Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations

Elizaveta Felsche[1,2,3] and Ralf Ludwig[2]

[1]Center for Digital Technology and Management, Munich, Germany
[2]Department of Geography, Ludwig Maximilians University of Munich, Munich, Germany
[3]Technical University of Munich, Munich, Germany

Correspondence: Elizaveta Felsche (felsche@cdtm.de)

Abstract.

There is a strong scientific and social interest ~~to understand~~ in understanding the factors leading to extreme events in order to improve the management of risks associated with hazards like droughts. In this study, artificial neural networks are applied to predict the occurrence of a drought in two contrasting European domains, Munich and Lisbon, with a lead time of one month. The approach takes into account a list of 28 atmospheric and soil variables as input parameters from a single-model initial condition large ensemble (CRCM5-LE). The data was produced in the context of the ClimEx project by Ouranos with the Canadian Regional Climate Model (CRCM5) driven by 50 members of the Canadian Earth System Model (CanESM2). Drought occurrence is defined using the Standardized Precipitation Index. The best performing machine learning algorithms manage to obtain a correct classification of drought or no drought for a lead time of one month for around ~~55-60~~55-57% of the events of each class for both domains. Explainable AI methods like SHapley Additive exPlanations (SHAP) are applied to ~~gain a better understanding of~~ understand the trained algorithms better. Variables like the North Atlantic Oscillation Index and air pressure one month before the event prove ~~to be of high importance~~ essential for the prediction. The study shows that seasonality ~~has a high influence on goodness~~ strongly influences the performance of drought prediction, especially for the Lisbon domain.

1 Introduction

Droughts remain to be one of the most dangerous hazards, having a serious and large-scale impact on environment, society and economy. Recent events like the Summer 2018 drought in huge parts of Central Europe led to severe forest fires and crop failures. The damage was estimated to several hundred millions euros solely in Germany (Federal Ministry of Food and Agriculture, 2018). Moreover the effect of global warming leads to major changes in the earth's climate system, having a direct influence on the frequency and severity of extreme events like droughts (Spinoni et al., 2016). An increase in frequency of drought occurrence is a major threat for current and future generations, and comprehensive knowledge on the phenomenon of drought is needed in order to take action early and to prevent humanitarian catastrophes. This goes in conjunction with drought prediction. Precise drought prediction would enable to mitigate the dangers connected to drought occurrences, such that e.g. stakeholders would be able to store the maximal possible amount of water in the endangered regions. This would help

to mitigate the water shortage when the drought arrives. Measures for demand reduction could like that be introduced earlier, and in better adjusted extent; this would help to reduce the economic and societal damage.

To mitigate the effects of droughts the information on the their onset is of crucial importance. This can be derived from a drought index. A variety of drought indices exist, which are typically defined according to statistical and physical measures. These are mostly taking into account atmospheric and soil variables. Among the most popular ones are the Standardized Precipitation Index (SPI), Standardized Precipitation Evaporation Index (SPEI), Soil Moisture Percentile (SMP), and Palmer Drought Severity Index (PDSI). Standardized Precipitation Index (SPI) is adopted as the standard meteorological index by World Meteorological Organization (2012). It is a measure of meteorological drought based on the probability of occurrence of certain precipitation amounts in the area of interest (Sheffield and Wood, 2011). Studies on drought prediction by Belayneh et al. (2016) and Bonaccorso et al. (2015) use SPI as a prediction variable for the forecast.

Forecasting of any physical phenomenon can either be done by a physical, conceptual or data-driven model. The latter ones are widely used due to their rapid development times and the flexibility in input parameters. McGovern et al. (2017) argues that AI-methods have a high potential for prediction of extremes due to the ability of machine learning methods to learn from past data, to handle large amounts of input variables, to integrate physical understanding into the models and to discover additional knowledge from the data.

A review on seasonal drought prediction given by Hao et al. (2018) identifies two typical predictor groups of variables: large-scale climate indices that reflect the atmosphere-ocean circulation patterns and local climate variables. The first ones are known to correlate with precipitation patterns in special regions and therefore are naturally correlated with the occurrence of drought. The teleconnection indices important for European precipitation include North Atlantic Oscillation (NAO), Scandinavian Oscillation (SCA), East Atlantic/Western Russia Oscillation (EA/WR), East Atlantic Oscillation (EA) and Atlantic Multidecadal Oscillation (AMO) (Hao et al., 2018). As shown by Folland et al. (2009) a positive NAO index in summer is associated with dry and warm conditions in the north-west of Europe, whereas southern Europe and the Mediterranean experience cooler and wetter conditions. More information on the influence of the NAO, SCA, EA, and EAWR on the European climate can be found in Folland et al. (2009), Bueh and Nakamura (2007), Mikhailova and Yurovsky (2016), Lim (2015), Barnston and Livezey (1987) and Sheffield et al. (2009). A positive phase of AMO is associated with humid conditions over Great Britain and parts of Scandinavia and with dry conditions in the Mediterranean (Sheffield and Wood, 2011, p. 26); the negative phase is associated with a reversed pattern: dry conditions in Great Britain and wet conditions in the Mediterranean. A study by Sheffield et al. (2009) showed a correlation between the amount of droughts and AMO of 62% with a significance at the 90% level. A recent study by Bonaccorso et al. (2015) uses NAO for prediction of probability of drought occurrence for Sicily. The local climate variables like precipitation, temperature, soil moisture were also used as inputs to reflect the conditions at the time the prediction occurs. Belayneh et al. (2016) and Bonaccorso et al. (2015) used SPI for the past months as input variable to the algorithm. A study by Morid et al. (2007) used precipitation as an input parameter.

This paper examines the possibilities of meteorological drought prediction with the lead time of one month applying artificial neural networks (ANN) for two domains with different climate: one with Mediterranean (Lisbon), one with continental climate (Munich) (Ceglar et al., 2019). Both sites experienced an increase of drought frequency when comparing 2015 and 1950 and

are projected to keep rising under RCP4.5 as well as RCP 8.5. (Spinoni et al., 2017). Observational data offers only a limited field for drought investigation as it can be seen from the following approximation. Systematical weather observations started in 1781 by Societas Meteorologica Palatina (Kington, 1980). In this study SPI1<-1 is used as a threshold for drought occurrence. It corresponds to the 15% driest months (John Keyantash, 2018) and can be estimated by a total amount of 430 observed events until the year 2020 (Eq. 1).

$$(2020 - 1781)\ \text{yr} \cdot 12\ \text{months/yr} \cdot 15\% = 430\ \text{events} \tag{1}$$

Compared to that CRCM5-LE offers a total amount of roughly 4500 events when using the first 50 years from the climate simulation data (1955-2005) (see Eq. 2).

$$50\ \text{yr/member} \cdot 50\ \text{members} \cdot 12\ \text{months/yr} \cdot 15\% = 4500\ \text{events} \tag{2}$$

This is a difference of an order of magnitude. The more data is available the better the predictions that can be derived by a drought predicting machine learning model and the more can be learned about drought formation. According to von Trentini et al. (2020) precipitation in summer and winter derived from the European gridded data set (E-OBS) does fall to a high percentage into the range produced by CRCM5-LE for the historic period. Therefore, the CRCM5-LE proves applicable to this study and its larger amount of extreme events can be used as input to the machine learning algorithms. In this study a variety of ANNs are trained. Best performing models are investigated to using explainable AI methods to understand the results.

While no comparable study exists for the Munich domain, Santos et al. (2014) performed a drought prediction based on SPI6 for Portugal for the months April, May and June using the following input variables: sea surface temperatures (JFM), NAO (DJFM) and cumulative precipitation (NDJFM for $SPI6_{April}$, DJFM for $SPI6_{May}$, JFM for $SPI6_{June}$). Best results were achieved for the prediction of SPI6 for April with a correlation coefficient of 0.98. SPI6 for May and June referred to a correlation coefficient of 0.78 and 0.77 respectively.

2 Data and Methods

2.1 Datasets

To investigate the predictability of droughts data from the single-model initial condition large ensemble (SMILE) consisting of 50 members, the Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE) is used. The data was produced within the scope of the ClimEx Project (Leduc et al. (2019), www.climex-project.org). The CRCM5-LE was generated by dynamical downscaling of the data provided by the 50-member initial condition Canadian Earth System Model 2 using the Canadian Regional Climate Model 5 (Martynov et al., 2013). The data has a resolution of $0.11\,\text{deg}$ ($12\ \text{km}$) and is produced for the years 1950-2099 for a European and an eastern North America domain. For the years 1950-2005 the historical greenhouse gas concentrations and aerosol emissions are being used. Starting from 2005, the model introduces the RCP8.5 (IPCC, 2013) forcing scenario. A total of 42 atmospheric variables is available in a temporal resolution of one to three hours. They are used on monthly basis as input to the machine learning algorithms. The list of variables is provided in ~~Tab.~~ Table 1.

clt	Total Cloud Fraction	%	*prw*	Water Vapor Path	$\mathrm{kgm^{-2}}$
dds	Near-Surface Dewpoint Depression	K	*ps*	Surface Air Pressure	Pa
evspsbl	Evaporation	$\mathrm{kgm^{-2}s^{-1}}$	*psl*	Sea Level Pressure	Pa
evspsblland	Water Evaporation from Land	$\mathrm{kgm^{-2}s^{-1}}$	rlds	Surface Downwelling Longwave Radiation	$\mathrm{Wm^{-2}}$
hfls	Surface Upward Latent Heat Flux	$\mathrm{Wm^{-2}}$	*rlus*	Surface Upwelling Longwave Radiation	$\mathrm{Wm^{-2}}$
hfss	Surface Upward Sensible Heat Flux	$\mathrm{Wm^{-2}}$	*rlut*	TOA Outgoing Longwave Radiation	$\mathrm{Wm^{-2}}$
hurs	Near-Surface Relative Humidity	%	*rsaa*	Shortwave Radiation Absorbed by Atmosphere	$\mathrm{Wm^{-2}}$
huss	Near-Surface Specific Humidity	1	*rsds*	Surface Downwelling Shortwave Radiation	$\mathrm{Wm^{-2}}$
mrfso	Soil Frozen Water Content	$\mathrm{kgm^{-2}}$	*rsdt*	TOA Incident Shortwave Radiation	$\mathrm{Wm^{-2}}$
mrlso	Soil Liquid Water Content	$\mathrm{kgm^{-2}}$	*rsus*	Surface Upwelling Shortwave Radiation	$\mathrm{Wm^{-2}}$
mrro	Total Runoff	$\mathrm{kgm^{-2}s-1}$	*rsut*	TOA Outgoing Shortwave Radiation	$\mathrm{Wm^{-2}}$
mrros	Surface Runoff	$\mathrm{kgm^{-2}s^{-1}}$	*sfcWindmax*	Daily Maximum Near-Surface Wind Speed	$\mathrm{ms^{-1}}$
mrso	Total Soil Moisture Content	$\mathrm{kgm^{-2}}$	*snc*	Snow Area Fraction	%
mrsos	Moisture in Upper Portion of Soil Column	$\mathrm{kgm^{-2}}$	*snd*	Snow Depth	m
prc	Convective Precipitation	$\mathrm{kgm^{-2}s^{-1}}$	*snw*	Surface Snow Amount	$\mathrm{kgm^{-2}}$
prdc	Deep Convective Precipitation	$\mathrm{kgm^{-2}s^{-1}}$	*tas*	Near-Surface Air Temperature	K
prfr	Freezing Rain	$\mathrm{kgm^{-2}s^{-1}}$	*tasmax*	Daily Maximum Near-Surface Temperature	K
pr	Precipitaiton	$\mathrm{kgm^{-2}s^{-1}}$	*tasmin*	Daily Minimum Near-Surface Temperature	K
prlp	Liquid Precipitation	$\mathrm{kgm^{-2}s^{-1}}$	*ts*	Surface Temperature	K
prrp	Refrozen Rain	$\mathrm{kgm^{-2}s^{-1}}$	*uas*	Eastward Near-Surface Wind	$\mathrm{ms^{-1}}$
prsn	Snowfall Flux	$\mathrm{kgm^{-2}s^{-1}}$	*vas*	Northward Near-Surface Wind	$\mathrm{ms^{-1}}$

Table 1. 42 monthly atmospheric and soil variables from CRCM5-LE

In the study we use monthly sea level pressure (*psl*) from the driving model CanESM2-LE (Kushner et al., 2018; Kirchmeier-Young et al., 2016) for the calculation of North Atlantic Oscillation (NAO), Scandinavian Oscillation (SCA), East Atlantic Oscillation (EA) and East Atlantic/Western Russia Oscillation (EA/WR) over the whole Atlantic basin ($20° - 80°\mathrm{N}$, $90°\mathrm{W} - 40°\mathrm{E}$). The Atlantic Multidecadal Oscillation (AMO) is calculated using the Sea Surface Temperature (SST) over the $0 - 60°\mathrm{N}, 0 - 80°\mathrm{W}$ from the CanESM2. Only the period 1955-2005 is considered in order to stay within the scope of historical climate. The CRCM5 domain is displayed in Fig. 1. For the machine learning training a gridpoint situated as $48.11°\mathrm{N}$ and $-9.17°\mathrm{W}$ is referenced as Munich and $38.67°\mathrm{N}$ and $11.91°\mathrm{W}$ is referenced as Lisbon.

2.2 Input variables for drought prediction

In order to calculate NAO, SCA, EA and EA/WR the method introduced by Hurrell et al. (2003) is used: a principal component analysis (PCA) of the monthly psl is performed over the $20° - 80°\mathrm{N}$, $90°\mathrm{W} - 40°\mathrm{E}$ domain. The leading eigenvectors, scaled by the amount of variance they explain, represent the leading circulation patterns of the atmospheric system. The first eigenvector corresponds to NAO, the second one to SCA, the third one to EA, the fourth one to EA/WR. To calculate the teleconnection indices (NAO, SCA, EA, EA/WR) the Eof package described in Dawson (2016) is used. The leading modes of the PCA corresponding to NAO, SCA, EA and EA/WR derived from the CanEsm2 dataset are shown in Fig. 2.

AMO is calculated by spatial averaging over the $0 - 60°\mathrm{N}, 0 - 80°\mathrm{W}$ area of the anomaly of sea surface temperature (E Trenberth, 2011). Additionally the 10-year running mean of AMO is calculated as an input variable, as it is widely used in various studies and was shown to be correlated with precipitation (Enfield et al., 2001).

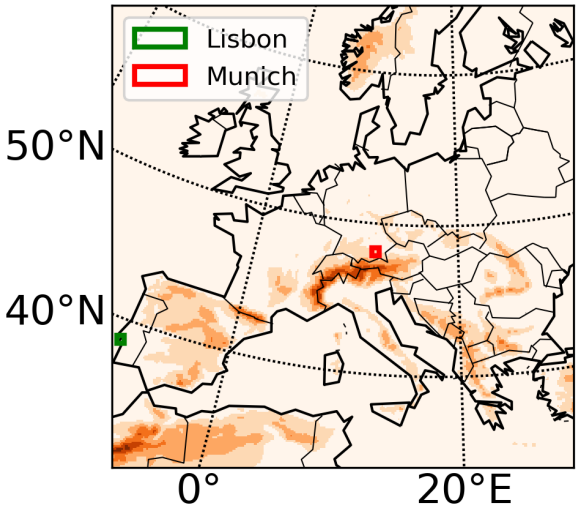

Figure 1. CRCM5 topography.

~~To limit the computation time a preselection of variables for the input data is performed~~Variable subset selection helps to limit the computational time and to improve predictive accuracy (Kumar and Minz, 2014). In order to eliminate redundant variables ~~Pearsons R between~~ Pearson's R between all the CRCM5 variables for the chosen domains is calculated. ~~Pearsons~~ Pearson's R ($\rho_{X,Y}$) is a measure of linear correlation between two variables X and Y. $\rho_{X,Y}$ equals 1 if the correlation is total positive, 0 if there is no linear correlation and -1 if the correlation is total negative (Guyon and Elisseeff, 2003). For two samples x and y the ~~Pearsons~~ Pearson's R is defined in the following way:

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3}$$

The bar refers to the average over the index i (Guyon and Elisseeff, 2003). Pearson's R is a popular and easy method for feature selection of continuous variables as introduced in Biesiada and Duch (2007). ρ is calculated for all possible permutations of the 41 input variables. The ones correlating to a high degree are examined and a threshold of 0.95 is chosen. In ~~Tab.~~ Table 2 a list of sorted out variables and the corresponding values of Pearson's R is given. The high correlation values can be explained by a physical relationship between the variables: e.g. the total evaporation (*evspsbl*) is almost the same as evaporation from land (*evspsblland*), as there are no relevant water bodies in the chosen domains. Out of the full list of 42 variables 14 are sorted out as being redundant.

2.3 Standardized Precipitation Index

The Standardized Precipitation Index (SPI) is a precipitation based index introduced by McKee et al. (1993). For the calculation of SPI a continuous monthly precipitation dataset is used. The index can be calculated on different timescales: typically, it is 1, 3, 6, 12 or 24 months. As a first step the precipitation values are accumulated for the needed timescale. The resulting

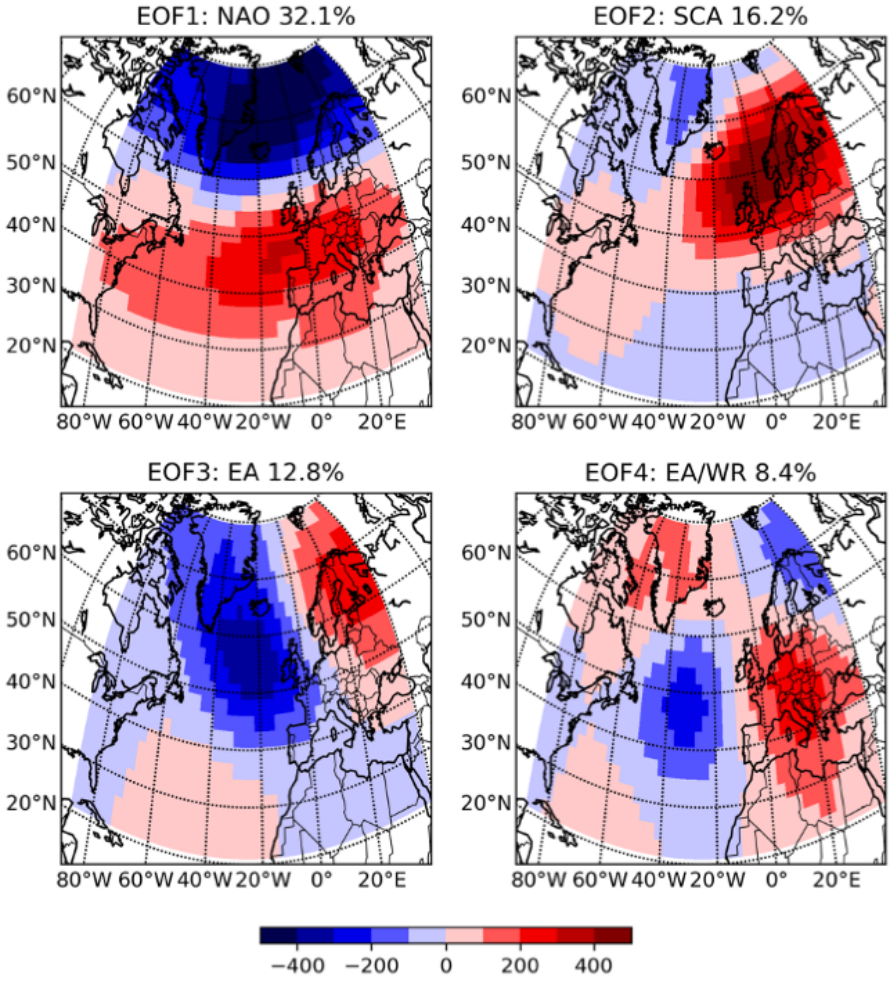

Figure 2. First four leading eigenfunctions of the mean sea level pressure in CanESM2. Percentage of variance the mode explains is given on top of the figures.

Kept variable	Sorted out variable	~~Pearsons~~ Pearson's R
hurs	*dds*	-0.9879
evspsbl	*evspsblland*	0.9994
evspsbl	*hfls*	0.9988
mrso	*mrlso*	0.9991
rlut	*rlaa*	-0.9549
tas	*rlds*	0.9550
tas	*rlus*	0.9960
rnt	*rns*	0.9954
rnt	*rsaa*	0.9831
rnt	*rsdt*	0.9970
rnt	*rss*	0.9872
rnt	*rst*	0.9926
tas	*tasmax*	0.9932
tas	*tasmin*	0.9864

Table 2. List of sorted out variables

dataset is fitted to a Gamma distribution for each month separately and then transformed to a normal distribution, such that the mean SPI is zero. The SPI value for a given precipitation is then the number of standard deviations from normal. Because of the normalization SPI is especially useful to represent wetter and drier climates, as well as to account for differences among seasons. ~~Here~~ As the two study sites are having different meteorological conditions, SPI provides a convenient and comparable measure (Zargar et al., 2011). As noted in Yoon et al. (2012) the accumulation period of the SPI value needs to be chosen equal or less to the prediction lead time, as otherwise the precipitation values needed for the mathematical calculation of SPI would be given as input to the machine learning algorithm. Therefore the accumulation period of one month is chosen. SPI1 is calculated for Lisbon and Munich each using the data from 1955-2005 from all members as reference.

2.4 Machine learning

This study investigates drought predictability applying the technique of supervised machine learning for this purpose. Machine learning is a promising tool for the analysis of complex and data-rich phenomena as droughts (McGovern et al., 2017). The python package Keras, a high-level neural network package, is used for the design of the machine learning models (Chollet et al., 2015), as it allows to design neural networks in an easy way by adding layers. Three crucial elements are needed to perform drought prediction by supervised machine learning: input data, a target variable to be predicted and a computation pipeline, which includes the machine learning algorithm.

The data from the years 1957 - 1999 is used as training data, the years 2000-2005 are used for the testing purpose. Each of the time periods is available 50 times as we are dealing with an ensemble of 50 members. This results in 2150 model years

for training and 250 years for testing. A small fraction of the training data is used for the validation of the machine learning algorithms. The target variable chosen for the prediction of droughts is SPI1. Two classes for the prediction are identified in the following way: $SPI1 < -1$ is defined as an event and is initialized with 1, $SPI1 > -1$ is initialized with 0 and corresponds to a non-drought event. The lead time of one month is chosen for the prediction - as it has been used in previous studies by Yoon et al. (2012) and Deo et al. (2017). Moreover shorter prediction lead times usually obtain better results when compared to longer periods, as seen in Bonaccorso et al. (2015). After the feature selection 28 variables originating directly from the CRCM5-LE dataset are used as input. In addition to those the teleconnection indices NAO, SCA, EA, EA/WR, AMO and AMO10 are used as input.

To predict e.g. a drought or non-drought in April of 1980, the data for twelve months before the event is used as input, this is NAO and other teleconnection and atmospheric variables for the period April 1989 – March 1980; for a prediction of an event in May 1980, May 1989 – April 1980 is used as input. The twelve months before the event are chosen in accordance with the study by Morid et al. (2007), which found that the best performing drought prediction model was the one including the value up to twelve months before the predicted one. We perform a time series prediction with no limitation on special months or seasons to be inspected.

For this analysis we use a supervised machine learning algorithm, an Artificial Neural Network (ANN). ANNs are algorithms which design is inspired by the architecture of the human brain with its neurons (Russell and Norvig, 2009).; they both consists of connected nodes. A link between the node i and the node j serves to propagate the activation a_i from i to j. To each connection a numeric weight $w_{i,j}$ is assigned. The output of the node is computed by:

$$a_i = g(in_j) = g\left(\sum_{i=0}^{n} w_{i,j} a_i\right) \tag{4}$$

(Russell and Norvig, 2009, p. 728). The activation function defines the output of the node. In order to have stable learners with confident predictions a function with a soft threshold is recommended (Russell and Norvig, 2009). In this study the following three activation functions are used: Sigmoid, Rectified Linear Unit (ReLU), Exponential Linear Unit (ELU). Sigmoid activation is especially useful for the output layer (Russell and Norvig, 2009), while ReLU and ELU both have the property of allowing very fast optimization (Maas, 2013) .

Sigmoid function, also called logistic function, is defined in the following way:

$$Logistic(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

(Russell and Norvig, 2009). This function has an output between 0 and 1. This can be interpreted as a probability of belonging to the class 1. One of the main disadvantages of the sigmoid activation function is the vanishing gradient problem: at higher, almost saturated layers with values of 1 or -1, the gradients become nearly 0 resulting in a slow optimization convergence (Russell and Norvig, 2009, p. 726).

ReLU refers to Rectified Linear Unit and shows better performance when dealing with the vanishing gradient problem (Maas, 2013). ReLU is defined in the following way:

$$f(x) = max(0, x) \tag{6}$$

ELU refers to the Exponential Linear Unit and was introduced by Clevert et al. (2016). Clevert et al. (2016) claim that in experiments the ELU activation led to faster learning and significantly better generalization performance than ReLU and sigmoid activation. The function is defined as:

$$f(x) = \begin{cases} x \text{ if } x > 0 \\ \alpha(exp(x) - 1) \text{ if } x \leq 0 \end{cases} \tag{7}$$

α controls the value to which an ELU saturates for negative inputs. Per default the value is set to 1 such that the function saturates at -1.

Two kinds of layers are used in this study: Dense and Dropout. *Dense* refers to a regular fully connected neural network layer. *Dropout* refers to a layer which is randomly setting a fraction of inputs to zero at each update. This technique is used to prevent overfitting and therefore improving the performance of the algorithm (Chollet et al., 2015). The first part of the study concentrates on the methodological search for the best performing algorithms. A pipeline to search for the best performing architecture, value for L2-Regularization and loss function is built up.

The model performance is evaluated using Accuracy and F1-score (Sasaki, 2007). The latter one is especially useful when training on datasets with an imbalanced class distribution, as it is in the case of our dataset. Accuracy is defined in the following way:

$$Accuracy = \frac{Number\ of\ right\ predictions}{Total\ number\ of\ samples} \tag{8}$$

F1-score is a harmonic measure between precision and recall. Precision is the amount of true positives with respect to the amount of positively classified data. Recall is the amount of true positives with respect to the total number of positives in the data. F1-score is defined in the following way:

$$F1 - score = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{9}$$

Due to the class imbalance within the dataset we require that the accuracy on each class is at least 50%. In that case given the distribution of the test dataset of 1803 non-drought events to 387 droughts for Lisbon and 1848 non-drought events to 352 drought events for Munich a marginal F1-score of 0.26 for Lisbon and 0.24 for Munich is given.

Best performing models are additionally evaluated using the Heidke Skill Score(HSS). The range of the HSS is $-\infty$ to 1. Values below zero indicate that the random forecast (a forecast which randomly assigns the labels) has a better performance than the trained model. HSS of 1 indicates a perfect forecast. HSS is defined in the following way:

$$HSS = 2\frac{ad - bc}{(a+c)(c+d) + (a+b)(b+d)} \tag{10}$$

where a is the number of true positives, b the number of false positives, c number of false negatives, d number of true negatives.

The second part of the study analyzes the best performing algorithms (one for Lisbon domain, one for Munich domain) by applying explainable AI methods. SHAP (SHapley Additive exPlanations) is a state of the art method for interpretation of

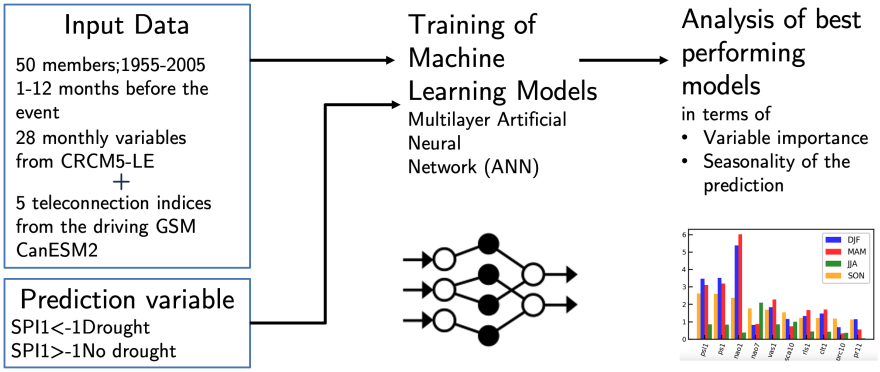

Figure 3. Overview of the proposed methodology

machine learning models, which was inspired by game theory (Lundberg and Lee, 2017). It estimates for each input feature an average marginal contribution to the prediction of the result and therefore allows a comparison of the contributions among different features. In addition to that the difference in predictability among the seasons is calculated and compared to gain a better understanding on the influence of seasonal weather patterns.

An overview of the proposed methodology can be found in Fig. 3.

3 Results

This study consists of two parts: the first parts deals with a systematical search for the best performing setup of the ANN model for the two domains of interest: Munich and Lisbon. A repeated training is conducted by varying the values of parameters like the architecture of the hidden layers, L2-Regularization and the loss function. In the second part of the analysis the best performing models for the two domains are analyzed using explainable AI methods.

3.1 Model training results

For the design of the ANN it is crucial to perform fine tuning of the model parameters to find the optimal setup. An architecture has to have enough layers and neurons to capture the complexity of the dataset (Goodfellow et al., 2016). In order to find the best architecture the learning curve of the algorithm is inspected. The learning curve shows the loss of the training and validation datasets on the weights during the training (Goodfellow et al., 2016). Two examples are shown in ~~figure~~ Fig. 4. The plot shown in the top refers to an architecture, which is not able to capture the complexity of the dataset: the loss is hardly decreasing on the training or validation data. The bottom figure refers to an architecture which overfits: in the last epochs the loss of the validation dataset is rising, while it decreases on the training dataset.

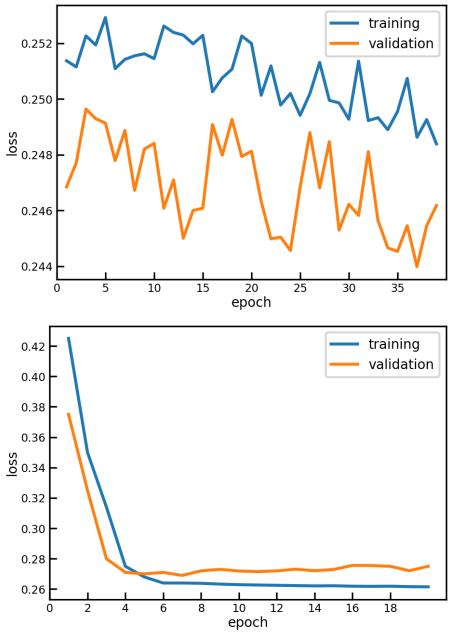

Figure 4. Learning curve for two chosen fitting examples: algorithm complexity insufficient (top) and overfitting (bottom).

In such way a network is searched which captures the given complexity of the dataset. This is reached with an algorithm consisting of at least five layers. Additionally two dropout layers, which are setting a specified number of nodes to zero in a random way, are introduced in order to fight overfitting.

3.1.1 L2-Regularization

L2-Regularization is a broadly applied method to prevent overfitting on the training data (Bishop, 2007). The main idea behind regularization is to add a penalty term to the loss function, which will punish the classifier for complexity and force some of the weights to zero (Russell and Norvig, 2009). In case of L2-Regularization the punishing term is proportional to L2-norm of the weight vector. The weight of the punishing term λ determines the relative importance of the regularization.

The results of the training with different values of λ for L2-Regularization are shown in ~~Tab.~~ Table 3. Training results are displayed in this particular case as the regularization is introduced to prevent overfitting. Generally the performance on the test dataset is more important and will be inspected in following experiments. If λ is set to zero the regularization term vanishes. Especially in those cases the overfitting is high. For Lisbon overall higher performance could be seen for values of λ around 0.01, 0.001 or 0.0001. Models that are trained on the Munich dataset perform better with the value of λ of 0.001. Since the performance of the model on the F1-score has a higher importance for an imbalanced dataset than the pure accuracy the value of 0.001 is chosen for the following ANN model training.

	Lisbon				Munich			
	Train		Test		Train		Test	
λ	Acc	F1	Acc	F1	Acc	F1	Acc	F1
0	0.961	0.861	0.733	0.206	0.959	0.865	0.787	0.176
0.1	0.495	0.233	0.373	0.294	0.506	0.241	0.536	0.215
0.01	0.517	0.245	0.460	0.269	0.519	0.268	0.431	0.275
0.001	0.572	0.261	0.540	0.288	0.490	0.288	0.563	0.266
0.0001	0.765	0.472	0.627	0.259	0.823	0.557	0.719	0.189

Table 3. Results of ANN training for different values for λ for L2-Regularization. λ of 0.001 (highlighted in grey) is chosen for both domains for following training, since the performance of the model on the F1-score has a higher importance for an imbalanced dataset than the pure accuracy.

3.1.2 Loss function

As a next step the influence of the different loss functions on the model performance is investigated. Loss function is a function to evaluate how well a specific algorithm manages to fit the training data (Janocha and Czarnecki, 2017). It is an important part of the optimization function which has a direct influence on the updating of the weights of the ANN (Russell and Norvig, 2009). In addition to overall accuracy and F1-metric, the accuracies on the non-drought and drought classes on the test dataset are displayed. The results are shown in ~~Tab.~~ Table 4. Binary cross-entropy, mean absolute error and hinge loss functions show the best performance for the Munich domain. In contrast to that for the Lisbon domain only the mean absolute error loss function has an accuracy of higher than 0.5. Also in the case of the Munich domain mean absolute error shows a higher performance on the F1-score. Therefore mean absolute error is used for further analysis.

	Lisbon				Munich			
Loss-Function	Acc nd	Acc d	Acc	F1	Acc nd	Acc d	Acc	F1
mean absolute error	0.511	0.516	0.540	0.288	0.500	0.582	0.512	0.276
mean squared error	0.440	0.655	0.479	0.312	0.562	0.509	0.553	0.267
binary crossentropy	0.436	0.610	0.467	0.292	0.589	0.440	0.565	0.245
hinge	0.229	0.753	0.323	0.287	0.568	0.486	0.555	0.259
squared hinge	0.486	0.501	0.489	0.261	1.000	0.000	0.840	0.000

Table 4. Performance of the model for different loss functions on the test dataset. Acc nd refers to the accuracy on the non-drought class and Acc d to the accuracy of the drought class. Mean absolute error (highlighted in grey) is chosen for following analysis, since for Munich and Lisbon it shows an Accuracy of at least 0.5 on both classes and a higher performance on the F1-score.

3.1.3 Model architecture

Lastly the models are trained on both domains using different architectures. Table 5 is displaying the model training results on the test dataset. The column "architecture" refers to the number of neurons in each Dense (De) layer separated by the *-sign. For Dropout (Dr) layers the fraction of weights which are randomly set to zero is given. The model architecture consists of overall seven layers. For example the architecture for the model in the first line of Table 5 is the following:

1. Dense layer with 4000 neurons
2. Dropout Layer randomly setting 50% of weights to zero
3. Dense layer with 1000 neurons
4. Dropout Layer randomly setting 50% of weights to zero
5. Dense layer with 500 neurons
6. Dense layer with 100 neurons
7. Dense layer with 5 neurons

We require the accuracy on both classes individually to be higher than 0.5 and search for an F1-score as high as possible. In case of the Lisbon domain three trained models are satisfying the criterion of at least 50% accuracy on each class: the model in the first, in the fourth and in the last row. Best performance in terms of F1-score is obtained for the last model with the following architecture: 5000*0.5*4000*0.5*1000*500*100. For the Munich domain only the first and the fourth models are satisfying the criterion of at least 50% accuracy on each class. For further analyses the first model is chosen, as it shows the highest F1-score. The following model architecture is used for the Munich domain: 4000*0.5*1000*0.5*500*100*5. For the best performing models HSS equals 0.06 for Lisbon and 0.04 for Munich. This results confirm that the obtained prediction is better than the one obtained by a random forecast and therefore does show a weak prediction skill. In the next step those models are analyzed using explainable AI methods.

		Lisbon				Munich			
Neurons	Architecture	Acc nd	Acc d	Acc	F1	Acc nd	Acc d	Acc	F1
De*Dr*De*Dr*De*De*De	4000*0.5*1000*0.5*500*100*5	0.511	0.516	0.540	0.288	0.562	0.509	0.553	0.267
De*Dr*De*Dr*De*De*De	5000*0.5*1000*0.5*500*100*5	0.581	0.496	0.566	0.292	0.378	0.693	0.428	0.279
De*Dr*De*Dr*De*De*De	5000*0.5*4000*0.5*500*100*5	0.457	0.602	0.483	0.296	0.725	0.338	0.663	0.243
De*Dr*De*Dr*De*De*De	5000*0.5*4000*0.5*1000*100*5	0.570	0.501	0.558	0.290	0.527	0.514	0.525	0.257
De*Dr*De*Dr*De*De*De	5000*0.5*4000*0.5*1000*500*5	0.402	0.635	0.444	0.292	0.683	0.409	0.640	0.266
De*Dr*De*Dr*De*De*De	5000*0.5*4000*0.5*1000*500*100	0.575	0.526	0.566	0.305	0.420	0.619	0.452	0.266

Table 5. Performance of models for the Lisbon and Munich domains for different variations of architecture on the test dataset. Acc nd refers to the accuracy on the non-drought class and Acc d to the accuracy of the drought class. Sixth model architecture is chosen for following analysis for Lisbon and first for Munich (highlighted in grey), due to an Accuracy of at least 0.5 on both classes and a higher performance on the F1-score.

3.2 Explainable AI methods for the analysis of best performing algorithms

3.2.1 Shapely values

For the Munich and Lisbon domain Shapely values are calculated using the results of the best performing models on the test dataset. For the calculation each of the 12 month used as input to the predicting algorithm for each variable is considered individually, resulting in 28 atmospheric variables *12 + 6 teleconnection indices * 12 = 408 variables. The number behind the variable name refers to the number of months before the event (NAO1 - NAO value one month before the predicted event). The results are shown in Fig. 5. Since the calculation of Shapely values is computationally expensive they are calculated 5 times on a subset of 500 data points. The error bars displayed in black on the plot indicate that the uncertainties are smaller than the nominal values of the variable contributions. The nominal Shapely values are normed and recalculated to a percentage of contribution to the prediction, e.g. the NAO1 value explains roughly 2.3% of the prediction for the Lisbon domain.

We see that for both domains the contribution to the prediction is broadly distributed among the many input variables. Between Lisbon and Munich Shapely values show a distinct difference in the nominal values of the feature contributions: values for Lisbon are about 6 times higher than those for Munich (e. g. the contribution of NAO1 for Munich is around 0.3% and for Lisbon around 1.9%).

For the Lisbon domain, the variables with a higher impact are sea level pressure (*psl*), surface pressure (*ps*) and NAO one month before the event. The first two variables are strongly autocorrelated for the Lisbon domain due to its location at the sea. The strong influence of *ps/psl* and NAO shows the influence of the atmospheric pressure system on drought formation in Lisbon. It is also striking that the influence of the local pressure seems to be higher than the influence of NAO. The next two variables for the Lisbon domain with the strongest contribution to the prediction are Northward Near-Surface Wind (*vas*) and Evaporation (*evspsbl*). The latter variable has a very direct influence on the formation of drought given that if evaporation is getting lower, also the probability of formation of rain clouds decreases (Sheffield and Wood, 2011). The contribution of *vas* to drought formation in Lisbon needs to be further studied. For the Munich domain the highest influence is found for NAO1, *psl1* , EAWR5 and *ps1*. The results indicate that NAO is the most influential drought predictor for Munich. Additionally the contribution of EAWR5 and SCA5 on the Munich domain cannot be neglected as they are found within the top five predictors. A further investigation of this relationship is of interest for the understanding of drought formation in Munich.

3.2.2 Seasonality

In order to evaluate the influence of seasonality on the prediction the performance of the model is calculated separately for the four seasons. Since the distribution between the drought and non-drought classes is different among the seasons (e.g. range of 17% to 19% of drought events for the Lisbon domain) a rescaling of the number of drought and non-drought events is performed to ensure comparability among the results. To compare the performance a precision recall plot is used (Saito and Rehmsmeier, 2015). Recall and precision are calculated for each of the four seasons (MAM, JJA, SON, DJF) and for the two half years (MAMJJA and SONDJF) using the estimated scaling factors. Results of the calculation are shown in Fig. 6. The dotted line is marking the line under which the classifier shows no skill. The line is defined as a proportion of drought

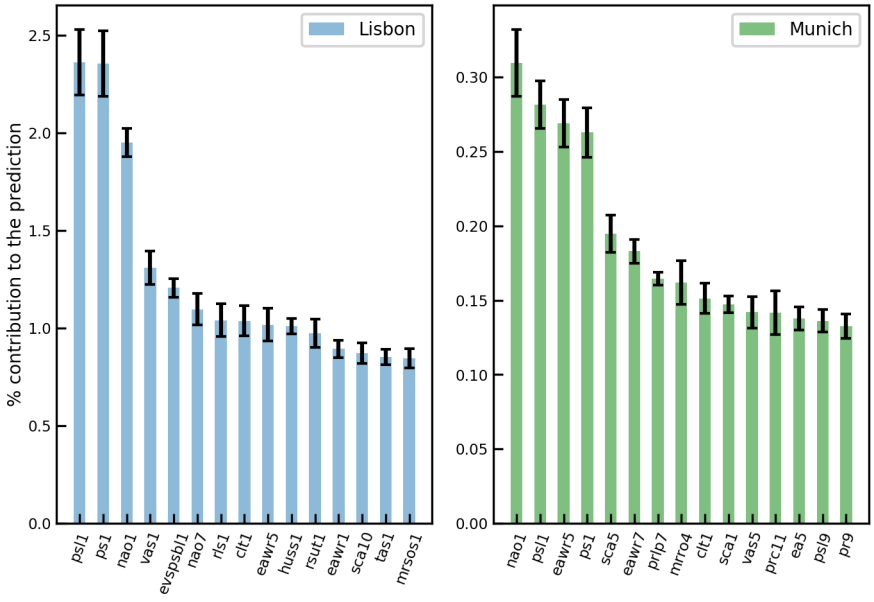

Figure 5. Mean Shapely values normalized to the contribution to the prediction for the top 15 variables with the highest importance for Lisbon (left) and Munich (right) on the test dataset. The number behind the variable name refers to the number of months before the event (NAO1 - NAO value one month before the predicted event). The results indicate that for the Lisbon domain *psl1* and *ps1* are the most influential drought predictors, for Munich this is NAO1.

events against overall amount of events (Saito and Rehmsmeier, 2015). For the Lisbon domain it becomes evident that the model performance is very different across seasons: higher precision of around 0.23 can be found during the winter half year. However for the spring season and summer half year the recall rises, while precision goes down. For the Munich classifier the results for the different seasons are closer together in terms of recall. It shows a worse performance for the winter months (DJF), while fall, spring and summer show an overall better model performance. This is an indication that for the Munich domain better drought predictability is possible in spring, fall and summer.

An additional analysis is conducted to calculate the Shapely values separately for the four season and the two domains in order to understand the influence of the different variables on the prediction. The results of the analysis can be seen in Fig. 7 and 8. The results for the Lisbon domain show that NAO1 is the strongest predictor in winter and spring season, while the contribution of pressure on drought predictability is higher in fall, followed by NAO1. On the contrary for the summer season NAO1 is not among the top 10 predictors, but other teleconnection indices like EAWR5, NAO7 and SCA7. Those teleconnection indices are originating from winter months where NAO showed to have the highest impact on the prediction. However, given the low performance of the model in the summer season, further investigation is needed. For the Munich domain NAO1 has one of the highest contributions for spring, summer and fall, while it cannot be found among the strongest predictors for winter. EAWR5 is one of the strongest predictors for summer, spring and fall. The feature contributions for

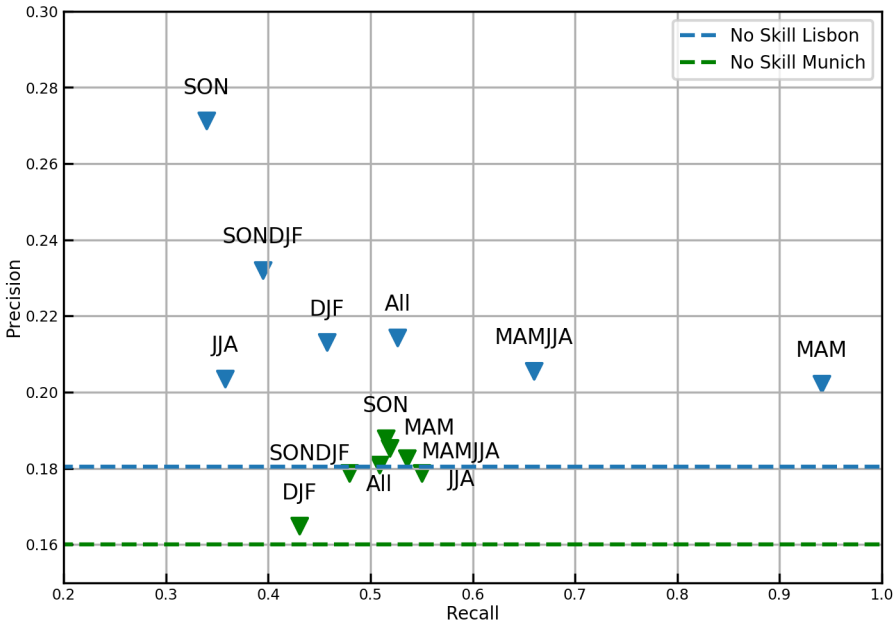

Figure 6. The effect of seasonality on precision and recall for Lisbon (blue) and Munich (green). The results indicate that for the Munich and Lisbon domain better drought predictability is possible in spring, fall and summer.

predictions in the winter season in Munich indicate that atmospheric variables 10 or 12 month before the event might be drought indicators.

4 Discussion and conclusion

Drought is a multiscale phenomenon and its formation and evolution is different to every climatology and season. In this study, we i) explored the possibilities of using the data provided by CRCM5-LE to predict droughts using ANN and ii) applied explainable AI methods to gain a better understanding of the results. A drought event is defined as a SPI1 less than -1 at the given site. The first half of the study deals with the systematic search for best performing models. For the Lisbon domain the model with L2-Regularization of 0.001, mean absolute error as loss function and the architecture 5000*0.5*4000*0.5*1000*500*100, where five layers are fully connected and two layers are Dropout layers, obtain the best results. For the Munich domain the model with L2-Regularization of 0.001, mean absolute error as loss function and the architecture 4000*0.5*1000*0.5*500*100*5, where five layers are fully connected and two layers are Dropout layers, obtain best results. Best performing models obtain accuracies of 57% for the Lisbon domain and 55% for the Munich domain.

The precision of the prediction in both cases is rather moderate, as a high percentage of data is misclassified. For Lisbon, classifier precision remains at around 22 %. This means that one out of four predicted drought events is an actual drought. For the Munich case, this ratio is even lower and amounts to 18 %. However, the models provide an important basis for the

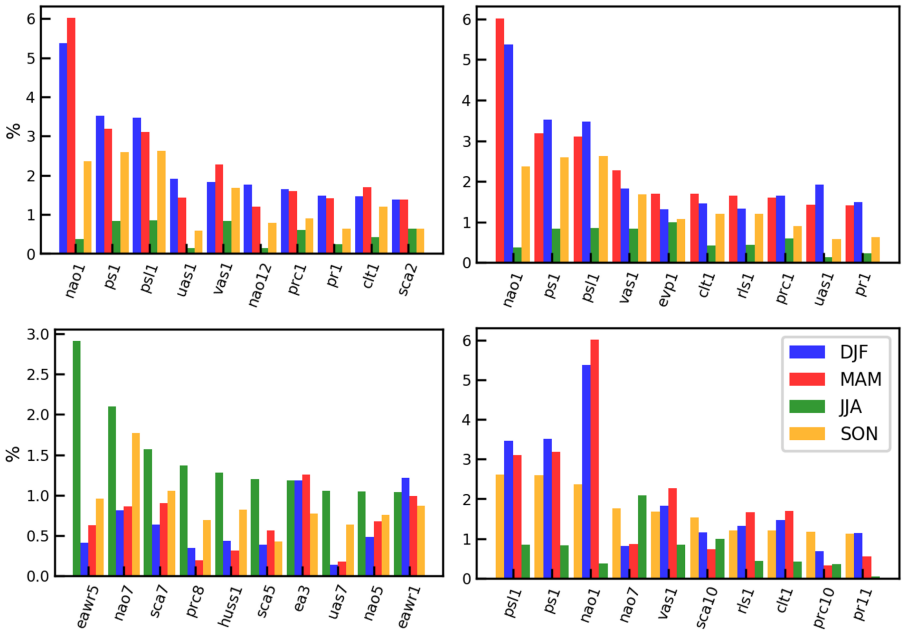

Figure 7. Shapely values for Lisbon calculated separately for the four seasons and sorted by the maximum contribution in DJF (top left), MAM (top right), JJA (bottom left) and SON (bottom right) for the test dataset. evspsbl abbreviated as evp.

development of future drought predicting models and offer a fruitful ground for the investigation of influence of single input variables during different seasons on drought formation.

Compared to the study by Santos et al. (2014), which investigated drought predictability in Portugal, the weak prediction accuracies of our study are not surprising. In Santos et al. (2014) SPI6 for April, May and June is predicted, however precipitation amounts for the months until March were also given as input. As SPI6 is calculated using the sum of 6 months precipitation, the model is receiving over the half of the information it needs for the calculation of the value. As no similar studies exist for the Munich domain, no comparison can be performed.

The second half of the study concentrates on the analysis of the obtained algorithms using explainable AI methods. Among the strongest predictors for the domains are NAO, *psl* and *ps* one month before the event. This underlines the importance of the atmospheric system on the drought formation. For the model trained for the Lisbon domain the variables of Northward Near-Surface Wind (*vas*) and Evaporation (*evspsbl*) followed. For the Munich domain, EAWR and SCA five month before the event are found among the strongest predictors. In general the percentages of the contribution of the strongest predictors for the Munich domain are around six times lower than those for the Lisbon domain.

This study indicates that seasonality is a crucial factor for drought predictions. Precision and recall of the prediction is getting lower in summer for the Lisbon domain and for winter for Munich domain. Moreover while for Munich domain the spread of precision and recall across the seasons is rather low, huge differences are found for Lisbon domain: the trained model obtained higher recall and lower precision for spring and higher precision and lower recall for fall when comparing to the baseline of

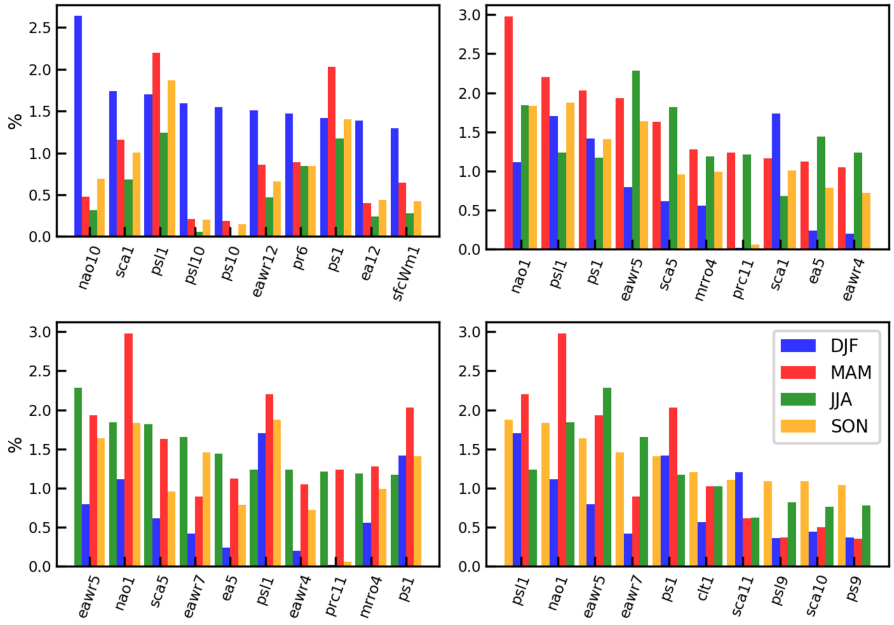

Figure 8. Shapely values for Munich calculated separately for the four seasons and sorted by the maximum contribution in DJF (top left), MAM (top right), JJA (bottom left) and SON (bottom right) for the test dataset. sfcWindmax abbreviated as sfcWm.

all data. The results show that for the Lisbon domain NAO1 is the strongest predictor in winter and spring season, while the contribution of pressure on drought predictability is higher in fall, followed by the contribution of NAO1. For Munich domain NAO1 is found to have one of the highest contributions for spring, summer and fall, while it could not be found among the ten strongest predictors for winter.

Further investigations are of interest for scientific research on both objectives. In terms of drought prediction, further research is possible within the same setting. The field of AI is evolving rapidly, showing new algorithms, methods and frameworks, such that there is a high potential for finding better suited algorithms (Hao, 2019). One of the main limitations of this study remains that an application of the obtained framework on observation data is not possible, due to the fact that observational data lacks a multitude of variables which are used as input in this study e.g. Heat Fluxes, radiation, etc. However the results obtained by shapely value calculation are of high importance for the choice of variables for a development of a future model which potentially could be applied to observational data. Given the high Shapely importance of NAO for drought prediction, other large scale variables, such as atmospheric blocking, can be added to the input variables. Moreover, the application to new domains is of interest to investigate the regionality of drought prediction possibilities. Explainable AI methods offer an important approach to improve the current limitations of machine learning models; their application is of high importance in the field of physical geography since it enables providing a physical interpretation to statistical results.

Data availability. Ensemble model data used in this study may be retrieved from the following sources: CanESM2-LE data are available via https://open.canada.ca/data/en/dataset/aa7b6823-fd1e-49ff-a6fb-68076a4a477c (Environment and Climate Change Canada, 2020). CRCM5-LE data can be retrieved at https://climex-data.srv.lrz.de/Public/ (Ouranos, 2020). The ERAInterim reanalysis data set was obtained at https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/ (European Centre for Medium-Range Weather Forecasts, 2020).

Author contributions. This study was conceptualized by EF under supervision of RL. Formal analysis, visualization of results and writing of the original draft was performed by EF. All authors contributed to the interpretation of the findings and revision of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The CRCM5-LE was created within the ClimEx project, which was funded by the Bavarian State Ministry for the Environment and Consumer Protection. Computations of the CRCM5-LE were made on the SuperMUC supercomputer at Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. We acknowledge Environment and Climate Change Canada for providing the CanESM2-LE driving data.

References

Barnston, A. G. and Livezey, R. E.: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns, MON WEATHER REV, 115, 1083–1126, https://doi.org/https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2, 1987.

Belayneh, A., Adamowski, J., Khalil, B., and Quilty, J.: Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction, ATMOS RES, 172, https://doi.org/10.1016/j.atmosres.2015.12.017, 2016.

Biesiada, J. and Duch, W.: Feature Selection for High-Dimensional Data — A Pearson Redundancy Based Filter, in: Computer Recognition Systems 2, edited by Kurzynski, M., Puchala, E., Wozniak, M., and Zolnierek, A., pp. 242–249, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

Bishop, C. M.: Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 1 edn., 2007.

Bonaccorso, B., Cancelliere, A., and Rossi, G.: Probabilistic forecasting of drought class transitions in Sicily (Italy) using Standardized Precipitation Index and North Atlantic Oscillation Index, J HYDROL, 526, 136–150, 2015.

Bueh, C. and Nakamura, H.: Scandinavian pattern and its climatic impact, Q J ROY METEOR SOC, 133, 2117–2131, https://doi.org/10.1002/qj.173, 2007.

Ceglar, A., Zampieri, M., Toreti, A., and Dentener, F.: Observed Northward Migration of Agro-Climate Zones in Europe Will Further Accelerate Under Climate Change, Earth's Future, 7, 1088–1101, https://doi.org/https://doi.org/10.1029/2019EF001178, 2019.

Chollet, F. et al.: Keras, https://keras.io, 2015.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S.: Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), 2016.

Dawson, A.: eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data, Journal of Open Research Software, 4, https://doi.org/10.5334/jors.122, 2016.

Deo, R. C., Kisi, O., and Singh, V. P.: Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model, Atmospheric Research, 184, 149–175, 2017.

E Trenberth, K.: Changes in Precipitation with Climate Change. Climate Change Research, CLIM RES, 47, 123–138, https://doi.org/https://doi.org/10.3354/cr00953, 2011.

Enfield, D. B., Mestas-Nuñez, A. M., and Trimble, P. J.: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S., GEOPHYS RES LETT, 28, 2077–2080, https://doi.org/10.1029/2000GL012745, 2001.

Federal Ministry of Food and Agriculture: Trockenheit und Dürre 2018 – Überblick über Maßnahmen, https://www.bmel.de/DE/Landwirtschaft/Nachhaltige-Landnutzung/Klimawandel/_Texte/Extremwetterlagen-Zustaendigkeiten.html, accessed: 25.07.2019, 2018.

Folland, C. K., Knight, J., Linderholm, H. W., Fereday, D., Ineson, S., and Hurrell, J. W.: The Summer North Atlantic Oscillation: Past, Present, and Future, J CLIMATE, 22, 1082–1103, https://doi.org/10.1175/2008JCLI2459.1, 2009.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, The MIT Press, 2016.

Guyon, I. and Elisseeff, A.: An Introduction to Variable and Feature Selection, JJ MACH LEARN RES, https://doi.org/https://doi.org/110.5555/944919.944968, 2003.

Hao, K.: We analyzed 16,625 papers to figure out where AI is headed next, https://www.technologyreview.com/2019/01/25/1436/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/, accessed: 21.03.2021, 2019.

Hao, Z., Singh, V. P., and Xia, Y.: Seasonal Drought Prediction: Advances, Challenges, and Future Prospects, REV GEOPHYS, 56, 108–141, https://doi.org/10.1002/2016RG000549, 2018.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M., eds.: The North Atlantic Oscillation: Climatic Significance and Environmental Impact, American Geophysical Union, 2003.

IPCC: Climate Change 2013 - The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the IPCC, Assessment report (Intergovernmental Panel on Climate Change).: Working Group, Cambridge University Press, 2013.

Janocha, K. and Czarnecki, W. M.: On Loss Functions for Deep Neural Networks in Classification, CoRR, abs/1702.05659, https://doi.org/https://doi.org/10.4467/20838476SI.16.004.6185, 2017.

John Keyantash, N.: The Climate Data Guide: Standardized Precipitation Index (SPI), https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi, accessed: 21.03.2021, 2018.

Kington, J. A.: Daily weather mapping from 1781:, CLIMATIC CHANGE, 3, 7–36, https://doi.org/10.1007/bf02423166, 1980.

Kirchmeier-Young, M., Zwiers, F., and Gillett, N.: Attribution of Extreme Events in Arctic Sea Ice Extent, J CLIMATE, 30, https://doi.org/10.1175/JCLI-D-16-0412.1, 2016.

Kumar, V. and Minz, S.: Feature selection: a literature review, SmartCR, 4, 211–229, 2014.

Kushner, P. J., Mudryk, L. R., Merryfield, W., Ambadan, J. T., Berg, A., Bichet, A., Brown, R., Derksen, C., Déry, S. J., Dirkson, A., Flato, G., Fletcher, C. G., Fyfe, J. C., Gillett, N., Haas, C., Howell, S., Laliberté, F., McCusker, K., Sigmond, M., Sospedra-Alfonso, R., Tandon, N. F., Thackeray, C., Tremblay, B., and Zwiers, F. W.: Canadian snow and sea ice: assessment of snow, sea ice, and related climate processes in Canada's Earth system model and climate-prediction system, The Cryosphere, 12, 1137–1156, https://doi.org/10.5194/tc-12-1137-2018, 2018.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), J APPL METEOROL CLIM, 58, 663–693, https://doi.org/10.1175/JAMC-D-18-0021.1, 2019.

Lim, Y.-K.: The East Atlantic/West Russia (EA/WR) teleconnection in the North Atlantic: climate impact and relation to Rossby wave propagation, CLIM DYNAM, Online first, https://doi.org/10.1007/s00382-014-2381-4, 2015.

Lundberg, S. and Lee, S.: A unified approach to interpreting model predictions, CoRR, abs/1705.07874, http://arxiv.org/abs/1705.07874, 2017.

Maas, A. L.: Rectifier Nonlinearities Improve Neural Network Acoustic Models, in: Proceedings of the 30thInternational Conference on Machine Learning, Atlanta, Georgia, USA, 2013.

Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., and Dugas, B.: Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation, CLIM DYNAM, 41, 2973–3005, https://doi.org/10.1007/s00382-013-1778-9, 2013.

McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., and Williams, J. K.: Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather, B AM METEOROL SOC, 98, 2073 – 2090, https://doi.org/https://doi.org/10.1175/BAMS-D-16-0123.1, 2017.

McKee, T., Doesken, N., and Kleist, J.: THE RELATIONSHIP OF DROUGHT FREQUENCY AND DURATION TO TIME SCALES, in: Proceedings of the 8th Conference on Applied Climatology. American Meteorological Society Boston, USA, 1993.

Mikhailova, N. and Yurovsky, A.: The East Atlantic Oscillation: Mechanism and Impact on the European Climate in Winter, Physical Oceanography, 4, https://doi.org/10.22449/1573-160X-2016-4-25-33, 2016.

Morid, S., Smakhtin, V., and Bagherzadeh, K.: Drought forecasting using artificial neural networks and time series of drought indices, INT J CLIMATOL, 27, 2103–2111, https://doi.org/10.1002/joc.1498, 2007.

Russell, S. and Norvig, P.: Artificial Intelligence: A Modern Approach, Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edn., 2009.

Saito, T. and Rehmsmeier, M.: The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, PLOS ONE, 10, 1–21, https://doi.org/10.1371/journal.pone.0118432, 2015.

Santos, J. F., Portela, M. M., and Pulido-Calvo, I.: Spring drought prediction based on winter NAO and global SST in Portugal, HYDROL PROCESS, 28, 1009–1024, https://doi.org/10.1002/hyp.9641, 2014.

Sasaki, Y.: The truth of the F-measure, Teach Tutor Mater, 2007.

Sheffield, J. and Wood, E. F.: Drought : past problems and future scenarios , London ; Washington, DC : Earthscan, 2011.

Sheffield, J., Andreadis, K. M., Wood, E. F., and Lettenmaier, D. P.: Global and Continental Drought in the Second Half of the Twentieth Century: Severity–Area–Duration Analysis and Temporal Variability of Large-Scale Events, J CLIMATE, 22, 1962–1981, https://doi.org/10.1175/2008JCLI2722.1, 2009.

Spinoni, J., Naumann, G., Vogt, J., and Barbosa, P.: Meteorological droughts in Europe events and impacts: past trends and future projections, OCLC: 1076014255, 2016.

Spinoni, J., Naumann, G., and Vogt, J. V.: Pan-European seasonal trends and recent changes of drought frequency and severity, GLOBAL PLANET CHANGE, 148, 113–130, https://doi.org/https://doi.org/10.1016/j.gloplacha.2016.11.013, 2017.

von Trentini, F., Aalbers, E. E., Fischer, E. M., and Ludwig, R.: Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe, EARTH SYST DYNAM, 11, 1013–1031, https://doi.org/10.5194/esd-11-1013-2020, 2020.

World Meteorological Organization: Standardized Precipitation Index User Guide, 2012.

Yoon, J.-H., Mo, K., and Wood, E. F.: Dynamic-model-based seasonal prediction of meteorological drought over the contiguous United States, Journal of Hydrometeorology, 13, 463–482, 2012.

Zargar, A., Sadiq, R., Naser, B., and Khan, F. I.: A review of drought indices, Environmental Reviews, 19, 333–349, 2011.