They are constructed according to the occurrence of specific configurations, such as the combination of specific changes in the drivers of EE or the fulfilment of particular constraints that are expected to influence the analysed EE. In this work storylines are constructed considering the future evolution of certain drivers of EEs, as simulated by CMIP6 GCMs (Section 2.3). The drivers can be weather or climate variables, often summarised as indices or coefficients influencing the occurrence or intensity of that type of EE. Drivers are assessed and identified by means of AI-based algorithms developed in WP2 and applied in WP3-5. These algorithms represent powerful tools to integrate the current knowledge about EEs and discover unreported drivers, teleconnections or non-linear relationships in the climate system.

The driver selection method is a two-step approach, comprising dimensionality reduction via spatial clustering of candidate drivers and feature extraction for optimised selection of drivers. A similar algorithm has been designed for forecasting in CLINT (D2.2). For the selection of drivers, a list of potential candidates has been created (Table 6) aiming to capture regional aspects of atmospheric and ocean variables that are potentially relevant for many types of EEs. Driver variables include SSTs, SLP or soil moisture (the selection can be customised based on process understanding of the considered EE). Then, a K-means spatial clustering is applied to each driver in order to reduce the dimensionality of the input data (and the number of candidates).

In a second step, the time series of the spatial clusters of all variables are used as input of the Coral Reef Optimisation (CRO) algorithm (Salcedo-Sanz et al. 2014; Pérez-Aracil et al. 2023), which performs an optimised feature extraction of the most relevant drivers of the EE. CRO is an evolutionary algorithm implementing several strategies for optimisation in feature selection problems. It has been developed in WP2. For technical details, see D2.2. Herein, CRO is used to find the combination of drivers that better explain the evolution of the target (e.g. HW occurrence in Po Valley) in a given dataset. The algorithm is able to detect combinations of drivers at different lags and with different durations. Further details about the algorithm are provided in D2.3.

Table 6 List of candidate drivers of EEs.

Variable	Abbreviation in CMIP6	Abbreviation in CLINT	Domains of clustering
Geopotential height at 500 hPa	Zg	Z500	Europe, World
Maximum temperature at 2m	Tasmax	Tmax	Europe
Mean sea level pressure	Psl	mslp	Europe, World
Outgoing longwave radiation	Rlut	olr	North Atlantic, World
Sea ice cover	Siconc	Sic	Arctic
Sea Surface Temperature	Tos	sst	North Atlantic, World
Soil moisture (mass of water in all phases in the top 10 cm layer)	Mrsos	Sm	Europe
Total precipitation	Pr	Tp	Europe

The K-means - CRO approach for the detection of drivers is first applied to ERA5, and the resulting drivers are considered the ground truth. Before the construction of the storylines, the GCMs are validated by applying the same detection algorithms to the set of GCMs described in Section 2.3, and retaining those GCMs that identified the same drivers as ERA5. That way, the validation of the GCMs is driver-based, meaning that for each driver there is a set of validated GCMs that will be employed for the construction of storylines. The best solutions are retained, thus assigning a probability of occurrence to each set of N drivers, which is combined with physical understanding to assist in the final selection of drivers. Due to the high number of storylines that would result from the use of a high number of drivers, all analyses are performed considering a maximum of two drivers at a time, which is expected to generate four storylines.

In the construction of storylines, the validated GCMs are sorted into groups according to the future changes in the selected drivers. Changes in the drivers are defined as the difference between future and current climates (i.e. CWS15 and CWS14.2; Section 3.2). For each driver, GCMs are then classified into those simulating higher or lower changes than the multi-model mean. Groups of GCMs are then formed in the multivariate driver space, by inspecting which GCMs follow specific combinations of changes in the drivers. For each combination of responses in the two drivers, it is then possible to identify the GCMs that follow that specific storyline, and evaluate accordingly the projected changes in the EE of interest (including best and worst scenarios of EEs). That way, the storylines assess the changes in EEs as a function of the responses in the selected drivers, trying to reduce the uncertainty related to traditional multi-model approaches.

3.6 ML algorithms for concurrent EEs

The ML algorithms aim to identify the dominant components and associated drivers that capture the characteristics of the system formed by concurrent HWs and DRs. PCA and canonical correlation analysis along with their combinations are commonly used as dimension reduction algorithms in climate science (e.g., Wilks, 2011). However, these methods have limitations because they can only discern linear relationships, are confined to two spatial fields, and are sensitive to pre-processing choices (e.g., retained number of components, orthogonality of input features, etc.). Tenenhaus et al. (2015) introduced KRGCCA, which can address these challenges by employing appropriate regularisation schemes to handle high-dimensional predictors and mitigate high collinearity or spatial dependencies, as observed in climate data. Moreover, it can be used to extend the methods to the analysis of non-linear relationships by making implicit calculations through the use of kernels, thereby avoiding the need to specify a non-linear aggregation function (Schölkopf 2002). More details in these algorithms can be found in D2.2, D3.2 (Preliminary AI-enhanced EE detection) and D2.3.

As described in Section 3.1, the KRGCCA is employed to generate CEEI as well as the associated drivers. Furthermore, we perform blockwise scaling (Garali et al. 2018) to ensure that the algorithm is not dominated by certain climate variables. Finally, spatial weighting (North et al. 1982) of the input features is applied to take the spatial information into account. In what concerns the drivers, variations in SSTs are a relevant predictor of EEs in seasonal forecasts. Thus, lagged SSTs, with a maximum lag of three months are considered as input variables. In particular, spring SSTs may serve as early indicators of summer HWs in central Europe (Beobide-Arsuaga et al. 2023). Moreover, we consider lagged effects of Z500 to account for potential driver interactions since large-scale atmospheric patterns may contribute to decreased precipitation and soil dryness, which can precede the onset of HWs (Miralles et al. 2019).

After the CEEI and the corresponding drivers have been extracted by the KRGCCA, additional AL-based models are employed to assess the relationships between those variables. To model the conditional distribution function (CDF) of the CEEI on the associated drivers, we use a recently proposed approach called Quantile Regression Using I-Spline Neural Network (QUINN) by Xu and Reich (2023). It uses a synergy of Bayesian NNs and I-splines to estimate the CDF. For more details, see D2.2 and Xu et al. (2022).

Accumulated local effect (ALE) plots (Apley and Zhu 2020) are employed to compensate for the lack of interpretability of NNs and to understand the marginal effect of the drivers on the CEEI. They approximate the (potentially non-linear) effect of the drivers on the CEEI. Furthermore, the standard deviation of these quantities can be used to define variable importance criteria (Xu and Reich 2023), which are useful to identify the drivers that are more important for the behaviour of the CEEI and thus for concurrent hot and dry EEs. Additionally, ALE plots can visualise higher-order effects of input variables, describing the predictive power of interactions among input variables. This allows us to explore how combinations of input variables affect specific parts of the distribution, which is valuable for characterising the compounded nature of concurrent EEs that often considered as pre-conditioned events (Zscheischler et al. 2020).

4. RESULTS

4.1 Attribution

4.1.1 Attribution with AE

First, we compare the ability of the AM and AE-AM approaches to reconstruct the intensity of the HWs considered (Table 4). To do so, the reconstructed variable is averaged over the targeted spatial domain and over the duration of the EE. The performance of the AE-AM compared to the AM is quantified with the Skill Score. A sensitivity test based on different sizes of the AE latent space (8, 64, 128, 256, 400, 600, 700 and 800 nodes) is also carried out in order to assess the influence of this hyper-parameter in the training of the AE.

Focusing on the observed magnitude of the EE, the performance of the AE-AM is better than that of the AM for all of the HWs considered, and for all the latent space dimensions (Figure 8), except for the smallest one (8 nodes), which does not seem to include enough information to yield a competitive reconstruction of the HW intensity. In the rest of the cases, the AE-AM distribution is closer to the target than that of the AM, indicating an AI-enhanced reconstruction of TX for the most severe European HWs. The improvement of the TX reconstructions obtained with the AE-AM is larger than 10% (exception made for the latent space of 8 nodes). Overall, the reconstructed error depends more on the HW analysed than on the latent space dimension, meaning little sensitivity to this hyper-parameter.

Both, AM and AE-AM, are able to reconstruct warmer-than-average conditions, indicating that the predictor (atmospheric circulation) is a major driver of these EEs (e.g. Jézéquel et al. 2018; Faranda et al. 2022). The reconstructions explain a large fraction of the observed TX, the remaining being attributed to non-dynamical processes (e.g., land-atmosphere feedbacks) and/or limited sampling (i.e. few historical analogues of the given severity). Despite this, the results demonstrate a clear advantage of pre-processing the data with the AE before applying the AM method.

Additional analyses confirm the superiority of AE-AM, also during non-extreme periods (not shown). However, the differences between the AE-AM and the AM reconstructions increase during HWs, suggesting that AE-AM is especially well suited for EEs. Furthermore, the reconstruction of TX with simpler linear methods of dimensionality reduction such as PCA does not outperform that obtained with the AE-AM, at least for the analysed HWs (see examples in Figure 9).

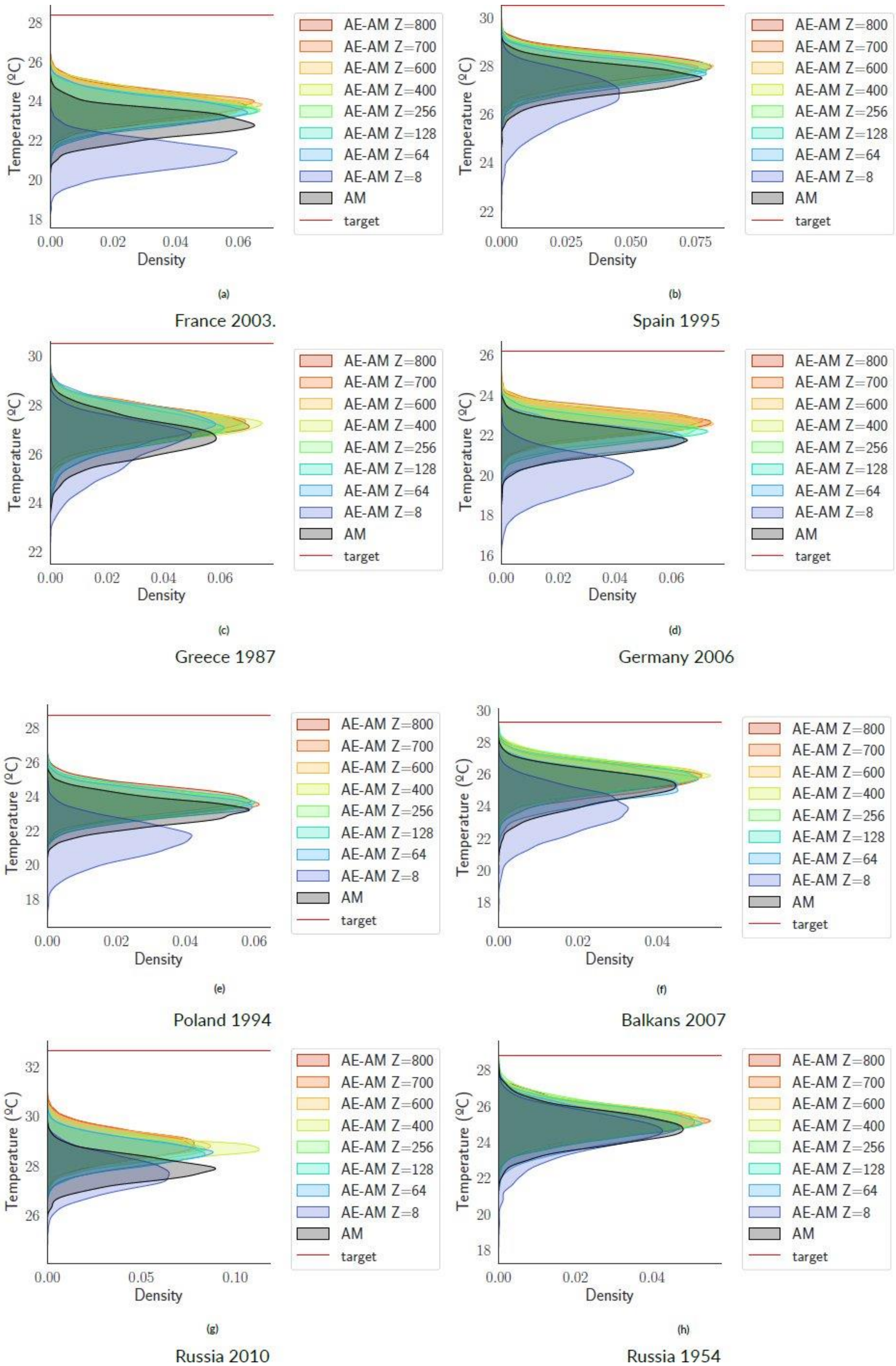

Figure 8 Comparison of the TX distributions reconstructed by the AM and the AE-AM using different dimensions of the latent space in the HWs of: a) France 2003; b) Spain 1995; c) Greece 1987; d) Germany 2006; e) Poland 1994; f) Balkans 2007; g) Russia 2010; h) Russia 1954.

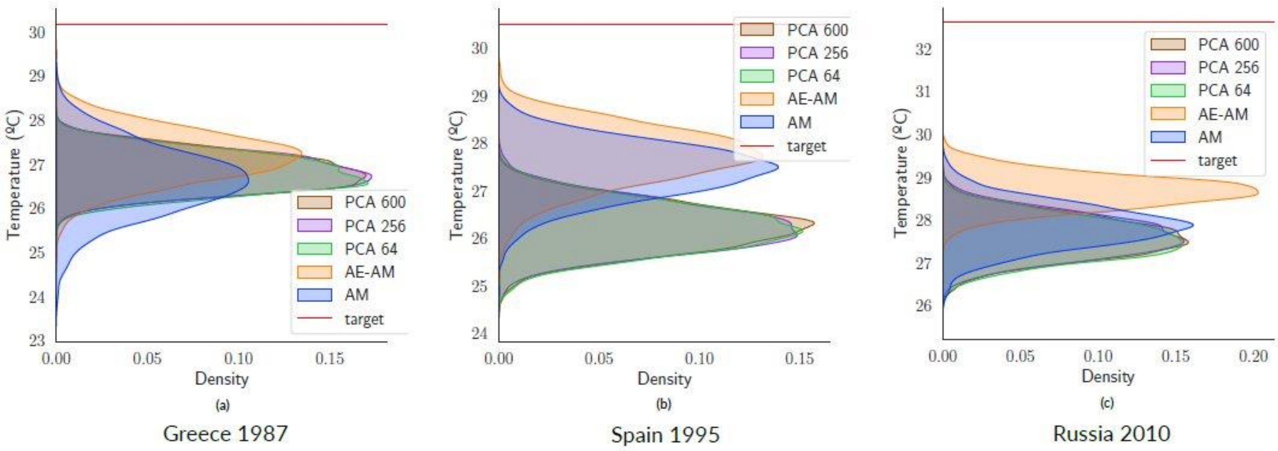

Figure 9 Comparison of the TX distributions reconstructed by PCA-AM for different number of components (64, 256 and 600), and the AE-AM and the AM methods in the HWs of: a) Greece 1987; b) Spain 1995; c) Russia 2010.

To understand the differences between the TX reconstructions retrieved from AM and AE-AM, we analysed the analogue days selected by both methods. For simplicity, the analysis will focus on the 2003 HW only. Many analogue days of that HW (188) are shared by the two methods. However, AM uses 118 days for reconstruction that AE-AM does not, and AE-AM has 99 days that are not analogues in AM. Using only these disjoint analogues, it is found that the AE-AM performs better than AM in terms of the magnitude and spatial pattern of the targeted TX field, reducing the bias of the AM reconstruction in all grid points (Figure 10). This shows that AE-AM is more efficient in selecting optimal analogues for the reconstruction of HWs than the classical AM method.

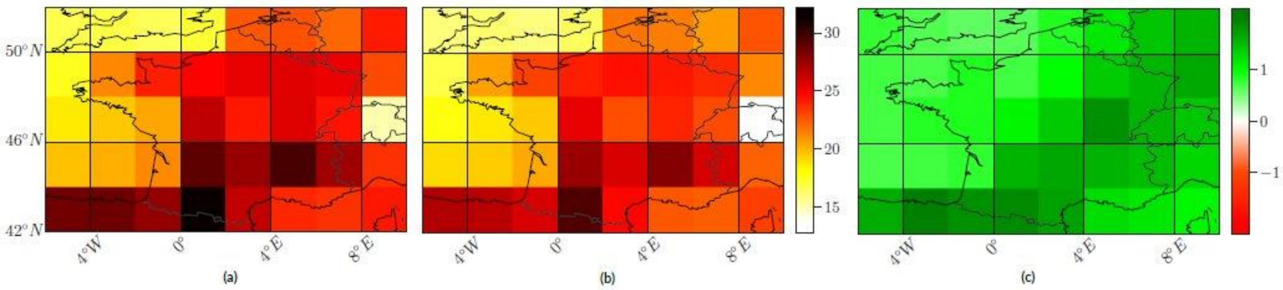

Figure 10 Comparison of TX (°C) reconstructions of the France 2003 HW obtained by analogue days that are unique of the: a) AE-AM and b) AM approaches; c) the difference between AE-AM and AM.

A pronounced difference between AE-AM and AM is the intraseasonal distribution of (disjoint) analogues. Although both methods tend to select days of the summer period (June-to-August), the AM also picks days of April, May, September and October, which contrasts with the AE-AM preference for July and August (in agreement with the timing of the observed 2003 HW). Therefore, seasonality is more marked in the latent space (which codifies the most important information of the input field) than in the original pressure field, suggesting that the AE-AM can learn seasonal (or other relevant) aspects of the target that may not be present in the predictor.

Having demonstrated the superiority of the AE-AM in reconstructing historical HWs, this hybrid approach has been employed for EE attribution, following the methodology described in Section 3.3. In an attribution mode, the AE-AM reconstructs the expected intensity of the EE in two different climates, given the observed atmospheric conditions that caused the event. As a testbed for the potential of the AE-AM algorithm to capture climate change signals, we have selected the results for two different types of EE: a HW and a HP event.

The left panel of Figure 11 shows the distribution of Iberian-mean TX averaged over the period of the 2018 HW, as inferred from flow analogues of the past (blue) and present (red) climate. The comparison reveals that similar atmospheric conditions trigger warmer conditions (~1.5°C) now than in the recent past (i.e., the observed circulation would have caused a less severe HW in the past). As the atmospheric circulation is constrained, the reported differences should be attributed to thermodynamic changes (warming trend between 1940-1980 and 1981-2022). Therefore, recent climate change made this HW at least ~1.5 °C warmer.

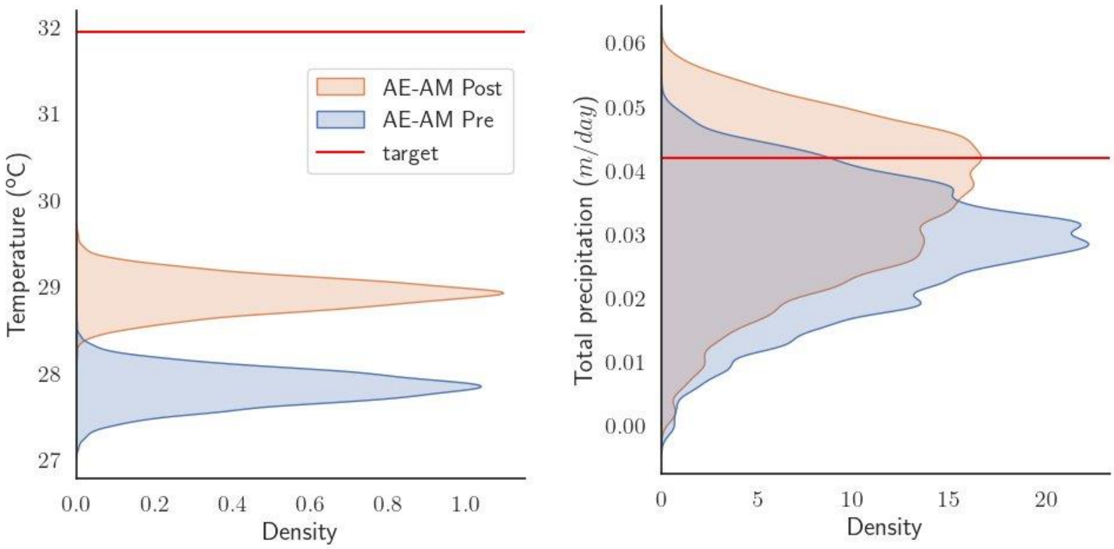

Figure 11 Flow-analogue distributions of (left) TX (°C) for Iberian HW of 1-8 August 2018; (right) precipitation anomalies (m day^{-1}) for the 2014 winter European heavy precipitation event (January-to-February).

On the other hand, the right panel of Figure 11 shows the reconstructions of daily mean precipitation anomaly averaged over northwestern Europe and derived from past and present analogues of the HP winter of 2015. The comparison reveals that the precipitation triggered by circulation would have been less severe in the past (i.e. the observed atmospheric conditions caused higher precipitation anomalies than those expected from past analogues). This is in agreement with the Clausius-Clapeyron relationship (a 7% increase in water holding capacity of the atmosphere per 1 °C global warming) and the reported tendency to total precipitation increases, particularly over wet regions of the continent (IPCC 2021). The present-past differences in precipitation are not high, though, reflecting the uncertainty in dynamical aspects of climate change. Overall the results of the 2018 Iberian HW and the 2014/2015 HP winter are in agreement with previous attribution studies based on classical methodologies (e.g. Barriopedro et al. 2020 and Yiou et al. 2017, respectively), supporting the usefulness of the AE-AM as an attribution tool.

4.1.2 Attribution with VAE

A VAE-anomaly detection method is being implemented for attribution of EEs, leveraging its potential to reproduce complex fields (Pang et al. 2021). The AI-based model has been trained with daily TX fields from natural simulations of the HadGEM3-GC31-LL GCM and applied for the reconstruction of the historical simulation of the same GCM (see Section 3.3). In what follows, the term anomaly will refer to the difference between the reconstruction and the original fields, following the standard nomenclature in AI, whereas the term departure will be employed for the differences with respect to mean climatic values. Figure 12 shows different metrics of performance corresponding to the training (natural simulations) and reconstruction (historical simulations) stages. The annual average of the mean squared error (MSE) does not often exceed ~1 °C in the train dataset (see L2 in left panel of Figure 12), whereas it can exceed 1.2 °C in the historical simulation (right panel in Figure 12), particularly after 1975 and the 21st century. For the last 20 years of the historical simulation, the metrics show a clear positive trend, which follows that of GSAT (Pearson's correlation coefficient of 0.6). This indicates that an AI-based model trained in a natural world captures well the climate conditions of the past (i.e. a world with reduced anthropogenic influences) but it cannot reconstruct the magnitude of the anomalies experienced in a recent period, when the impact of anthropogenic climate change is most evident, because these anomalies would have been virtually impossible in a natural world.

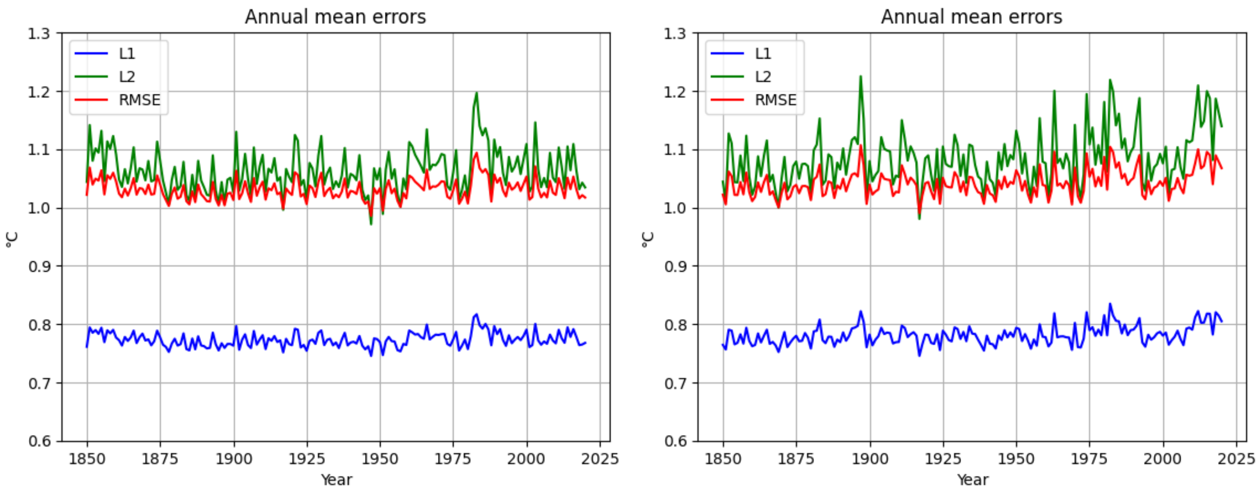

Figure 12 Time series of annual mean MAE, MSE and RMSE of natural simulations (train dataset; left) and historical simulations (target dataset; right) of the HadGEM3-GC31-LL GCM.

Figure 13 (top) shows the distribution of TX (expressed as percentiles of the empirical cumulative distribution function of the train dataset) for two groups, comprising those days and grid points of the reconstructed historical simulation with MSE values above and below a given threshold. These two groups identify TX values that would be detected as an anomaly (i.e. a climate change signal) by the VAE and those that could have occurred in the natural climate, respectively. Different thresholds are employed for the detection of anomalies, including the median, the percentiles 75th and 95th and the maximum anomaly detected in the train dataset. These thresholds are defined for each gridpoint and calendar day. Using the maximum error as a threshold (left top panel of Figure 13), most of the detected anomalies would correspond to TX extremes (of either the upper or lower tail of the distribution). As expected, the less demanding the threshold, the more anomalies fall within the interquartile range. Nevertheless, irrespectively of the threshold, the population of detected TX anomalies is consistently larger in the tails than in the rest of the distribution, and the opposite behaviour is observed for non-anomalous values. Therefore, the detection of highly unlikely TX values by the VAE is largely confined to HWs or cold spells that are unexpectedly warm.

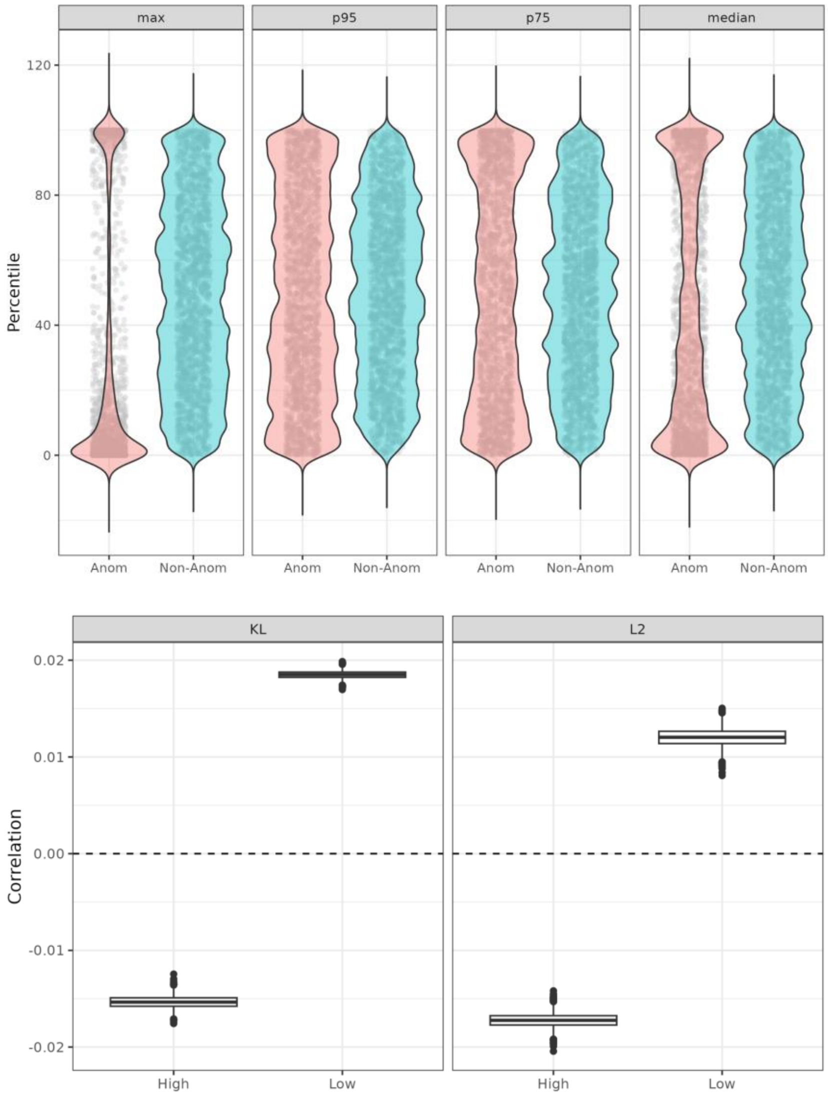

Figure 13 Top: Violin plots of simulated TX values for those cases detected as anomalies (i.e. TX values above the threshold, pink) and non-anomalies (TX values below the threshold, blue) by the VAE. TX values are expressed as percentiles of the empirical cumulative distribution function of the train dataset. VAE anomalies are detected using different thresholds (columns): maximum, 95th percentile, 75th percentile and 50th percentile of MSE. Bottom: Index of plausibility based on correlations of the spatial patterns for the 100 days with the highest and lowest anomaly, as defined by (right) MSE and (left) Kullback-Leibler divergence error.

To further assess the relationship between the VAE-detected anomaly and the plausibility of an historical TX field in the natural world, the 100 days with the highest and lowest anomalies were considered. For the classification of days two types of errors were assessed: the MSE (L2, right bottom panel in Figure 13) and the Kullback-Leibler divergence error (left bottom panel in Figure 13). The plausibility of a given daily TX pattern is quantified as the time-mean spatial correlation of the TX field of that day with that of the remaining days of the record. Then, a bootstrapping procedure with 10,000 trails was performed to estimate the mean correlation and associated uncertainty for the group of high and low anomalies. For both types of error, the mean correlation is positive for low errors and negative for high errors. This means a higher probability of detecting anomalies (i.e. high errors) in days with spatial patterns that tend to oppose those occurring in non-anomalous days.

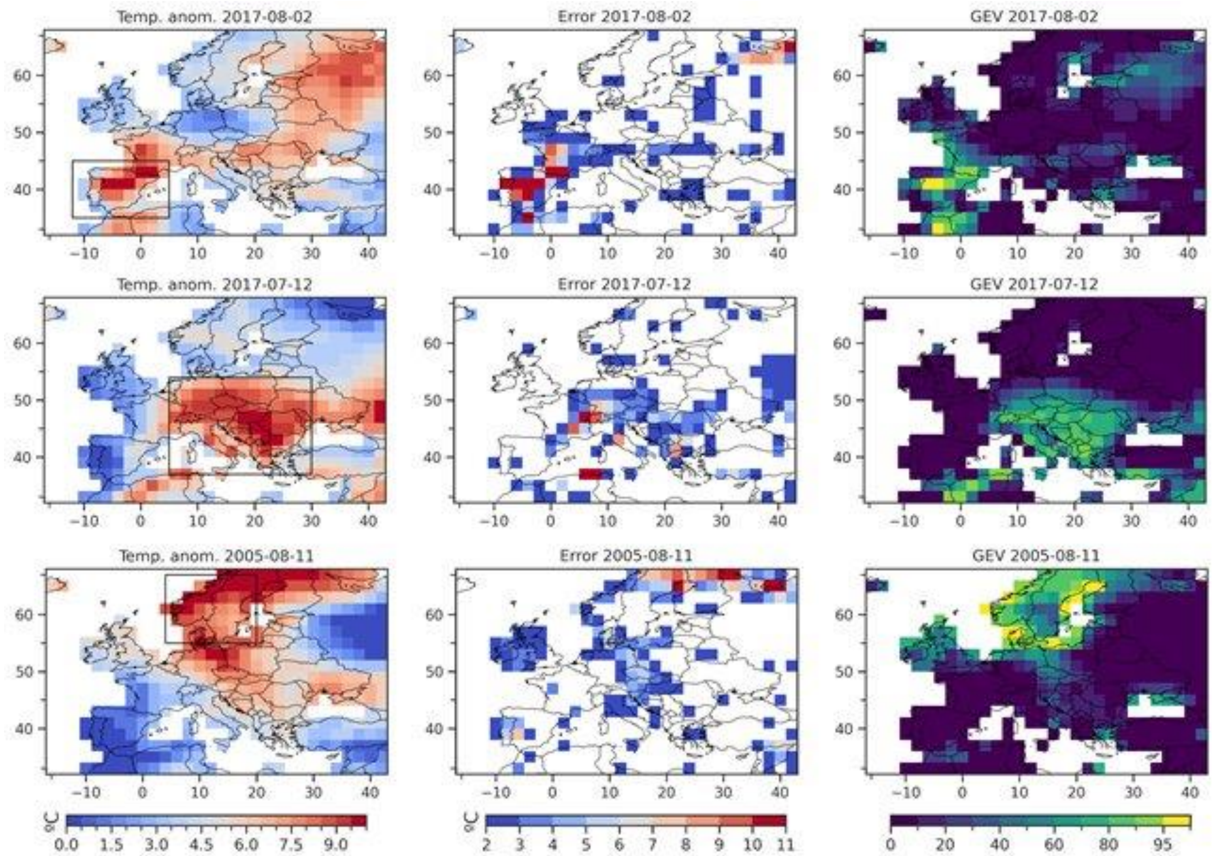

Figure 14 Daily fields for HWs simulated in the historical period of HadGEM3-GC31-LL GCM affecting (top) the Iberian Peninsula, (middle) central Europe and (bottom) Scandinavia. Columns denote the temperature departure (°C), the VAE anomaly (only grid points where the MSE values are greater than the median MSE of the training dataset are shown) and the percentile of the TX value computed with a GEV fitted to the training dataset.

The results indicate that the VAE-anomaly detection can be used to detect field structures that have never been seen in a climate without anthropogenic climate change. As an illustration, the method is applied to outstanding HWs of the historical simulation over different regions of Europe (all of them occurring towards the end of the simulated period, when anthropogenic influences are large): Iberian Peninsula, Central Europe and Scandinavia (Figure 14). In all cases, the largest TX anomalies detected by the VAE (middle panels) coincide with the regions of marked HW intensity (left panels) and high percentiles of a Generalised Extreme Value (GEV) distribution (right panels), the latter meaning that the simulated TX values would have been extremely unlikely in a pre-industrial climate. The detection of VAE anomalies can also extend to regions that are not directly affected by the HW (e.g. eastern Europe in the case of the Iberian HW). The correspondence between the GEV percentile and the VAE anomaly patterns is very good in all cases (also on daily scales). The results indicate that the VAE-anomaly detection is capable of detecting spatially resolved climate change signals on daily scales during EEs, and their temporal evolution, thus providing the basis for future development of ultrafast attribution methods of EEs.

4.2 Trends in extreme indices

To reconstruct the HadEX-CAM dataset, we have trained 12 infilling models for each extreme index: warm days and nights, and cool days and nights (Section 3.1). These models underwent evaluations on unseen data using the RMSE and spatial correlation metrics (Section 3.4). To explore the ability for generalisation, the infilling models were applied and evaluated systematically on different types of dataset: i) a multi-GCM dataset (test set); ii) the ERA5 reanalysis dataset, and iii) an observational dataset (HadEX-CAM). For the test and ERA5 datasets, artificial missing values were introduced in each input sample by applying a mask of missing values derived from the corresponding month of the HadEX-CAM dataset. In the case of the HadEX-CAM dataset, additional missing values were created by applying a mask corresponding to the month with the highest prevalence of missing values (January 1901). In all three cases, evaluation metrics were computed by comparing the reconstructed values with the original values that have been masked out.

The results of the evaluation are shown in Tables 7 and 8. The performance of the reconstruction is also compared with the results obtained using the Kriging method on the same data. Apart from the correlation for the cool nights' index of HadEX-CAM, the AI model outperforms Kriging for all datasets, extreme indices and metrics. We notice a general improvement of the metrics for the ERA5 dataset compared to the test set, which is likely due to the differences in time span. Indeed, the proportion of valid values within the ERA5 (1940-2018) period is higher than in the test set (1901-2018) period (see Figure 2), hence providing more statistical information for the reconstruction. In contrast, the results obtained for the HadEX-CAM show a deterioration of the accuracy in comparison with the test and ERA5 datasets, due to the additional reduction in the amount of valid values for the AI and Kriging reconstructions of the HadEX-CAM experiments. Despite the narrowing disparities in evaluation metrics between the two reconstruction methods, the AI models remain superior overall.

Table 7 RMSE for the reconstruction of the extreme temperature indices of the test, ERA5 and HadEX-CAM datasets.

	RMSE	Warm days	Cool days	Warm nights	Cool nights
Test set	**Kriging**	5.12	6.96	4.77	7.12
	AI	4.29	5.79	4.17	6.20
ERA5	**Kriging**	5.08	5.31	4.95	5.81
	AI	4.39	4.70	4.33	5.24
HadEX-CAM	**Kriging**	6.85	7.34	5.72	6.91
	AI	6.73	6.99	5.67	6.91

Table 8 Spearman rank correlation coefficient for the reconstruction of the extreme temperature indices of the test, ERA5 and HadEX-CAM datasets.

		Warm days	Cool days	Warm nights	Cool nights
Test set	**Kriging**	0.81	0.85	0.81	0.84
	AI	0.85	0.88	0.84	0.86
ERA5	**Kriging**	0.84	0.86	0.84	0.85
	AI	0.87	0.88	0.87	0.87
HadEX-CAM	**Kriging**	0.63	0.65	0.64	0.67
	AI	0.65	0.66	0.65	0.66

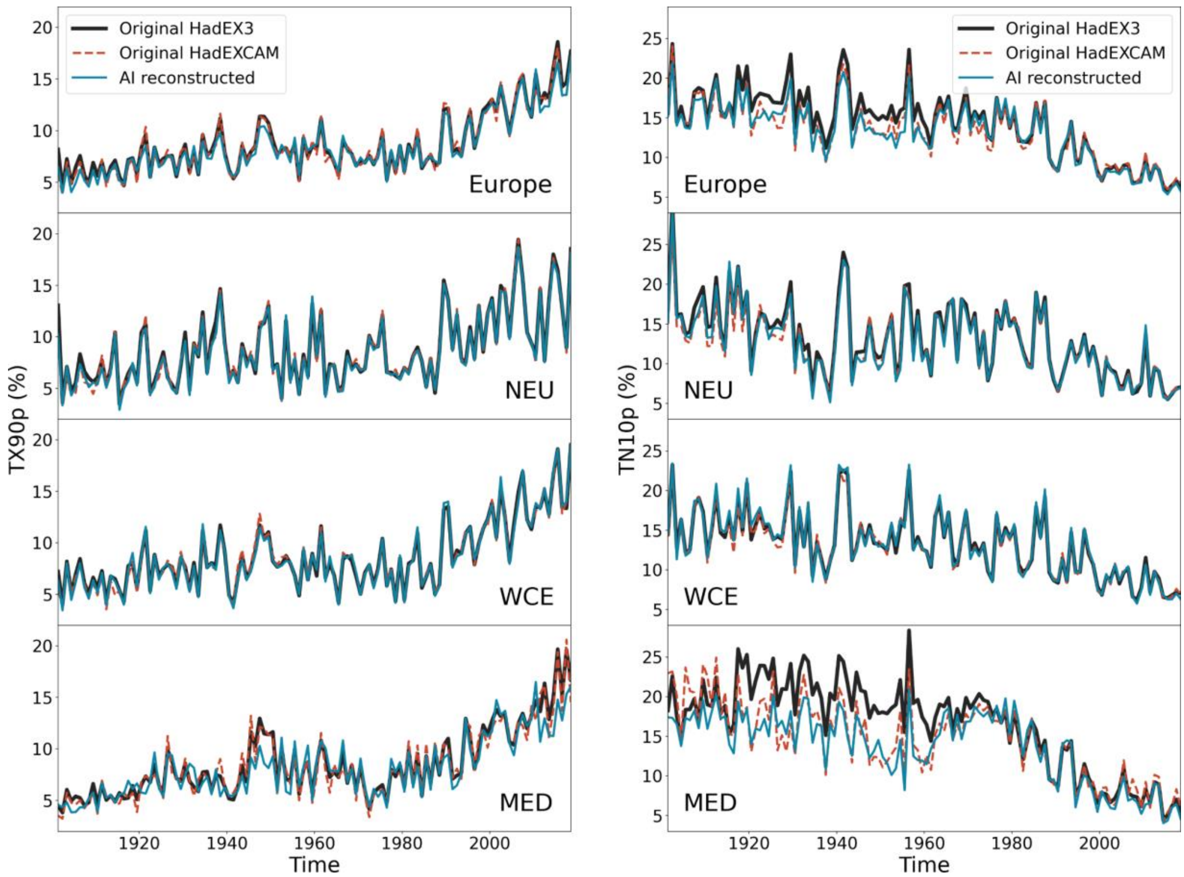

Figure 15 Regional means of warm days (left) and cool nights (right) for the original HadEX3 dataset, the original HadEX-CAM dataset and its AI reconstruction over the full grid (Europe) and three European regions defined by the IPCC report (IPCC 2021): NEU, WCE, MED. Units are percentage of warm days and cold nights in a month.

Having established the effectiveness of our AI method through successful evaluation, we can confidently proceed with the AI reconstruction of the HadEX-CAM dataset. Figure 15 shows the spatial mean of the reconstructed warm days and cool nights compared with the original HadEX3 and HadEX-CAM datasets. The three curves present a good overall agreement for the entire European region and align with the findings of the IPCC regarding the increase (decrease) in the frequency of warm days (cool nights) (Seneviratne et al. 2021). Discrepancies across datasets are slightly more pronounced for the cool nights, for which the original HadEX-CAM dataset (and its AI reconstruction) display lower values than HadEX3. These differences are particularly noticeable during the first half of the 20th century in the Mediterranean region (Figure 15), where the amount of missing values is the largest.

Due to the temporal constraints of available observational datasets, the published spatial trend analyses of ETCCDI indices for European regions primarily cover recent periods, typically the second half of the 20th century (Chervenkov et al. 2019, Squintu et al. 2021). Our AI methodology offers the opportunity to expand the temporal context of recent changes and improve the characterisation of spatial patterns and regional means compared to HadEX3. While continental mean values remain similar between HadEX3 and our AI model (Figure 15), significant disparities arise at smaller spatial scales. For instance, the long-term linear trends of warm days (Figure 16) and cool nights (Figure 17), calculated for the entire period (1901-2018) using the median of pairwise slopes estimator (Sen 1968; Theil 1992) reveal complex patterns and pronounced regional contrasts, with the AI reconstruction exhibiting higher spatial heterogeneity compared to HadEX3.

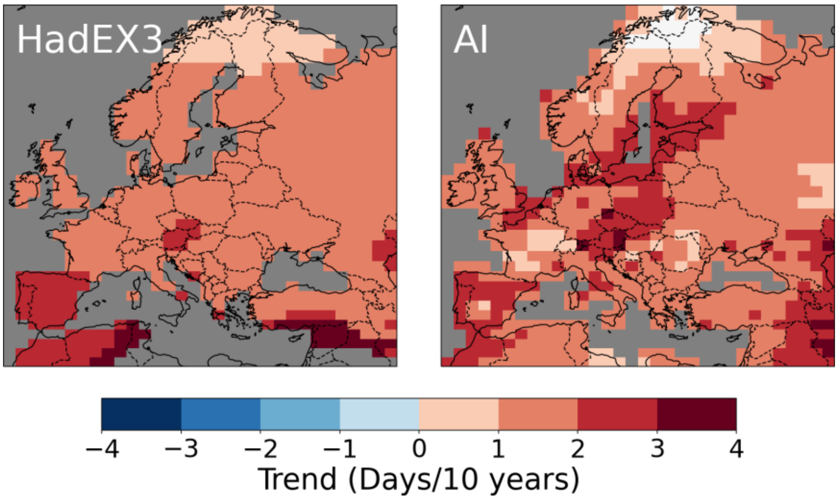

Figure 16 Linear trends (in days decade^{-1}) of warm days for the period 1901-2018. Left panel: original HadEX3 dataset (considering only grid boxes with at least 66% of valid data across the whole time period). Right panel: AI reconstruction.

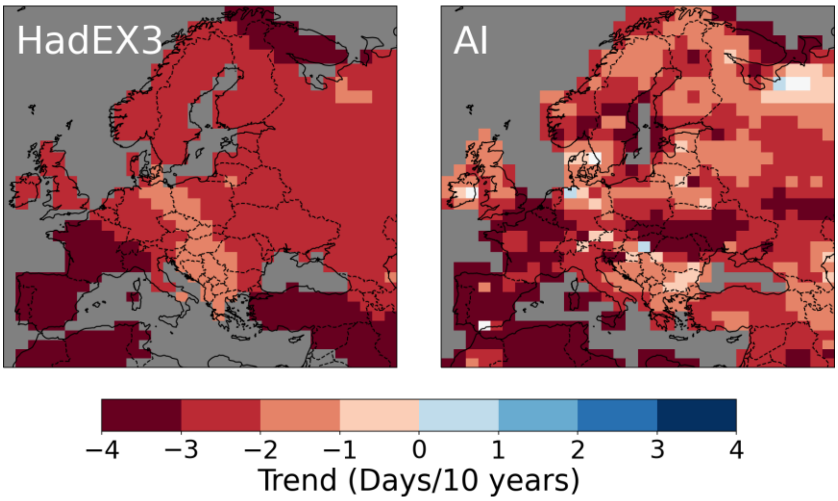

Figure 17 Same as Figure 16 for cool nights.

For instance, the AI model indicates small trends in warm days over North Africa and southern Turkey/Syria coastlines, alongside increased warm day frequencies in central Europe and the Baltic Sea. The original and reconstructed datasets suggest larger absolute trends in cool nights than in warm days, with notable decreased frequencies of cool nights in North Africa and western Europe. The AI model depicts a more varied landscape with regional variations, including significant negative trend values in cool nights over Ukraine and Romania not captured by HadEX3. Notable discrepancies are also visible in the Middle East, where the AI model suggests smaller trends compared to HadEX3. Results for a more recent period (1980-2018) also exhibit detailed spatial structures in the AI reconstruction (see Figure 18 and 19). For this recent period, it is possible to compare our results with those of ERA5. The spatial variability observed in the ERA5 trends closely resembles that of the AI model. The similarities between the AI and ERA5 trends are particularly evident in certain regions with pronounced changes, such as the Black Sea coasts.

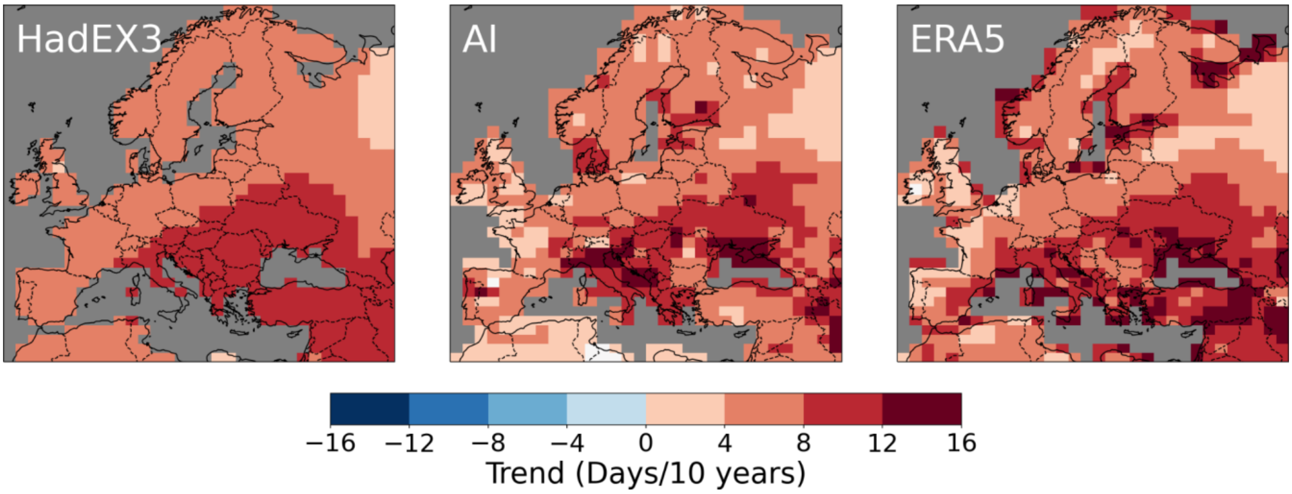

Figure 18 Linear trends (in days decade^{-1}) of warm days for the period 1980-2018. Left panel: original HadEX3 dataset (considering only grid boxes with at least 66% of valid data across the whole time period). Central panel: AI reconstruction. Right panel: ERA5 dataset.

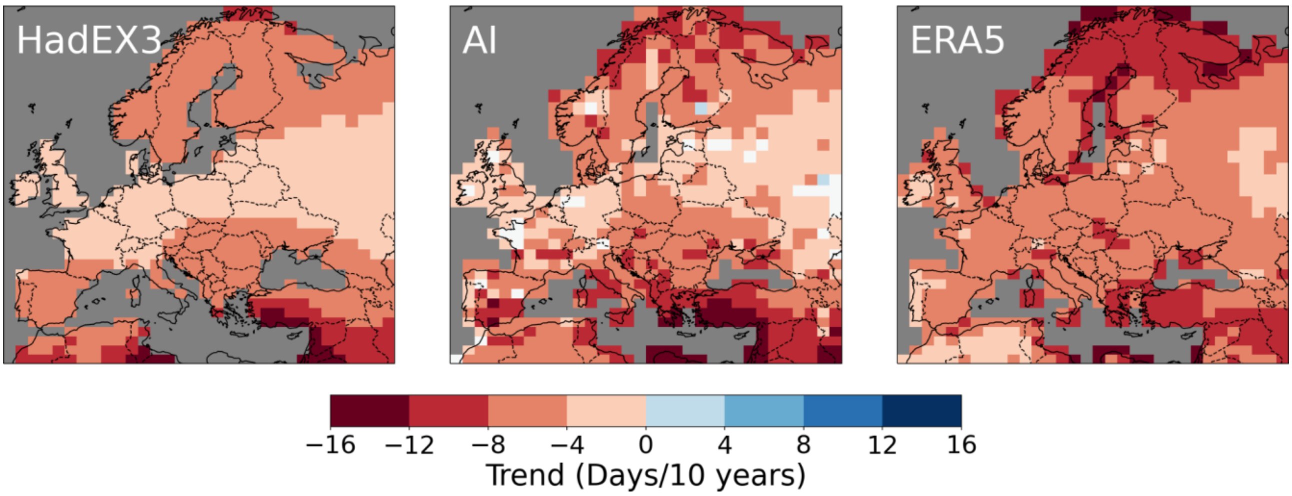

Figure 19 Same as Figure 18 for cool nights.

4.3 Storylines

4.3.1 Storylines of EE

This section illustrates the application of ML algorithms for the construction of storylines of EEs, using HWs in the Po Valley as a pilot. Figure 20 highlights the regions considered for the calculation of HW-related indices (TX90P, HWMI; Section 3.1). The drivers of HWs were obtained by considering a scrutinised list of candidate variables at daily resolution (see Table 6), and applying a two-step driver-detection procedure (see Section 3.5). First, for each variable, geographical areas with similar variability were determined by applying k-means clustering to the daily series of anomalies of ERA5 (defined with respect to 1981-2010). These clusters are calculated over different domains (Europe, World, North Atlantic, Arctic, etc.), depending on the spatio-temporal characteristic of the variables (Table 6). Some examples of the resulting clusters are shown in Figure 21.

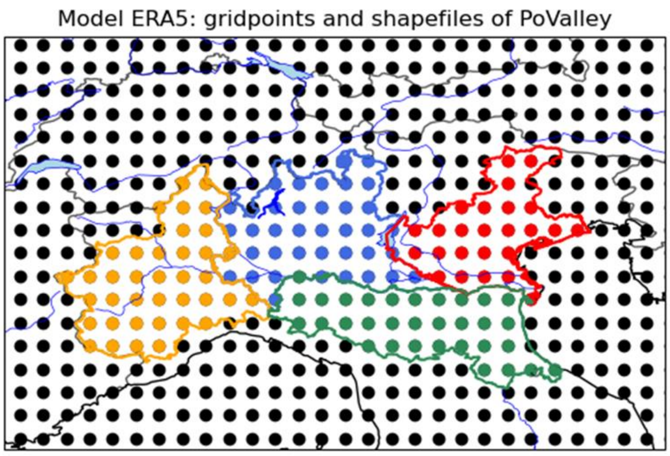

Figure 20 Administrative regions of Piedmont (yellow), Lombardy (blue), Veneto (red) and Emilia-Romagna (green) and the corresponding grid-points of ERA5.

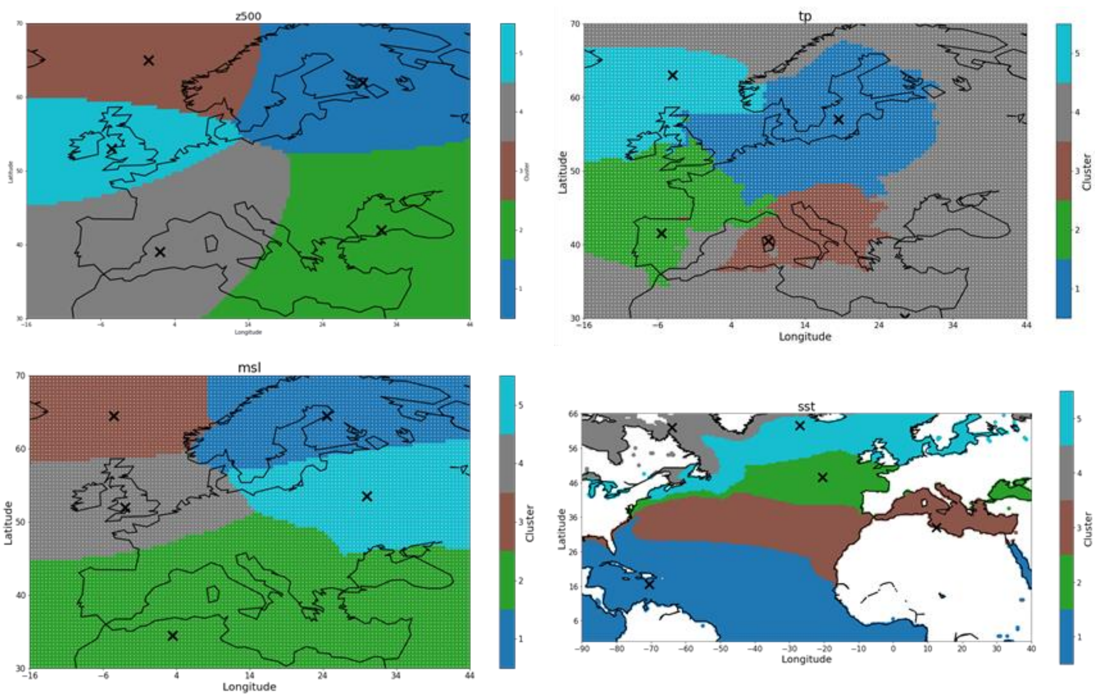

Figure 21 Examples of geographical clusters obtained for some of the variables: Z500, total precipitation, SLP and SST.

This process is repeated for each GCM simulation, using the clusters calculated on ERA5, and considering the CWS14.2 period as baseline for the computation of anomalies. For each cluster three daily time series are calculated from geographical summary statistics (spatial average, and the 25th and 75th percentiles) of the local variables over the grid points embedded in that cluster. In a second step, the set of daily series anomalies related to all the clusters are used as candidate drivers for an optimised feature selection. The selection of the drivers is performed with the CRO algorithm (Section 3.5), using summer HW occurrence as the target series and the aforementioned cluster-based daily series of candidate drivers as predictors.

As a first step, the CRO algorithm is applied to the current climate of ERA5 (1981-2020 period), considering the mean statistics of each cluster as candidate predictors, with the aim of determining benchmark drivers. The algorithm is run 20 times with a maximum of 20,000 evaluations each. The 10% solutions with the highest cross validation scores are selected and analysed. Figure 22 shows the selection rate among the top 10% solutions, indicating how many times each candidate driver is selected at different lags. The candidates with a selection rate of at least 0.67 at any lag are identified as the benchmark drivers of HWs in the Po Valley. They are listed in Table 9. These findings are consistent with the work on EE detection performed within WP3 (see D3.2) and the use of this information for the developing of hybrid or data-driven sub-seasonal and seasonal forecasts is currently under inspection.

Table 9 Clusters selected during the application of CRO to ERA5 (1981-2010).

Feature selection code	Short Name	Variable	Cluster domain	Cluster number	Covered area	Selection rate
tasmax_Europe_cllow01	TASMAX-CEU	TASMAX	Europe	2	Central Europe	84.7%
psl_Europe_cllow02	PSL-NS	PSL	Europe	3	Norwegian Sea	96.0%
zg_World_cllow04	ZG-GL	ZG	World	5	Greenland	82.6%
pr_Europe_cllow02	PR-IWB	PR	Europe	3	Italy & W Balkans	98.7%
mrsos_Europe_cllow00	MRSOS-CEU	MRSOS	Europe	1	Central Europe	96.4%
mrsos_Europe_cllow02	MRSOS-IBP	MRSOS	Europe	3	Iberian Peninsula	80.2%
pr_Europe_cllow03	PR-NAS	PR	Europe	4	N. Africa & Scandinavia	76.1%
tos_North_Atlantic_cllow00	TOS-Trop	TOS	N Atlantic	1	[0 – 30]°N	89.0%
zg_Europe_cllow03	ZG-SWEU	ZG	Europe	4	SW Europe	99.7%

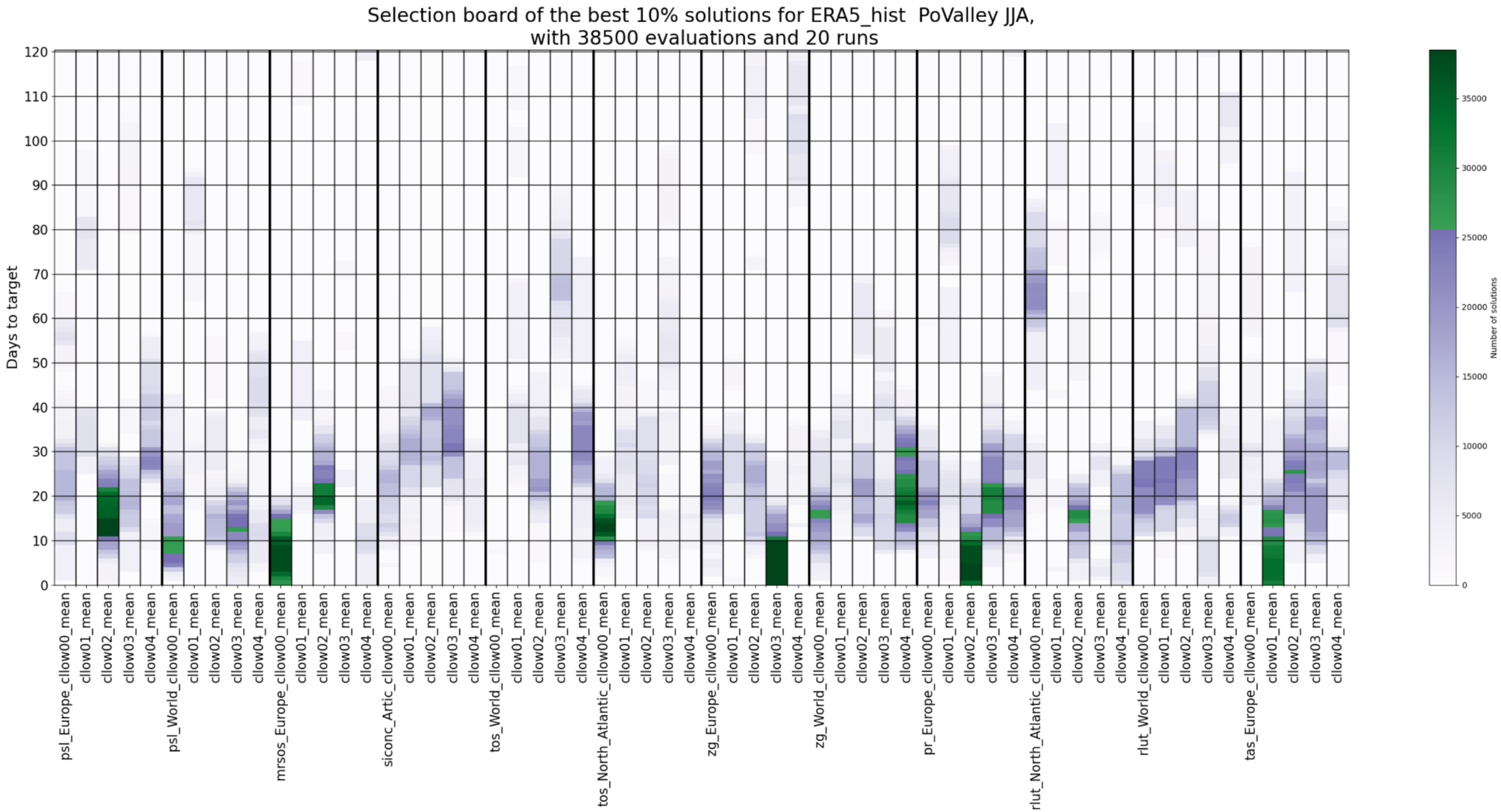

Figure 22 Selection rate related to the ERA5 clusters (columns, only spatial average statistics are presented) for different lags (rows). Purple colour denotes selection rates below 0.67, green stands for values above the threshold, indicating the selected benchmark drivers.

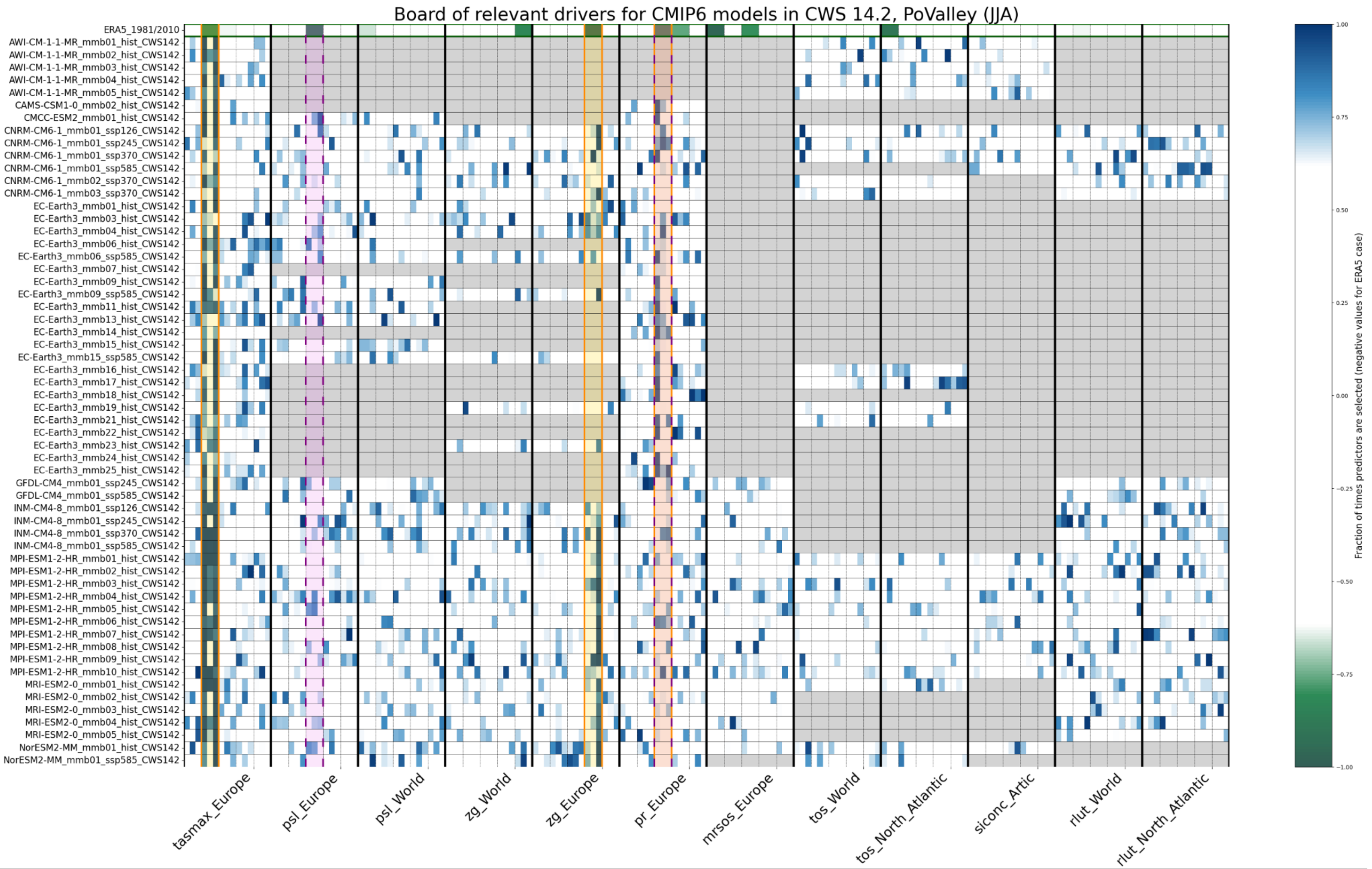

Figure 23 Selection rates for the same drivers as in Figure 22 (rows) and the three statistics (squares in each cell). Columns indicate the analysed simulations. Green is used for ERA5 benchmark drivers and blue for the selection rates of the simulations. Orange vertical bands indicate the on-site benchmark drivers. Purple vertical bands highlight the two clusters used for the construction of storylines.

Once the benchmark drivers are found, a validation of GCMs is performed by applying the same procedure (CRO model) to their simulations and identifying those that selected the benchmark drivers. Such a process generates the list of simulations capable of reconstructing the observed relationship between the driver and the EE previously found in reanalysis data. For the evaluation process, all cluster-related statistics (including the mean and the 25th and 75th percentiles of the local predictors) are included. This allows to empower the search of drivers in conditions where relationships among the variables simulated by the GCMs could be statistically similar, but not identical, to those of the reanalysis. The board in Figure 23 shows the results of this feature selection for the CWS14.2 period of each simulation (rows). For a given driver (columns), each cell is composed of three squares, each standing for one of the statistics. Blue shading indicates the maximum selection score among all the lags inspected by the CRO. It can be seen that feature selection gives some noisy results on simulations. Nevertheless, the benchmark drivers are often selected in the simulations, especially the on-site drivers (orange highlighted columns). The three statistics can perform differently from each other, in part because they measure different statistical properties but also because the algorithm tends to neglect redundant drivers. Finally, the validation of the GCM simulations is performed with a driver-based approach. For each driver, a simulation is validated if a score larger than 0.67 is reached by at least one of the considered statistics.

The storyline approach is based on the inspection of pairs of drivers whose interaction may affect HW occurrence. As a consequence, all the possible pairs of benchmark drivers are tested. Figure 24 is a heatmap illustrating the number of simulations that select each pair. Different combinations of drivers can provide a satisfactory description of HW evolution. However, some of them are more frequent than others. As expected, combinations including information from local conditions (i.e. clusters of TX, Z500 and total precipitation over the Po Valley region, orange shaded columns) are selected by the majority of the GCM simulations. These drivers are widely acknowledged, since high Z500 and low precipitation are often associated with HW conditions (e.g. Barriopedro et al. 2023 and references therein).

The storyline approach also allows one to explore unreported drivers of summer HWs in the Po Valley by focusing on remote variables affecting on-site conditions, such as SLP over the Norwegian Sea (PSL-NS) or northern tropical SST anomalies (TOS-Trop). In particular, 13 GCM simulations selected both PSL-NS and precipitation in central-east southern Europe (PR-IWB) as a driver combination affecting summer HWs in the Po Valley. This is one of the highest counts among those pairs that include at least one remote driver, as shown in Figure 24.

Once the pair of drivers has been chosen, the next step involves the selection of GCM projections that will be considered for the construction of the storylines of future changes of EEs. This process consists of three phases:

a) preliminary filter of projections: the selection rates related to the three statistics of each driver are checked; if at least one of them is larger than 0.5, the projection is selected;

b) statistics selection: considering the projections that passed (a), the mean selection rate of each one of the three statistics is calculated, allowing to assess the best-performing statistics on average;

c) final selection of projections: once the best-performing statistics is chosen, the projections with a selection rate of at least 0.33 in that statistics are selected, discarding those simulations that fail this criterion.

This procedure provides a good compromise between finding statistics with good scores and identifying a reasonable number of GCM projections. In the design of the process, the following considerations were taken into account: i) mean, 25th percentile and 75th percentile statistics are correlated with each other; ii) a strict threshold on one single statistics (without considering the scores of the others) could cause an excessive reduction in the number of selected simulations, compromising the statistical robustness of the storylines. Finally, it is important to stress that this process is specifically related to each pair of drivers. A different pair will result in different selected GCM projections, according to the simulated features of each model, member and scenario.

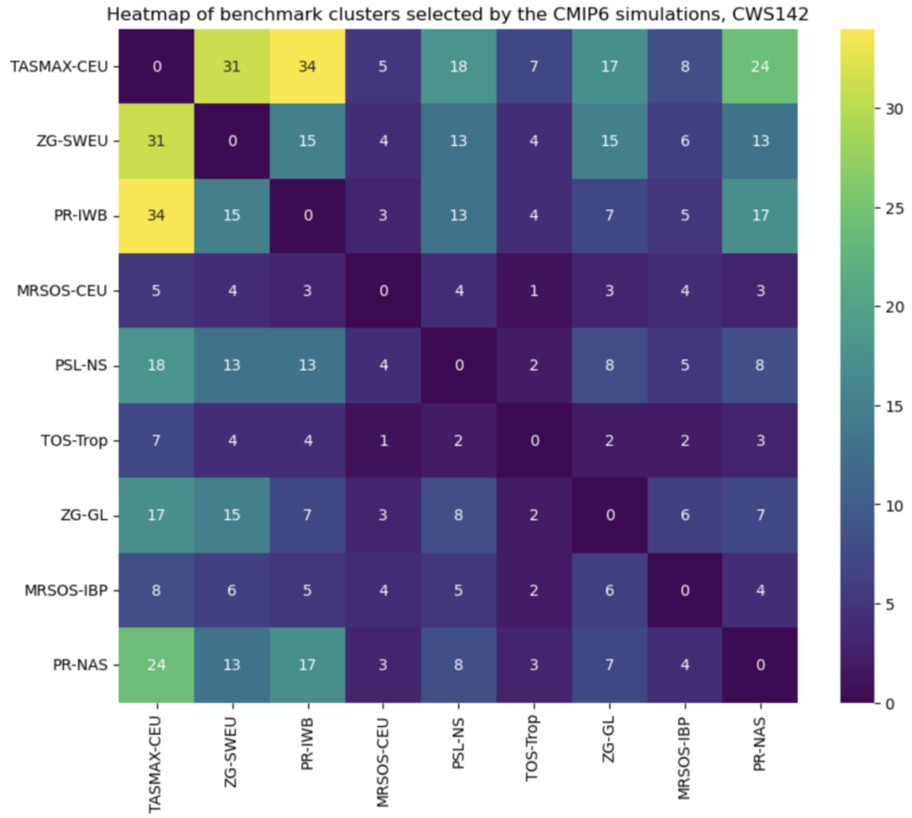

Figure 24 Number of simulations validating each pair of ERA5 benchmark drivers.

As described above, the storylines constructed here focus on the interaction between two drivers and the EE of interest. As an illustration, we have considered the influence of PR-IWB and PSL-NS on the evolution of summer HWs in the Po Valley. The list of validated GCM simulations for this pair of drivers can be found in the legend of Figure 25. For each simulation, changes in the drivers are computed from the daily series as the difference of 30-yr averages between CWS15.0 (future climate) and CWS14.2 (current climate).

The resulting changes in the drivers are displayed in the scatterplot of Figure 25. There, the x-axis represents the climate change responses in PR-IWB and the y-axis summarises the simulated changes of PSL-NS in the multi-model ensemble. The quadrants delimited by the two multi-model means (vertical and horizontal lines) determine four groups of GCM simulations, each one defining a storyline. These are named low-low (LL), low-high (LH), high-low (HL), and high-high (HH), with LH denoting low changes in PR-IWB and high changes in PSL-NS. To avoid including simulations whose drivers' changes are too far from (or too close to) the multi-model mean, a bivariate gaussian regression (without rotation of the axes, i.e. no diagonalization of the correlation matrix) is applied. The resulting distribution was used to determine thresholds containing the 80% and 5% of the GCM simulations (outer and inner ellipses in Figure 25). The simulations with drivers' responses laying outside (inside) of the outer (inner) ellipse are discarded.

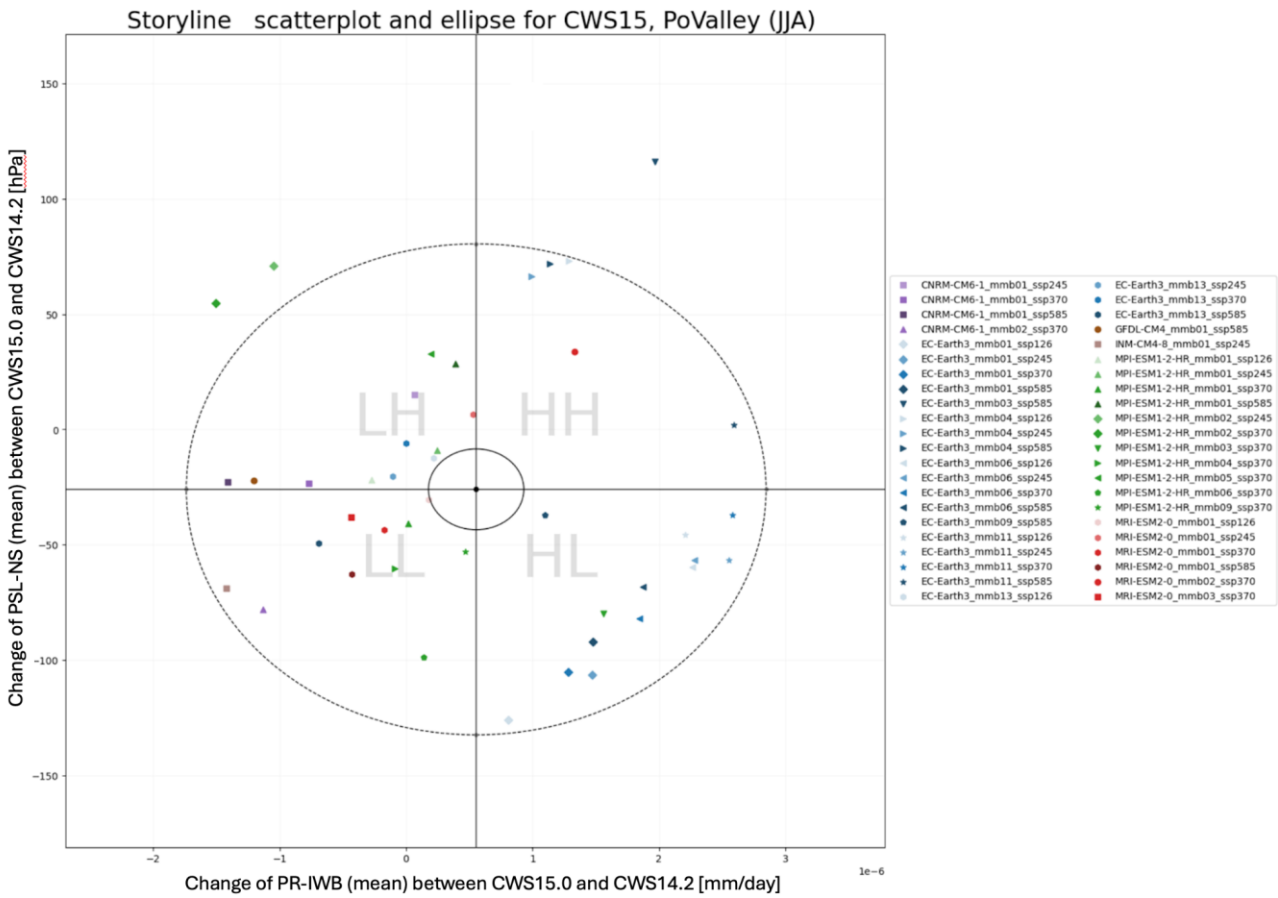

Figure 25 Scatter plot of the change of the drivers (x axis: PR-IWB, y axis: PSL-NS) between CWS15 and CWS14.2. Each dot represents a validated simulation for this pair of drivers. A bivariate Gaussian regression is applied to the spatial distribution and is here displayed. 80% and 5% confidence levels of this regression are represented by the ellipses.

In each storyline, the yearly indices related to HWs (Section 3.1) are averaged for each year of the CWS15.0 period, using the selected GCM simulations. Figure 26 shows the results for TX90P, which gives an indication of the distribution of TX in each summer. For a global warming of 0.8 °C (difference of GSAT between the two CWS), the LL storyline projects a mean increase of 1.72 °C in TX90P. This storyline is related to changes in PR-IWB and PSL-NS that lay below those projected by the multi-model mean. While the first driver describes the well-known regional feedbacks between precipitation deficits and HWs, changes in the second driver can be interpreted as a deepening of low pressure systems over the area enclosed by Scotland, Iceland and Norway.

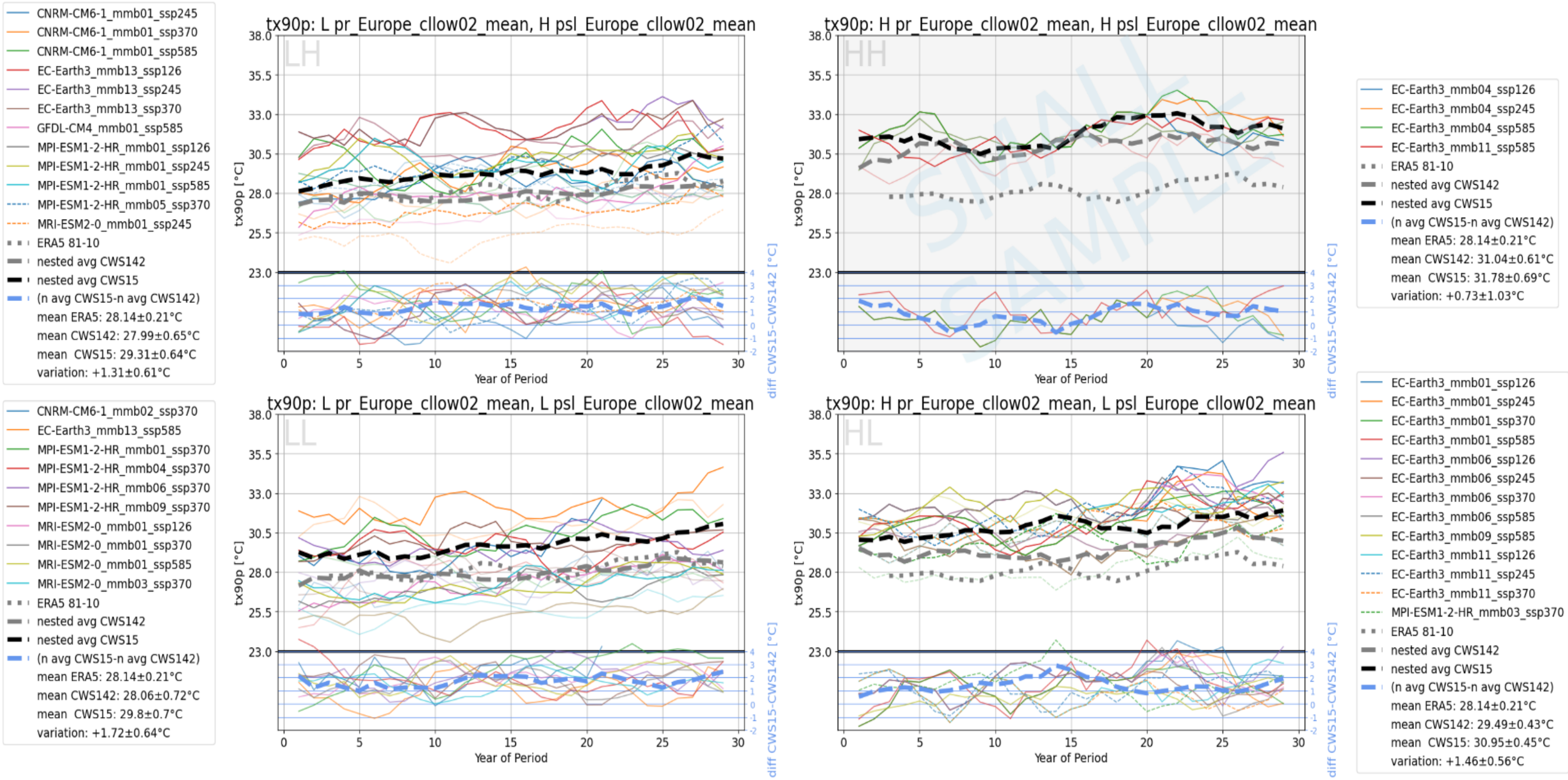

Figure 26 Series of annual TX90p (with a 4-tr running mean) for each selected simulation (lines) and storyline (panels). Lines are full colour (CWS15) or shaded (CWS14.2). Multi-model means are displayed with thick dashed lines (black: CWS15, grey: CWS14.2). Thick grey dotted lines represent ERA5. Differences between CWS15 and CWS14.2 are shown in the lower sectors of each graph. Averages and uncertainties are provided in the legends

To conclude, we stress the agreement between the indices calculated on ERA5 and those obtained on the CWS14.2 of simulations, which indicates a correct relationship between large-scale GSAT and the representation of local temperature patterns. This agreement is also observed in almost all storylines, especially those with low changes in PR-IWB, suggesting that the overestimation of precipitation might be connected with biases in temperature.

4.3.2 Storylines of concurrent events

This section focuses on concurrent EEs (hot-dry days) over Central Europe, as defined in Figure 3 (Section 3.1). The potential drivers of concurrent EEs have been thoroughly examined in D4.1 (EE causation analysis). Here the focus is on Z500 and SSTs due to their relatively smooth nature (Section 3.6). The identification of the drivers is done using the 1981-2010 reference period of ERA5 reanalysis. KRGCCA is employed to derive the CEEI. In order to capture the non-linear interactions of the local variables, we use non-linear kernels namely the first order arc-cosine kernel. For the selected drivers (SST and Z500) a linear kernel was found to be sufficient.

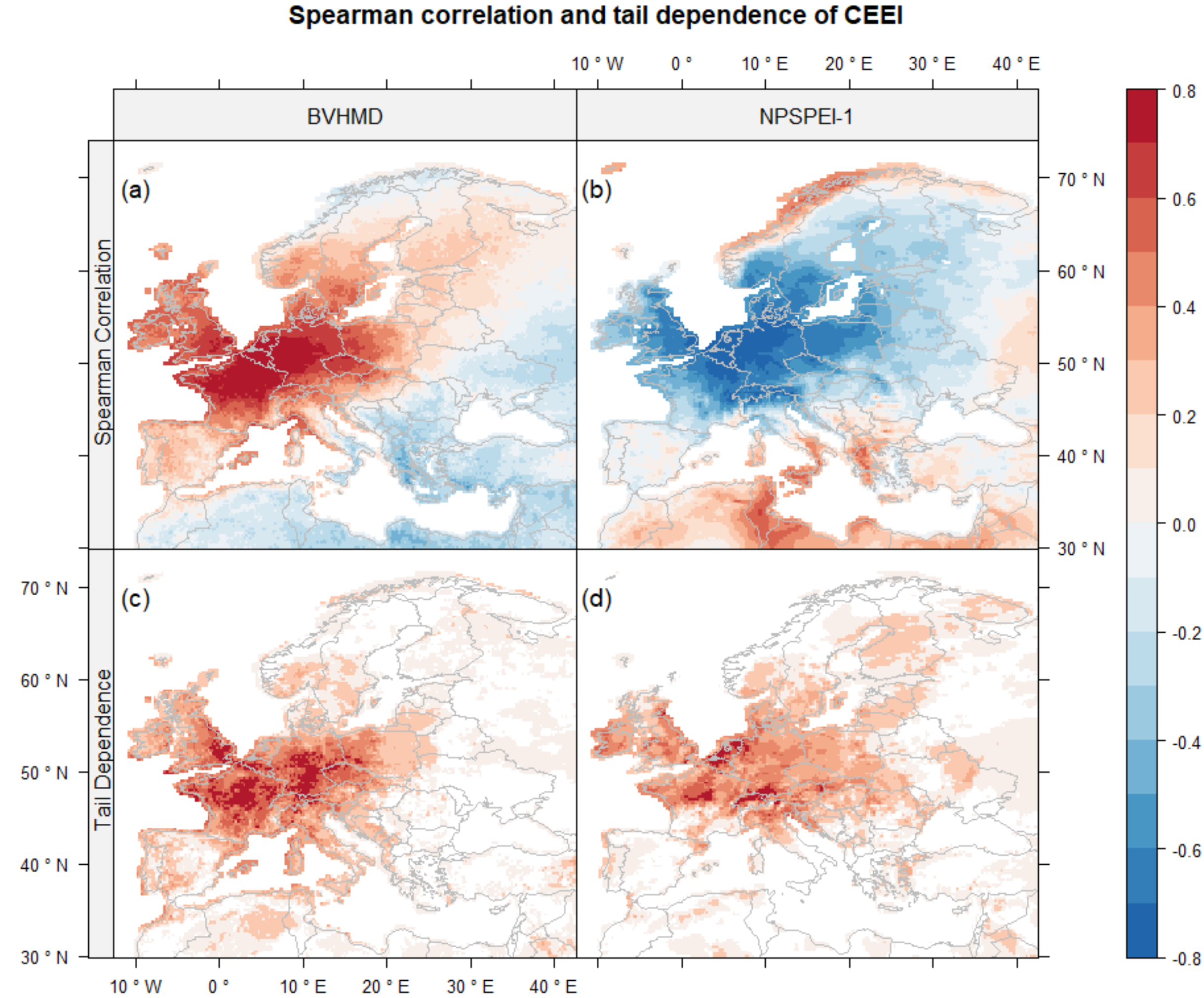

Figure 27 Summary statistics of the univariate CEEI at each grid point. Spearman correlation of CEEI with: (a) BVHMD; (b) NPSPEI-1. Coefficient of tail dependence (using the 90th percentile) for: (c) BVHMD; (d) NPSPEI-1.

The suitability of the CEEI is validated at each grid point by calculating the local Spearman correlation coefficient of this index with both the NPSPEI and BVHMD, as well as the tail dependence (e.g. Coles 2004) (Figure 27). The correlations and tail dependencies are notably high for the central region of Europe, indicating a strong correspondence between high values of the CEEI and hot-dry conditions.

Furthermore, the explained variances of NPSPEI, CEEI, and BVHMD are 72.3%, 68.5%, and 74.9%, respectively (not shown). This indicates that CEEI can account for ~70% of the variability of the HW and DR indices over the study region. In conjunction with Figure 27, this suggests that the non-linear KRGCCA successfully captures meaningful non-linear influences of NPSPEI and BVHMD on CEEI, demonstrating its suitability for describing large-scale concurrent HWs and DRs.

Moving on to the dominant drivers of CEEI, we use the QUINN framework (Section 3.6) to estimate the CDF of CEEI, incorporating SST and Z500 as input variables with lags up to three months. As a prior, we use the Automatic Relevance Determination (MacKay 1992), which is able to learn the relevance of each driver. The model's quality can be evaluated by examining the values of the CDF, which should follow a standard uniform distribution. The model shows a satisfactory performance, as illustrated in Figure 28.

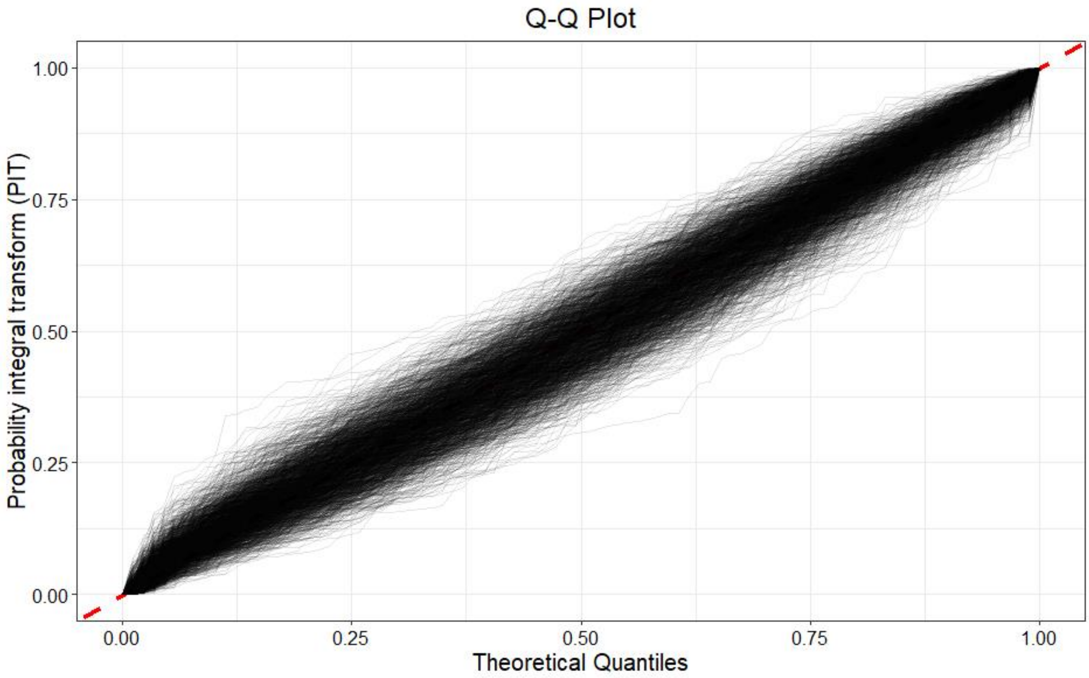

Figure 28 Q-Q-Plot for all ensemble members of the derived QUINN model.

Next, we address the most important variables for learning the CDF of CEEI by QUINN. Notably, high values of the CEEI coincide with concurrent EEs, prompting us to focus on the (conditional) 90th percentile. Using ALE plots as a variable importance criterion, it is revealed that Z500 in June-to-August and April-June SSTs are the two most influential variables (Figure 29). The associated patterns indicate an omega-like atmospheric blocking pattern over central Europe (Figure 30). This finding is physically consistent, since blockings are typically linked with subsidence, resulting in clear skies, increased solar radiation, and reduced precipitation, which leads to warm and dry conditions

(Sousa et al. 2018). These mechanisms can be further influenced by soil drying and land-atmosphere feedbacks (Miralles et al. 2019). On the other hand, the SST-related patterns reveal a dipole in the western part of Europe, with negative anomalies in the north and warm anomalies towards western and southwestern Europe.

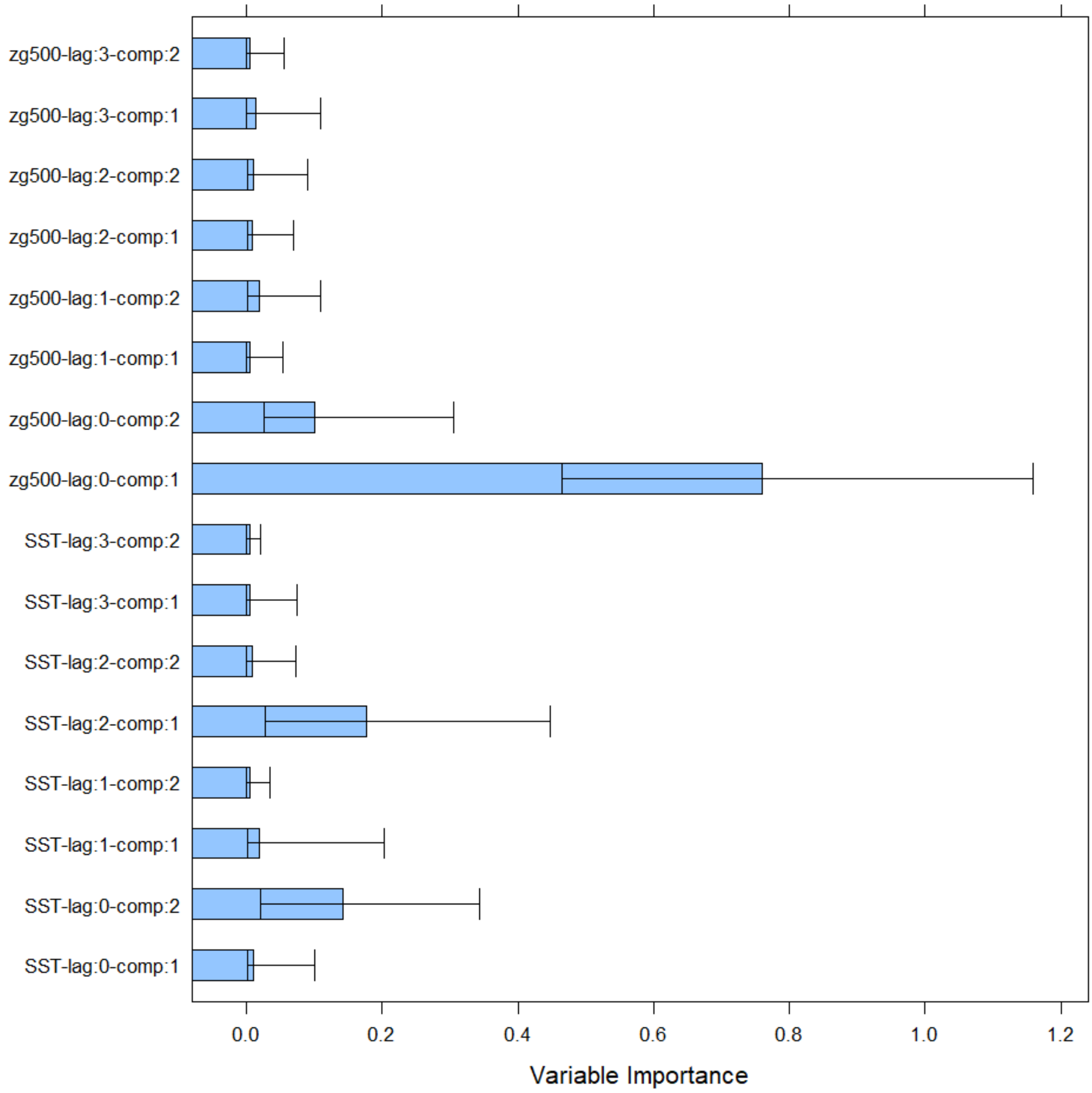

Figure 29 Variable importance for the input variables of QUINN. Bar charts reflect the median of the ensemble of the variable importance criterion and whiskers correspond to the 95% confidence intervals.

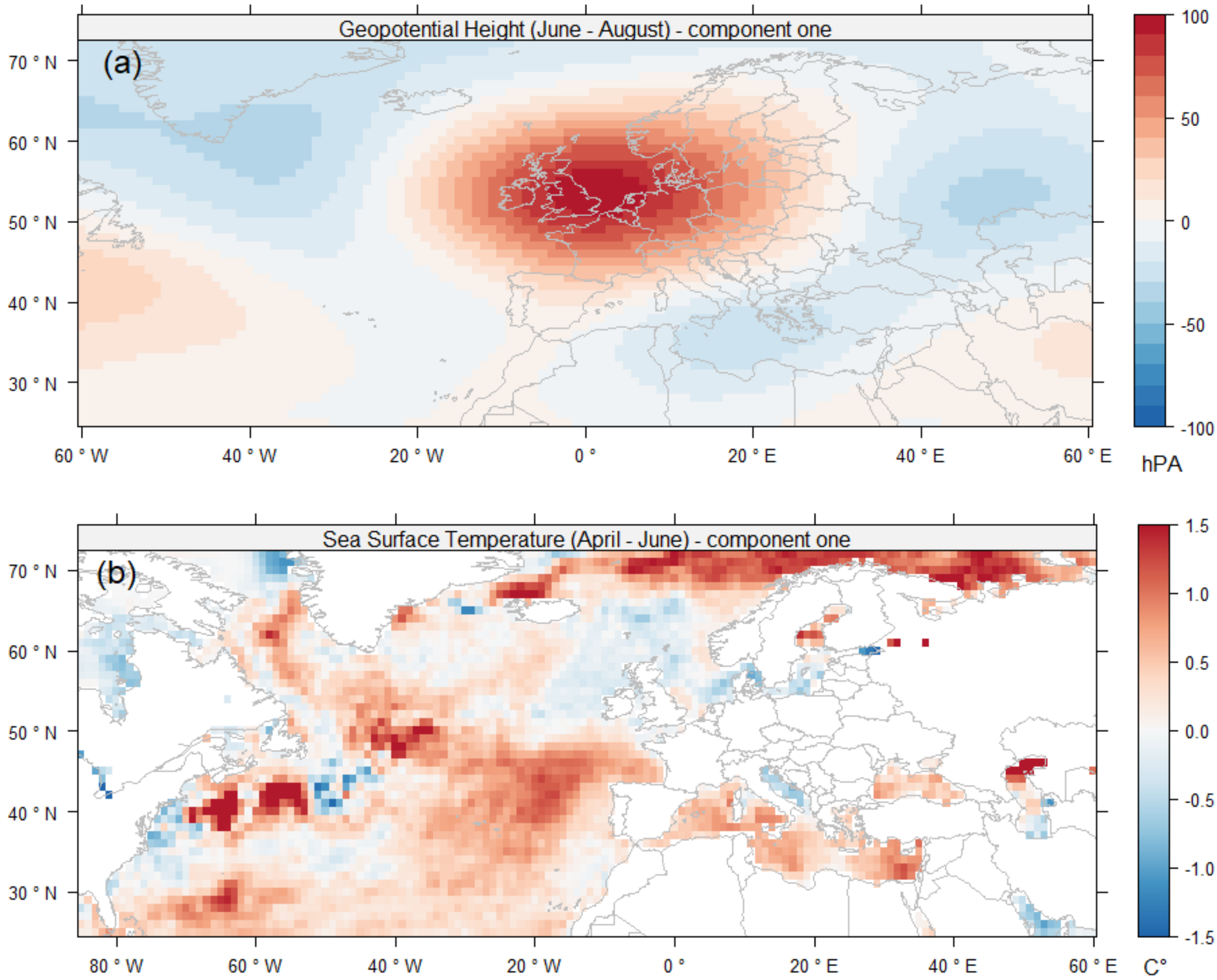

Figure 30 Extracted spatial patterns from the KRGCCA analysis for: (a) June Z500 (m); (b) April-June SST (°C) anomalies corresponding to a time lag of two months. Anomalies are computed with respect to the 1981-2010 reference period.

Finally, the marginal effects of these two variables are inspected through ALE plots. The results are shown in Figure 31. For both drivers, ALEs tend to shift towards positive values (thus increasing values of CEEI and magnitude of concurrent EEs), suggesting a monotonic relationship. This effect dampens for Z500 after reaching a certain level, with a similar but weaker effect for SST. The greatest ALE, and the highest likelihood of observing hot and dry conditions, occurs when both variables are in a positive state. If the SST pattern is opposite to that of Figure 30 the ALE still remains positive, indicating that warm and dry conditions still persist in the presence of the omega blocking. On the other hand, when SSTs are in a strong positive phase but Z500 anomalies are negative (thus corresponding to a trough), the chances for the occurrence of concurrent events are lower, likely due to rainier conditions, which could be reinforced by the positive anomalies of SST. Therefore, the omega blocking dominates the joint effect of the variables on CEEI, with the SSTs further accelerating the effect of the atmospheric circulation.

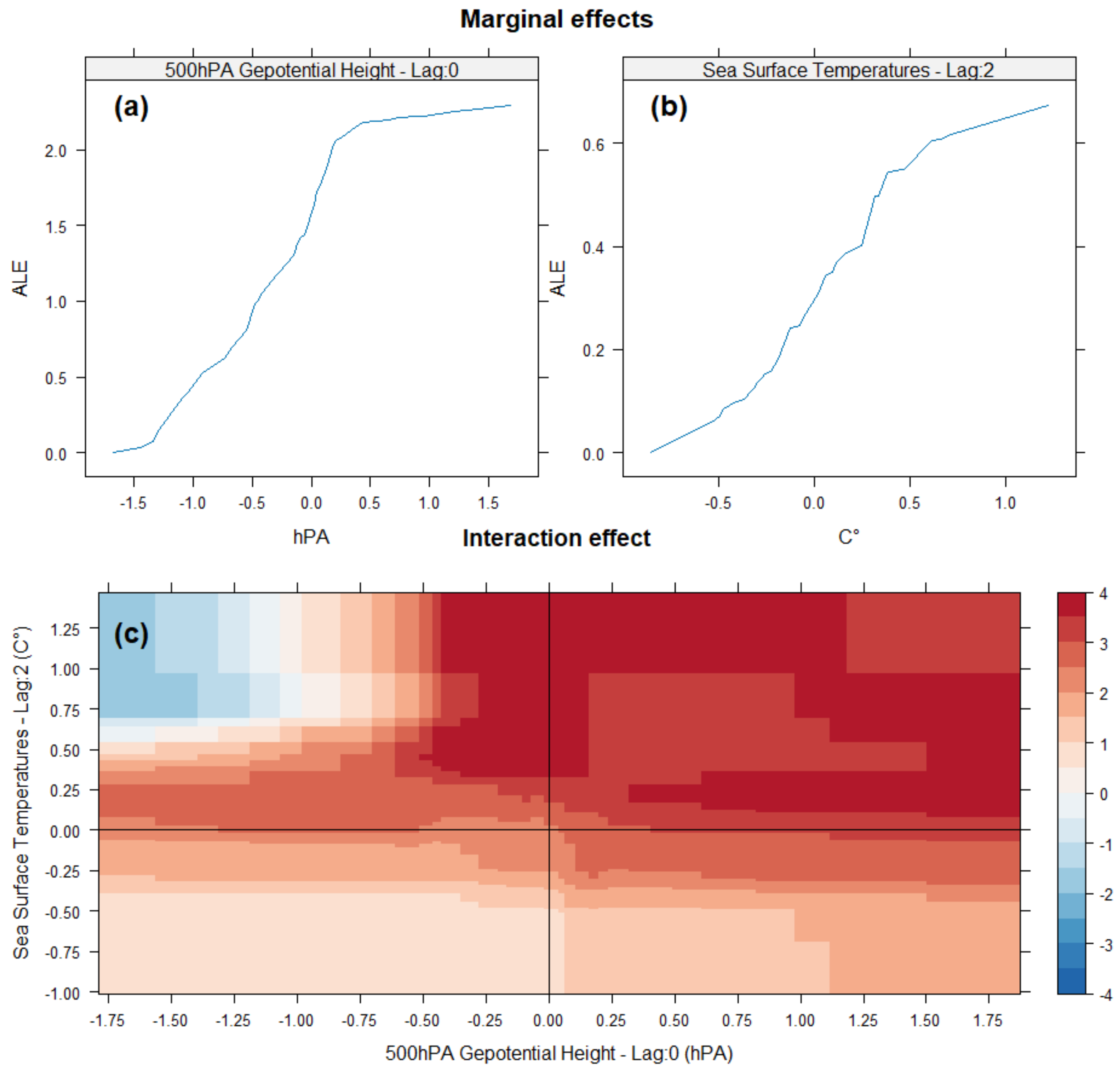

Figure 31 Estimated marginal effects of the QUINN model based on the ensemble mean for the most important input features on the CEEI. (a) and (b) correspond to the individual effects of June-August Z500 and April-June SSTs, respectively. c) displays the joint effect or interaction of the two variables.

The evaluation of CMIP6 GCMs is carried out by applying the KRGCCA to construct the CEEI in the CWS14.2 reference period of each simulation. We assess the similarity of the spatial patterns extracted from KRGCCA with those obtained from ERA5 (Figure 30) by using the Perkins-Score. Figure 32 illustrates the Z500 patterns for MPI-ESM1-2-HR and their corresponding scores. As a criterion, a given simulation is retained if the mean Perkins-Score (Perkins et al. 2007) for the SST and Z500 is above 0.75, and both scores are above 0.6. The simulations meeting the specified criteria are selected to form the storylines.

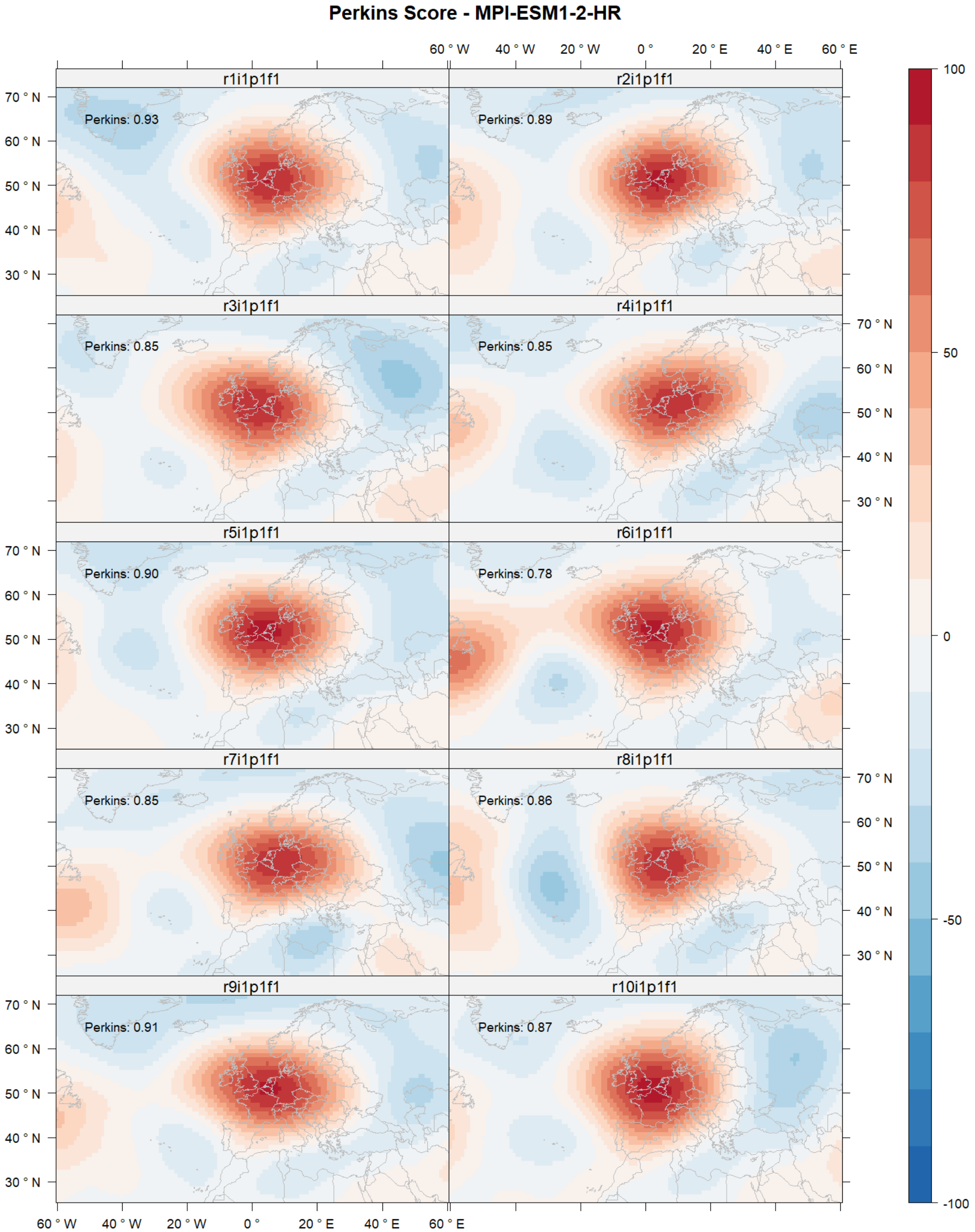

Figure 32 Spatial patterns of June-August Z500 anomalies extracted from the MPI-ESM1-2-HR GCM. Anomalies are defined with respect to CWS14.2.

To develop the storylines, we consider the CWS15 period of each validated simulation and construct scatter plots similar to those of Figure 25. To display these plots, simulated data are projected onto the extracted spatial patterns of Figure 30 in order to obtain their associated time series. Then, the 30-yr mean values for the CWS14.2 and CWS15 periods of each simulation are obtained (note that the mean of the considered drivers over CWS14.2 is zero). Figure 33 shows the scatter plots of the drivers for the CWS15 period, with positive values of the drivers denoting a higher frequency of positive phases than in current climate conditions. Most GCMs simulate a positive shift of the Z500 pattern, which was also the most important driver of CEEI in present-day climate, as well as a shift towards positive phases of the SST patterns. Together, with the observed marginal effects (Figure 31), the results suggest disproportionate increases towards more frequent hot-dry EEs.

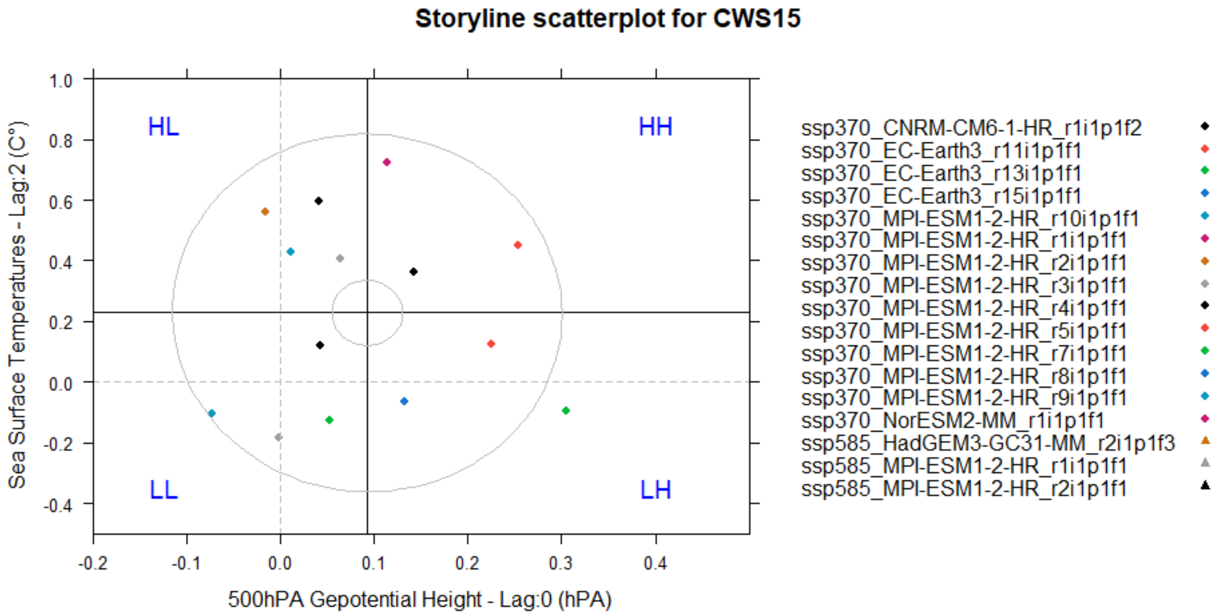

Figure 33 Scatter plots with the change in the drivers of hot-dry conditions between CWS15 and CWS14.2, together with the 5 and 80 % confidence bands for the multi-model mean.

To verify it, we analyse the evolution of the variables (TX, TN and WB) used to construct the CEEI in order to provide a more comprehensive view of its evolution for different storylines. For this purpose, we calculate the standardised anomalies of the variables with respect to the CWS14.2. The four quadrants of Figure 33 are determined by the multi-model mean and define the storylines. The results of each storyline are presented in the form of boxplots in Figure 34.

We notice that, except for LL, all storylines indicate a shift towards warmer and drier conditions, albeit the magnitude of the changes is uncertain. Accordingly, increases in hot and dry events are very likely. These shifts can be pronounced for storylines involving increases in Z500, in agreement with Z500 being the most strongly influenced variable (Figure 29). However, large changes in SSTs may also trigger similar changes in these variables. TN shows stronger shifts than TX in all storylines, showing the added value of its inclusion into the used indices. It also implies a marked increase of warm nights, which are impactful phenomena for the health sector. Most of the WB distribution shifts towards drier conditions than in CWS14.2, except for LL, for which the WB variability would remain similar to that in CWS14.2.

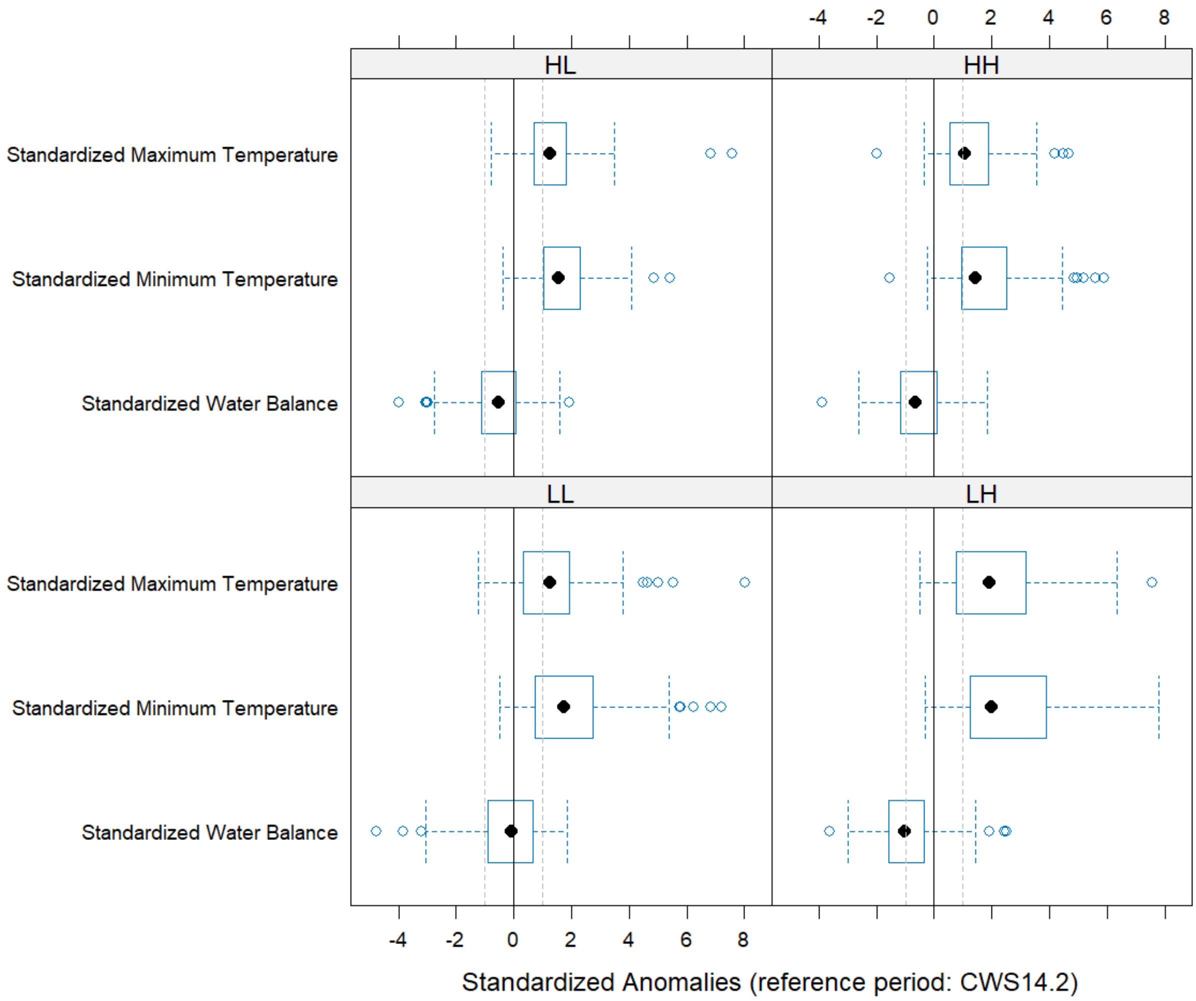

Figure 34 Boxplots of standardised anomalies of TX, TN and WB (with respect to CWS14.2). Storylines are based on changes with respect to the multi-model mean (Figure 33). Grey lines indicate an interval of one standard deviation and the black line is the mean value.

5. CONCLUSIONS

This report describes Artificial Intelligence (AI) and Machine Learning (ML) applications for the topics addressed in Work Package 5 (WP5), which include the attribution of Extreme Events (EEs), the detection of observed trends, and the quantification of future changes in EEs and concurrent EEs. The deliverable covers different types of EEs, with special emphasis on high-impact heatwaves (HWs) and droughts (DRs) at continental and regional hotspot scales of interest for WP6 and WP7. The key findings and outlook for next steps are summarised as follows:

A novel hybrid approach combining AI-based Autoencoders and the classical Analogue Method (AE-AM) has been developed in CLINT for a probabilistic reconstruction of meteorological fields during EEs. This algorithm uses a deep AE trained with predictor fields of the EE (herein sea-level pressure, SLP, although other input variables may be considered), which are encoded into a reduced latent space. Then, the AM is directly applied to the states of this latent space in order to find similar situations in the historical record. These analogue days are finally employed to reconstruct the targeted field. The AE-AM approach has shown better performance than the classical AM in reconstructing the daily maximum temperature (TX) during the major historical European HWs. The results indicate that the AE can condense important information in the latent space that the AM may exploit for reconstruction in a more efficient way than using the explicitly resolved field. This illustrates a clear advantage of pre-processing the data with the AE before the application of the AM method. Next steps foresee the inclusion of additional inputs (information channels), which could lead to reduced reconstruction errors and the development of highly conditioned attribution of EEs. The issues arising from the increased complexity in the model, such as the definition of distances in the multivariate space, or the relative importance of the predictors chosen for HW reconstruction are amenable to deep learning techniques.

In addition to AE-AM, a pure AI-based model for attribution of EEs has been developed within CLINT. It uses a Variational Autoencoder (VAE) anomaly method for the detection of climate change signals. To do so, the model is trained in natural climate conditions, as simulated by Global Climate Models (GCMs), and this learning is applied to reconstruct a naturalised version of historical fields associated with EEs. The performance of the model is tested in simulated HWs. The results indicate that the VAE-anomaly detection is capable of detecting climate change signals in the spatio-temporal patterns of EEs (i.e. field structures that have never been seen in a climate without anthropogenic climate change). Once trained, the reconstruction is very fast and hence the method can be exploited for near real-time attribution of EEs (i.e. reconstructing how the EE could have evolved in a preindustrial climate without human influences). Future developments include: 1) cross-validation experiments in other (out-of-training) GCMs and/or reanalyses to address the influence of model biases and sensitivity to training datasets; 2) the transference of the VAE-anomaly method to the real world (observed EEs) and; 3) its application to other types of EEs.

An AI-based method for the reconstruction of observational extreme indices over Europe has been designed. The infilling model employs a U-Net with partial convolutional layers. It is trained with historical GCM runs that were artificially masked with the missing values of the observational product to be completed, so that the infilled dataset can be compared with the original one. The trained model is then applied for the reconstruction of extreme indices of HadEX-CAM for each month of 1901-2018. For extreme temperature indices, the method produces consistent results when trained with different datasets, and outperforms the Kriging method, being very effective in the infilling of a large amount of missing information with irregular distribution. The model provides a longer temporal context for the assessment of trends of extreme indices and improves the characterisation of spatial patterns and regional means. Our AI-enhanced reconstruction reveals complex and more heterogeneous patterns of trend in European extreme temperature indices than HadEX3. The analyses demonstrate the capability of our AI method to investigate observed trends in a continuous observational dataset with a spatial resolution similar to that of modern reanalysis products but with a much longer temporal coverage. The reconstructed dataset has been prepared for the climate community to foster investigations that could further improve the understanding of EEs and their changes at local and regional scales. Additionally, it seeks to contribute to the development of nuanced climate-related policy at the regional level. The infilling model has also been adapted successfully for the reconstruction of global monthly mean temperature (HadCRUT5) and the Global Precipitation Climatology Centre dataset operated by the World Meteorological Organization, therefore allowing the infilling of derived indices related to multivariate climate EEs such as DRs. Next steps include: 1) the reconstruction of other extreme indices (e.g. those related to precipitation); 2) the infilling of other observational datasets (e.g. E-OBS); 3) the extension of the analysis to global scales, including challenging regions characterised by severe data scarcity (e.g. Africa).

Novel algorithms based on multiple AI-based strategies have been developed for the construction of storylines of EEs. These methods combine different AI-based approaches for the identification of candidate drivers (spatial clustering for dimensionality reduction and evolutionary algorithms for the optimisation of extracted features) with physical understanding and model evaluation supporting the final choice of drivers. AI-based methods deal with several limitations of classical methods, such as the large number of potential predictors (high-dimensionality), and their non-linear interactions and dependencies among the drivers. Drivers are uncovered by applying these AI-based methods to reanalysis data, and subsequently verified on the multi-model ensemble. Then, the storyline technique makes it possible to construct physically-consistent evolutions of EEs based on the combined effect of climate change responses in these drivers. Such an approach allows managing the intrinsic multi-model uncertainty in future projections of EEs, reducing the large spread to a manageable number of storylines. The use of Common Warming States (CWSs), which select specific periods in the GCM simulations, further circumvents eventual model biases by focusing on periods with the same climate conditions. As an illustration of the AI-enhanced identification of climate drivers of future changes of EEs, the CLINT approach has been applied to summer HWs in the Po Valley, pinpointing its dependence on two main discovered drivers linked to regional precipitation in southeastern Europe and large-scale atmospheric circulation. In particular, an enhanced reduction of precipitation over Italy and the western Balkans and a deepening of North Atlantic Oscillation (both with respect to the multi-model mean) would increase the 90th percentile of maximum temperatures over the Po Valley by +1.72°C, which is approximately double than the simulated global warming (+0.8°C).

Similar storyline experiments have been conducted for summer hot and dry events in central Europe (as represented by the Concurrent Extreme Event Index, CEEI), which can affect multiple socio-economic sectors such as agriculture and health. In this case, Kernel Regularized Canonical Correlation Analysis method, combined with expert-based pre-selection of candidate drivers (mid-tropospheric geopotential height and sea surface temperature, SST), has been employed to identify the dominant predictors and encapsulate their influences. The most important drivers of hot-dry days were atmospheric blocking in summer and a dipole of SSTs from April to June. Most of the GCMs project these drivers to occur more frequently in the future. As a consequence, storylines overall indicate that hot and dry conditions in central Europe will become more intense in the future. However, there are non-linear responses in the magnitude of hot-dry EEs to the changes in these drivers. Given the plethora of existing kernel functions, this CLINT approach can be adapted to different types of compound EEs. This will allow investigating non-linear feedbacks of these drivers and their evolution in future scenarios for a wide range of EEs. The storyline approach developed in CLINT dissects the interdependencies between the considered EEs and a set of selected drivers. Furthermore, it enables to identify the set of GCMs that follow a specific driver's or combination of drivers' responses (storyline), which can be used to run impact-based models (WP6 and WP7) and construct best and worst case scenarios. Additional future developments include the application of storylines to other hotspots (e.g. Douro Basin or Rhine Delta) and EEs (e.g. Tropical Cyclones and

DRs), and the exploration of processes and causal links between the drivers and the EEs. This will be the object of near future research, in cooperation with other WPs.

Some of the applications carried out in WP5 have been developed at regional scales (e.g., storylines of HWs in the Po Valley; hot-dry days in central Europe), but with methods that are easily portable and adaptable to other regions. Other developments provide continental-wide results with information at the grid point scale (e.g. trends in extreme indices, attribution in spatially resolved fields), which would allow applications at different spatial aggregations, from the local scales relevant to WP7 to the pan-European scales addressed in WP6. As a note of caution, it is worth mentioning that the use of datasets and methods developed in WP5 are fit for purpose. Therefore, they might not be optimal for other regions or EEs and hence we strongly encourage testing the sensitivity of AI models before its application, and tuning the architecture and hyperparameters to the specific problem at hand, if needed.

REFERENCES

Allen, R., Pereira, L., Raes, D., & Smith, M. (1998). Crop evapotranspiration guidelines for computing crop requirements. FAO Irrig. Drain. Report modeling and application. Journal of Hydrology, 285, 19–40.

Apley, D.W., & Zhu, J. (2020). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82, 1059–1086.

Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2019). Viewing forced climate patterns through an AI Lens. Geophysical Research Letters, 46, 13,389–13,398.

Barriopedro, D., Sousa, P. M., Trigo, R. M., García-Herrera, R., & Ramos, A. M. (2020). The exceptional Iberian heatwave of summer 2018. Bulletin of the American Meteorological Society, 101(1), S29–S34.

Barriopedro, D., García-Herrera, R., Ordóñez, C., Miralles, D. G., & Salcedo-Sanz, S. (2023). Heat waves: Physical understanding and scientific challenges. Reviews of Geophysics, 61(2), e2022RG000780.

Beobide-Arsuaga, G., Düsterhus, A., Müller, W.A., Barnes, E.A., & Baehr, J. (2023). Spring Regional Sea Surface Temperatures as a Precursor of European Summer Heatwaves. Geophysical Research Letters, 50, e2022GL100727.

Callaghan, M., Schleussner, CF., Nath, S., Lejeune, Q., Knutson, T. R., Reichstein, M., Hansen, G., Theokritoff, E., Andrijevic, M., Brecha, R. J., Hegarty, M., Jones, C., Lee, K., Lucas, A., van Maanen, N., Menke, I., Pfleiderer, P., Yesil B., & Minx, J. C. (2021). Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. Nature Climate Change, 11, 966–972.

Camps-Valls, G., Tuia, D., Xiang Zhu, X. & Reichstein, M. (2021). Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences. 1st Edition, John Wiley & Sons, 432 pp., https://doi.org/10.1002/9781119646181

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., & Codron, F. (2010). Winter 2010 in Europe: A cold extreme in a warming climate. Geophysical Research Letters, 37(20),1–6.

Chervenkov, H., Slavov, K. & Ivanov, V. (2019). STARDEX and ETCCDI climate indices based on E-OBS and CARPATCLIM: Part one: General description, 11189. Springer International Publishing, 360–367, https://doi.org/10.1007/978-3-030-10692-8 40.

Ciavarella, A., Christidis, N., Andrews, M., Groenendijk, M., Rostron, J., Elkington, M., Burke, C., Lott, F. C., & Stott, P. A. (2018). Upgrade of the HadGEM3-A based attribution system to high resolution and a new validation framework for probabilistic event attribution. Weather and Climate Extremes, 20, 9-32.

Coles, S. (2004). An introduction to statistical modeling of extreme values. 4th Edition. Springer, London, Berlin, Heidelberg, 208 pp.

Dong, J., Yin, R., Sun, X., Li, Q., Yang, Y., & Qin, X. (2019). Inpainting of Remote Sensing SST Images with Deep Convolutional Generative Adversarial Network. IEEE Geoscience and Remote Sensing Letters, 16 (2), 173–177.

Dunn, R. J., & Coauthors (2020). Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. Journal of Geophysical Research: Atmospheres, 125 (16), 1–28.

Faranda, D., Bourdin, S., Ginesta, M., Krouma, M., Noyelle, R., Pons, F., Yiou, P., & Messori, G. (2022). A climate-change attribution retrospective of some impactful weather extremes of 2021. Weather and Climate Dynamics, 3(4), 1311–1340.

Garali, I., Adanyeguh, I.M., Ichou, F., Perlbarg, V., Seyer, A., Colsch, B., Moszer, I., Guillemot, V., Durr, A., Mochel, F., & Tenenhaus, A. (2018). A strategy for multimodal data integration: application to biomarkers identification in spinocerebellar ataxia. Briefings in Bioinformatics, 19, 1356–1369.

Garrido-Pérez, J.M., Barriopedro, D., Trigo, R.M., Soares, P.M.M., Zappa, G., Álvarez-Castro, M.C., & García-Herrera, R. (2024). Storylines of summer warming in Iberia: Roles of the atmospheric circulation and the preceding conditions of soil moisture and Mediterranean sea surface temperature. Atmospheric Research (under review).

Geenens, G., Charpentier, A. & Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. Bernoulli, 23 (3), 1848-1873.

Geiss, A., & Hardin, J.C. (2021). Inpainting radar missing data regions with deep learning. Atmospheric Measurement Techniques, 14 (12), 7729–7747.

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., & Tebaldi, C. (2016). The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. Geoscientific Model Development, 9, 3685–3697.

Hao, Z., Hao, F., Xia, Y., Feng, S., Sun, C., Zhang, X., Fu, Y., Hao, Y., Zhang, Y., & Meng, Y. (2022). Compound droughts and hot extremes: Characteristics, drivers, changes, and impacts. Earth-Science Reviews, 235, 104241.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049.

Horton, D.E., Johnson, N.C., Singh, D., Swain, D.L., Rajaratnam, B., & Diffenbaugh, N.S. (2015). Contribution of changes in atmospheric circulation patterns to extreme temperature trends. Nature, 522, 465–469.

Hu, Y., Wang, G., Wei, X., Zhou, F., Kattel, G., Amankwah, S. O. Y., Hagan, D. F. T., & Duan, Z. (2023). Reconstructing long-term global satellite-based soil moisture data using deep learning method. Frontiers in Earth Science, 11, 1–14.

Hutchinson, M. F. (1995). Interpolating mean rainfall using thin plate smoothing splines. International Journal of Geographical Information System, 9 (4), 385–403.

IPCC (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., …, & Zhou, B. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, In press, doi:10.1017/9781009157896.

Jézéquel, A., Yiou, P., Radanovics, S., & Vautard, R. (2018). Analysis of the exceptionally warm December 2015 in France using flow analogues. Bulletin of the American Meteorological Society, 99(1), S76–S79.

Jones, P. D. (1994). Hemispheric surface air temperature variations: a reanalysis and an update to 1993. Journal of Climate, 7 (11), 1794–1802.

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. Nature Geoscience, 13 (6), 408–413.

Kingma, D.P., & Welling, M. (2019). An introduction to variational autoencoders. Foundations and Trends in Machine Learning, 12(4), 307–392.

Klampanos, I.A., Davvetas, A., Andronopoulos, S., Pappas, C., Ikonomopoulos, A. & Karkaletsis, V. (2018). Autoencoder-driven weather clustering for source estimation during nuclear events. Environmental Modelling & Software, 102, 84–93.

Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. Nature, 529 (7584), 84–87.

Liu, G., Reda, F.A., Shih, K. J., Wang, T. C., Tao, A., & Catanzaro, B. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, 11215. Springer, https://doi.org/10.1007/978-3-030-01252-6_6.

Loader, C.R. (1996). Local likelihood density estimation. Annals of Statistics, 24(4), 1602–1618.

MacKay, D.J.C. (1992). A Practical Bayesian Framework for Backpropagation Networks. Neural Computation 4, 448–472.

Mindlin, J., Shepherd, T.G., Vera, C.S., Osman, M., Zappa, G., Lee, R.W., & Hodges, K.I. (2020). Storyline description of Southern Hemisphere midlatitude circulation and precipitation response to greenhouse gas forcing. Climate Dynamics, 54, 4399–4421.

Miralles, D. G., Gentine, P., Seneviratne, S. I., & Teuling, A. J. (2019). Land–atmospheric feedbacks during droughts and heat waves: state of the science and current challenges. Annals of the New York Academy of Sciences, 1436(1), 19–35.

Monerie, P.A., Biasutti, M., Mignot, J., Mohino, E., Pohl, B., & Zappa, G. (2023). Storylines of Sahel Precipitation Change: Roles of the North Atlantic and Euro-Mediterranean Temperature. Journal of Geophysical Research: Atmospheres, 128, 1–17.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., & Simpson, I. R. (2021). An updated assessment of near-surface temperature change from 1850: The hadcrut5 data set. Journal of Geophysical Research: Atmospheres, 126 (3), e2019JD032361.

National Academies of Sciences, Engineering, and Medicine, Committee on Extreme Weather Events and Climate Change Attribution, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies (2016). Attribution of Extreme Weather Events in the Context of Climate Change. National Academies Press, Washington, D.C. https://doi.org/10.17226/21852

North, G., Bell, T., Cahalan, R., & Moeng, F. (1982). Sampling Errors in the Estimation of Empirical Orthogonal Functions. Monthly Weather Review, 110, 699–706.

Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System, 4 (3), 313–332.

Olonscheck, D., Suarez-Gutierrez, L., Milinski, S., Beobide-Arsuaga, G., Baehr, J., Fröb, F.,… & Brune, S. (2023). The New Max Planck Institute Grand Ensemble With CMIP6 Forcing and High-Frequency Model Output. Journal of Advances in Modeling Earth Systems, 15 (10), 1–21.

Otto, F. E. L. (2017). Attribution of Weather and Climate Events. Annual Review of Environment and Resources, 42, 627–646.

Pang, G., Shen, C., Cao, L. & van Den Hengel, A. (2021). Deep Learning for Anomaly Detection: A Review. Association for Computing Machinery Computing Surveys, 54, 2, 38.

Pasini, A., Racca, P., Amendola, S., Cartocci, G., & Cassardo, C. (2017). Attribution of recent temperature behaviour reassessed by a neural-network method. Scientific Reports, 7, 17681.

Pérez-Aracil, J., Camacho-Gómez, C., Lorente-Ramos, E., Marina, C. M., Cornejo-Bueno, L. M., & Salcedo-Sanz, S. (2023). New Probabilistic, Dynamic Multi-Method Ensembles for Optimization Based on the CRO-SL. Mathematics, 11(7), 1666.

Pérez-Aracil, J., Marina, C.M., Zorita, E., Barriopedro, D., Zaninelli, P., García-Herrera, R., Giuliani, M., Castelletti, A., Gutiérrez, P.A., & Salcedo-Sanz, S. (2024). Variational autoencoder-based flow-analogue reconstruction of heat waves from pressure fields. Annals of the New York Academy of Sciences (under review).

Perkins, S.E., Pitman, A.J., Holbrook, N.J., & McAneney, J. (2007). Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions. Journal of Climate, 20, 4356–4376.

Perkins, S.E., Alexander, L.V. (2013). On the Measurement of Heat Waves. Journal of Climate 26, 4500–4517.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv, https://doi.org/10.1109/ACCESS.2021.3053408, 1505.04597.

Russo, S., Sillmann, J., & Fischer, E.M. (2015). Top ten European heatwaves since 1950 and their occurrence in the coming decades. Environmental Research Letters, 10(12), 124003.

Salcedo-Sanz, S., Del Ser, J., Landa-Torres, I., Gil-López, S., & Portilla-Figueras, J. A. (2014). The coral reefs optimization algorithm: a novel metaheuristic for efficiently solving optimization problems. The Scientific World Journal, 2014, 739768.

Salcedo-Sanz, S., Ascenso, G., Del Ser, J., Casillas-Pérez, D., Kadow, C., Fister, D., Barriopedro, D., García-Herrera, R., Giuliani, M., & Castelletti, A. (2024). Analysis, characterization, prediction and attribution of extreme atmospheric events with machine learning: a review. Theoretical and Applied Climatology, 155, 1-44.

Schölkopf, B. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge, Mass., 626 pp.

Schulzweida, U. (2022) Cdo user guide. https://doi.org/10.5281/zenodo.7112925.

Sen, P. K. (1968). Estimates of the regression coefficient based on kendall's tau. Journal of the American Statistical Association, 63 (324), 1379–1389.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., …, & Zhou, B. (2021). Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson–Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., … & Zhou, B. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766.

Shepard, D. (1968). A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. Proceedings of the 1968 23rd Association for Computing Machinery National Conference, 517–524.

Shepherd, T. G. (2016). A common framework for approaches to extreme event attribution. Current Climate Change Reports, 2, 28–38.

Shepherd, T.G., Boyd, E., Calel, R.A., Chapman, S.C., Dessai, S., Dima-West, I.M., Fowler, H.J., James, R., Maraun, D., Martius, O., Senior, C.A., Sobel, A.H., Stainforth, D.A., Tett, S.F.B., Trenberth, K.E., van den Hurk, B.J.J.M., Watkins, N.W., Wilby, R.L., & Zenghelis, D.A. (2018). Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. Climatic Change, 151, 555–571.

Shibata, S., Iiyama, M., Hashimoto, A., & Minoh, M. (2018). Restoration of Sea Surface Temperature Satellite Images Using a Partially Occluded Training Set. Proceedings – International Conference on Pattern Recognition, 2771–2776.

Sousa, P.M., Trigo, R. M., Barriopedro, D., Soares, P. M. M., & Santos J. A. (2018). European temperature responses to blocking and ridge regional patterns. Climate Dynamics, 50, 1-2, 457-477.

Squintu, A. A., van der Schrier, G., van den Besselaar, E., van der Linden, E., Putrasahan, D., Roberts, C., Roberts, M., Scoccimarro, E., Senan, R. & Klein Tank, A. (2021). Evaluation of trends in extreme temperatures simulated by HighResMIP models across Europe. Climate Dynamics, 56 (7-8), 2389–2412.

Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J. - P., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P. & Zwiers, F. W. (2016). Attribution of extreme weather and climate-related events. Wiley Interdisciplinary Reviews: Climate Change, 7, 23–41.

Tenenhaus, A., Philippe, C., & Frouin, V. (2015). Kernel Generalized Canonical Correlation Analysis. Computational Statistics & Data Analysis 90, 114–131.

Tenenhaus, M., Tenenhaus, A., & Groenen, P.J.F. (2017). Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods. Psychometrika 82, 737–777.

Theil, H. (1992). A Rank-Invariant Method of Linear and Polynomial Regression Analysis. In: Raj, B., Koerts, J. (eds) Henri Theil's Contributions to Economics and Econometrics. Advanced Studies in Theoretical and Applied Econometrics, 23. Springer, Dordrecht.

Toreti, A., Belward, A., Perez-Dominguez, I., Naumann, G., Luterbacher, J., Cronie, O., Seguini, L., Manfron, G., Lopez-Lozano, R., Baruth, B., Berg, M., Dentener, F., Ceglar, A., Chatzopoulos, T. & Zampieri, M. (2019). The Exceptional 2018 European Water Seesaw Calls for Action on Adaptation. Earth's Future, 7, 652–663.

Vicente-Serrano, S., Beguería, S., & López-Moreno, J.I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. Journal of Climate, 23, 1696–1718.

Wilks, D.S. (2011). Statistical methods in the atmospheric sciences, 3rd ed. Academic Press, Oxford, Waltham, MA, 676 pp.

Woollings, T. (2010). Dynamical influences on European climate: An uncertain future. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368, 3733–3756.

Wu, X., Hao, Z., Tang, Q., Singh, V. P., Zhang, X., & Hao, F. (2021). Projected increase in compound dry and hot events over global land areas. International Journal of Climatology, 41(1), 393–403.

Xu, S.G., Majumder, R., & Reich, B.J. (2022). SPQR: An R Package for Semi-Parametric Density and Quantile Regression, 25 pp. http://arxiv.org/pdf/2210.14482.

Xu, S.G. & Reich, B.J. (2023). Bayesian nonparametric quantile process regression and estimation of marginal quantile effects. Biometrics, 79, 151–164.

Yao, Z., Zhang, T., Wu, L., Wang, X., & Huang, J. (2023). Physics-Informed Deep Learning for Reconstruction of Spatial Missing Climate Information in the Antarctic. Atmosphere, 14 (4), 1–16.

Yiou, P., Jézéquel, A., Naveau, P., Otto, F. E. L., Vautard, R., & Vrac, M. (2017). A statistical framework for conditional extreme event attribution. Advances in Statistical Climatology Meteorology and Oceanography, 3(1), 17–31.

Zampieri, M., Ceglar, A., Dentener, F., & Toreti, A. (2017). Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. Environmental Research Letters, 12 (6), 064008.

Zappa, G. (2019). Regional Climate Impacts of Future Changes in the Mid–Latitude Atmospheric Circulation: a Storyline View. Current Climate Change Reports, 5, 358–371.

Zappa, G., & Shepherd, T.G. (2017). Storylines of atmospheric circulation change for European regional climate impact assessment. Journal of Climate, 30, 6561–6577.

Zorita, E. & von Storch, H. (1999). The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. Journal of Climate, 12(8), 2474–2489.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R.M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M.D., Maraun, D., Ramos, A.M., Ridder, N.N., Thiery, W., & Vignotto, E. (2020). A typology of compound weather and climate events. Nature Reviews Earth & Environment, 1, 333–347.

CLINT

CLIMATE INTELLIGENCE

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