The value of large-scale climatic indices for monthly forecasting severity of widespread flooding using dilated convolutional neural networks

Larisa Tarasova[1], Bodo Ahrens[2], Amelie Hoff[3], and Upmanu Lall[4]

[1]Helmholtz Centre for Environmental Research - UFZ
[2]IAU, Goethe University Frankfurt am Main
[3]Deutscher Wetterdienst
[4]Columbia University

September 19, 2023

The value of large-scale climatic indices for monthly forecasting severity of widespread flooding using dilated convolutional neural networks

Tarasova, Larisa[1,2], Ahrens, Bodo[3], Hoff, Amelie[3,4], and Upmanu Lall[2,5]

[1]Department Catchment Hydrology, Helmholtz Centre for Environmental Research - UFZ, Halle (Saale), Germany
[2]Department of Earth and Environmental Engineering, Columbia University in the City of New York, New York, USA
[3]Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt am Main, Frankfurt am Main, Germany
[4]Deutscher Wetterdienst, Offenbach, Germany
[5]Columbia Water Center, Columbia University in the City of New York, New York, USA

Key Points:

- An adapted Wavenet model conditioned by climatic indices is applied to forecast the monthly severity of widespread flooding in Germany
- Predictability of rain-on-snow floods is higher than of rainfall-generated floods showing the potential for forecasting severe winter floods
- Feature attribution reveals variable importance of large-scale climatic indices and antecedent wetness conditions across German regions

Corresponding author: Larisa Tarasova, `larisa.tarasova@ufz.de`

Abstract

Spatially co-occurring floods pose a threat to the resilience and recovery of the communities. Their timely forecasting plays a crucial role for increasing flood preparedness and limiting associated losses. In this study we investigated the potential of a dilated Convolutional Neural Network (dCNN) model conditioned on large-scale climatic indices and antecedent precipitation to forecast monthly severity of widespread flooding (i.e., spatially co-occurring floods) in Germany with one month lead time. The severity was estimated from 63 years of daily streamflow series as the sum of concurrent exceedances of at-site two-year return periods within a given month across 172 mesoscale catchments (median area 516 km^2). The model was trained individually for the whole country and three diverse hydroclimatic regions to provide insights on heterogeneity of model performance and flood drivers. Our results showed a considerable potential for forecasting widespread flood severity using dCNN especially as the length of training series increases. However, event-based evaluation of model skill indicates large underestimation for rainfall-generated floods during dry conditions despite overall lower severity of these events compared to the rain-on-snow floods. Feature attribution and wavelet coherence analyses both indicated considerable difference in the major flood drivers in three regions. While the flooding in North-Eastern region is strongly affected by the Baltic Sea, the North-Western region is affected more by global patterns associated with the El-Niño activity. In the Southern region in addition to global patterns we detected the effect of the Mediterranean Sea, while antecedent precipitation plays a less important role in this region.

Plain Language Summary

Floods that occur simultaneously at different locations in space are associated with higher damages and are more difficult for communities to cope. Forecasting of such events still remains a challenge. Here we investigate the potential of recently developed deep learning networks for forecasting widespread flooding one month ahead. We test this model on a large dataset of long streamflow time series in Germany and its three geographical regions using climatic indices and past precipitation as supporting variables. We show that the model has a considerable potential for flood forecasting, especially for rain-on-snow floods, but its performance is limited for rain-on-dry events. Finally, using explainable machine learning we provide insights on contrasting differences in climatic flood drivers in different German regions.

1 Introduction

River floods are one of the most devastating natural hazards in the world, annually causing considerable fatalities and socio-economical losses that are expected to increase with advancing climate change (B. Merz et al., 2021). Their spatial co-occurrence, i.e., when river flooding occurs simultaniously at multiple locations, is an increasingly recognized threat for efficient flood adaptation (Brunner et al., 2020; Zscheischler et al., 2020). It limits preparedness and the resilience of the communities (Kreibich et al., 2017), and poses a major challenge for the insurance and re-insurance industries (Zanardo et al., 2019). Spatial co-occurrence of floods in Europe and globally is often associated with the activity of large scale climatic patterns and teleconnections, such as North Atlantic Oscillation (Zanardo et al., 2019) and El Niño Southern Oscillation (Ward et al., 2014). At the same time changing spatial extent of floods (i.e., number of simultaneously affected river gauges) in Europe (Berghuijs et al., 2019; Kemter et al., 2020) is strongly related to changes in dominant flood generation processes.

In Germany, as in the USA (Brunner et al., 2020), spatially co-occurring floods (i.e., widespread flooding) occur in winter season (Uhlemann et al., 2010) and most frequently caused by the combination of intensive snowmelt and rainfall that are modulated by the dynamics of North Atlantic Oscillation and Scandinavian Pattern (Krug et al., 2020).

However, around 36% of all trans-basin floods in Germany in the second half of the 20th century occurred in summer (Uhlemann et al., 2010). The two recent very severe summer widespread floods in 2002 and 2013 were associated with the north-east propagating cyclone originating from the Mediterranean Sea (Blöschl et al., 2013). Moreover, there is evidence that spatial co-occurrence of flood-rich periods in Germany is mostly caused by elevated antecedent wetness conditions (B. Merz et al., 2018) indicating that additional information on land surface conditions or on prior rainfall that could be related to the same atmospheric circulation mechanism might be useful for predicting widespread flooding (Nakamura et al., 2013; Nied et al., 2017).

Timely forecasting of catastrophic widespread flooding is very important for successful flood adaptation in changing climate. Currently global climate models are able to issue medium-range and seasonal forecasts (i.e, from several weeks to several months) of the large-scale climatic patterns and teleconnections, such as North Atlantic Oscillation and El Niño Southern Oscillation, with relatively high fiedelity (Ludescher et al., 2014; Feng et al., 2021), compared to the limited forecasting skill of precipitation and floods (Slater et al., 2019). Therefore, large-scale climatic indices might become useful for improving accuracy and increasing lead time for forecasting of widespread flooding events using data-driven models.

Climatic indices are based on the large scale (i.e., regional to global) geophysical states and atmospheric circulations that are often used to characterize and investigate the teleconnections between large scale circulation patterns and local scale phenomena (e.g., precipitation, floods). In this study we investigate the value of ten standard climatic indices for forecasting the severity of widespread flooding in Germany. Since there are clear indications that the emergence of floods (R. Merz et al., 2020) and particularly the occurrence of spatially coherent floods (Nied et al., 2017; B. Merz et al., 2018) are related to antecedent wetness conditions we also investigate the role of antecedent precipitation as a proxy of wetness in the forecasting of severity of widespread flooding.

In recent years deep learning neural networks have emerged as a promising tool for time series forecasting, especially for medium range climate and weather forecast (Rasp & Thuerey, 2020; Schultz et al., 2021). Currently, Recurrent Neural Networks (RNN), particularly the Long Short-Term Memory (LSTM) models are recognized as effective neural networks for time series forecasting (Hsu, 2017). Their recurrent connections that allow the network to use the entire history of time series and to capture recurrent patterns or dynamic structures at different time scales (Hochreiter & Schmidhuber, 1997). Recent work shows that a more parsimonious and interpretable representations of such dynamics can be obtained by using multiple dilated convolutional layers (Yu & Lin, 2015). Comparison of the dilated Convolutional Networks (dCNN) to the stat-of-the-art RNNs (particularly to LSTM) using synthetic and real world examples showed comparable and even superior performance of dCNN models for conditional time series forecasting in cases when availability of long series is limited (Borovykh et al., 2017; Y. Chen et al., 2019; Benhaddi & Ouarzazi, 2021). Moreover, dCNN models are much easier to train, faster to converge and require less memory compared to the state-of-the-art RNN models (Benhaddi & Ouarzazi, 2021) since the convolutional structure of the network reduces the number of trainable parameters (Borovykh et al., 2017).

In this study, we apply the convolutional model Wavenet (van den Oord et al., 2016) originally developed for audio forecasting and adapted by Borovykh et al. (2017) for conditional forecasting of the severity of widespread flooding. The model is based on the dilated convolutions applied to both the input time series (i.e., index of widespread flood severity) and to the covariates (e.g., large-scale climatic indices, antecedent wetness) allowing the model not only learn the inherent patterns from the input time series, but also to learn its dependence on the covariates at different time scales. Previously the Wavenet model was successfully applied for conditional forecasting of financial time series (Borovykh et al., 2017), online retail time series (Y. Chen et al., 2019) and air quality time series

(Benhaddi & Ouarzazi, 2021). However, to the best of our knowledge adapted Wavenet was not previously used for the conditional forecasting of any hydrological time series. Moreover, the utility of large-scale climatic patterns and antecedent wetness for forecasting of widespread flooding in Germany, as well as the relative predictability of floods in different hydroclimatic regions and of floods generated by different processes (e.g., rain-on-snow vs rainfall) is still unclear. Given this background, this study has the following objectives:

1. test adapted Wavenet for forecasting of widespread floods in Germany
2. examine variability of its performance across hydroclimatologically distinct regions in Germany
3. examine variability of its performance for forecasting floods of different generation processes
4. compare the importance of climatic indices and antecedent wetness conditions obtained using a model-based feature attribution method (explainable machine learning techniques) with the results of the model-independent wavelet coherence analysis

2 Data

2.1 Severity index of widespread flooding and flood generation processes

To quantify the severity of widespread flooding across multiple gauges in Germany, we construct a severity index (Figure 1a) in line with the index used by Nied et al. (2014). First, we identify all streamflow events regardless of their peak magnitude from continuous daily time series of streamflow recorded at each gauge station using the method of Tarasova et al. (2018). From these, we select events whose peak magnitude exceeds the threshold of a local (i.e., at-site) 2-year return period (often associated with the full river bank) defined using annual maxima observed for the period from 1951 to 2013 (the records have minimum length 37 years, median length 61 years, maximum length 63 years). Finally, we sum up the identified exceedences (i.e., ratio of peak magnitude and the magnitude of 2-year return period) for all catchments for a given month and normalize it by the number of available gauges to account for a variable length of the observation time series (Figure 1b). Therefore, the derived index comprises the information about two features, (a) the number of gauges that in a given month report an exceedence of the 2-year return period (i.e., flooding), and (b) the magnitude of the exceedence of the at-site 2-year return period flow. In case none of the gauges report an exceedence above at-site 2-year return period the resulting severity index corresponds to 0 (Figure S1).

We compute the index of widespread flood severity for the whole country by combining the observations in all 172 streamflow stations across the country and we derive regional indices for each of the three hydroclimatic regions in Germany: North Western (NW, 76 catchments), North Eastern (NE, 28 catchments) and Southern (South, 68 catchments) regions (Figure 1a). The North-Western region (the Lower Rhine, Weser Rivers) is associated with winter flooding and is strongly affected by the Atlantic storms (Conticello et al., 2020). The North-Eastern region (the Elbe and Oder Rivers) is often associated with moderate winter floods with rare but severe summer floods (Petrow et al., 2007; Tarasova, Basso, & Merz, 2020). Finally, the Southern region (the Upper Rhine, Main and Danube Rivers) is associated with a mixed seasonality of floods, while in the right tributaries of the Danube summer floods are even more frequent compared to all other seasons (Beurton & Thieken, 2009). For the analysis we only consider mesoscale catchments with area range from 31 to 23,700 km^2 (median 516 km^2).

Additionally, to every value of the monthly severity of widespread flooding we assign a corresponding type of flood generation process that is identified as the most frequent event type among all affected catchments (Figure 1b, color-coding). Flood gen-

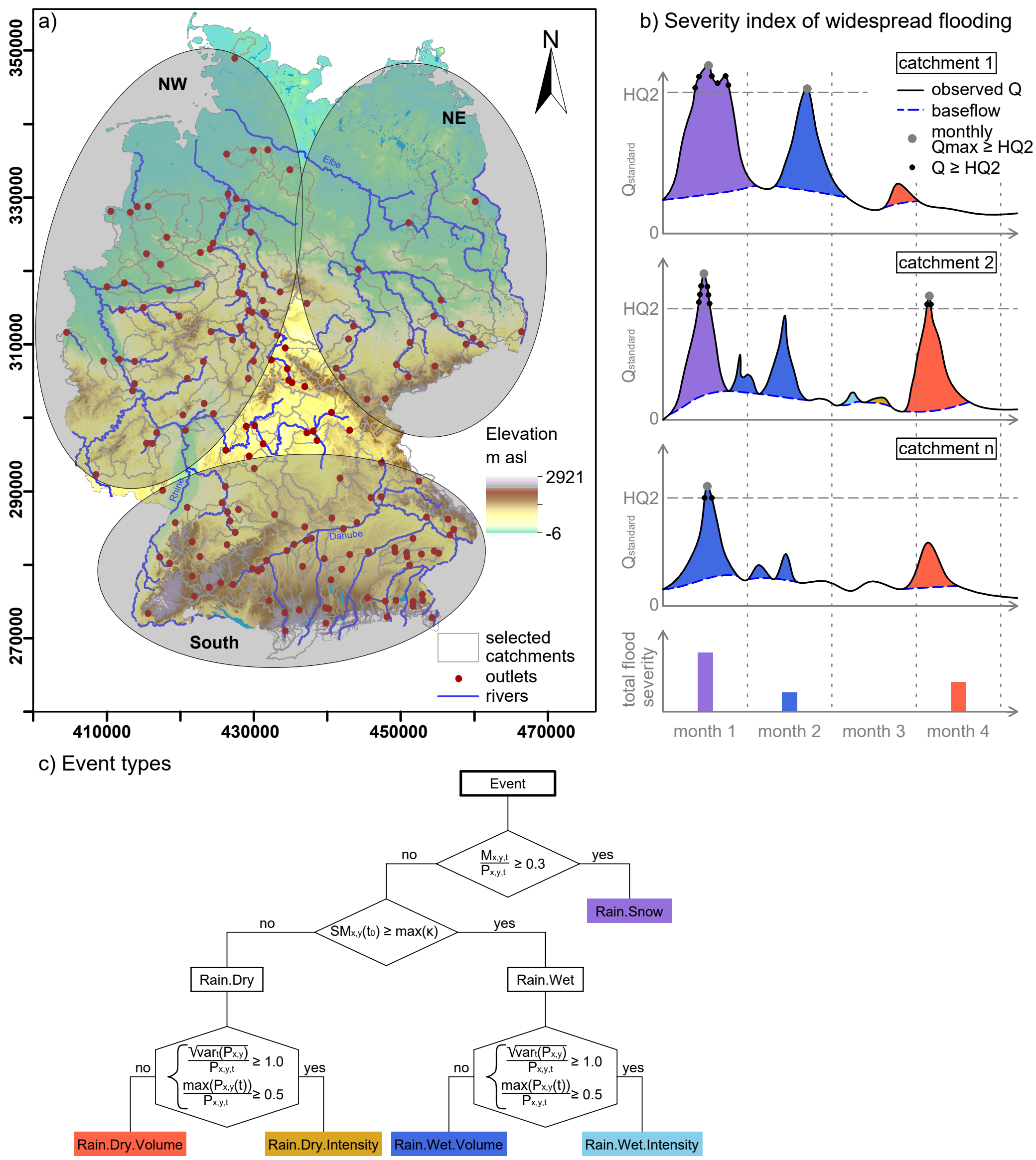

Figure 1. a) Study area and three distinct hydroclimatic regions (NW - North-Western, NE - North-Eastern and South - Southern regions); b) Derivation of the schematic index for widespread flooding (HQ2 stands for a local flood magnitude associated with 2-year return period); c) a hierarchical decision tree for classification of event types based on the event classification framework of Tarasova, Basso, Wendi, et al. (2020) (modified from Tarasova et al. (2023)), $M_{x,y,t}$ and $P_{x,y,t}$ stand for catchment- and event-averaged snowmelt and total precipitation respectively, $P_{x,y}$ represents a set of daily catchment-averaged total precipitation during the event, $SM_{x,y}(t_0)$ stands for antecedent soil moisture on the day prior to the event begin, $\max(\kappa)$ indicates the value of soil moisture that corresponds to the point of maximum curvature of the function that describes the relation between event runoff coefficients and antecedent soil moisture (see Tarasova et al. (2018) for more details). Final event types are indicated as colored boxes. For a detailed description of indicators, classification thresholds and event types please refer to Tarasova, Basso, Wendi, et al. (2020)

eration processes are identified (see Figure 1c for the assignment process) using the information on the proportion of catchment- and event-averaged snowmelt ($M_{x,y,t}$) in the total volume of precipitation ($P_{x,y,t}$), catchment-averaged soil moisture on the day before the event has started ($SM_{x,y}(t_0)$), proportion of maximum precipitation intensity during the event ($\max(P_{x,y}(t))$) and event-averaged total precipitation ($P_{x,y,t}$) and variability of daily catchment-averaged precipitation sums during the event (var(Px,y)) (Tarasova, Basso, Wendi, et al., 2020). Daily observed gridded (1 km) precipitation data are obtained from the REGNIE dataset (Rauthe et al., 2013), daily observed temperature gridded (4 km) is obtained from Zink et al. (2017). Daily gridded (4 km) soil moisture and snowmelt are simulated by the mesoscale Hydrological Model (mHM) (Samaniego et al., 2010; Kumar et al., 2013; Zink et al., 2017). All these datasets are available for the whole study period from 1951 to 2013. Using this information we distinguish five different event types (Figure 1c): Rain.Snow (combination of rainfall and snowmelt when the proportion of the latter in the total precipitation sum is at least 30%), Rain.Wet.Intensity (intensity-dominated (i.e., most of precipitation has occurred in a single time step during event) rainfall-induced event with wet antecedent conditions), Rain.Wet.Volume (volume-dominated rainfall-induced event with wet antecedent conditions), Rain.Dry.Intensity (intensity-dominated rainfall-induced event with dry antecedent conditions), Rain.Dry.Volume (volume-dominated rainfall-induced event with dry antecedent conditions (Tarasova et al., 2023). The threshold between wet and dry conditions ($\max(\kappa)$) is defined as the value of catchment-averaged soil moisture that corresponds to the point of maximum curvature of the non-linear function that capture the increase in event runoff coefficients with the increase of soil moisture (Tarasova et al., 2018). The sensitivity analysis performed in Tarasova, Basso, Wendi, et al. (2020) indicates that the parametric uncertainty and the choice of the hydrological model have only minor effect on classification results. For more details on the indicators and thresholds used for streamflow event classification refer to Tarasova, Basso, Wendi, et al. (2020) and for the details on its application for classification of flood events to Tarasova et al. (2023).

2.2 Covariates: large-scale climatic indices and antecedent precipitation

Climatic indices are a low dimensional proxy of the dynamics of the atmosphere and the ocean from seasonal to multi-decadal scales (Domeisen et al., 2018). We selected ten different climatic indices that capture the effects of global climatic variability (e.g., Pacific North American Pattern, Southern Oscillation Index, North Pacific, Antarctic Oscillation), patterns that capture more local effect of the Baltic Sea (Scandinavian Pattern, East Atlantic) and the effect of blocking conditions (Greenland Blocking Index, North Atlantic Oscillation, Arctic Oscillation) that were previously related to flood occurrence in Europe (e.g., Zanardo et al. (2019)). We also include the Mediterranean Oscillation Index that captures the effect of the Mediterranean Sea (Trigo et al., 2002). We provide a short description of each index that was used in this study.

Pacific North American Pattern (PNA) is one of the strongest extratropical teleconnections (Wallace & Gutzler, 1981) and is recognized as a major mode of planetary-scale atmospheric variability evident at all time scales over North America during boreal winter (Leathers et al., 1991). PNA modes are defined by the anomalies in the geopotential height fields over the North Pacific Ocean and North American continent and are strongly influenced by the El Niño-Southern Oscillation (Li et al., 2019).

Southern Oscillation index (SOI) is based on the sea level pressure differences between Tahiti and Darwin, Australia, and represents air pressure differences between western and eastern tropical Pacific. Periods of negative index correspond to El Niño episodes and abnormally warm ocean in the eastern tropical Pacific (Harrison & Larkin, 1998). The opposite is representative of La Niña episodes. These irregular occurring episodes have a major effect of the Earth's climate system and result in devastating hydrometeorological extremes in different regions of the world (Ward et al., 2016). Although El Niño-

Southern Oscillation teleconnections show only weak effect on European region compared to tropical regions and Pacific-North American sector, at seasonal time scales it remains one of the strongest predictors of European climate (Brönnimann, 2007).

North Pacific (NP) Oscillation describes the seesaw pattern of sea level pressure over subtropical and mid-latitude North Pacific Ocean (Wallace & Gutzler, 1981). It impacts temperature and precipitation over Eurasia and North America (S. Chen et al., 2021) and is believed to be a trigger of El Niño-Southern Oscillation events (Park et al., 2013).

Antarctic Oscillation (AAO) also known as Southern Annual Mode strongly affects the climate in high-latitudes of the the Southern Hemisphere (Thompson And & Wallace, 2000). However, several studies have indicates that seasonally variable effect of the El Niño-Southern Oscillation on extratropical circulation is strongly modulated by AAO (Silvestri & Vera, 2003; L'heureux & Thompson, 2006).

Arctic oscillation (AO) variability depends strongly on the strength of the Icelandic and Aleutian Lows that are in turn correspond to the variability of the North American Oscillation and Pacific North American patterns respectively (Thompson & Wallace, 1998). Although Arctic Oscillation has a strong correlation with North Atlantic Oscillation, it does not show the same changes in summer period as the latter in the last few decades (Hanna et al., 2015).

North Atlantic Oscillation (NAO) is a hemispheric meridional oscillation represented as the surface pressure fluctuation between the Azores and Iceland, which governs major weather patterns in Europe (Visbeck et al., 2001). NAO is strongly related to the occurrence and intensity of the blocking high pressure over Greenland (Hanna et al., 2015).

East Atlantic (EA) is structurally similar to NAO and is characterized by north-south dipole anomaly centers over North Atlantic. Compared to NAO the centers of the anomalies are shifted southeastward. Differently from NAO, EA has a strong subtropical link (Barnston & Livezey, 1987). EA affects the temperature and precipitation in Europe and North America (Comas-Bru & Mcdermott, 2014).

Greenland Blocking Index (GBI) is the mean 500 hPa geopotential height for the 60-80°N, 20-80°W region. GBI represent the blocking over Greenland which has an impact on the climate and weather in the Northern Hemisphere (Hanna et al., 2016). Variability of the GBI is related to the North Atlantic Oscillation variations, as well as to changes in East Atlantic and Scandinavian patterns (Scherrer et al., 2006; Hanna et al., 2015).

Scandinavian Pattern (SCA) is a low-frequency teleconnection that represents geopotential height anomalies over the extratropical Northern Hemisphere, particularly centered around the Scandinavian Peninsula, northeastern Atlantic and central Siberia (Bueh & Nakamura, 2007). The positive phase is related to the major blocking anticyclones over Scandinavia and Russia. This teleconnection is often related to temperature and precipitation in Europe with the contrasting effect on the Southern and Northern parts (Liu et al., 2014).

Mediterranean Oscillation Index (MOI) is a patterns that captures pressure differences between the West and East Mediterranean (i.e., Algiers and Cairo). The MOI is related to the activity of cyclogenesis over the Mediterranean Sea (Trigo et al., 2002). Since this process is triggered by the cold fronts from the Atlantic, there is a relation between MOI variability and the North Atlantic Oscillation, as well as Arctic Oscillation (Dünkeloh & Jacobeit, 2003).

The monthly time series of all climatic indices, their wavelet and power spectrums (see Section 3.3.2) used in this study are displayed in Figures S2-S3. The links to the time series of climatic indices sources is provided in the Open Research Section.

The time series of antecedent wetness conditions for the whole study period are also considered as a possible covariate and approximated as the monthly mean of catchment-averaged precipitation of all study catchments for the forecasting of the country-wide severity index and of catchments belonging to the respectful region (i.e., North-Western, North-Eastern and Southern) for forecasting of the regional severity indices (Figure S4). Catchment-averaged precipitation was obtained from daily observed gridded precipitation dataset REGNIE (Rauthe et al., 2013).

3 Methods

3.1 Architecture of the dilated Convolutional Neural Networks

Convolutional Neural Networks (CNN) are deep learning models that consist of the stacked convolutional layers where the output is connected to local regions of the input (Figure 2b) by means of a convolution (i.e., dot product) between the input and sliding filters (i.e., matrix of weights). This architecture results in much smaller number of trainable parameters compared to the fully-connected networks (Figure 2a). Due to their ability to effectively recognize patterns in the input series, CNN networks are particularly successful in image recognition and time series classification (Krizhevsky et al., 2012; Lecun et al., 2015; Wang et al., 2016). To allow the CNN network to access larger portion of time series beyond local nodes and to learn recurring patterns at both shorter and longer time scales, a dilation can be added to each convolutional layer (Figure 2c). The dilation essentially means applying the same sliding filter, but skipping input values with a certain step (van den Oord et al., 2016). At each subsequent convolutional layer the dilation progressively increases by a factor of two (Figure 2c), hence allowing model to efficiently learn connections even between far-apart points while keeping number of trainable parameters low (Borovykh et al., 2017). In case of time series forecasting the underlying idea of the dilated convolutional architecture is to learn repeating patterns in the time series and use them to predict future values. Multi-layer structure of these networks acts similar to wavelet transformation and is effective for discerning low- and high-frequency components from time series while reducing noise propagation with each layer (Aussem & Murtagh, 1997).

Thanks to the convolutional structure of the dCNN model (Figure 2b), it is much easier to train, faster to converge and requires less memory compared to the state-of-the-art RNN models (Benhaddi & Ouarzazi, 2021). Moreover, the adapted Wavenet model performs dilated convolutions on both the response time series (i.e., index of widespread flood severity) and on the covariates (e.g., large-scale climatic indices, antecedent precipitation) (Figure 2d) allowing the model to learn not only the inherent patterns from the input time series, but also to learn its dependencies with the covariates at different time scales. The adapted version uses rectified linear unit (ReLU) instead of the gated activation function (Figure 2d,e). The parametrized skip connection at the first layer (Figure 2d) ensures that the Wavenet network can learn to discard a covariate if it does not improve the forecast. The result from the first layer is the input for the next convolutional layer (Figure 2e), this step is repeated till the last convolutional layer that corresponds to the selected depth of the network. The output of the last convolutional layer is passed through a 1x1 convolution to flatten the output and produce one-dimentional forecast (Figure 2f). For a detailed description of the adapted Wavenet model refer to Borovykh et al. (2017).

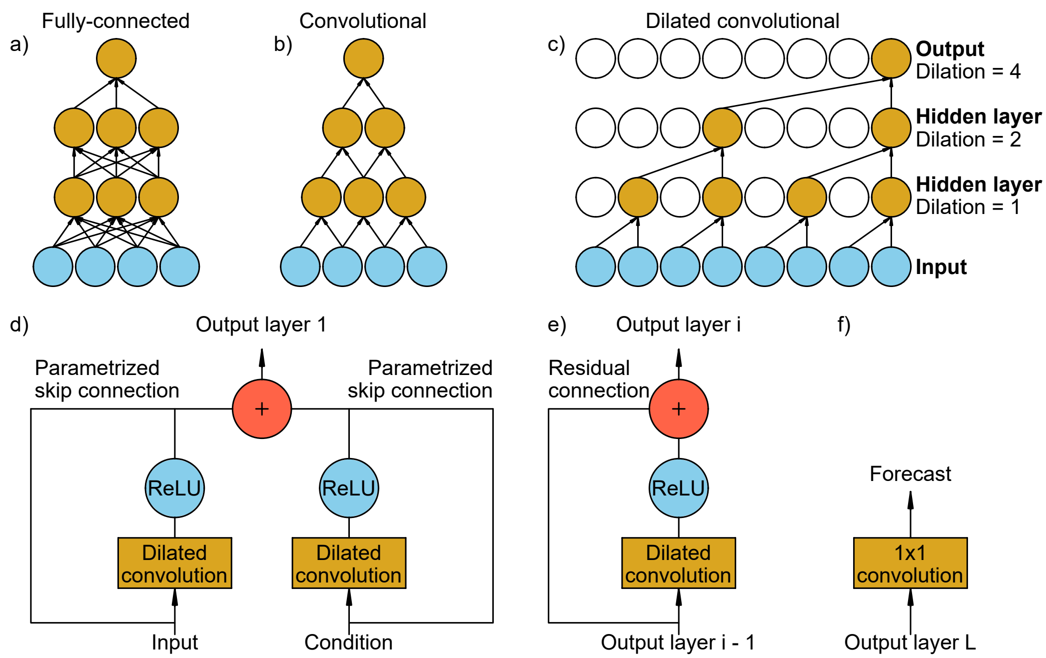

Figure 2. Architecture of dilated Convolutional Neural Networks and the structure of the adapted Wavenet model: a) example of a fully-connected network where all nodes of the output layer are connected to all nodes of the input resulting in many trainable parameters; b) example of a convolutional network where each node is connected only to a local region of the input and the weights are shared, i.e., reducing number of trainable parameters compared to the fully-connected network and allowing more efficient training, but only learning local (i.e., short-term) dependencies); c) example of a dilated convolutional network where regular local convolutions as in panel b are substituted by dilations to allow output to be influenced by more nodes from the input and efficiently learn the connections even between far-apart data points, i.e., in case of time series this allows to account for short-term (i.e., local) and long-term structures in the data; d) network structure of the adapted Wavenet model: dilation of the input (i.e., autoregressive part) and covariate time series with ReLU activation function (non-linearity term) and parametrized skip connection at the first convolutional layer that allows the model learn to discard covariates if they do not improve the forecast; e) dilation at each subsequent convolutional layer i of the network; f) 1x1 convolution of the last convolutional layer L to reduce dimentionality and obtain one-dimentional forecast. Differently from fully-connected networks (panel a) where the data is passed through all layers sequentially, residual connections at each convolutional layer allow data to bypass some layers resulting in very efficient training (modified from Borovykh et al. (2017))

Table 1. Model set-ups and corresponding covariates to forecast severity of widespread floods at time step t

Model	Structure	# covariates
auto	$[severity_1, severity_{t-1}]^{a}$	0
season	$[severity_1, severity_{t-1}] + [sin_1, sin_{t-1}] + [cos_1, cos_{t-1}]^{b}$	2
precip	$[severity_1, severity_{t-1}] + [sin_1, sin_{t-1}] + [cos_1, cos_{t-1}] + [P_1, P_{t-1}]^{c}$	3
climate	$[severity_1, severity_{t-1}] + [sin_1, sin_{t-1}] + [cos_1, cos_{t-1}] + [Clim_1, Clim_{t-1}]^{d}$	12
full	$[severity_1, severity_{t-1}] + [sin_1, sin_{t-1}] + [cos_1, cos_{t-1}] + [P_1, P_{t-1}] + [Clim_1, Clim_{t-1}]$	13

[a] Forecast of flood severity for time step t using only past observations of flood severity index from time step 1 to $t-1$ (i.e., autoregressive part)

[b] Sine and cosine of the month of the year as a proxy of seasonality

[c] Mean monthly country-averaged or region-averaged observed precipitation

[d] Ten different climatic indexes are considered simultaneously: PNA, SOI, NP, AAO, AO, NAO, EA, GBI, SCA, MOI

3.2 Setup of the adapted Wavenet model

We use a dCNN network with four dilation layers, which means that 32 input time steps (i.e., 2^5 months) are used to produce a single output time step (i.e., forecast) (Figure 2c). This choice leverages the length of the available time series and the number of trainable weights. The total length of monthly time series (756 months, i.e., 63 years from January 1951 to December 2013) was split into training (32+640 months, i.e., ca. 2.7+53.3 years) and test samples (32+52 months, i.e., 2.7+4.3 years). The test sample includes one of the most severe floods (June 2013) recorded in instrumental observation period in Germany that affected the Elbe River and the Danube River catchments (Blöschl et al., 2013) (the North-Eastern and the Southern regions respectively). We use time-series-split cross-validation that resembles K-fold cross-validation for ordered time series and is designed specifically for time series forecasting (i.e., validation split never preceeds training splits). In our case we use 7 time series splits (i.e., folds) for the training set. We use adaptive model estimation (Adam) optimizer (Kingma & Ba, 2014) with the initial learning rate of 0.005 to train the model and mean absolute error as a measure of model performance. The maximum number of epochs was set to 150. Additionally, we used an early stopping strategy to avoid overfitting.

To efficiently analyze the value of different covariates used in this study we examine models of different complexity that correspond to a different number of covariates used for conditioning the forecast of widespread flood severity. The simplest model (auto model) does not use any covariates and makes the forecast purely based on the structure of the patterns identified from the past time series of flood severity index (Table 1). The seasonal model uses two covariates, namely the sine and cosine of the month of the year to allow the model learning of the seasonality dynamics (Jiang et al., 2022). The precipitation model (precip model) adds the mean monthly precipitation time series, i.e., it can use observed precipitation from previous months for the forecast (Table 1). Essentially, it only uses antecedent precipitation since the model structure is causal (i.e., only past data is used for generating the forecast (Y. Chen et al., 2019)). The climate model does not use precipitation as a covariate (Table 1), but instead includes ten different climatic indices listed in the Section 2.2. Again, similar to precipitation only climatic indexes from previous months (Table 1) are used for the forecast given the causal structure of the model. The full model uses all above mentioned information (i.e., total of 13 covariates, Table 1).

To account for the stochastic nature of the training of the deep learning models we initialize training of each of the above mentioned models 10 times independently of each other (Jiang et al., 2022). Therefore, we obtain an ensemble of 10 realizations for training and also for test period in forecasting mode for each examined model complexity. To avoid overfitting of the model during training we use L1 type of regularization that reduces the weights of the less influential covariates to zero (i.e., it only selects covariates that are essential for the forecasting of flood severity). We identify an optimal regularization parameter value for each model complexity by systematically varying L1 parameter value and examining corresponding model performance (Figure S13). Too low a value of L1 parameter results in model overfitting (i.e., higher errors for test samples compared to train samples in cross-validation), while too high values of L1 parameter results in a very strict penalty and can prevent model from learning the variability beyond the long-term mean and seasonality (Figure S13).

Although from the theoretical point of view dCNN models are specifically designed to effectively learn patterns from short time series (Borovykh et al., 2017), the length of the input data available for training can still affect model performance. To evaluate the effect of the length of time series available for training on model performance we perform an experiment by systematically reducing the length of the training series, while keeping the length of the test series unchanged. In all examined cases we compare the performance with the baseline (i.e., mean monthly flood severity index derived from the whole observed series) to identify minimum requirements for the duration of training data for flood forecasting using dCNN models.

3.3 Explainable machine learning

Recently, explainable machine learning methods (XAI) (Buhrmester et al., 2021) have been shown to be useful for understanding the structure of the otherwise black-box deep learning models. Interpretability and explainability of deep learning models rests on the ability of the post hoc explanation of model prediction strategy by understanding the importance of the features used for predictions (Mamalakis et al., 2022b) and by evaluating the physical plausibility of the identified importances (Ebert-Uphoff & Hilburn, 2020). In our case, this means evaluating the role of different climatic indices and antecedent precipitation for the forecasting of widespread flood severity using a model dependent feature attribution method (i.e., Shapley values, see Section 3.3.1) and comparing the results with the model-independent wavelet coherence analysis between these variables (see Section 3.3.2), as well as analyzing the results on the importance of different features obtained from this study with the reports from previous studies (see Section 5).

3.3.1 Feature attribution: Shapley value

To evaluate the usefulness of different covariates used in this study (i.e., large-scale climatic indices, antecedent precipitation) we used the SHAP method (Lundberg et al., 2017) implemented in DeepExplainer library that approximates Shapley values (Shapley, 1952). Shapley values quantify the average marginal contribution of a feature value (e.g., antecedent precipitation) across all possible coalitions (i.e., all possible combinations of the features). The Shapley value of each feature shows how the forecast will change when this feature is added to the set of all other considered features. We implemented SHAP method in a global setting, i.e., by computing Shapley values not for a single forecast (i.e., month or event), but for the full sample to understand the overall effect of each covariate on model decisions (Buhrmester et al., 2021).

3.3.2 Wavelet transformation and wavelet coherence

Given the inherent similarity of the dCNN and wavelet transform (Aussem & Murtagh, 1997) we compare the results of the model-dependent feature attribution method described in the previous Section with the model-independent continuous wavelet transform and biwavelet analysis of the flood severity index and the covariates used in this study. Wavelet analysis is an effective tool to transform 1D-time series into 2D-time-frequency spectrum. We performed wavelet analysis using a continuous wavelet transform with the Morlet wavelet function (Torrence & Compo, 1998) of the monthly time series of climatic indices (Figure S2-S3), as well as the flood severity index (Figure S1) and precipitation (Figure S4) for the whole country and for the three hydroclimatic regions individually. The significance level of the wavelet spectrum is computed in relation to red noise (Torrence & Compo, 1998).

Biwavelet analysis is performed using Wavelet Coherence Transform (R package 'biwavelet') to identify the relation between two time series at different scales. In this case the coherence is quantified by a cross-wavelet power (the measure of the coherence in a time frequency domain) that resembles standard correlation coefficient (Torrence & Compo, 1998).

4 Results

4.1 Forecasting performance in Germany and its three hydroclimatic regions

All investigated models reduce the mean absolute error (MAE) of the forecast when compared to the baseline (i.e., mean monthly observed flood severity index) (Figure 3). However, more complex models that include climatic indices (i.e., climate and full model) are able to predict the widespread flood severity index more accurately when compared to the models with fewer covariates (R^2 increases from 0.04 to 0.32, Figure 4). Particularly, more complex models reduce the phase shift that is apparent for simpler models (Figure 4), although there is still a considerable underestimation of flood severity, especially for the most extreme events (Figure S14). The forecasting skill of the best model (i.e., climate or full model depending on the hydroclimatic region) varies across different regions when compared to the corresponding baseline: the models show higher skill in the North-Western and the North-Eastern regions (Figure 3), indicating that the predictability of widespread flooding in these regions from climate indices and antecedent precipitation might be higher.

4.2 The role of antecedent precipitation for forecast performance

There is a clear reduction of mean absolute error with increasing model complexity when forecasting flood severity index in Germany, although it is worth to mention that the full model is associated with the highest level of performance variability as may be expected given the increase in the number of effective parameters (Figure 3). When comparing the performance of the climate and full models we observe considerable regional differences. As for the country-wide forecast, in the North-Eastern region the full model outperforms the climate model that does not account for antecedent precipitation. In the Southern and North-Western region the full model shows much higher uncertainty (i.e., high variability in the performance among different realizations) compared to the climate model, with considerably better median performance of the climate model in the Southern region (Figure 3). Variable regional performance of the model that accounts for antecedent wetness conditions (i.e., full model, Table 1) confirms a variable role of antecedent wetness for generation of widespread flooding in different German regions suggested in R. Merz et al. (2020) who showed that especially in the drier North-Eastern parts of Germany antecedent soil moisture is an even more important feature for generation of annual maximum floods than the volume of precipitation events that induces the corresponding flood.

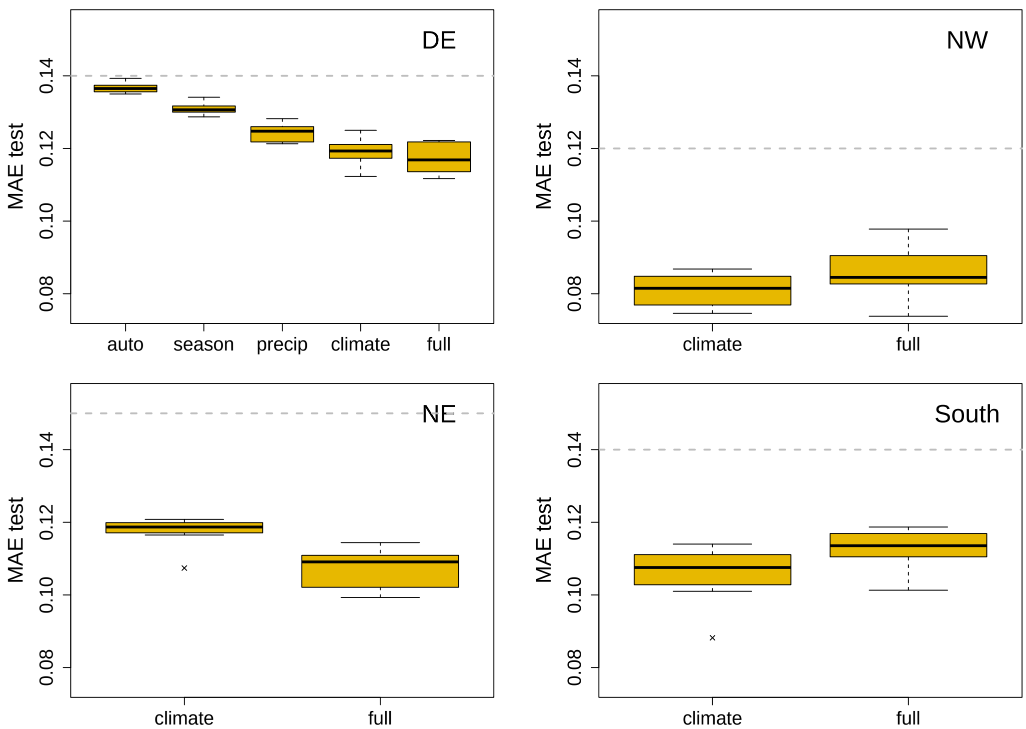

Figure 3. Mean absolute error (MAE) for the forecast of monthly widespread flood severity in Germany (DE) and its three hydroclimatic regions (NW: North-Western region; NE: North-Eastern region; South: Southern region) for the test period (August 2009-December 2013) using models of different complexity (ensemble of 10 realizations for each case) (see Table 1). Grey dashed line indicates MAE for the baseline case (i.e., mean monthly index of flood severity for the whole study period from 1951 to 2013). The tendency in performance improvement of the auto, season and precip models in Germany and individual regions are very similar, therefore, only the performance of climate and full models are shown for the three hydroclimatic regions

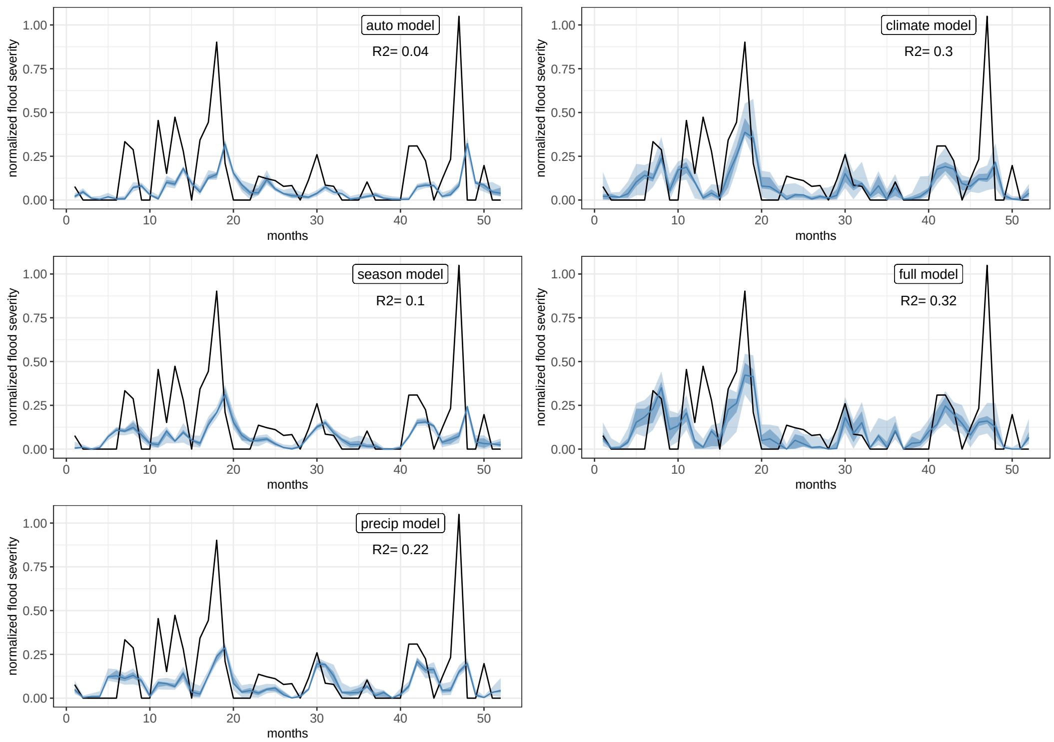

Figure 4. One month ahead forecast of monthly severity of widespread flooding in Germany for test period (August 2009-December 2013) using models of increasing complexity (see Table 1). The black solid line indicates observed flood severity index. Blue solid line represent the mean of the ensemble of 10 realizations for each model. Light shaded area shows the spread of the ensemble simulations with darker color indicating the spread between 25th and 75th percentiles.

4.3 Temporal variability of performance: forecasting floods generated by different processes

We also detect considerable differences in the forecast skill in time (Figure 4). For example, the forecast for summer 2013 when one of the most extreme flood events in Germany occurred was very poor for all models (Figure 4). On the other hand, the forecast of monthly severity corresponding to the winter-spring flood of 2011 was fairly accurate (Figure 4), especially in individual regions (Figure S15). This indicates that there might be considerable differences in the predictability of floods in the examined sample.

As floods can be generated by different atmospheric and hydrological processes (Hirschboeck, 1987; Tarasova et al., 2019) and might be characterized by considerably different contribution of different ingredients (e.g., volume and intensity of inducing precipitation events, wetness states) (R. Merz et al., 2020), their intrinsic predictability from climatic indices and antecedent precipitation might indeed be different. Therefore, we compare the ability of the full model to forecast events that correspond to different event types according to Tarasova, Basso, Wendi, et al. (2020). The results show that during the test phase there is on average higher underestimation (in terms of percent bias) of rainfall-generated events compared to the events that are generated by mixture of rainfall and snowmelt (Rain.Snow) (Figure 5d), although the variability in the performance for individual events is rather high. Despite their rarity only rainfall intensity-dominated events (Rain.Dry.Intensity, Figure 5b) are associated with smaller bias than mixtures of rainfall and snowmelt in the cross-validation phase (Figure 5c). However, these events also correspond to the lowest severity (in terms of the number of affected gauges and the magnitude of the events) (Figure 5a), while Rain.Snow events together with the Rain.Wet.Volume are associated with the highest severity. Interestingly, although Rain.Wet.Intensity and Rain.Dry.Volume that usually occur in autumn and summer (Tarasova, Basso, Wendi, et al., 2020) are associated with smaller severity, they are strongly underestimated (Figure 5). Therefore, it seems that winter/spring processes can be better predicted using monthly climatic indices than summer processes when convection might contribute considerably to the generation of rainfall (Conticello et al., 2020).

4.4 Model-independent regional wavelet coherence of the covariates and flood severity index

To increase interpretability of modeling results, first we analyze wavelet coherence between model covariates and flood severity index that is independent from model performance. As expected, we observe a very strong coherence between precipitation and flood severity index in Germany, especially at longer time scales (up to several years, Figure 6) that corresponds well with the previously reported dependence between the occurrences of multi-year flood-rich and flood-poor periods and antecedent wetness conditions in Germany (B. Merz et al., 2018). Interestingly, at the shorter time scales the coherence is rather intermittent. We also detect considerable differences in coherence among different regions (Figure 6). Particularly in the Southern region the coherence with precipitation is lower than in the rest of the country, which corresponds well with inferior performance of the full model in this region (Figure 3).

Among climatic indices that reflect global-scale climatic patterns and are related with the El Niño Southern Oscillation, there is particularly persistent coherence between PNA and NP and flood severity index in Germany (Figure 7, Figure S5). A very similar coherence pattern is observed in the North-Western region (Figure 7, Figure S6), while in the North-Eastern the coherence is very intermittent indicating weaker effect of these global patterns on the occurrence of widespread flooding in the region (Figure 7, Figure S7). Generally, there is also a very weak coherence between flooding in the North-Eastern region and all considered global patterns compared to the North-Western region (Figure S6 and S7). Interestingly, PNA seems to be very coherent with flood severity in

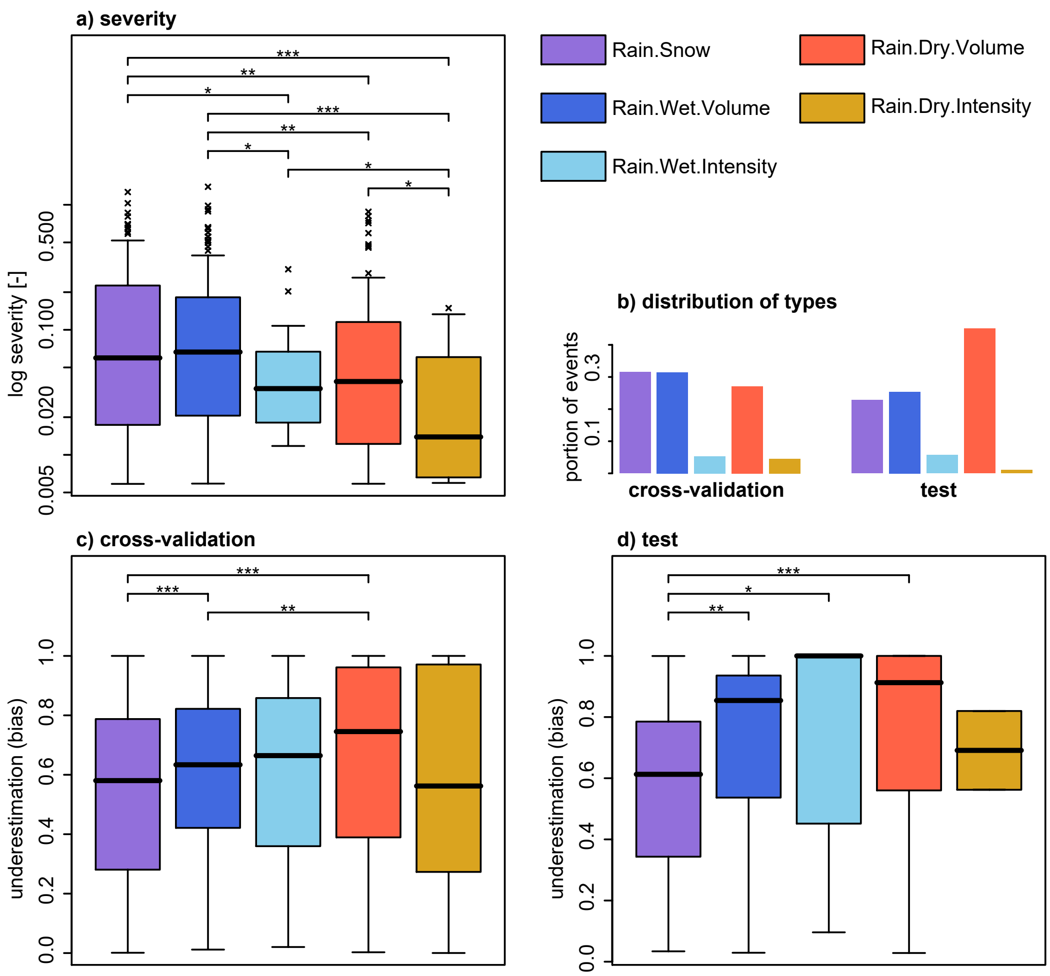

Figure 5. Severity of flood events generated by different processes in Germany (a) and underestimation (expressed as $bias = (severity_{obs} - severity_{mod})/severity_{obs}$) of their corresponding magnitudes during: c) cross-validation and d) test periods. Forecast is based on the ensemble of 10 full models that consider all covariates. Significance between event types is evaluated using pairwise Kruskal-Wallis test with false discovery rate correction. Significance levels: *** for p<0.01; ** for p<0.05; * for p<0.1. Panel b shows the proportions of events generated by different processes during cross-validation and test periods. Panel a is shown in log space. Panels b, c and d are shown in linear space.

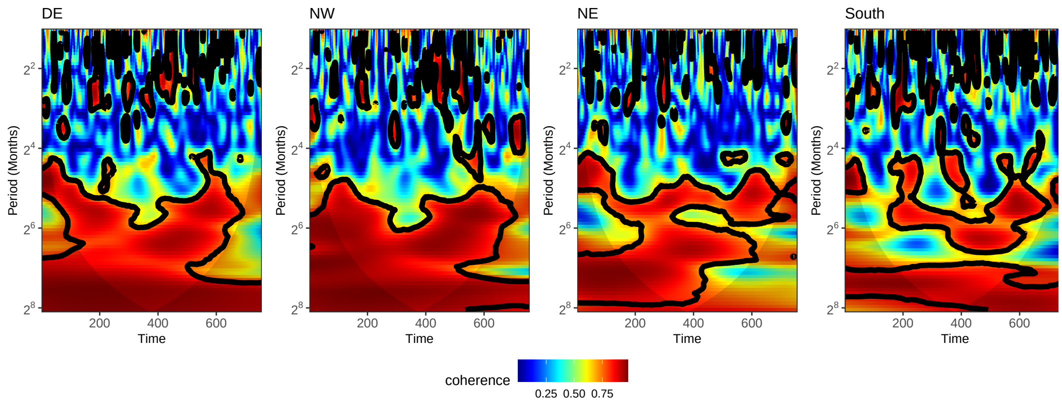

Figure 6. Wavelet coherence (WC) between monthly flood severity index and antecedent precipitation in Germany (DE) and its three hydroclimatic regions (NW: North-Western region; NE: North-Eastern region; South: Southern region). x axis is a time axis in months from the beginning to the end of the time series. y axis indicates periodicity or scale at which two time series are coherent at a given moment in time of the x axis. Black outline highlights regions of significant coherence. Red colors correspond to higher coherence values, blue colors correspond to lower coherence values. The shaded part outlines the cone of influence. Wavelet spectrum of the corresponding monthly flood severity index and precipitation series can be found in Figure S1 and Figure S4.

the Southern region especially at multi-annual time scales and this relation is even more persistent than in the North-Western region (Figure 7, Figure S8).

Previous study indicated the importance of the climatic patterns associated with the activity Atlantic Ocean, particularly of the Baltic Sea, and with the occurrence of blocking situations for precipitation and flooding in the Northern Europe (Zanardo et al., 2019; Comas-Bru & Mcdermott, 2014). Indeed, our results also indicate coherence between widespread flooding and SCA in all regions (Figure 7). There is also considerable coherence with NAO at annual and multi-annual scales in the North-Western and North-Eastern regions and with EA at longer time scales (Figure 7). Instead, in the Southern region the flooding is coherent with AO at longer time scales (Figure S8).

The coherence between widespread flooding and MOI seems to be rather similar between North-Western and North-Eastern regions where the dependence is strong at longer time scales. Differently, in the Southern region, the coherence is stronger at shorter time scales (Figure 7). Generally, wavelet coherence analysis between climatic indices and widespread flood severity shows clear differences in hydroclimatic controls across three German regions.

4.5 Model-dependent feature attribution using Shapley values

Since climate model showed a comparable or even better performance than the full model in the North-Western and Southern regions we performed feature attribution using Shapley values for these model complexity versions. We compare the differences between the feature rankings of two model versions to provide insights on the importance of different covariates and on potential of substituting antecedent precipitation solely by the information available from climatic indices in different hydroclimatic regions of Germany (Figure S9-S12).

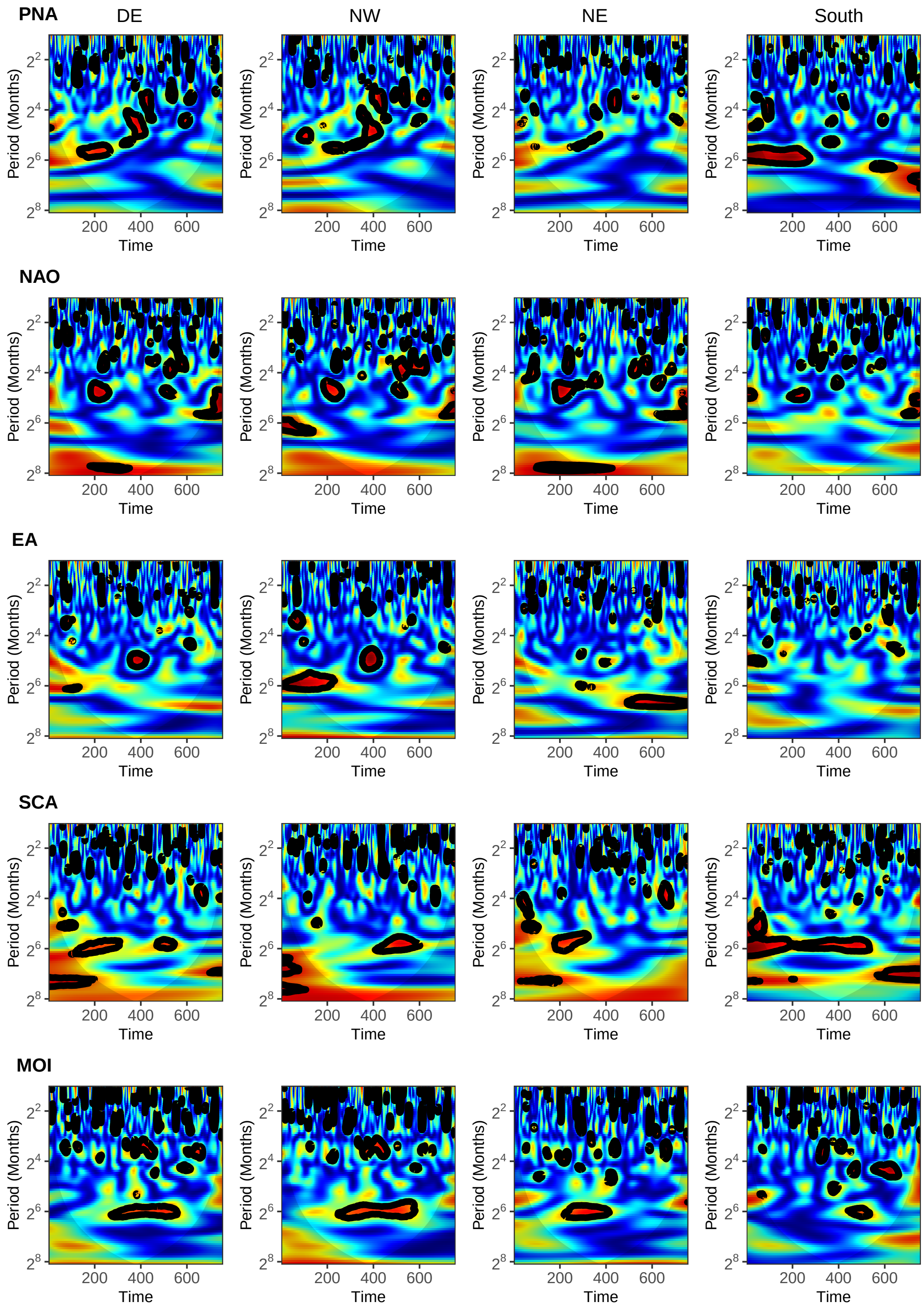

Figure 7. Wavelet coherence (WC) between monthly flood severity index and large scale climatic indices in Germany (DE) and its three hydroclimatic regions (NW: North-Western region; NE: North-Eastern region; South: Southern region): PNA - Pacific North American Pattern; NAO - North Atlantic Oscillation; EA - East Atlantic; SCA - Scandinavian Pattern; MOI - Mediterranean Oscillation Index. x axis is a time axis in months from the beginning to the end of the time series. y axis indicates periodicity or scale at which two time series are coherent at a given moment in time of the x axis. Black outline highlights regions of significant coherence. Red colors correspond to higher coherence values, blue colors correspond to lower coherence values (see colorbar in Figure 6). The shaded part outlines the cone of influence. Wavelet spectrum of the corresponding monthly flood severity and climate indices can be found in Figures S1-S3.

Regardless of model complexity, sine and cosine of the month of the year corresponds to high Shapley values in all regions (Figure 8), as widespread flooding is associated with very pronounced seasonality, especially in the Northern regions (Figure S1). In the Southern region where floods occur throughout the year (Beurton & Thieken, 2009) and the seasonality of widespread flooding is not as pronounced (Figure S1), sine and cosine correspondingly are ranked lower (Figure 8).

4.5.1 Germany-wide model

For the forecasting of the Germany-wide flood severity index the full model that includes antecedent precipitation performed better than climate model. Feature attribution shows that apart from antecedent precipitation and seasonality, global patterns, such as PNA and AAO, patterns associated with the Baltic Sea (EA and SCA) and with the Mediterranean Sea (MOI) are found important for the model performance (Figure 8). This corresponds well with the wavelet coherence analysis (Figure 7) that indicates a coherence between these indices and the flood severity index at different time scales. Interestingly, despite a clear coherence between Germany-wide flood severity indices and NP pattern (Figure S5), it was not identified as an important predictor by the Wavenet model (Figure 8). This might be explained by a strong relation between NP and other global atmospheric patterns related to the El Niño Southern Oscillation (Park et al., 2013). There is also a strong relation to the GBI index at annual scale for Germany-wide flood severity (Figure S5), although GBI does not seem to be an important covariate for model predictions (Figure 8). A very pronounced seasonality of GBI (Figure S3) might be already accounted for by sine and cosine of month of the year that are used as support covariates to improve seasonality learning (Jiang et al., 2022). Interestingly, when the precipitation are not considered (i.e., in the case of climate model), the importance of the NP and GBI indices that have strong coherence with precipitation at annual scale (Figure S9) increases (Figure S16).

4.5.2 North-Western region

The climate model (i.e., the model without the antecedent precipitation) performs slightly better than the full model in the North-Western region (Figure 3), indicating that the information on antecedent precipitation can be efficiently substituted by monthly variability of other covariates in this region. In fact, although antecedent precipitation is ranked first among covariates of the full model, feature attribution of the climate model shows that SCA index gains more importance in the absence of precipitation (Figure 8), as they are both strongly related to each other in this region according to the model independent wavelet coherence analysis (Figure S10). Also the importance of the dynamics of the historical flood severity index itself increases for climate model compared to the full model of the North-Western region.

Apart from antecedent precipitation and seasonality, global patterns that are particularly related to the El Niño Northern Oscillation (e.g., PNA, NP, AAO and SOI), as well as the East Atlantic pattern that indicates the strong effect of the Baltic Sea, are ranked high for model performance (Figure 8). Interestingly, according to the wavelet coherence analysis there is a strong relation between PNA and SOI patterns and flood severity index particularly at medium time scale (annual to bi-annual) (Figure 7, Figure S6), at which the relation between the flood severity index and precipitation is the lowest (Figure 6). AAO is related to flood severity index at much shorter scales, while the coherence with the East Atlantic pattern is instead detected at longer (ca. 3 to 5 years) time scales (Figure 7, Figure S6).

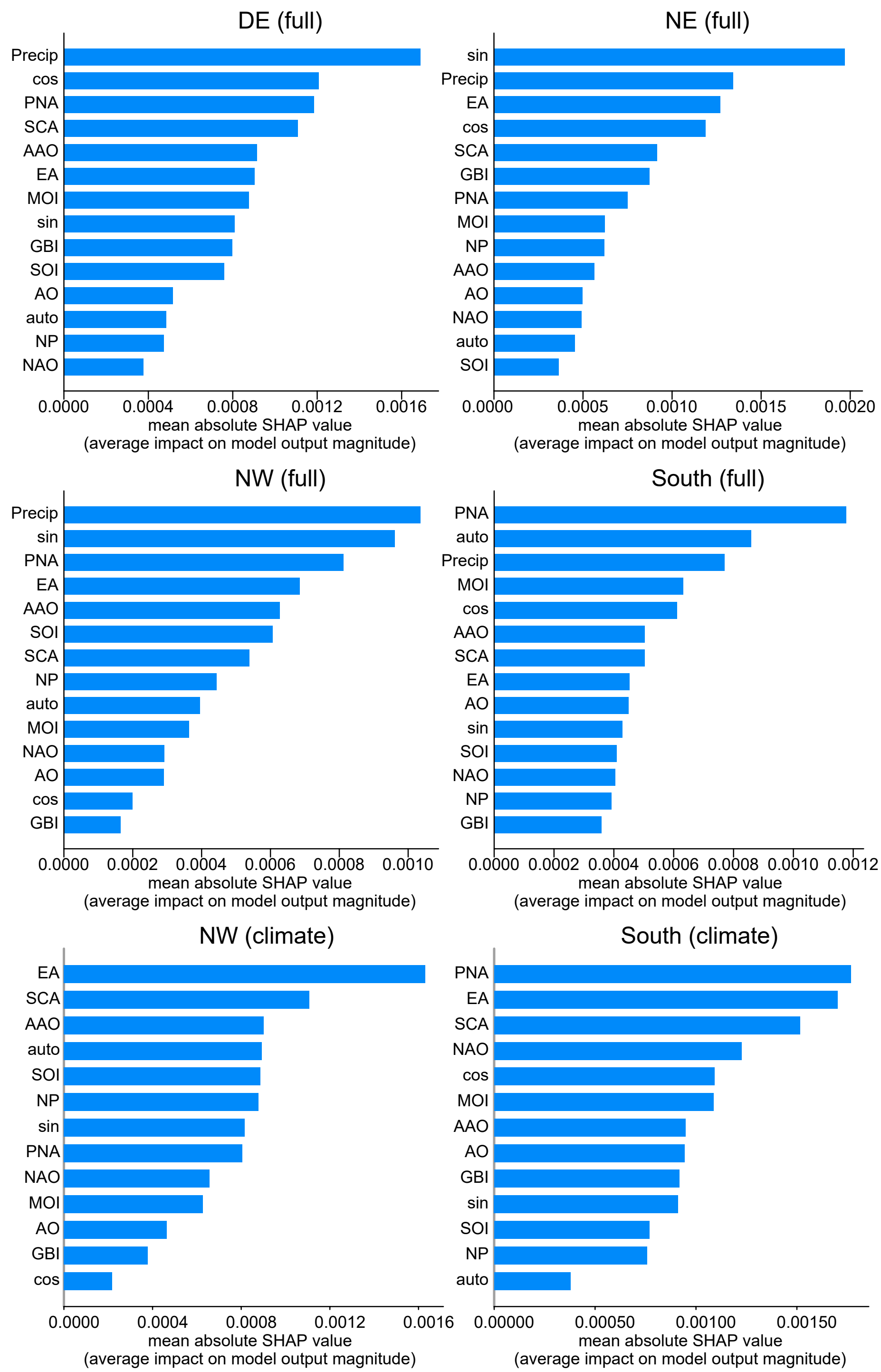

Figure 8. Shapley values of the covariates of the full model (for Germany and all regions) and for climate model (for the North-Western and Southern region where climate model performed comparable or better than the full model) computed for the whole time series from 1951 to 2013 (i.e., global settings) ordered by decreasing variable importance: auto - the historical time series of flood severity; sin - sine of the month of the year; cos - cosine of the month of the year; Precip - antecedent mean monthly precipitation; NP - North Pacific Patterns; PNA - Pacific North American Pattern; SOI - Southern Oscillation Index; AAO - Antarctic Oscillation; AO - Arctic Oscillation; NAO - North Atlantic Oscillation; GBI - Greenwich Blocking Index; SCA - Scandinavian Pattern; EA - East Atlantic; MOI - Mediterranean Oscillation Index

4.5.3 North-Eastern region

In contrast to the North-Western region, feature attribution analysis does not indicate any major effect of global patterns on the occurrence of widespread flooding in the North-Eastern region, where apart from very strong effect of seasonality, only antecedent precipitation and East Atlantic pattern contribute considerably to the performance of the full model (Figure 8). The strong effect of the latter on the flood severity index especially at longer time scales is also apparent from the corresponding wavelet coherence analysis (Figure 7). Interestingly, there is no difference in feature importance of climatic indices between climate and full model of the North-Eastern region (Figure S16), but the full model performs considerably better (Figure 3).

4.5.4 Southern region

In the Southern region, the PNA pattern has the highest contribution to model performance for both climate and full models (Figure 8), and is strongly related to the flood severity index at longer time scales (Figure 7). At shorter time scales the effect of the MOI pattern is apparent that is also associated with high Shapley values (Figure 8) and might indicate the importance of the Mediterranean Sea for the occurrence of flooding in this region (Figure 7). In this region the climate model performs considerably better than the full model and the feature attribution indicates that the importance of SCA, EA and NAO indices increases in the absence of precipitation covariate (Figure 8). According to wavelet coherence analysis precipitation in the Southern region show strong dependencies with SCA at longer time scales and with EA and NAO at shorter time scales (Figure S12), indicating that the the information on antecedent precipitation can be effectively extracted from the climatic indices in this region.

5 Discussion

5.1 Adapted Wavenet conditioned by climatic indices for monthly flood forecasting

Dilated convolutional neural networks provide a powerful tool for time series forecasting when only limited length of time series is available for training (Borovykh et al., 2017), especially compared to the recurrent neural networks (Y. Chen et al., 2019). The consistent improvement compared with the baseline (i.e., mean monthly values of widespread flood severity index) for all model complexities investigated in our study shows the potential of dilated convolutional models for flood forecasting, although even the most complex model tends to underestimate the most extreme events (Figure S14). Moreover, unfortunately, even for the adapted Wavenet there is a clear dependence between model performance and the length of time series available for training. When less than 40 years of the data is available the model is not able to outperform the baseline based on mean monthly flood severity for the whole study period (Figure S13). This indicates that extending forecast lead times to seasonal and longer scales using the Wavenet model alone does not seem feasible yet. However, using high-fidelity seasonal forecasts of climate indices produced by the global climate models (Feng et al., 2021) as the input for the Wavenet model might be a promising avenue to extend lead times also for flood forecasting.

Apart from that, wavelet coherence analysis clearly shows that there are strong dependences between severity of widespread flooding and climatic indices even at long time scales (i.e., longer than 5 years) (Figure 7, Figure S5-S8). Rather shallow dilation depth (i.e., four layers) selected in our study due to the limited time series available for training might have impacted the ability of the adapted Wavenet model to account for such dependencies.

Despite this, a good agreement between the model-independent wavelet coherence analysis and feature attribution analysis using Shapley values indicate that the model was able to identify the primary climatic drivers and might become a valuable forecasting tool especially as the length of flood observations increases.

Even with good agreement between wavelet coherence analysis and feature attribution method, it might be difficult to pre-select the features beforehand using wavelet coherence analysis alone due to high inter-relation between the indices (Figure S9-S12). Despite the comprehensiveness of monthly climatic indices they do not capture the full picture of large scale climatic patterns that might be important for spatially co-occurring floods. An alternative design of the Wavenet model including daily indices that resolve shorter phenomena such as synoptic waves (Lu et al., 2013) might provide an opportunity for a more detailed near-time event-based widespread flood forecasting. Moreover, a more comprehensive and flexible patterns (e.g., based on integrated vapor transport) (Nakamura et al., 2013; Lima et al., 2017; Conticello et al., 2020), in combination with the convolutional neural networks might also be a promising avenue for a more accurate event-based forecasting of spatially co-occurring floods.

5.2 Predictability of floods generated by different processes

The predictability seems to vary depending on generation processes of individual widespread flooding events (Figure 5). Rain-on-snow events (i.e., mixture of rainfall and snowmelt) are associated with the highest predictability despite their severity. In Germany these events are associated with westerly and north-westerly circulation patterns (Nied et al., 2014), and are often the cause of the widespread transboundary floods particularly during negative NAO and SCA phases (Krug et al., 2020). Our results also indicate that SCA is one of the most important covariates for the forecasting of widespread flooding in Germany (Figure 8). Moreover, Conticello et al. (2020) notes that high flow events in the North-Western part of Germany, where floods almost exclusively occur in winter and are usually generated by rain-on-snow processes (Tarasova, Basso, & Merz, 2020), are associated with higher predictability from large scale circulation patterns than the events in the Eastern and Southern parts of the country.

In contrast, volume-dominated floods generated by rainfall during dry conditions often occur in late summer and the beginning of the autumn (Tarasova, Basso, & Merz, 2020) and have limited predictability (Figure 5). On one hand it can be limited by the complex patterns of soil moisture that modulate local occurrence of floods and hence directly affect the spatial extent of flooding (Nied et al., 2017), while on the other hand the meteorological phenomenon that are not well captured by the climatic indices, such as local convective processes (R. Merz & Blöschl, 2003; Kemter et al., 2020) and the activity of the Vb cyclones that is associated with heavy precipitation and recent summer floods in Central Europe and particularly in Germany (Hofstätter et al., 2018; Krug et al., 2021), might play a more important role for the generation of these events.

Changing climatic conditions and changing flood generation processes indicate that the predictability of widespread flooding might decrease as the number of dry rainfall generated events increases in Central, Eastern and Southern Germany (Tarasova et al., 2023; Winter et al., 2022), which might deteriorate the efficiency of early warning systems in these regions and urges the development of the methods that might integrate the atmospheric phenomena and soil moisture patterns relevant for their forecasting.

5.3 Regionally variable importance of climatic indices and antecedent wetness conditions for forecasting widespread floods

Feature attribution analysis implemented in our study shows clear differences in hydroclimatic drivers of widespread flooding across different regions in Germany. Particularly, there is apparent difference in the value of global patterns associated with El Niño Southern Oscillation between Western and Eastern parts of Germany. While in the North-West and in the Southern parts a global Pacific North American pattern is a very important feature for model predictions (Figure 8), it has only secondary effect in the North-Eastern part. Although there is few evidence about the effect of El Niño in Europe (Brönnimann, 2007), there is a clear coherence between widespread floods in the North-West of Germany and PNA patterns (Figure 7) that is also coherent with the findings of Conticello et al. (2020), who found that high discharge events in the Western part of Germany are associated with intensive moisture transport from Tropics that is governed by global patterns, such as PNA.

Feature attribution analysis indicates that AAO pattern might be important for forecasting widespread flood severity in the North-Western region as well. However, although AAO is able to modulate the effect of El Niño Southern Oscillation on the extratropical circulations (L'heureux & Thompson, 2006), the exact process chain of its effect on flooding in the Northern Hemisphere is not understood.

In the drier North-Eastern region we see a very pronounced role of seasonality and antecedent wetness conditions (Figure 8). This is in line with R. Merz et al. (2020), who noted that in this region the antecedent moisture conditions is even more important for the emergence of local floods than the volume of inducing precipitation events. Also Nied et al. (2017) found that in this region the spatial extent of moderate floods (i.e., 2-year return periods) are modulated more by the specifics of the soil moisture patterns than by weather patterns. This was later corroborated by B. Merz et al. (2018) highlighting the ultimate role of antecedent wetness conditions in this region for the occurrence of spatially coherent flood-rich periods. It is also important to note the effect of the East Atlantic pattern for widespread flooding in this region (Figure 8) indicating the primary role of the Baltic Sea for moisture transport and generation of flood-producing precipitation in this region.

Interestingly, the results of feature attribution indicates relatively low importance of the North Atlantic Oscillation on model predictions, that is also confirmed by the model-independent wavelet coherence analysis that shows rather low coherence between the severity of widespread flooding and NAO (Figure 7-8). Previous studies have shown strong dependence between NAO and winter precipitation in Europe (Haylock & Goodess, 2004; Villarini et al., 2011). However, particularly for flood losses the effect of NAO is notably weaker in Germany compared to the Atlantic coast of Great Britain and Norway and the Mediterranean region (Zanardo et al., 2019).

In the Southern region in addition to the effect of global PNA pattern, the effect of the Mediterranean Sea in terms of the Mediterranean Oscillation Index also becomes apparent (Figure 8). Severe floods in this region are often associated with the activity of the Vb cyclones propagating north-eastward from the Mediterranean Sea to Central Europe (Blöschl et al., 2013; Hofstätter et al., 2018; B. Merz et al., 2018) that might be additionally intensified by the Mediterranean Sea through pre-moistening of the continental moisture (Krug et al., 2021). Therefore, adding more indices or patterns related to the activity of the Mediterranean Sea might improve the forecasting of widespread flooding, especially in the regions where summer floods occur.

It is worth to note that the relevance of the climatic indices discussed above is conditioned on the selected study period. The attribution results depend greatly on the chosen reference period (Mamalakis et al., 2022a) (i.e., in our case the period from 1951 to 2013) and the relevance of climatic indices might change for a different period of reference.

6 Conclusions

In this study we investigate the potential of the adapted Wavenet model to forecast the monthly severity of the widespread flooding in Germany and its three distinct hydroclimatic regions individually using various large-scale climatic indices and antecedent precipitation as covariates. We observe a consistent increase in model performance with increasing model complexity (i.e., increasing number of covariates), but there is a clear dependence between the length of the available time series for training and model skill. This dependence indicates that even the dilated structure of the Wavenet model is not able to reconcile the limited data availability and the number of trainable parameters, which might be also a limiting factor for adapting the Wavenet model for flood forecasting at longer (e.g., annual) time scales.

The results of the model-dependent feature attribution based on Shapley values are consistent with the model-independent wavelet coherence analysis indicating the reliability of trained Wavenet model for forecasting of widespread floods. Both methods highlight considerable differences in the large-scale hydroclimatic drivers of widespread flooding across different regions in Germany. While in the North-Western and the Southern regions the effect of the global patterns associated with El Niño Southern Oscillation (i.e., Pacific North American pattern) is apparent, the North Eastern region is affected by more local processes related to the Baltic Sea activity (East Atlantic pattern). In the Southern region, where summer floods are as frequent as winter events, we also detect the effect of the Mediterranean Sea (Mediterranean Oscillation Index).

Moreover, the feature attribution based on the trained Wavenet model indicates variable importance of antecedent wetness conditions across German regions. Apart from the Southern region antecedent precipitation is ranked as one of the most important covariates. Especially in drier North-Eastern region model performance increases considerably when antecedent precipitation is added as a covariate.

We also see considerable differences in the predictability of flood events generated by different processes. Despite their superior severity, events generated by the mixture of rainfall and snowmelt are associated with low biases, while the predictability of volume-dominated rainfall events generated during dry conditions is low. Increasing frequency of the latter in several regions might lead to the deterioration of the predictability of widespread flooding in Germany when only using large scale climatic indices for the forecast. Therefore, more efforts should focus on developing sub-seasonal to seasonal forecasting approaches based on convolutional neural networks conditioned on more flexible spatial climatic patterns that might be more relevant for flood generation in warmer and drier periods.

Open Research Section

Time series of SOI, NP, AAO, NAO, AO, PNA and GBI used in this study are available from https://psl.noaa.gov/data/climateindices/list/ (no registration required, individual download for each climatic index), MOI data (Algiers and Cairo) are available from https://crudata.uea.ac.uk/cru/data//moi/ (no registration required, individual download for each computational variant of the index) and SCA and EA data are available in Comas-Bru and Hernández (2018). Mean monthly precipitation per catchment were obtained from the daily interpolated precipitation dataset available in Rauthe et al. (2013). Severity index of spatial flooding is computed using streamflow events (available in Tarasova et al. (2018)) identified from daily streamflow observations available for download from Global Runoff Dataset (https://portal.grdc.bafg.de/applications/ no registration required, individual download for each streamflow gauge or selected region) and Bavarian Ministry for Environment (https://www.gkd.bayern.de/de/fluesse/abfluss no registration required, individual download for each streamflow gauge). The post-processed data used in this study: the severity index of widespread flooding, flood generation processes of each flood event and monthly precipitation for Germany and its three hydroclimatic regions, are available in Tarasova (2023).

Acknowledgments

We thank Munir Ahmad Nayak and one anonymous reviewer for insightful comments that helped to improve our original manuscript. The financial support of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in terms of the Research Unit FOR2416 'Space-time Dynamics of Extreme Floods (SPATE)' is gratefully acknowledged.

References

Aussem, A., & Murtagh, F. (1997). Combining Neural Network Forecasts on Wavelet-transformed Time Series. *Connection Science*, *9*(1), 113–122. doi: 10.1080/095400997116766

Barnston, A., & Livezey, R. E. (1987). Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Monthly Weather Review*, *115*.

Benhaddi, M., & Ouarzazi, J. (2021). Multivariate Time Series Forecasting with Dilated Residual Convolutional Neural Networks for Urban Air Quality Prediction. *Arabian Journal for Science and Engineering*, *46*(4), 3423–3442. doi: 10.1007/s13369-020-05109-x

Berghuijs, W. R., Allen, S. T., Harrigan, S., & Kirchner, J. W. (2019). Growing Spatial Scales of Synchronous River Flooding in Europe. *Geophysical Research Letters*, *46*(3), 1423–1428. doi: 10.1029/2018GL081883

Beurton, S., & Thieken, A. H. (2009). Seasonality of floods in Germany. *Hydrological Sciences Journal*, *54*(1), 62–76. doi: 10.1623/hysj.54.1.62

Blöschl, G., Nester, T., Komma, J., Parajka, J., & Perdigão, R. A. (2013). The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. *Hydrology and Earth System Sciences*, *17*(12), 5197–5212. doi: 10.5194/hess-17-5197-2013

Borovykh, A., Bohte, S., & Oosterlee, C. W. (2017). Conditional Time Series Forecasting with Convolutional Neural Networks. Retrieved from `http://arxiv.org/abs/1703.04691`

Brönnimann, S. (2007). *Impact of El Niño-Southern Oscillation on European climate* (Vol. 45) (No. 3). doi: 10.1029/2006RG000199

Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., & Clark, M. (2020). Spatial Dependence of Floods Shaped by Spatiotemporal Variations in Meteorological and Land-Surface Processes. *Geophysical Research Letters*, *47*(13), 1–13. doi: 10.1029/2020GL088000

Bueh, C., & Nakamura, H. (2007). Scandinavian pattern and its climatic impact. *Quarterly Journal of the Royal Meteorological Society*, *133*(629 B), 2117–2131. doi: 10.1002/qj.173

Buhrmester, V., Münch, D., & Arens, M. (2021). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *Machine Learning and Knowledge Extraction*, *3*(4), 966–989. doi: 10.3390/make3040048

Chen, S., Yu, B., Wu, R., Chen, W., & Song, L. (2021). The dominant North Pacific atmospheric circulation patterns and their relations to Pacific SSTs: historical simulations and future projections in the IPCC AR6 models. *Climate Dynamics*, *56*(3-4), 701–725. doi: 10.1007/s00382-020-05501-1

Chen, Y., Kang, Y., Chen, Y., & Wang, Z. (2019). Probabilistic Forecasting with Temporal Convolutional Neural Network. Retrieved from `http://arxiv.org/abs/1906.04397`

Comas-Bru, L., & Hernández, A. (2018). Reconciling North Atlantic climate modes:

Revised monthly indices for the East Atlantic and the Scandinavian patterns beyond the 20th century [Dataset]. *Earth System Science Data*, *10*(4), 2329–2344. doi: 10.5194/essd-10-2329-2018
Comas-Bru, L., & Mcdermott, F. (2014). Impacts of the EA and SCA patterns on the European twentieth century NAO-winter climate relationship. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 354–363. doi: 10.1002/qj.2158
Conticello, F. R., Cioffi, F., Lall, U., & Merz, B. (2020). Synchronization and Delay Between Circulation Patterns and High Streamflow Events in Germany. *Water Resources Research*, *56*(4), 1–16. doi: 10.1029/2019WR025598
Domeisen, D. I., Badin, G., & Koszalka, I. M. (2018). How predictable are the Arctic and North Atlantic Oscillations? Exploring the variability and predictability of the Northern Hemisphere. *Journal of Climate*, *31*(3), 997–1014. doi: 10.1175/JCLI-D-17-0226.1
Dünkeloh, A., & Jacobeit, J. (2003). Circulation dynamics of Mediterranean precipitation variability 1948-98. *International Journal of Climatology*, *23*(15), 1843–1866. doi: 10.1002/joc.973
Ebert-Uphoff, I., & Hilburn, K. (2020). Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society*, *101*(12), E2149–E2170. doi: 10.1175/BAMS-D-20-0097.1
Feng, P. N., Lin, H., Derome, J., & Merlis, T. M. (2021). Forecast skill of the NAO in the subseasonal-to-seasonal prediction models. *Journal of Climate*, *34*(12), 4757–4769. doi: 10.1175/JCLI-D-20-0430.1
Hanna, E., Cropper, T. E., Hall, R. J., & Cappelen, J. (2016). Greenland Blocking Index 1851–2015: a regional climate change signal. *International Journal of Climatology*, *36*(15), 4847–4861. doi: 10.1002/joc.4673
Hanna, E., Cropper, T. E., Jones, P. D., Scaife, A. A., & Allan, R. (2015). Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index. *International Journal of Climatology*, *35*(9), 2540–2554. doi: 10.1002/joc.4157
Harrison, D. E., & Larkin, N. K. (1998). El Nino-Southern Oscillation sea surface temperature and wind anomalies, 1946-1993. *Reviews of Geophysics*, *36*(3), 353–399. doi: 10.1029/98RG00715
Haylock, M. R., & Goodess, C. M. (2004). Interannual variability of European extreme winter rainfall and links with mean large-scale circulation. *International Journal of Climatology*, *24*(6), 759–776. doi: 10.1002/joc.1033
Hirschboeck, K. K. (1987). Hydroclimatically-defined mixed distributions in partial duration flood series. In *Hydrologic frequency modeling: Proceedings of the international symposium on flood frequency and risk analyses, 14-17 may 1986, louisiana state university, baton rouge, u.s.a.* (pp. 199–212). doi: 10.1007/978-94-009-3953-0_13
Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. doi: 10.1162/neco.1997.9.8.1735
Hofstätter, M., Lexer, A., Homann, M., & Blöschl, G. (2018). Large-scale heavy precipitation over central Europe and the role of atmospheric cyclone track types. *International Journal of Climatology*, *38*(December 2017), e497–e517. doi: 10.1002/joc.5386
Hsu, D. (2017). Multi-period Time Series Modeling with Sparsity via Bayesian Variational Inference. Retrieved from `http://arxiv.org/abs/1707.00666`
Jiang, S., Bevacqua, E., & Zscheischler, J. (2022). River flooding mechanisms and their changes in Europe revealed by explainable machine learning. *Hydrology and Earth System Sciences*, 6339–6359. doi: 10.5194/hess-26-6339-2022
Kemter, M., Merz, B., Marwan, N., Vorogushyn, S., & Blöschl, G. (2020). Joint

Trends in Flood Magnitudes and Spatial Extents Across Europe. *Geophysical Research Letters*, *47*(7), 1–8. doi: 10.1029/2020GL087464

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. Retrieved from `http://arxiv.org/abs/1412.6980`

Kreibich, H., Di Baldassarre, G., Vorogushyn, S., Aerts, J. C., Apel, H., Aronica, G. T., ... Merz, B. (2017). Adaptation to flood risk: Results of international paired flood event studies. *Earth's Future*, *5*(10), 953–965. doi: 10.1002/2017EF000606

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Krug, A., Pothapakula, P. K., Primo, C., & Ahrens, B. (2021). Heavy Vb-cyclone precipitation: a transfer entropy application showcase. *Meteorologische Zeitschrift*. doi: 10.1127/metz/2021/1071

Krug, A., Primo, C., Fischer, S., Schumann, A., & Ahrens, B. (2020). On the temporal variability of widespread rain-on-snow floods. *Meteorologische Zeitschrift*, *29*(2), 147–163. doi: 10.1127/metz/2020/0989

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, *49*(1), 360–379. doi: 10.1029/2012WR012195

Leathers, D. J., Yarnal, B., & Palecki, M. A. (1991). The Pacific/North American Teleconnection Pattern and United States Climate Part 1: Regional Temperature and Precipitation Associations. *Journal of Climate*.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning* (Vol. 521) (No. 7553). Nature Publishing Group. doi: 10.1038/nature14539

L'heureux, M. L., & Thompson, D. W. J. (2006). *Observed Relationships between the El Niño-Southern Oscillation and the Extratropical Zonal-Mean Circulation* (Tech. Rep.).

Li, X., Hu, Z.-Z., Liang, P., & Zhu, J. (2019). American Meteorological Society Contrastive Influence of ENSO and PNA on Variability and Predictability of North American Winter Precipitation. *Source: Journal of Climate*, *32*(19), 6271–6284. doi: 10.2307/26831709

Lima, C. H., Aghakouchak, A., & Lall, U. (2017). Classification of mechanisms, climatic context, areal scaling, and synchronization of floods: The hydroclimatology of floods in the Upper Paraná River basin, Brazil. *Earth System Dynamics*, *8*(4), 1071–1091. doi: 10.5194/esd-8-1071-2017

Liu, Y., Wang, L., Zhou, W., & Chen, W. (2014). Three Eurasian teleconnection patterns: Spatial structures, temporal variability, and associated winter climate anomalies. *Climate Dynamics*, *42*(11-12), 2817–2839. doi: 10.1007/s00382-014-2163-z

Lu, M., Lall, U., Schwartz, A., & Kwon, H. (2013). Precipitation predictability associated with tropical moisture exports and circulation patterns for a major flood in france in 1995. *Water Resources Research*, *49*, 6381-6392. doi: 10.1002/wrcr.20512

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., & Schellnhuber, H. J. (2014). *Very early warning of next El Niño* (Vol. 111) (No. 6). doi: 10.1073/pnas.1323058111

Lundberg, S. M., Allen, P. G., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *31st conference on neural information processing systems.*

Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022a). Carefully choose the baseline: Lessons learned from applying xai attribution methods for regression tasks in geoscience. *Artificial Intelligence for the Earth Systems*, *2*, 1-12. doi: 10.1175/aies-d-22-0058.1

Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022b). Investigating the Fidelity of Explainable Artificial Intelligence Methods for Applications of Convolutional Neural Networks in Geoscience. *Artificial Intelligence for the Earth Systems*, *1*(4). doi: 10.1175/AIES-D-22-0012.1

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C., Bates, P., ... Macdonald, E. (2021). Causes, impacts and patterns of disastrous river floods. *Nature Reviews Earth and Environment*, *2*(9), 592–609. doi: 10.1038/s43017-021-00195-3

Merz, B., Dung, N. V., Apel, H., Gerlitz, L., Schröter, K., Steirou, E., & Vorogushyn, S. (2018). Spatial coherence of flood-rich and flood-poor periods across Germany. *Journal of Hydrology*, *559*, 813–826. doi: 10.1016/j.jhydrol.2018.02.082

Merz, R., & Blöschl, G. (2003). A process typology of regional floods. *Water Resources Research*, *39*(12), 1–20. doi: 10.1029/2002WR001952

Merz, R., Tarasova, L., & Basso, S. (2020). The flood cooking book: ingredients and regional flavors of floods across Germany. *Environmental Research Letters*, *15*(11), 114024. doi: 10.1088/1748-9326/abb9dd

Nakamura, J., Lall, U., Kushnir, Y., Robertson, A. W., & Seager, R. (2013). Dynamical structure of extreme floods in the U.S. Midwest and the United Kingdom. *Journal of Hydrometeorology*, *14*(2), 485–504. doi: 10.1175/JHM-D-12-059.1

Nied, M., Pardowitz, T., Nissen, K., Ulbrich, U., Hundecha, Y., & Merz, B. (2014). On the relationship between hydro-meteorological patterns and flood types. *Journal of Hydrology*, *519*, 3249–3262. doi: 10.1016/j.jhydrol.2014.09.089

Nied, M., Schröter, K., Lüdtke, S., Nguyen, V. D., & Merz, B. (2017). What are the hydro-meteorological controls on flood characteristics? *Journal of Hydrology*, *545*, 310–326. doi: 10.1016/j.jhydrol.2016.12.003

Park, J. Y., Yeh, S. W., Kug, J. S., & Yoon, J. (2013). Favorable connections between seasonal footprinting mechanism and El Niño. *Climate Dynamics*, *40*(5-6), 1169–1181. doi: 10.1007/s00382-012-1477-y

Petrow, T., Merz, B., Lindenschmidt, K.-E., & Thieken, a. H. (2007). Aspects of seasonality and flood generating circulation patterns in a mountainous catchment in south-eastern Germany. *Hydrology and Earth System Sciences Discussions*, *4*(2), 589–625. doi: 10.5194/hessd-4-589-2007

Rasp, S., & Thuerey, N. (2020). Purely data-driven medium-range weather forecasting achieves comparable skill to physical models at similar resolution.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., & Gratzki, A. (2013). A Central European precipitation climatology - Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS)[Dataset]. *Meteorologische Zeitschrift*, *22*(3), 235–256. doi: 10.1127/0941-2948/2013/0436

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5), 1–25. doi: 10.1029/2008WR007327

Scherrer, S. C., Croci-Maspoli, M., Schwierz, C., & Appenzeller, C. (2006). Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *International Journal of Climatology*, *26*(2), 233–249. doi: 10.1002/joc.1250

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., ... Stadtler, S. (2021). *Can deep learning beat numerical weather prediction?* (Vol. 379) (No. 2194). Royal Society Publishing. doi: 10.1098/rsta.2020.0097

Shapley, L. S. (1952). A value for n-Person games. In *Kuhn, h. and tucker, a., eds., contributions to the theory of games ii, princeton university press, princeton* (p. 307-317). doi: 10.1515/9781400881970-018

Silvestri, G. E., & Vera, C. S. (2003). Antarctic Oscillation signal on precipitation anomalies over southeastern South America. *Geophysical Research Letters*, *30*(21). doi: 10.1029/2003GL018277

Slater, L. J., Villarini, G., & Bradley, A. A. (2019). Evaluation of the skill of North-American Multi-Model Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA. *Climate Dynamics*, *53*(12), 7381–7396. doi: 10.1007/s00382-016-3286-1

Tarasova, L. (2023). *Widespread flood severity index time series for germany [dataset].* doi: 10.5281/zenodo.7688006

Tarasova, L., Basso, S., & Merz, R. (2020). Transformation of Generation Processes From Small Runoff Events to Large Floods. *Geophysical Research Letters*, *47*(22). doi: 10.1029/2020GL090547

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., & Merz, R. (2020). A Process-Based Framework to Characterize and Classify Runoff Events: The Event Typology of Germany. *Water Resources Research*, *56*(5), 1–24. doi: 10.1029/2019WR026951

Tarasova, L., Basso, S., Zink, M., & Merz, R. (2018). Exploring Controls on Rainfall-Runoff Events: 1. Time Series-Based Event Separation and Temporal Dynamics of Event Runoff Response in Germany [Dataset]. *Water Resources Research*, *54*(10), 7711–7732. doi: 10.1029/2018WR022587

Tarasova, L., Lun, D., Merz, R., Blöschl, G., Basso, S., Bertola, M., . . . Kumar, R. (2023). Shifts in flood generation processes exacerbate regional ood anomalies in Europe. *Communications Earth & Environment*. doi: https://doi.org/10.21203/rs.3.rs-2238190/v1

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., . . . Wietzke, L. (2019). Causative classification of river flood events. *Wiley Interdisciplinary Reviews: Water*, *6*(4), e1353. doi: 10.1002/wat2.1353

Thompson, D. W., & Wallace, J. M. (1998). The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, *25*(9), 1297–1300. doi: 10.1029/98GL00950

Thompson And, D. W. J., & Wallace, J. M. (2000). Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability*. *Journal of Climate*, *13*.

Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*. doi: 10.4324/9780429311369-6

Trigo, I. F., Bigg, G. R., & Davies, T. D. (2002). Climatology of Cyclogenesis Mechanisms in the Mediterranean. *Monthly Weather Review*, *130*.

Uhlemann, S., Thieken, A. H., & Merz, B. (2010). A consistent set of trans-basin floods in Germany between 1952-2002. *Hydrology and Earth System Sciences*, *14*(7), 1277–1295. doi: 10.5194/hess-14-1277-2010

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., . . . Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. , 1–15. Retrieved from `http://arxiv.org/abs/1609.03499` doi: https://doi.org/10.48550/arXiv.1609.03499

Villarini, G., Smith, J. A., Ntelekos, A. A., & Schwarz, U. (2011). Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria. *Journal of Geophysical Research Atmospheres*, *116*(5). doi: 10.1029/2010JD015038

Visbeck, M. H., Hurrell, J. W., Polvani, L., & Cullen, H. M. (2001). *The North Atlantic Oscillation: Past, present, and future* (Vol. 6; Tech. Rep. No. 23).

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, *109*.

Wang, Z., Yan, W., & Oates, T. (2016). Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. Retrieved from `http://arxiv.org/abs/1611.06455`

Ward, P. J., Jongman, B., Kummu, M., Dettinger, M. D., Weiland, F. C., & Winsemius, H. C. (2014). Strong influence of El Niño Southern Oscillation on flood risk around the world. *Proceedings of the National Academy* *of Sciences of the United States of America*, *111*(44), 15659–15664. doi: 10.1073/pnas.1409822111

Ward, P. J., Kummu, M., & Lall, U. (2016). Flood frequencies and durations and their response to El Niño Southern Oscillation: Global analysis. *Journal of Hydrology*, *539*, 358–378. doi: 10.1016/j.jhydrol.2016.05.045

Winter, C., Tarasova, L., Lutz, S. R., Musolff, A., Kumar, R., & Fleckenstein, J. H. (2022). Explaining the Variability in High-Frequency Nitrate Export Patterns Using Long-Term Hydrological Event Classification. *Water Resources Research*, *58*(1). doi: 10.1029/2021WR030938

Yu, H. L., & Lin, Y. C. (2015). Analysis of space-time non-stationary patterns of rainfall-groundwater interactions by integrating empirical orthogonal function and cross wavelet transform methods. *Journal of Hydrology*, *525*, 585–597. doi: 10.1016/j.jhydrol.2015.03.057

Zanardo, S., Nicotina, L., Hilberts, A. G., & Jewson, S. P. (2019). Modulation of Economic Losses From European Floods by the North Atlantic Oscillation. *Geophysical Research Letters*, *46*(5), 2563–2572. doi: 10.1029/2019GL081956

Zink, M., Kumar, R., Cuntz, M., & Samaniego, L. (2017). A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrology and Earth System Sciences*, *21*(3), 1769–1790. doi: 10.5194/hess-21-1769-2017

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., … Vignotto, E. (2020). A typology of compound weather and climate events. *Nature Reviews Earth and Environment*, *1*(7), 333–347. doi: 10.1038/s43017-020-0060-z