

Predicting agricultural drought indicators: ML approaches across wide-ranging climate and land use conditions

Jung-Ching Kan^{a,b}, Carla S.S. Ferreira^{b,c,*}, Georgia Destouni^b, Pan Haozhi^d, Marlon Vieira Passos^e, Karina Barquet^e, Zahra Kalantari^{a,b}

^a Department of Sustainable Development, Environmental Science and Engineering, Sustainability Assessment and Management, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

^b Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, SE-106 91 Stockholm, Sweden

^c Research Centre for Natural Resources, Environment and Society (CERNAS), Polytechnic Institute of Coimbra, Coimbra Agrarian Technical School, Bencanta 3045 Coimbra, Portugal

^d School of Public and International Affairs, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, China

^e Stockholm Environment Institute (SEI), Stockholm, Sweden

ARTICLE INFO

Keywords:

Drought
Soil moisture
Palmer drought severity index
Climate
Land use
Machine learning
Sweden

ABSTRACT

Agricultural drought can severely reduce crop yields, lead to large economic losses and health impacts. Combined climate and land use variations determine key indicators of agricultural drought, including soil moisture and the Palmer drought severity index (PDSI). This study investigated the use of machine learning (ML) methods for predicting these indicators over Sweden, spanning steep climate and land use gradients. Three data arrangement methods (multi-features, temporal, and spatial) were used and compared in combination with seven ML/deep learning (DL) models (random forest (RF), decision tree, multivariate linear regression, support vector regression, autoregressive integrated moving average (AMIRA), artificial neural network, and convolutional neural network). Seven investigated features, obtained from Google Earth Engine, were used in the ML/DL modeling (soil moisture, PDSI, precipitation, evapotranspiration, elevation, slope and soil texture). The temporal ARIMA model (found most suitable for local scale prediction) and the multi-features RF model (more suitable for national-scale prediction) emerged as best performing for soil moisture prediction (with MAE of 9.1 and 11.95, and R² of 0.79 and 0.59, respectively). All models generally performed better in predicting the soil moisture than the PDSI indicator of drought. For drought indicator prediction and mapping, previous-year average monthly soil moisture emerged as the most important feature, combined with the four additional corresponding features of PDSI, precipitation, evapotranspiration and elevation.

1. Introduction

Drought is a hazard with devastating impacts on agriculture and the ecological environment (Jiang et al., 2021; Zhong et al., 2019). The frequency and severity of extreme drought events are increasing worldwide due to global warming (Jiang et al., 2021; Mokhtar et al., 2021). For example, Europe has experienced several extreme droughts in recent decades, with one of the worst to date being recorded during summer 2022 (Copernicus, 2022). Drought is also associated with land

use and land cover variations (Hishe et al., 2021). In combination, climate and weather variability (Orth and Destouni, 2018) and variations and changes in land use (e.g., in rainfed and/or irrigated agriculture) affect the hydrological cycle (Destouni et al., 2013; Kåresdotter et al., 2022) and associated soil moisture and drought events (Destouni and Verrot, 2014; Orth et al., 2020) with disruptive societal and ecosystem impacts (Aguilos et al., 2021; Deng et al., 2016; Yohannes et al., 2021).

Agricultural drought, which is the focus of this study, has major

Abbreviations: ANN, Artificial neural network; AR, Autoregressive; ARIMA, Autoregressive integrated moving average; CNN, Convolutional neural network; DL, Deep learning; DT, Decision tree; GEE, Google earth engine; MA, Moving average; MAE, Mean absolute error; ML, Machine learning; MLP, Multilayer perceptrons; MLR, Multivariate linear regression; PDSI, Palmer Drought Severity Index; RF, Random forest; RFE, Recursive feature elimination; SPEI, Standardised precipitation-evapotranspiration index; SPI, Standardised precipitation index; SVR, Support vector regression.

* Corresponding author.

E-mail address: carla.ferreira@natgeo.su.se (C.S.S. Ferreira).

<https://doi.org/10.1016/j.ecolind.2023.110524>

Received 5 January 2023; Received in revised form 13 April 2023; Accepted 12 June 2023

Available online 21 June 2023

1470-160X/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

impacts on plant development and is one of the major causes of productivity losses in agroecosystems. Agricultural drought accounts for >40% of all economic losses due to global meteorological disasters (Liu et al., 2016), and can lead to social and political conflicts, especially in developing regions such as Africa (Martínez-Fernández et al., 2016; WFP, 2022).

Research on agricultural drought has increased significantly in the past decade (Liu et al., 2016), with a main focus on arid- and semi-arid regions (Aghelpour et al., 2021; Feng et al., 2019), while cold-humid regions often have been overlooked (Blyverket et al., 2019). However, drought in the early summer has become a recurring event also in, e.g., the cold-humid Nordic region, along with snowfall reduction during winter and heatwaves during summer (Mustafa et al., 2022). For example, the Nordic region experienced extreme drought in summer 2018, causing a 40% reduction in cereal yield (Beillouin et al., 2020). Accurate prediction of such drought occurrences is essential, for example in early warning systems that can support drought preparedness and decision-making for impact prevention and mitigation.

Remote sensing has been proposed as a good data source for agricultural drought monitoring, as it provides spatial and temporal information at the regional scale (Liu et al., 2016). Various indices have been used to quantify drought and its magnitude, such as standardised precipitation index (SPI), standardised precipitation-evapotranspiration index (SPEI) and Palmer drought severity index (PDSI) (Yihdego et al., 2019). Among these, PDSI is widely used (Yihdego et al., 2019) and considered suitable for assessing impacts of, e.g., climate change on future drought hazard (Mishra and Singh, 2010). Soil moisture is also a main indicator of drought that could be used for agricultural drought monitoring (Samaniego et al., 2013; Orth and Destouni, 2018; Orth et al., 2020).

In recent years, artificial intelligence (AI), using recent rapid increases in open data availability, computational power, and Machine Learning (ML) algorithms, has been widely applied for prediction of natural hazards (Lei et al., 2021; Panahi et al., 2022), as well as droughts (Zhang et al., 2019; Rahmati et al., 2020). Multivariate linear regression (MLR) has been reported to achieve high accuracy in estimating meteorological drought index in Korea (Kim et al., 2020). Support vector regression (SVR) has been reported to achieve reliable prediction of, e.g., SPEI in Australia (Deo et al., 2018) and China (Tian et al., 2018), and SPI in Morocco (El Ibrahimi & Baali, 2018). Decision tree (DT) was effective in weather prediction in Hong Kong (Chauhan & Thakur, 2014). Random forest (RF) has been widely applied in remote sensing research and has been reported to exhibit good performance in predicting drought at different sites such as Korea and Iran (Park et al., 2019; Lotfirad et al., 2022). Artificial neural network (ANN) has shown high accuracy in predicting SPI in Ethiopia (Mokhtarzad et al., 2017). Convolutional neural network (CNN) is reported to outperform other deep learning algorithms (AlexNet and VGGNet) in drought prediction in Ethiopia (Chaudhari et al., 2021). ML and deep learning (DL) approaches have emerged as effective methods for revealing complex relationships and hidden nonlinear trends in this context, such as causal inference (LeCun et al., 2015).

However, the ML and DL application in drought prediction is still at early stages, for example with the following two main problems remaining to be resolved. First, data-driven models, such as ML and DL ones, are sensitive to dataset structure, with different arrangement methods leading to different results (Gudivada et al., 2017). Existing data arrangement methods include the multi-features method (Hanadé Houmma et al., 2022), spatial method (Sardar et al., 2021) and temporal method (Xu et al., 2022), with their efficiency remaining to be comparatively assessed for drought prediction. Second, research so far has been insufficient in comparatively analyzing the importance and selecting the most appropriate drought prediction features (Sundararajan et al., 2021).

The overall aim of this study was such comparative analysis and assessment for identifying appropriate ML-based methodology for

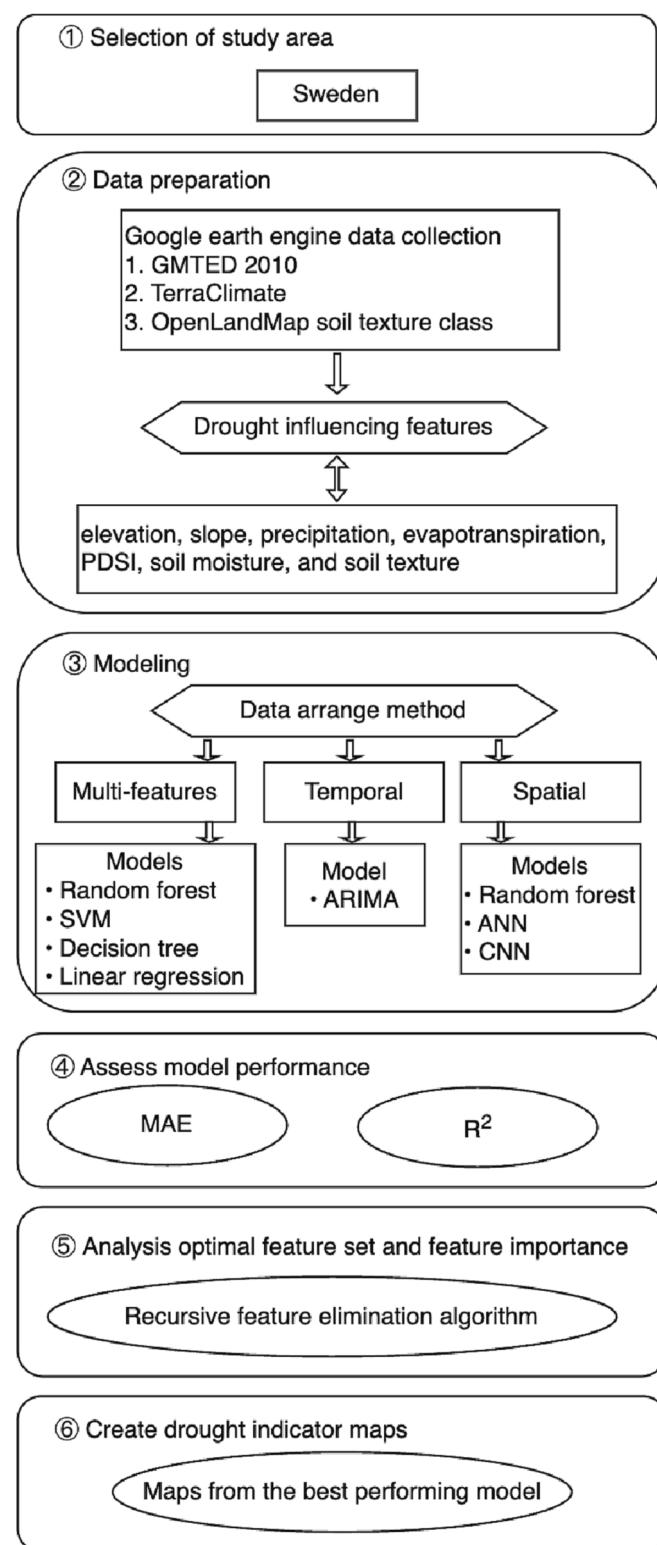


Fig. 1. Flowchart of the six steps in the methodological framework.

predicting PSDI and soil moisture as key indicators of agricultural drought and associated hazard. As a relevant regional case study for comparative application and assessment of ML methods and models, we used Sweden considering the steep and wide-ranging hydro-climatic and land use gradients spanning the country (Van der Velde et al., 2013). Specific objectives were to (i) build ML and DL models based on existing remote sensing data; (ii) identify differences between data arrangement

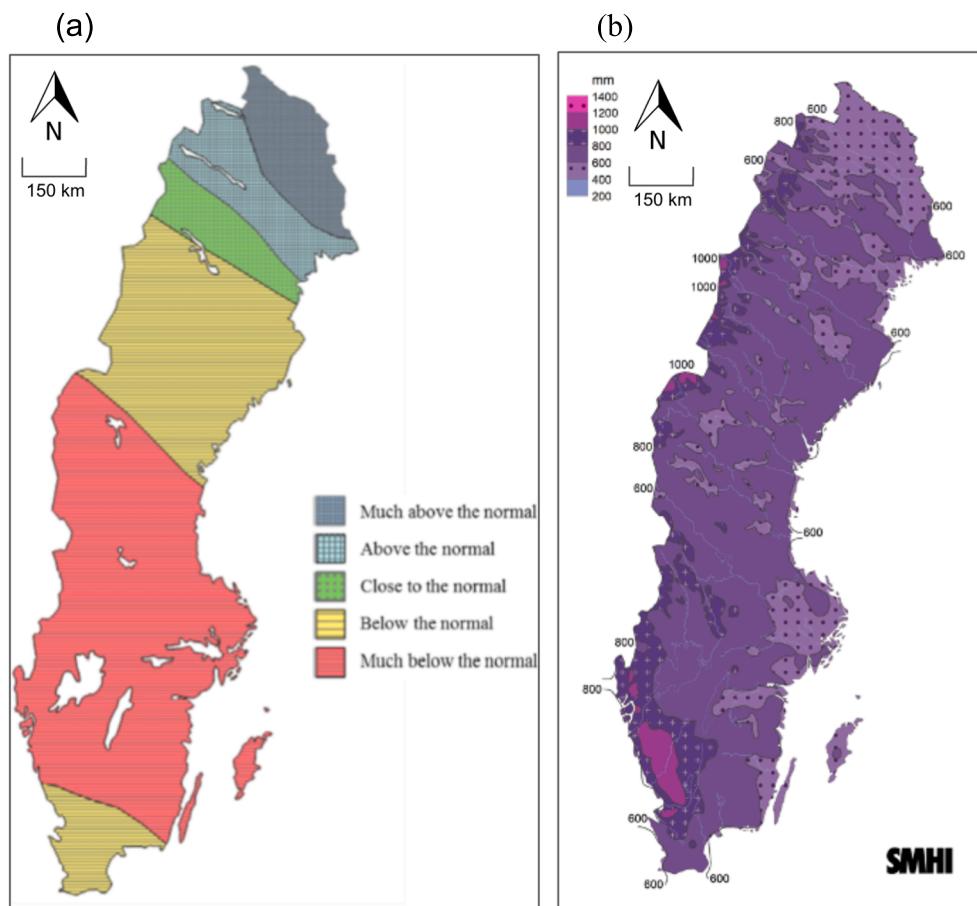


Fig. 2. Maps of Sweden showing (a) groundwater levels in August 2017 compared with a typical year (adapted from Campana et al., 2018) and (b) normal annual precipitation in the period 1991–2020 (adapted from SMHI, 2022).

methods (i.e., multi-features, spatial, temporal) in the ML/DL models; (iii) evaluate model performance in predicting the considered agricultural drought indicators; and (iv) assess the relative importance of various features in the predictions of PDSI and soil moisture as target indicators for agricultural drought.

2. Materials and methods

The method applied comprised six steps (Fig. 1): 1) selection of study area; 2) data preparation on drought-related factors; 3) drought indicator modeling using three data arrangement methods; 4) assessment of model performance; 5) investigation of best feature set and feature importance; and 6) production of drought indicator maps.

2.1. Study area

The selected study area was Sweden, with its wide coverage and steep gradients of hydro-climatic and land use conditions (Van der Velde et al., 2013) in northern Europe ($55\text{--}69^{\circ}\text{N}$; $11\text{--}24^{\circ}\text{E}$). Sweden is a country with commonly sufficient water resources, in terms of both surface water and groundwater, but the extremely dry summer of 2017 led to a severe groundwater decline (Fig. 2a), and municipalities had to limit water consumption and provide emergency water supply (Barthel et al., 2021). The Swedish agricultural system is vulnerable to increasing weather variations along with climate change since the majority of agriculture is rainfed (Fig. 2b) and only around 2% of arable land is irrigated (Grusson et al., 2021). The Swedish Meteorological and Hydrological Institute (SMHI) has projected that future precipitation will be concentrated to autumn and winter, while spring and summer face a

Table 1

Drought features and prediction targets considered in this study, along with associated resolution and data source for each feature/target. PDSI: Palmer drought severity index.

Data	Usage	Resolution	Google Earth Engine dataset
Elevation	Feature	225 m	GMTED2010: Global Multi-resolution Terrain Elevation Data 2010
Slope	Feature	225 m	GMTED2010: Global Multi-resolution Terrain Elevation Data 2010
Precipitation	Feature	4638.3 m	TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces
Evapotranspiration	Feature	4638.3 m	TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces
Soil texture	Feature	250 m	OpenLandMap Soil Texture Class (USDA System)
Soil moisture	Feature/Target	4638.3 m	TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces
PDSI	Feature/Target	4638.3 m	TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces

precipitation reduction (Rummukainen et al., 2004). A precipitation deficit in the growing season and heavy precipitation in the harvesting season may have significant impacts on reduced crop yields (Horn et al., 2022).

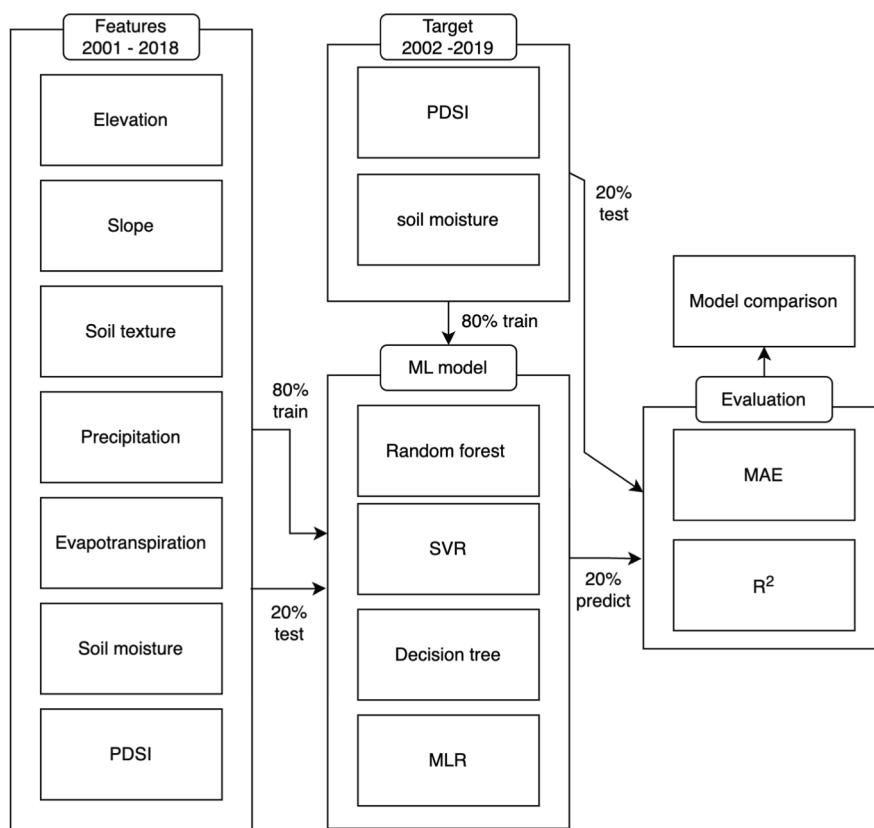


Fig. 3. Flowchart of model training and evaluation in the multi-features method.

2.2. Input data for the models

Seven possible main factors were considered and investigated regarding their contributions to drought conditions in this study (Table 1). Elevation, land slope and soil texture (represented by categorical data) are static features (characteristic parameters), while precipitation, evapotranspiration, soil moisture and PDSI are temporally dynamic features (variables), related to both weather/climate and land use conditions and their variations and changes. Soil moisture and PDSI are not only used as features but also targets with a year time shift in between, for example, soil moisture in 2001 is used as feature to predict the soil moisture in 2002 target. Data on these features for the period 2001–2020 were taken from Google Earth Engine (GEE) (<https://developers.google.com/earth-engine/datasets>). GEE is an open-source cloud-computing web-based platform that has been widely applied in drought studies and enables users to easily acquire, process, and visualize data in a specific area (Sazib et al., 2018).

2.3. Data arrangement methods

Three widely applied data arrangement methods (multi-features, spatial, temporal) were used to arrange the data derived from GEE as comparative alternatives for predicting two drought indicators targets (soil moisture and PDSI). The multi-features method uses all selected features in Table 1 to predict the target (Rahmati et al., 2020). The spatial method uses one selected feature value from one pixel and adjacent pixels to predict the corresponding target (Trebing et al., 2021). The temporal method focuses on the history pattern of oscillation and the input training data includes the selected feature time series data for prediction of corresponding target (Box et al., 2015). The detailed explanation of each method is described further in the following subsections.

2.3.1. Multi-features method

The multi-features method was used to extract correlations in each spatial pixel between the considered seven features (elevation, slope, soil texture, precipitation, evapotranspiration, soil moisture, PDSI) and two drought indicator targets for later-time prediction (PDSI or soil moisture). The model input feature was the previous-year monthly pixel values of all features listed in Table 1, and the target output value was the same pixel's value in the same month of the following year. For example, feature input was one pixel's elevation, slope, soil texture, precipitation, evapotranspiration, soil moisture, and PDSI in January 2001 and the corresponding output target would be the same pixel's PDSI or soil moisture value in January 2002. The feature value inputs comprised data between 2001 and 2018, and the corresponding output target from 2002 to 2019. These data groups (2001–2018 and 2002–2019) were split into 80% for training the selected multi-features ML models and 20% for testing the models (Fig. 3).

2.3.2. Spatial method

The spatial method was used to extract the potential spatial value relationships around nearby pixels (Alzubaidi et al., 2021). The input feature values are the previous-year monthly values of the target in the surrounding 5×5 pixels (in total 25 values) and the target output value is the central pixel in the same month of the following year. For instance, feature input was one pixel's and 24 adjacent pixels' soil moisture in January 2001 and the corresponding output target was the center pixel's soil moisture in January 2002. The time scale and data group splitting ratio for training and testing in the multi-features method was the same used in section 2.3.1.

2.3.3. Temporal method

The temporal method was used to extract relevant temporal variation patterns for each pixel (Box et al., 2015). Each pixel has a model, and the model is trained with selected feature's whole time series data,

Table 2

Overview of the data arrangement methods and models used in the study.

Multi-features method	Extract correlations between features in each pixel	MLR: linear regression with predictors SVR: provides a general optimal solution DT: creates data subsets based on features RF: averages values of various tree predictions RF: averages values of various tree predictions ANN: simplified model of neural network CNN: explores spatial relationship ARIMA: predicts time-related information
Spatial method	Extract the potential value relationships of nearby pixels	
Temporal method	Extract the temporal variation pattern in a time series	

and output the corresponding target. For example, the time series model was trained with soil moisture data recorded at a pixel from 2001 to 2019 and predicted the soil moisture values in 2020.

2.4. Machine learning and deep learning models

Seven ML/DL models were selected based on the literature review (see section 1) and the suitability of their characteristics to fit the data arrangement method. See the overview in Table 2 and a brief summary of each model given below.

2.4.1. Models used with the multi-features method

The multi-features method is suitable for models that can deal with several features (Hanadé Houmma et al., 2022). The selected models were firstly fitted with the preprocessing training dataset and then evaluated with the test set (see section 2.3.1).

2.4.1.1. Multivariate linear regression (MLR). MLR is one of the most common statistical method models using multiple features to predict a target (Tajik et al., 2012). It defines the coefficient for each feature to find the linear equation that can describe the feature-target relationship.

2.4.1.2. Support vector regression (SVR). SVR grounds in statistical learning theory (Besalatpour et al., 2012). It maps the input feature in a high dimensional space and finds the hyperplane that fits the data (Smola & Schölkopf, 2004). SVR minimizes the regression error which is the distance between true value predicted value, and the margin error which is the distance between hyperplane and the data (Smola & Schölkopf, 2004).

2.4.1.3. Decision tree (DT). DT is a tree-like model consisting of nodes and branches breaking down datasets into subsets based on features (Navada et al., 2011; Nourani & Molajou, 2017; Song and Ying, 2015). DT selects the feature that maximizes the reduction of variance from target at each node in the tree (Navada et al., 2011).

2.4.1.4. Random forest (RF). RF is an ensemble model based on DT that operates by randomly constructing a multitude of independent and uncorrelated DT (Breiman, 2001; Azizi et al., 2022). RF increases the overall stability by randomly sampled subsets, and the predicted value is the average prediction of trees (Azizi et al., 2022).

2.4.2. Models used with the spatial method

The spatial method is suitable for models that can interpret spatial information (Sardar et al., 2021). Three algorithms were tested with the spatial method, namely ANN, CNN, and RF (see section 2.4.1.4.). The preprocessing training data (see section 2.3.2) was used to train the selected model, and then testing data was used to test model's

performance.

2.4.2.1. Artificial neural network model (ANN). ANN is an information process system that structures and operates analogously to a human brain (Zolfaghari et al., 2015). Multilayer perceptron neural network (MLP), a type of ANN model, was applied in this study. It has a feed-forward structure that neuron receives information from the previous layer and then passes to the next layer after applying the activation function (Khan et al., 2020; Maier et al., 2010).

2.4.2.2. Convolutional neural network model (CNN). CNN is a deep learning model considering spatial information, and thus commonly used for image and video recognition (Alzubaidi et al., 2021). CNN consists of convolutional layers with filters to extract the local features (Danandeh et al., 2022), and the pooling layers downsample the outputs of convolutional layers to reduce dimensions and increase compute efficiency (Danandeh et al., 2022).

2.4.3. Model used with the temporal method

The temporal method requires availability of data time series (Box et al., 2015). The data between 2001 and 2019 at each pixel (see section 2.3.3) was used to train the selected time series model (ARIMA). Trained model predicted the target value in 2020 and is compared with the observed value in 2020.

2.4.3.1. Autoregressive integrated moving average model (ARIMA). ARIMA is a popular time series model based on statistical theory (Han et al., 2010). ARIMA is the integration of autoregressive model (AR) and moving average model (MA). AR depends on the past value while MA captures the moving average of the history value (Box et al., 2015). ARIMA requires stationary data, and the integration component makes the nonstationary data stationary (Box et al., 2015).

2.5. Performance evaluation of all ML/DL models

Model performance was assessed using two metrics, mean absolute error (MAE) and coefficient of determination (R^2). MAE measures the mean of model-produced errors (Elavarasan et al., 2018), whereas R^2 examines the correlation and difference between the observed and the predicted values, with values ranging between 0 and 1. The R^2 coefficient has the advantage of indicating the goodness of fit between two datasets, but both evaluation metrics have been widely used and reported as relevant in drought forecasting studies (e.g., Sundararajan et al., 2021).

$$MAE = \frac{1}{N} \sum_{i=1}^N |D_p^i - D_A^i| \quad (1)$$

whereas D_p^i is the actual value collected from GEE, D_A^i is model predicted value and N the number of data.

The models used with the multi-features method and the spatial method were validated by 20% of the data group with the other 80% used to train the model, while those used with the temporal method model were validated with predictions for the year 2020. In order to compare model performance between these three methods, the best-performing model used with the multi-features and spatial methods was trained with previous year data groups (2001–2018 and 2002–2019) and validated with the prediction for 2020.

2.6. Analysis of the importance of features and the optimal set of features

To analyze the importance of the considered features, the recursive feature elimination (RFE) algorithm was applied to the best-performing ML model with the multi-features method. RFE is a recursive process that ranks features based on their importance (Granitto et al., 2006). It

Table 3

Model performance in prediction of Palmer drought severity index (PDSI) and soil moisture based on the 20% test data groups. MAE: mean absolute error, RF: Random Forest, SVR: Support vector regression, MLR: Multivariate linear regression, DT: Decision tree, ANN: Artificial neural network, CNN: Convolutional neural network.

Data arrangement method	Model	PDSI		Soil moisture	
		MAE	R ²	MAE	R ²
Multi-features	RF	1.40	0.46	6.62	0.87
	SVR	1.94	0.10	9.32	0.74
	MLR	2.12	0.02	11.77	0.74
	DT	1.65	0.05	7.78	0.75
Spatial	ANN	2.05	0.02	11.8	0.62
	CNN	2.07	0.01	17.9	0.40
	RF	1.93	0.14	10.46	0.71

starts with the complete set of features, then recursively prunes the least important feature and runs with the smaller pruned set of features (Zhou et al., 2014). It identifies redundant and weak features so as to retain the strong and independent ones (Chen & Jeong, 2007). The analysis improves model performance by removing weak, noisy and redundant features.

2.7. Drought indicator mapping

Drought indicator maps were created for the summer months of 2020 (June, July and August) using the best-performing model. The summer season is associated with the highest drought risk in the study region and predictions were made for 2020 to best utilize the collected data. To create the maps, the best-performing model was fed with the specific monthly data for Sweden in 2019 and produced indicator predictions for the corresponding months in 2020 as outputs.

3. Results

3.1. Model performance

The performance of the ML/DL models using the multi-features and spatial methods is shown in Table 3. Of the four models used with the multi-features method, RF achieved the best predictive performance, with the lowest MAE and highest R², for both PDSI (MAE = 1.40, R² =

0.46) and soil moisture (MAE = 6.62, R² = 0.87). DT was the second-best model and MLR showed the worst performance. Of the three models used with the spatial method, RF again achieved the best predictive performance based on MAE (1.93 for PDSI prediction, 10.46 for soil moisture prediction) and R² (0.14 for PDSI and 0.71 for soil moisture), while CNN had the worst performance. Overall, all models exhibited much better performance in predicting soil moisture than PDSI, with the multi-features method having higher accuracy (MAE < 10, R² > 0.7) than the spatial method (MAE > 10, R² ≤ 0.7) in the soil moisture prediction. Both data arrangement methods had equally low performance in PDSI prediction (R² < 0.46). Fig. 4 shows the results of the soil moisture models with the bottom right corner has lower MAE and higher R². Overall, the multi-features RF was the best performing model among all multi-features and spatial arrangement models tested (Fig. 4).

Table 4 further shows the model test results for predicted drought indicators in 2020, with much better performance exhibited in predicting soil moisture than PDSI also for the models considered in this comparison. For soil moisture prediction, the ARIMA model used with the temporal method (MAE = 9.10 and R² = 0.79) outperformed the multi-features RF (MAE = 11.95 and R² = 0.59); this also applies for the overall less good PDSI prediction (with MAE = 1.85 and R² = 0.24 for ARIMA, and MAE = 2.03 and R² = 0.13 for RF). Fig. 5 exemplifies the ARIMA model predictions, showing that they do not follow well the variations in PDSI implied by the observed data (Fig. 5a, in consistency with the corresponding low R² in Table 4), while the predicted data fit the observed data well for soil moisture (Fig. 5b, in consistency with the corresponding relatively high R² in Table 4). Overall, ARIMA thus captures relatively well the temporal patterns in the soil moisture time series.

Table 4

Model performance based on predictions of PDSI and soil moisture for 2020. MAE: mean absolute error, PDSI: Palmer drought severity index, RF: Random Forest, ARIMA: Autoregressive integrated moving average model.

Data arrangement method	Model	PDSI		Soil moisture	
		MAE	R ²	MAE	R ²
Multi-features	RF	2.03	0.13	11.95	0.59
Temporal	ARIMA	1.85	0.24	9.10	0.79

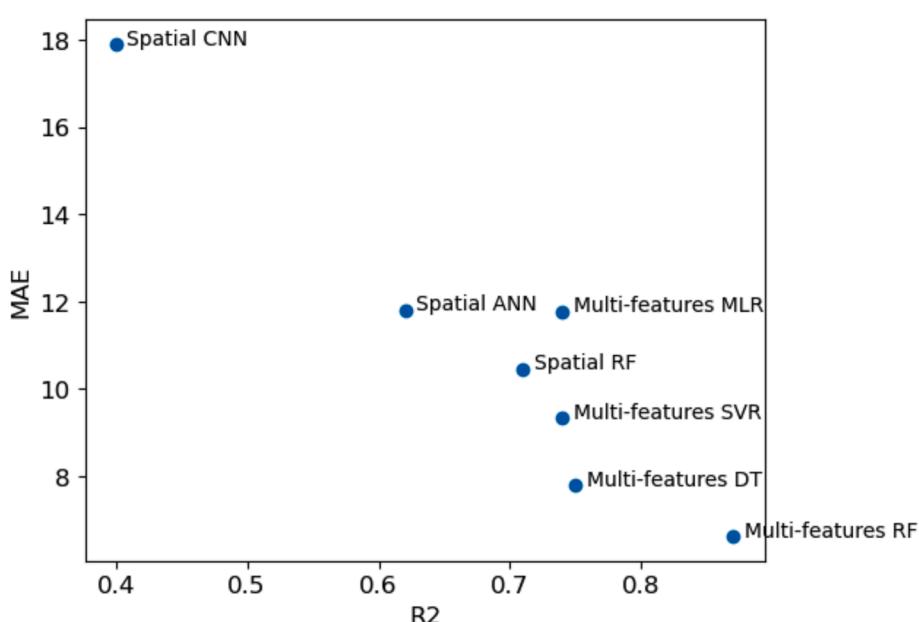


Fig. 4. Comparison between results from the seven soil moisture models: mean absolute error (MAE) vs. R².

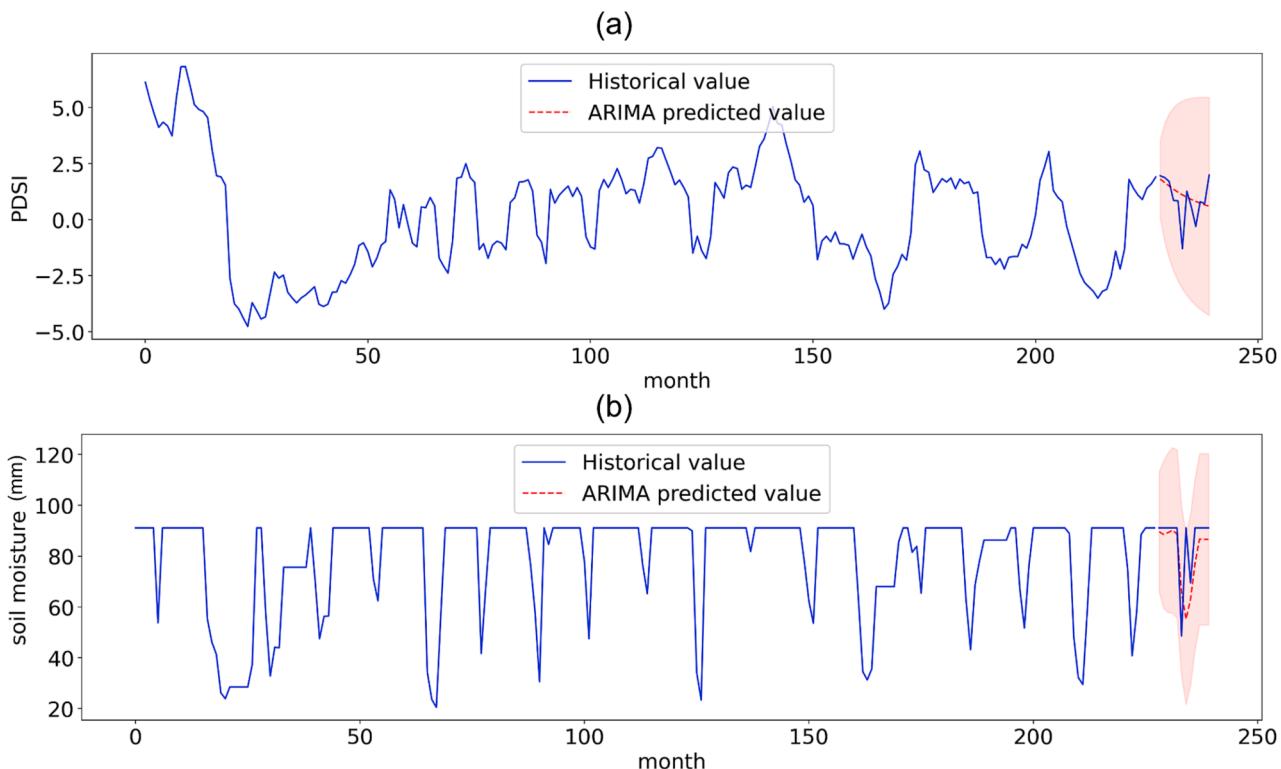


Fig. 5. ARIMA model predictions for 2020 of (a) Palmer drought severity index (PDSI) and (b) soil moisture. The blue line shows historical data and the red line predicted values. The pink area indicates the 95% probability range. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5

Feature ranking obtained using recursive feature elimination (RFE) applied to the multi-features random forest (RF) model for prediction of soil moisture and the Palmer drought severity index (PDSI).

Feature	Rank
Soil moisture	1
Palmer drought severity index, PDSI	2
Precipitation	3
Evapotranspiration	4
Elevation	5
Slope	6
Soil texture	7

3.2. Optimal set of features and feature importance

The feature ranking results of using RFE for the multi-features RF model prediction of both soil moisture and PDSI as targets are the same which summarizes in Table 5. Soil moisture on the same month of the previous year emerges as the most important feature for prediction of the considered drought indicators, while corresponding PDSI is the second most important, and the remaining monthly variable features, i.e., precipitation and evapotranspiration, rank third and fourth, the topographical features, i.e., elevation and slope, rank fifth and sixth, and soil texture ranks as the least important feature (Table 5).

To further identify the optimal feature set among those investigated, the feature inputs (numbered according to ranking) are further plotted against the values of the evaluation metrics (MAE and R^2) for PDSI and soil moisture prediction by the multi-features RF model (Fig. 6). All plots show a similar pattern in which the model reaches the highest performance with five features, lower performance with six features and also somewhat lower performance with all seven features. Soil moisture, PDSI, precipitation, evapotranspiration, and elevation constitute the optimal feature set for the multi-features RF, with maximized model

performance.

3.3. Drought indicator mapping

The best-performing model, ARIMA, required high computational power to account for the whole data time series at each pixel and was therefore not suitable for spatial mapping over the whole of Sweden. Therefore, the second best-performing model, the multi-features RF, was used to predict the spatial distribution of soil moisture over Sweden in the summer months of 2020. The maps created (Fig. 7) present the distribution of soil moisture as the overall best model-predicted indicator of agricultural drought (Tables 3–4) based on the optimal set of five features (see section 3.2.). As seen in Fig. 7, the predicted soil moisture maps show good visual agreement with actual soil moisture maps based on remote sensing data.

4. Discussion

4.1. Comparison of data arrangement methods

All three data arrangement methods were applied to the same dataset in this study. The temporal method showed the best performance in ML/DL prediction of soil moisture as a key drought indicator, followed by the multi-features method and then the spatial method. The superior performance of the temporal method reveals the importance of time series correlations for mid-term predictions. Time series data have been reported previously as effective for short-term and mid-term prediction, e.g., of sea surface (Xiao et al., 2019) and electric power load (Lee and Hong, 2015). However, computational costs can be a major limitation in generating spatial drought indicator maps using the temporal method (Al-Turjman and Baali, 2022), since each pixel needs to have its own predictive model and, in this study, there were 89,964 pixels over Sweden. Hence, the temporal method may be more appropriate for

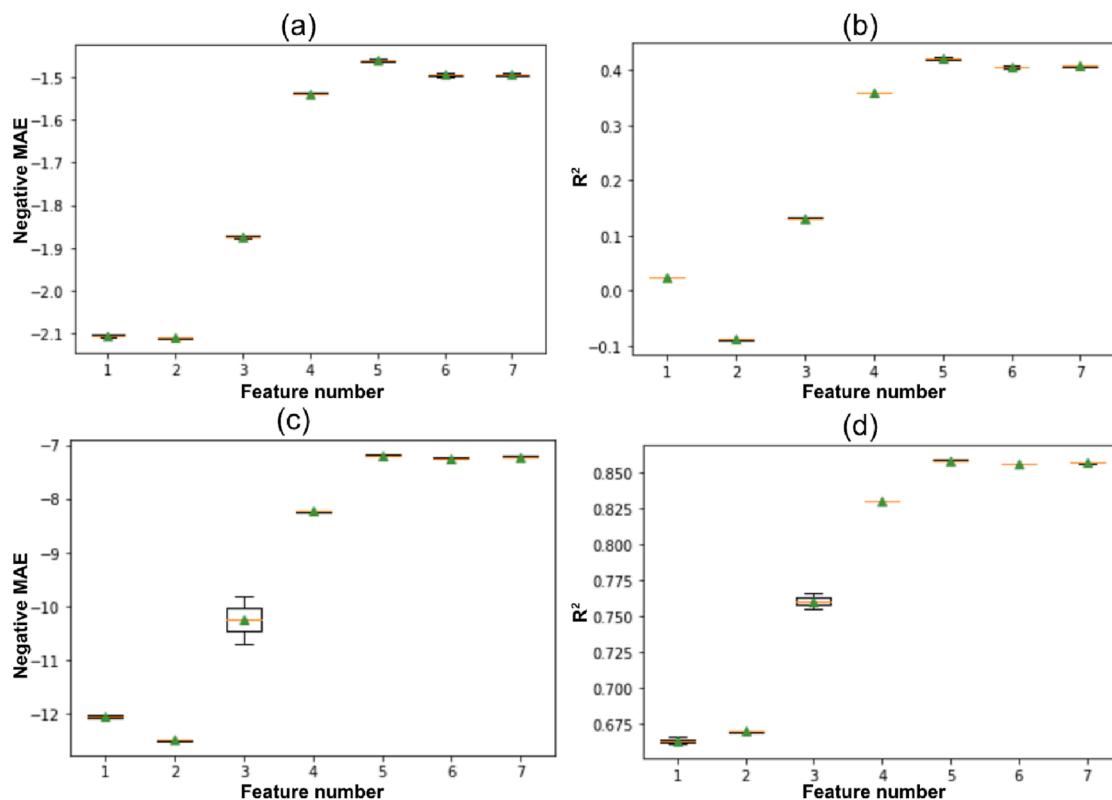


Fig. 6. Plots of feature number (as ranked in Table 5) in the multi-features random forest (RF) model against: (a) negative mean absolute error (MAE) for the Palmer drought severity index (PDSI) prediction; (b) R^2 values for PDSI prediction; (c) negative MAE values for soil moisture prediction; and (d) R^2 values for soil moisture prediction.

prediction at local rather than whole national scale.

The multi-features RF method has also emerged as effective in forecasting the overall best-modeled soil moisture indicator of drought. It has the advantages of good spatial prediction ability and relatively good predictive skill (over 70% fitness for soil moisture prediction), in consistency with previous findings of the multi-features method as effective for extreme event prediction with various ML models (Hanadé Houmoua et al., 2022).

The spatial method models show weaker predictive ability, possibly because spatial correlations between pixels is small at 5-km resolution. In previous studies, CNN has been reported to be effective for drought prediction based on high-resolution images (Chaudhari et al., 2021) and soil moisture prediction with 1-km resolution data (Leonarduzzi et al., 2022). Raziei et al. (2013) also found that finer spatial data resolution could better capture spatial drought variability in Iran. Further research is needed for how spatial resolution affects the performance of the spatial method in drought forecasting.

The drought features used in the study were based on remote sensing data derived from GEE, so the methodology can also be transferred and applied in other world regions, with relevant associated feature modification. The good performance of the temporal and multi-features methods in the study region provides guidance for the application of such ML/DL models elsewhere.

4.2. Assessment of model performance

The performance of seven ML/DL models with three data arrangement methods for predicting agricultural drought indicators in Sweden was tested using MAE and R^2 as evaluation metrics. Some similarities laid in between the models applied in the study. RF and DT were both tree-based models. MLR and SVR were linear regression models while ANN and CNN were deep learning models based on neural network. The

results obtained suggested ARIMA as the best-performing among all the models in predicting the considered drought indicators. This confirms previous findings by Mossad and Alazba (2015), who tested the performance of ARIMA in drought forecasting using SPEI and found good drought prediction ability over different time scales. The poor performance of ARIMA in predicting PDSI may be due to the non-stationary PDSI historical record data, with trend and pattern changes over time difficult to be captured by ARIMA. Furthermore, ARIMA has its limitation in capturing the complex interactions between various factors that is crucial for PDSI as a drought index.

RF showed the second-best performance in drought indicator prediction. This model has advantages of simplicity, robustness and versatility (Lotfirad et al., 2022). The good performance of RF is also consistent with conclusions in the latest review of multivariate agricultural drought models, which identified RF as the best model (Hanadé Houmoua et al., 2022).

Overall, much higher R^2 values were obtained here for the soil moisture (>0.5) than the PDSI (<0.3) prediction, indicating soil moisture as the best and most readily model-predicted indicator of drought in the Swedish region, spanning steep climate and land use gradients. The PDSI drought index is more complex and other features than those investigated here may be needed for better model performance in predicting this drought indicator. PDSI is normally calculated based on precipitation, temperature and available water capacity data (Tufaner & Özbelayaz, 2020), with our study representing temperature and available water capacity only indirectly, as reflected in the evapotranspiration and soil texture features, respectively.

4.3. Relevance of features in drought indicator prediction

Further analysis of feature-target relationships can improve understanding of model behavior and enhance model performance. In the RFE

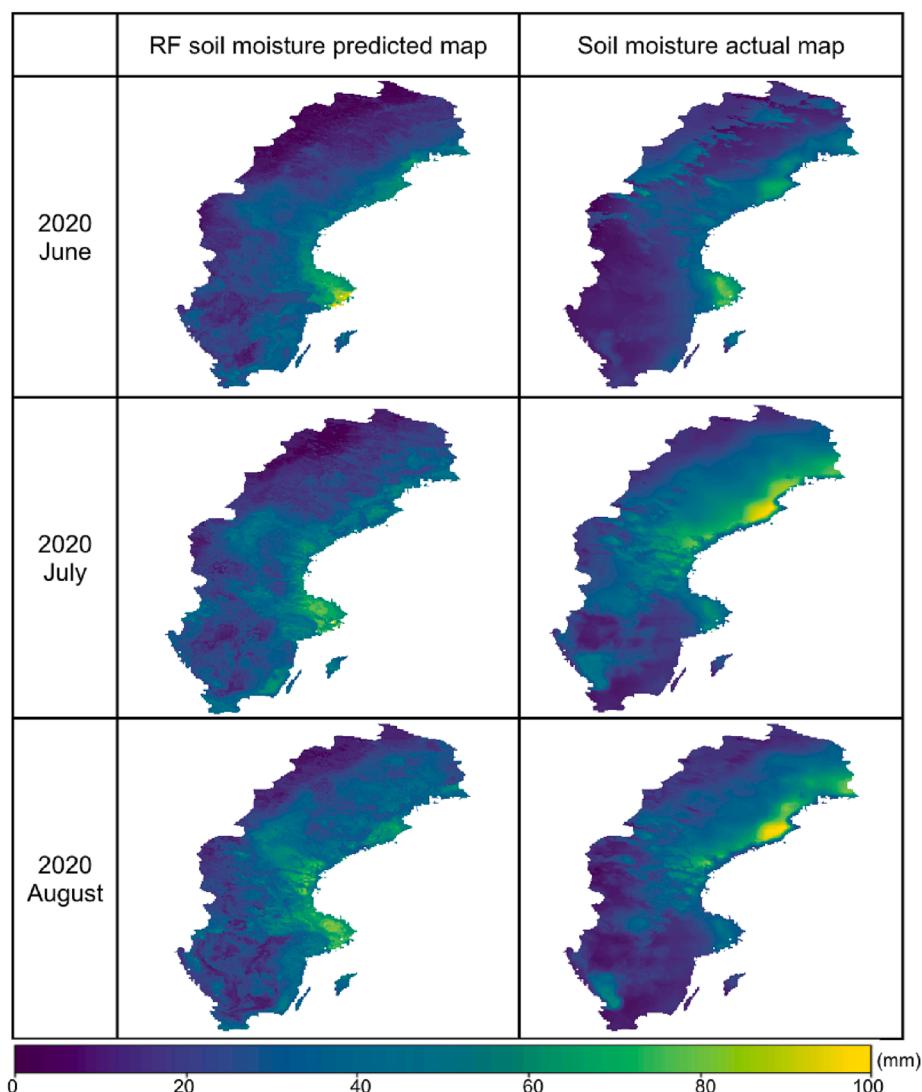


Fig. 7. Monthly soil moisture maps over Sweden for the summer months (June, July, August) of 2020 based on: (left) predictions obtained using the multi-features random forest (RF) model; and (right) actual remote sensing data for soil moisture from Google Earth Engine.

ranking results, the variable input features (precipitation, evapotranspiration, soil moisture, PDSI) emerged as more important than the static ones (elevation, slope, soil texture). Soil moisture was the most crucial feature and soil texture was the least important among the seven features investigated. Soil moisture and associated droughts relate to both weather/climate and land use variations and changes (Orth et al., 2020), which can substantially affect vegetation growth (Su & Shangguan, 2019), as mentioned depending on prevailing climate and land use conditions (Orth et al., 2020), and has also been regarded as a central variable for drought and drought-impact assessment in other recent studies (Orth and Destouni, 2018; Zhang et al., 2022). Even though the spatial variations of soil texture do not affect much model prediction of the dynamic temporal and spatial variations of soil moisture and PDSI in the study, soil texture, having strong implication on land use, relates to soil fragility and average soil water availability (Vinhal-Freitas et al., 2017).

The feature selection approach (RFE) used in this study is reported to be more accurate and robust than other feature selection techniques, such as correlation analysis (Kavhu et al., 2021). The RFE approach is effective in identifying the most informative feature set (Chen et al., 2018) and has been applied in feature selection for soil water storage mapping and agricultural crop yield prediction (Nketia et al., 2022; Elavarasan et al., 2020). The RF model achieved its best performance

with five features (soil moisture, PDSI, precipitation, evapotranspiration, elevation) out of the seven investigated. Model performance has slightly decreased by adding slope feature. Infiltration rate and groundwater recharge, connected through soil moisture conditions, are typically higher in flatter and lower-elevation regions than in more steeply sloping and higher-elevation areas (Shekhar & Pandey, 2015), with elevation and slope then having similar and overlapping topographical effects (Danasingh & Epiphany, 2020). This may have made the slope feature redundant on top of the elevation feature in the ML models.

The uncertainty of ML model performance was related to three major aspects which were model selection, feature selection, and overfitting. In order to address the uncertainty of model, seven popular ML/DL models have been tested. The performance of selected models was validated with independent data in 2019 to test model's ability to generate new data. Seven commonly used features were selected based on literature review and tested by RFE to minimize feature uncertainty. However, feature selection process in this study was limited to the remote sensing data available on the GEE platform and by the computing power needed to calculate large datasets. Further research should address a general lack of a standardized feature selection system for multi-features modeling of agricultural drought, with >84 different features currently used in this context (Hanadé Houmما et al., 2022). An important

feature selection development could be to apply RFE for identification of an optimal feature set among all such currently used features data.

5. Conclusions

This study showed the temporal data arrangement method in the ARIMA model as most effective for prediction of local agricultural drought indicators in the Swedish region. The required high computing power required for this model makes it less suitable for use over the whole national scale. The second best performing multi-features RF model is an appropriate choice for national-scale prediction of the spatial distribution of drought indicators over Sweden, with a set of five features (soil moisture, PDSI, precipitation, evapotranspiration, elevation) maximizing the performance of this model for drought indicator prediction.

The contributions of this study are novel in directly comparing state-of-the-art data arrangement methods in ML/DL modeling and their performance in predicting key agricultural drought indicators for wide-ranging climate and land use conditions prevailing across Sweden. With global data widely available in GEE platform and open source of ML models, the proposed methodology is applicable in other countries to test the suitability of drought indicator prediction method. Accurate drought prediction is needed to support water resource management and decision making from local to national scale. The task is challenging, however, with drought events being determined by complex relationships of both dynamic (temporally and spatially variable) and static (spatially variable) factors, and their short-term co-variation and long-term co-evolution under ongoing and future climate and land use changes. Conventional drought forecasting methods, such as linear regression, are limited and cannot fully meet the challenge and handle related data and model uncertainties. In support of overcoming these barriers, the present study has investigated and identified a well-performing and transferable ML/DL-based methodology for predicting key indicators for agricultural drought under spatiotemporally variable climate and land use conditions.

CRediT authorship contribution statement

Jung-Ching Kan: Formal analysis, Methodology, Writing – original draft. **Carla S.S. Ferreira:** Methodology, Supervision, Validation, Writing – review & editing. **Georgia Destouni:** Writing – review & editing. **Pan Haozhi:** Writing – review & editing. **Marlon Vieira Passos:** Writing – review & editing. **Karina Barquet:** Writing – review & editing. **Zahra Kalantari:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

We gratefully acknowledge funding from the Swedish Research Council (VR) for projects: Science for a secure society: Hydro-climatic hazard, risk, and crisis management in Sweden (CrisAct) (grant 2021-06309); and Coupled freshwater system variations, trends and their drivers around the world (grant 2022-04672).

References

- Aghelpour, P., Mohammadi, B., Mehdizadeh, S., Bahrami-Pichaghchi, H., Duan, Z., 2021. A novel hybrid dragonfly optimization algorithm for agricultural drought prediction. *Stoch. Env. Res. Risk A.* 35 (12), 2459–2477. <https://doi.org/10.1007/s00477-021-02011-2>.
- Aguilos, M., Sun, G.e., Noormets, A., Domec, J.-C., McNulty, S., Gavazzi, M., Minick, K., Mitra, B., Prajapati, P., Yang, Y., King, J., 2021. Effects of land-use change and drought on decadal evapotranspiration and water balance of natural and managed forested wetlands along the southeastern US lower coastal plain. *Agric. For. Meteorol.* 303, 108381.
- Al-Turjman, F., Baali, I., 2022. Machine learning for wearable IoT-based applications: A survey. *Trans. Emerg. Telecommun. Technol.* 33 (8), e3635.
- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y.e., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 8 (1). <https://doi.org/10.1186/s40537-021-00444-8>.
- Azizi, K., Ayoubi, S., Nabibollahi, K., Garosi, Y., Gislum, R., 2022. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. *J. Geochem. Explor.* 233, 106921. <https://doi.org/10.1016/j.geexplo.2021.106921>.
- Barthel, R., Stangefelt, M., Giese, M., Nygren, M., Seftigen, K., Chen, D., 2021. Current understanding of groundwater recharge and groundwater drought in Sweden compared to countries with similar geology and climate. *Geogr. Ann. Ser. B* 103 (4), 323–345. <https://doi.org/10.1080/04353676.2021.1969130>.
- Beillouin, D., Schauberger, B., Bastos, A., Ciais, P., Makowski, D., 2020. Impact of extreme weather conditions on European crop production in 2018. *Philos. Trans. R. Soc. B* 375 (1810), 20190510. <https://doi.org/10.1098/rstb.2019.0510>.
- Besalatpour, A., Hajabbasi, M.A., Ayoubi, S., Gharipour, A., Jazi, A.Y., 2012. Prediction of soil physical properties by optimized support vector machines. *Int. Agrophys.* 26 (2) <https://doi.org/10.2478/v10247-012-0017-7>.
- Blyverket, J., Hamer, P.D., Schneider, P., Albergel, C., Lahoz, W.A., 2019. Monitoring soil moisture drought over northern high latitudes from space. *Remote Sens. (Basel)* 11 (10), 1200. <https://doi.org/10.3390/rs11101200>.
- Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time series analysis: forecasting and control. John Wiley & Sons.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:101093404324>.
- Campana, P.E., Zhang, J., Yao, T., Andersson, S., Landelius, T., Melton, F., Yan, J., 2018. Managing agricultural drought in Sweden using a novel spatially-explicit model from the perspective of water-food-energy nexus. *J. Clean. Prod.* 197, 1382–1393. <https://doi.org/10.1016/j.jclepro.2018.06.096>.
- Chaudhari, S., Sardar, V., Rahul, D.S., Chandan, M., Shivakale, M.S., Harini, K.R., 2021. Performance analysis of CNN, alexNet and VGGNet models for drought prediction using satellite images. In: 2021 Asian Conference on innovation in Technology (ASIANCON). IEEE, pp. 1–6. <https://doi.org/10.1109/ASIANCON51346.2021.9545068>.
- Chauhan, D., Thakur, J., 2014. Data mining techniques for weather prediction: A review. *Int. J. Recent Innov. Trends Comput. Commun.* 2 (8), 2184–2189.
- Chen, X.W., Jeong, J.C., 2007. Enhanced recursive feature elimination. In: Sixth International Conference on Machine Learning and Applications (ICMLA 2007). IEEE, pp. 429–435. <https://doi.org/10.1109/ICMLA.2007.35>.
- Chen, Q., Meng, Z., Liu, X., Jin, Q., Su, R., 2018. Decision variants for the automatic determination of optimal feature subset in RF-RFE. *Genes* 9 (6), 301. <https://doi.org/10.3390/genes9060301>.
- Copernicus, Copernicus: Summer 2022 Europe's hottest on record <https://climate.copernicus.eu/copernicus-summer-2022-europe-s-hottest-record> 2022 accessed 19 October 2022.
- Danandeh Mehr, A., Rikhtehgar Ghiasi, A., Yaseen, Z.M., Sorman, A.U., Abualigah, L., 2022. A novel intelligent deep learning predictive model for meteorological drought forecasting. *J. Ambient Intell. Hum. Comput.* 1–15 <https://doi.org/10.1007/s12652-022-03701-7>.
- Danasingh, A.A.G.S., Epiphany, J.L., 2020. Identifying redundant features using unsupervised learning for high-dimensional data. *SN Appl. Sci.* 2 (8), 1–10. <https://doi.org/10.1007/s42452-020-3157-6>.
- Deng, L., Yan, W., Zhang, Y., Shangguan, Z., 2016. Severe depletion of soil moisture following land-use changes for ecological restoration: Evidence from northern China. *For. Ecol. Manage.* 366, 1–10.
- Deo, R.C., Salcedo-Sanz, S., Carro-Calvo, L., Saavedra-Moreno, B., 2018. Drought prediction with standardized precipitation and evapotranspiration index and support vector regression models. In: Integrating disaster science and management. Elsevier, pp. 151–174. <https://doi.org/10.1016/B978-0-12-812056-9.00010-5>.
- Destouni, G., Verrot, L., 2014. Screening long-term variability and change of soil moisture in a changing climate. *J. Hydrol.* 516, 131–139. <https://doi.org/10.1016/j.jhydrol.2014.01.059>.
- Destouni, G., Jaramillo, F., Prieto, C., 2013. Hydroclimatic shifts driven by human water use for food and energy production. *Nat. Clim. Change.* 3 (3), 213–217. <https://doi.org/10.1038/nclimate1719>.
- El Ibrahim, A., Baali, A., 2018. Application of several artificial intelligence models for forecasting meteorological drought using the standardized precipitation index in the saïss plain (Northern Morocco). *Int. J. Intell. Eng. Syst.* 11 (1), 267–275. <https://doi.org/10.22266/ijies2018.0228.28>.
- Elavarasan, D., Vincent PM, D. R., Srinivasan, K., & Chang, C. Y. 2020. A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling. *Agriculture*, 10(9), 400. <https://doi.org/10.3390/agriculture10090400>.

- Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y., Srinivasan, K., 2018. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Comput. Electron. Agric.* 155, 257–282. <https://doi.org/10.1016/j.compag.2018.10.024>.
- Feng, P., Wang, B., Li Liu, D., Yu, Q., 2019. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agr. Syst.* 173, 303–316. <https://doi.org/10.1016/j.agry.2019.03.015>.
- Granitto, P.M., Furlanello, C., Biasioli, F., Gasperi, F., 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemom. Intel. Lab. Syst.* 83 (2), 83–90. <https://doi.org/10.1016/j.chemolab.2006.01.007>.
- Grusson, Y., Wessström, I., Joel, A., 2021. Impact of climate change on Swedish agriculture: Growing season rain deficit and irrigation need. *Agric. Water Manage.* 251, 106858. <https://doi.org/10.1016/j.agwat.2021.106858>.
- Gudivada, V., Apon, A., Ding, J., 2017. Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *Int. J. Adv. Software* 10 (1), 1–20.
- Han, P., Wang, P.X., Zhang, S.Y., Zhu, D.H., 2010. Drought forecasting based on the remote sensing data using ARIMA models. *Math. Comput. Model.* 51 (11–12), 1398–1403.
- Hanadé Houmma, I., El Mansouri, L., Gadal, S., Garba, M., Hadria, R., 2022. Modelling agricultural drought: a review of latest advances in big data technologies. *Geomat. Nat. Haz. Risk* 13 (1), 2737–2776. <https://doi.org/10.1080/19475705.2022.2131471>.
- Hishe, H., Giday, K., Van Orshoven, J., Muys, B., Taheri, F., Azadi, H., Witlox, F., 2021. Analysis of land use land cover dynamics and driving factors in Des'a forest in Northern Ethiopia. *Land Use Policy* 101, 105039.
- Horn, B., Ferreira, C., Kalantari, Z., 2022. Links between food trade, climate change and food security in developed countries: A case study of Sweden. *Ambio* 51 (4), 943–954. <https://doi.org/10.1007/s13280-021-01623-w>.
- Jiang, W., Wang, L., Zhang, M., Yao, R., Chen, X., Gui, X., Sun, J., Cao, Q., 2021. Analysis of drought events and their impacts on vegetation productivity based on the integrated surface drought index in the Hanjiang River Basin, China. *Atmos. Res.* 254, 105536.
- Kåresdotter, E., Destouni, G., Ghajarnia, N., Lammers, R. B., & Kalantari, Z. 2022. Distinguishing Direct Human-Driven Effects on the Global Terrestrial Water Cycle. *Earth's Future*, 10(8), e2022EF002848. <https://doi.org/10.1029/2022EF002848>.
- Kavhu, B., Mashimbye, Z.E., Luvuno, L., 2021. Climate-based regionalization and inclusion of spectral indices for enhancing transboundary land-use/cover classification using deep learning and machine learning. *Remote Sens. (Basel)* 13 (24), 5054. <https://doi.org/10.3390/rs13245054>.
- Khan, M.M.H., Muhammad, N.S., El-Shafie, A., 2020. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *J. Hydrol.* 590, 125380 <https://doi.org/10.1016/j.jhydrol.2020.125380>.
- Kim, S.W., Jung, D., Choung, Y.J., 2020. Development of a multiple linear regression model for meteorological drought index estimation based on landsat satellite imagery. *Water* 12 (12), 3393. <https://doi.org/10.3390/w12123393>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444.
- Lee, W.J., Hong, J., 2015. A hybrid dynamic and fuzzy time series model for mid-term power load forecasting. *Int. J. Electr. Power Energy Syst.* 64, 1057–1062. <https://doi.org/10.1016/j.ijepes.2014.08.006>.
- Lei, X., Chen, W., Panahi, M., Falah, F., Rahmati, O., Uuemaa, E., Kalantari, Z., Ferreira, C.S.S., Rezaie, F., Tiefenbacher, J.P., Lee, S., Bian, H., 2021. Urban flood modeling using deep-learning approaches in Seoul, South Korea. *J. Hydrol.* 601, 126684.
- Leonarduzzi, E., Tran, H., Bansal, V., Hull, R.B., De la Fuente, L., Bearup, L.A., Melchior, P., Condon, L.E., Maxwell, R.M., 2022. Training machine learning with physics-based simulations to predict 2D soil moisture fields in changing climate. *Front. Water* 4, 927113. <https://doi.org/10.3389/frwa.2022.927113>.
- Liu, X., Zhu, X., Pan, Y., Li, S., Liu, Y., Ma, Y., 2016. Agricultural drought monitoring: Progress, challenges, and prospects. *J. Geog. Sci.* 26 (6), 750–767. <https://doi.org/10.1007/s11442-016-1297-9>.
- Lotfifrad, M., Esmaeili-Gisavandani, H., Adib, A., 2022. Drought monitoring and prediction using SPI, SPEI, and random forest model in various climates of Iran. *J. Water Clim. Change* 13 (2), 383–406. <https://doi.org/10.2166/wcc.2021.287>.
- Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* 25 (8), 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>.
- Martínez-Fernández, J., González-Zamora, A., Sánchez, N., Gumuzzio, A., Herrero-Jiménez, C.M., 2016. Satellite soil moisture for agricultural drought monitoring: Assessment of the SMOS derived Soil Water Deficit Index. *Remote Sens. Environ.* 177, 277–286. <https://doi.org/10.1016/j.rse.2016.02.064>.
- Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. *J. Hydrol.* 391 (1–2), 202–216. <https://doi.org/10.1016/j.jhydrol.2010.07.012>.
- Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., Abdo, H.G., Sammen, S.S., Gyasi-Agyei, Y., Rodrigo-Comino, J., 2021. Estimation of SPEI meteorological drought using machine learning algorithms. *IEEE Access* 9, 65503–65523.
- Mokhtarzad, M., Eskandari, F., Jamshidi Vanjani, N., Arabasadi, A., 2017. Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environ. Earth Sci.* 76 (21), 1–10. <https://doi.org/10.1007/s12665-017-7064-0>.
- Mossad, A., Alazba, A.A., 2015. Drought forecasting using stochastic models in a hyper-arid climate. *Atmos.* 6 (4), 410–430. <https://doi.org/10.3390/atmos6040410>.
- Mustafa, S.M.T., Ghag, K., Panchanathan, A., Dahal, B., Ahrari, A., Liedes, T., Haghghi, A.T., 2022. Smart drainage management to limit summer drought damage in Nordic agriculture under the circular economy concept. *Hydrol. Process.* e14560. <https://doi.org/10.22541/au.164743416.62268619/v1>.
- Navada, A., Ansari, A.N., Patil, S., Sonkamble, B.A., 2011. Overview of use of decision tree algorithms in machine learning. In: 2011 IEEE control and system graduate research colloquium. IEEE, pp. 37–42. <https://doi.org/10.1109/ICSGRC.2011.5991826>.
- Nketia, K.A., Asabere, S.B., Ramcharan, A., Herbold, S., Erasmi, S., Sauer, D., 2022. Spatio-temporal mapping of soil water storage in a semi-arid landscape of northern Ghana—A multi-tasked ensemble machine-learning approach. *Geoderma* 410, 115691. <https://doi.org/10.1016/j.geoderma.2021.115691>.
- Nourani, V., Molajou, A., 2017. Application of a hybrid association rules/decision tree model for drought monitoring. *Global Planet. Change* 159, 37–45. <https://doi.org/10.1016/j.gloplacha.2017.10.008>.
- Orth, R., Destouni, G., 2018. Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe. *Nat. Commun.* 9 (1), 1–8. <https://doi.org/10.1038/s41467-018-06013-7>.
- Orth, R., Destouni, G., Jung, M., Reichstein, M., 2020. Large-scale biospheric drought response intensifies linearly with drought duration in arid regions. *Biogeosciences* 17 (9), 2647–2656. <https://doi.org/10.5194/bg-17-2647-2020>.
- Panahi, M., Rahmati, O., Kalantari, Z., Darabi, H., Rezaie, F., Moghaddam, D.D., Ferreira, C.S.S., Foody, G., Aliramaee, R., Bateni, S.M., Lee, C.-W., Lee, S., 2022. Large-scale dynamic flood monitoring in an arid-zone floodplain using SAR data and hybrid machine-learning models. *J. Hydrol.* 611, 128001.
- Park, H., Kim, K., Lee, D.K., 2019. Prediction of severe drought area based on random forest: Using satellite image and topography data. *Water* 11 (4), 705. <https://doi.org/10.3390/w11040705>.
- Rahmati, O., Falah, F., Dayal, K.S., Deo, R.C., Mohammadi, F., Biggs, T., Moghaddam, D. D., Naghibi, S.A., Bui, D.T., 2020. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Sci. Total Environ.* 699, 134230.
- Rummukainen, M., Bergström, S., Persson, G., Rodhe, J., Tjernström, M., 2004. The Swedish regional climate modelling programme, SWECLIM: a review. *AMBIENT: A Journal of the Human. Environ.* 33 (4), 176–182. <https://doi.org/10.1579/0044-7447-33.4.176>.
- Samaniego, L., Kumar, R., Zink, M., 2013. Implications of parameter uncertainty on soil moisture drought analysis in Germany. *J. Hydrometeorol.* 14 (1), 47–68. <https://doi.org/10.1175/JHM-D-12-075.1>.
- Sardar, V.S., Yindumathi, K.M., Chaudhari, S.S., Ghosh, P., 2021. Convolution Neural Network-based Agriculture Drought Prediction using Satellite Images. In: 2021 IEEE Mysore Sub Section International Conference (MysuruCon). IEEE, pp. 601–607. <https://doi.org/10.1109/MysuruCon52639.2021.9641531>.
- Sazib, N., Mladenova, I., Bolten, J., 2018. Leveraging the google earth engine for drought assessment using global soil moisture data. *Remote Sens. (Basel)* 10 (8), 1265. <https://doi.org/10.3390/rs10081265>.
- Shekhar, S., Pandey, A.C., 2015. Delineation of groundwater potential zone in hard rock terrain of India using remote sensing, geographical information system (GIS) and analytic hierarchy process (AHP) techniques. *Geocarto Int.* 30 (4), 402–421. <https://doi.org/10.1080/10106049.2014.894584>.
- SMHI. Normal annual precipitation for the period 1991–2020. https://www.smhi.se/pd_klimat/time_period_maps/normal/Nbd_Periodnormal/Nbd_Periodnormal_1991_2020.ar.png (accessed 10 November 2022).
- Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. *Stat. Comput.* 14 (3), 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.
- Song, Y.Y., Ying, L.U., 2015. Decision tree methods: applications for classification and prediction. *Shanghai Arch. Psychiatry* 27 (2), 130. <https://doi.org/10.11919/j.issn.1002-0829.215044>.
- Su, B., Shangguan, Z., 2019. Decline in soil moisture due to vegetation restoration on the Loess Plateau of China. *Land Degrad. Dev.* 30 (3), 290–299.
- Sundararajan, K., Garg, L., Srinivasan, K., Bashir, A.K., Kalaiappan, J., Ganapathy, G.P., Meena, T., 2021. A contemporary review on drought modeling using machine learning approaches. *CMES-Computer Modeling in Engineering and Sciences* 128 (2), 447–487. <https://doi.org/10.32604/cmes.2021.0155>.
- Tajik, S., Ayoubi, S., Nourbakhsh, F., 2012. Prediction of soil enzymes activity by digital terrain analysis: comparing artificial neural network and multiple linear regression models. *Environ. Eng. Sci.* 29 (8), 798–806. <https://doi.org/10.1089/ees.2011.0313>.
- Tian, Y., Xu, Y.P., Wang, G., 2018. Agricultural drought prediction using climate indices based on Support Vector Regression in Xiangjiang River basin. *Sci. Total Environ.* 622, 710–720. <https://doi.org/10.1016/j.scitotenv.2017.12.025>.
- Trebing, K., Stańczyk, T., Mehrkanooon, S., 2021. SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recogn. Lett.* 145, 178–186. <https://doi.org/10.1016/j.patrec.2021.01.036>.
- Van der Velde, Y., Lyon, S.W., Destouni, G., 2013. Data-driven regionalization of river discharges and emergent land cover–evapotranspiration relationships across Sweden. *J. Geophys. Res. Atmos.* 118 (6), 2576–2587. <https://doi.org/10.1002/jgrd.50224>.
- Vinhal-Freitas, I.C., Corrêa, G.F., Wendling, B., Bobulská, L., Ferreira, A.S., 2017. Soil textural class plays a major role in evaluating the effects of land use on soil quality indicators. *Ecol. Ind.* 74, 182–190.
- WFP, 2022. Horn of Africa 'Cannot wait': WFP scales up assistance as historic drought raises famine threat. accessed 19 October 2022. <https://www.wfp.org/news/horn-africa-cannot-wait-wfp-scales-assistance-historic-drought-raises-famine-threat>.
- Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., Chen, Z., 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost

- combination approach. *Remote Sens. Environ.* 233, 111358 <https://doi.org/10.1016/j.rse.2019.111358>.
- Xu, D., Zhang, Q., Ding, Y., Zhang, D., 2022. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. *Environ. Sci. Pollut. Res.* 29 (3), 4128–4144. <https://doi.org/10.1007/s11356-021-15325-z>.
- Yihdego, Y., Vaheddoost, B., Al-Weshah, R.A., 2019. Drought indices and indicators revisited. *Arab. J. Geosci.* 12 (3), 1–12. <https://doi.org/10.1007/s12517-019-4237-z>.
- Yohannes, H., Soromessa, T., Argaw, M., Dewan, A., 2021. Impact of landscape pattern changes on hydrological ecosystem services in the Beressa watershed of the Blue Nile Basin in Ethiopia. *Sci. Total Environ.* 793, 148559.
- Zhang, X., Hao, Z., Singh, V.P., Zhang, Y.u., Feng, S., Xu, Y., Hao, F., 2022. Drought propagation under global warming: Characteristics, approaches, processes, and controlling factors. *Sci. Total Environ.* 838, 156021.
- Zhong, R., Chen, X., Lai, C., Wang, Z., Lian, Y., Yu, H., Wu, X., 2019. Drought monitoring utility of satellite-based precipitation products across mainland China. *J. Hydrol.* 568, 343–359. <https://doi.org/10.1016/j.jhydrol.2018.10.072>.
- Zhou, Q., Zhou, H., Zhou, Q., Yang, F., Luo, L., 2014. Structure damage detection based on random forest recursive feature elimination. *Mech. Syst. Sig. Process.* 46 (1), 82–90. <https://doi.org/10.1016/j.ymssp.2013.12.013>.
- Zolfaghari, Z., Mosaddeghi, M.R., Ayoubi, S., 2015. ANN-based pedotransfer and soil spatial prediction functions for predicting Atterberg consistency limits and indices from easily available properties at the watershed scale in western Iran. *Soil Use Manage.* 31 (1), 142–154. <https://doi.org/10.1111/sum.12167>.