Focal-TSMP: Deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation

Mohamad Hakam Shams Eddin[1] and Juergen Gall[1,2]

[1]Institute of Computer Science, University of Bonn, Friedrich-Hirzebruch-Allee 5, 53115 Bonn, Germany
[2]Lamarr Institute for Machine Learning and Artificial Intelligence, 53115 Bonn, Germany.

Correspondence: Mohamad Hakam Shams Eddin (shams@iai.uni-bonn.de)

Abstract. Satellite-derived agricultural drought indices can provide a complementary perspective of terrestrial vegetation trends and their integration for drought assessments under future climates is beneficial for providing more comprehensive assessments. However, satellite-derived drought indices are only available for observed periods. In this study, we aim to improve the agricultural drought assessments under future climate change by applying deep learning (DL) to predict satellite-derived vegetation indices from a regional climate simulation. The simulation is produced by the Terrestrial Systems Modelling Platform (TSMP) and performed in a free evolution mode over Europe. TSMP simulations incorporate variables from underground to the top of the atmosphere (Ground to Atmosphere G2A) and are widely used for research studies related to water cycle and climate change. We leverage these simulations for long-term forecasting and DL to map the forecast variables into Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) images that are not part of the simulation model. These predicted images are then used to derive different vegetation and agricultural drought indices, namely NDVI anomaly, BT anomaly, Vegetation Condition Index (VCI), Thermal Condition Index (TCI), and Vegetation Health Index (VHI). The developed DL model could be integrated with data assimilation and used for down-stream tasks, i.e., for estimating NDVI and BT for periods where no satellite data are available and for modelling the impact of extreme events on vegetation responses with different climate change scenarios. Moreover, our study could be used as a complimentary evaluation framework for TSMP-based climate change simulations. To ensure reliability and to assess the model's applicability to different seasons and regions, we provide an analysis of model biases and uncertainties across different regions over the Pan-Europe domain. We further provide an analysis about the contribution of the input variables from the TSMP model components to ensure a better understanding of the model prediction. A comprehensive evaluation on the long-term TSMP simulation using reference remote sensing data showed sufficiently good agreements between the model predictions and observations. While model performance varies on the test set between different climate regions, it achieves a mean absolute error (MAE) of 0.027 and 1.90 K with coefficient of determination (R^2) scores of 0.88 and 0.92 for NDVI and BT, respectively, at $0.11°$ resolution for sub-seasonal predictions. In summary, we demonstrate the feasibility of using DL on a TSMP simulation to synthesize NDVI and BT satellite images, which can be used for agricultural drought forecasting. Our implementation is publicly available at the project page (https://hakamshams.github.io/Focal-TSMP).

1 Introduction

According to recent studies on historical trends and current projections, different regions of the Earth would be under a changing climate more vulnerable to extreme events such as flash droughts (Christian et al., 2021, 2023; Yuan et al., 2023), meteorological and agricultural droughts (Essa et al., 2023), forest wildfires (Patacca et al., 2023), and water storage deficiency (Pokhrel et al., 2021). The expected increase in concurrence of agricultural droughts would cause crop production losses and vegetation mortality. In particular, people in regions with fragile adaptation and mitigation strategies will be more effected. Therefore, forecasting the vegetation responses and their evolving patterns conditioned on climate scenarios is a requirement to form better mitigation and adaptation strategies.

In relation to this, there has been a growing line of research over the past in improving and deploying climate modeling that attempt to simulate the underlying processes of the Earth system (Shrestha et al., 2014; Gasper et al., 2014; Lawrence et al., 2019). These modelling platforms are essential to realize and forecast climatic extreme events such as droughts in a model simulation (Miralles et al., 2019). For instance, the simulated outputs of modeling systems can be used to derive agricultural drought indices based on a deficiency in precipitation (McKee, 1995; Vicente-Serrano et al., 2010) or soil moisture (Martínez-Fernández et al., 2015). Nowadays, satellite observations around the world provide a near real-time global monitoring of vegetation and drought conditions. Vegetation products derived from satellite land surface reflectances can be used as proxies for vegetation health and consequently as agricultural drought indicators (Qin et al., 2021; Vreugdenhil et al., 2022). While historical trends in satellite-based droughts have been extensively studied, satellite-based agricultural drought assessment and its relation to climate simulations under climate change remains not fully explored. In this study, we propose to use deep learning (DL) to improve the agricultural drought analysis by predicting satellite-derived vegetation indices that can be combined with meteorological or hydrological indices which are often used in studies for drought assessment to provide more comprehensive assessments. In fact, some studies highlighted inconsistencies in the long-term drought trends (Sheffield et al., 2012; Kew et al., 2021; Vicente-Serrano et al., 2022). Meanwhile others showed a different perspective of trends related to terrestrial vegetation from remote sensing products (Zhu et al., 2016; Kogan et al., 2020). This is usually explained as assessments are highly dependent on drought definition (Satoh et al., 2021; Reyniers et al., 2023) and extreme event attribution (Van Oldenborgh et al., 2021), i.e., the drought indicator that was chosen in the methodology and the variations in modelling platforms. In addition, prescribed vegetation assumptions exist in climate simulations which limit the modeling of atmospheric carbon effects or soil moisture deficiency on vegetation (Pirret et al., 2020; Pokhrel et al., 2021; Reyniers et al., 2023). If we add to this the complex spatio-temporal response of vegetation to climate variability (Seneviratne et al., 2021; Jin et al., 2023), i.e., regional responses to climate have different dynamics and are more complicated than those at a global scale, we can conclude that predicting the vegetation state in response to drought under climate conditions still poses a major challenge. More precisely, in this study we predict satellite-based vegetation products from a free evolving simulation based on the Terrestrial Systems Modelling Platform (TSMP) (Furusho-Percot et al., 2019). TSMP simulations integrate variables from groundwater to the top of the atmosphere (Ground to Atmosphere G2A) and are primarily employed in studies on the water cycle and climate change (Ma et al., 2021; Furusho-Percot et al., 2022; Naz et al., 2023; Patakchi Yousefi and Kollet, 2023). In particular, we predict from

the TSMP simulation the Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) as they would have been observed from Advanced Very High Resolution Radiometer (AVHRR) from the National Oceanic and Atmospheric Administration (NOAA) satellite systems. NDVI is computed from the reflectance in visible red (ρ_R) and near-infrared bands (ρ_{NIR}). It is a standard product that is extensively used in applications for vegetation health and crop yield (Tucker, 1979). BT is a calibrated spectral radiation derived from the thermal band (ρ_{IR}) and can be used for temperature-related vegetation stress monitoring (Kogan, 1995a). We assume that a climate simulation (i.e., TSMP simulation) that is close to the true state of the Earth should be able to model vegetation products (i.e., NDVI and BT) regardless of the target satellite platform (in this study AVHRR NOAA). Recently, DL models have become popular to build a predictive model for tasks that include complex or intractable cause and effect relations within the Earth system (Bergen et al., 2019; Tuia et al., 2023; de Burgh-Day and Leeuwenburg, 2023). In addition, DL can be used to handle biases implicitly, thus simplifying the entire workflow (Schultz et al., 2021). For instance, DL was recently used in climate modelling for bias correction and down-scaling to project extremes (Blanchard et al., 2022), weather forecasting (Lam et al., 2022; Chen et al., 2023; Bi et al., 2023; Ben-Bouallegue et al., 2023), supporting data assimilation systems (Düben et al., 2021; Valmassoi et al., 2022; Yu et al., 2023), and generalized multi-task learning (Nguyen et al., 2023; Lessig et al., 2023). In this work, we thus propose a DL approach based on focal modulation networks (Yang et al., 2022) to simultaneously predict NDVI and BT from the model simulation. In this way, we leverage a climate simulation for long-term forecasting and DL for mapping the forecast variables to vegetation related indices that are not part of the simulation model.

Forward operators like radiative transfer solvers are normally used to synthesize spectral band satellite images from the output of a numerical weather model (Scheck et al., 2016; Geiss et al., 2021; Li et al., 2022). In this paper, we investigate the use of DL to predict products of atmospherically corrected albedo/emissivity on land (atmospherically corrected bottom of atmosphere) like NDVI and BT simultaneously rather than training the neural network to serve as an emulator for a predefined physical-based radiative transfer model. In other words, our training data for DL are derived from real-world satellite observations (empirical operator) without assimilating data or assumptions about radiations. Besides, there exist climate-vegetation models which directly simulate the vegetation dynamic based on ecological processes and statistical modeling. Nevertheless, they are limited by the complexity of the processes and poor generalization (Chen et al., 2021). Unlike hydro-meteorological variables that can be predicted or forecast using numerical weather model, vegetation products demand an extended modeling representation of the surface and sub-surface (Lees et al., 2022). Recently, Salakpi et al. (2022a, b) predicted short-term vegetation products based on previous vegetation conditions and observational anomaly indices in a Bayesian auto-regressive approach. However, the interaction between vegetation and climate variability exhibits a strong non-linear behavior. In this respect, many studies explored the applicability of DL for vegetation health prediction using climate models and remote sensing data (Das and Ghosh, 2016; Adede et al., 2019; Ferchichi et al., 2022; Wu et al., 2020; Kraft et al., 2019; Prodhan et al., 2021). A common approach is to use past vegetation conditions to predict the short-term future variations (Nay et al., 2018; Yu et al., 2022; Hammad and Falchetta, 2022; Lees et al., 2022; Vo et al., 2023). In a related work, Requena-Mesa et al. (2021) addressed the problem of optical satellite imagery forecasting as a guided video prediction task. In their framework, vegetation dynamics approximated by NDVI is modeled at high resolution using past satellite images as initial conditions and static and reanalysis

data as a model guidance. Similar approaches with this framework were presented in (Robin et al., 2022; Kladny et al., 2022; Diaconu et al., 2022) and on a continental scale in (Benson et al., 2023). While these works differ in their methodologies, i.e. in the predicted vegetation products, model architectures, and spatio-temporal resolutions, they have overall a good performance for short-term forecasting. Although short-term forecasting, i.e., for a few weeks, is very useful for short-term planning, a more significant contribution could be achieved with a much longer forecasting time (Marj and Meijerink, 2011). Nonetheless, only few studies addressed long-term vegetation conditions forecasting (Marj and Meijerink, 2011; Miao et al., 2015; Patil et al., 2017; Chen et al., 2021; Wei et al., 2023). In addition, most studies focused only on a single indicator. The combination of different indicators like NDVI and BT with their corresponding drought indices provides complementary information on the vegetation state and is beneficial for vegetation monitoring (Yang et al., 2020). As mentioned before, we aim to use DL to predict vegetation products like NDVI and BT from a regional climate simulation at a continental scale. We also focus on long-term forecasting without using an initial state, i.e., satellite images from previous time steps. Unlike aforementioned works, we use more input data for the neural network from the surface and sub-surface to account for a more detailed representation of the reflectance/emissivity on ground. In addition, we built the neural network on Vision Transformers (Dosovitskiy et al., 2021) and Convolutional Neural Networks (CNN) models taking into account the spatial context around each input pixel and operating on the whole scene at once. This was motivated by previous studies that indicate that an effective model of the environment should consider the spatial-correlation within the domain. Previous works train and evaluate DL models on biased-corrected reanalysis data. In contrast, we evaluate the approach with real-world observations using a run of the simulation in the past. It is worth to note that this evaluation is more consistent with real-world deployment schemes, since it is questionable how a model that has been trained and evaluated on reanalysis data will perform on a biased climate projection simulations. Thus, we opt for a simulation that mimics a climate projection of the past and train and evaluate the model on it to internally correct biases and predict vegetation products.

To showcase the potential of our approach, we apply the predicted NDVI and BT for long-term agricultural drought forecasting, where we derive Vegetation Condition Index (VCI), Thermal Condition Index (TCI), and Vegetation Health Index (VHI) (Yang et al., 2020) as agricultural drought indicators from the predicted NDVI and BT. As part of this, we analyze whether a DL model trained on simulation produced by TSMP can be used for vegetation health forecasting at a continental-scale by identifying regions and periods of uncertainty in the model prediction. Moreover, we analyze the importance of the input explanatory variables. We achieve an overall mean absolute error (MAE) of 0.027 and 1.90 K with coefficient of determination (R^2) scores of 0.88 and 0.92 in predicting NDVI and BT, respectively for sub-seasonal predictions at $0.11°$ resolution. Our results indicate that a direct prediction of vegetation products from a TSMP simulation with DL is an effective way for scenario-based assessments of vegetation response to climate change.

The rest of this article is organized as follows. Section 2 describes the datasets that are used in the experiments. The methodology is described in Sect. 3. Experimental results and an analysis about variable importance are given in Sect. 4. Finally, conclusions are provided in Sect. 5.

2 Datasets and data preprocessing

In this section, we describe the datasets used in the experiments. The TSMP simulation is presented in Sect. 2.1, the observational remote sensing data for model training and evaluation are presented in Sect. 2.2, and the preprocessing framework of the data is described in Sect. 2.3.

2.1 Regional Earth system simulation

For this study, we use the simulation produced by Terrestrial System Modelling Platform version 1.1. (TSMP) at the Research Centre Jülich (FZJ) IBG-3 Institute and originally described in (Shrestha et al., 2014) and (Gasper et al., 2014). The simulation used in this study is introduced in (Furusho-Percot et al., 2019). TSMP is a physics-based integrated simulation representing a realization of the terrestrial hydrologic and energy cycles that cannot be directly obtained from measurements. Its setup consists of three main interconnected model components:

- The Consortium for Small Scale Modelling (COSMO) version 5.01 is a numerical weather model to simulate the diabatic and adiabatic atmospheric processes (Baldauf et al., 2011).

- The Community Land Model (CLM) version 3.5 is used to simulate the bio-geophysical processes on the land surface (Oleson et al., 2004, 2008).

- ParFlow version 3.2. is a hydrological model to explicitly simulate the 3D dynamic processes of water in the land surface and underground (Jones and Woodward, 2001; Kollet and Maxwell, 2006; Jefferson and Maxwell, 2015; Maxwell et al., 2015; Kuffour et al., 2020).

ECMWF ERA-Interim data (Dee et al., 2011) were used to define the initial and boundary conditions for the simulation. Based on this setup, a spinup of 10 years (1979-1988) was conducted to initialize the surface and subsurface hydrologic and energy conditions and to reach the dynamic equilibrium with the atmosphere before the actual run (1989-2019). We selected variables available within the period applicable for the analysis. This results in 29 main variables from COSMO, 8 variables from CLM, and 2 main variables from ParFlow. Additionally, we used 3 static variables from the analysis (Poshyvailo-Strube et al., 2022). An analysis about the explanatory variables is provided in Sect. 4 and variable descriptions are listed in Tables A1 and A2. The three model components were fully coupled via the OASIS3 coupler (Valcke, 2013) to form a unified soil–vegetation–atmosphere model. This scheme was built without nudging allowing the free-running of the simulated variables. Thus, TSMP is ideal for representing the heterogeneity of the water cycle from the subsurface to the top atmosphere in a free evolution. In addition, the long-term simulation is performed for a historical time period from January 1989 until summer in September 2019 with output variables aggregated on a daily basis and extending over the Europe EURO-CORDEX EUR-11 domain (Giorgi et al., 2009; Gutowski Jr. et al., 2016; Jacob et al., 2020). The grid specification for TSMP is a standardized rotated coordinate system ($\phi_{(rotated\ pole)} = 39.5°$ N, $\lambda_{(rotated\ pole)} = 18°$ E) with a spatial resolution of $\sim 0.11°$ (~ 12.5 km) and 412×424 grid cells in the rotated latitudinal and longitudinal direction, respectively. These spatio-temporal dimensions and the model setup make TSMP suitable for climatological studies at a continental scale.

2.2 Observational remote sensing data

Satellite-based vegetation health products were obtained from NOAA[1]. The blended version (Yang et al., 2020) is composed of long-term remote sensing data derived from two systems of satellites: AVHRR from 1981 to 2012 and its successor Visible Infrared Imaging Radiometer Suite (VIIRS) from 2013 onward. The dataset includes two essential products, namely NDVI and BT (Table A3). NDVI is computed from the red (ρ_R) and near-infrared (ρ_{NIR}) bands:

$$\text{NDVI} = \frac{(\rho_{NIR} - \rho_R)}{(\rho_{NIR} + \rho_R)}. \tag{1}$$

The NDVI is unitless and given in the range [-0.1, 1]. Same NDVI values should not be interpreted similarly for different ecosystems. In other words, the interpretation is highly dependant on the location and ecosystem productivity (Kogan, 1995b). BT is derived from the infrared (ρ_{IR}) band and given in Kelvin (K) within the range [0, 400]. To handle high frequency noise caused by clouds, aerosol, and atmospheric variation along with different random error sources, NDVI and BT were temporally aggregated into smoothed noise reduced weekly products. In addition, post-launch calibration coefficients and solar/sensor zenith angles are applied to account for sensor degradation and orbital drift. The outlier removal is essential to exclude invalid measurements. Additionally, this weekly temporal resolution is enough to capture the phenological phases of vegetation and adequate for satellite data application (Kogan et al., 2011; Yang et al., 2020). Based on NDVI, BT and their long-term climatologies, the upper and lower bounds of the ecosystem (minimum and maximum values for NDVI and BT) can be estimated. Hence, VCI, TCI, and VHI can be derived pixel-wise (Kogan, 1995a, 1990). The vegetation condition index is given by:

$$\text{VCI} = 100 \frac{(\text{NDVI} - \text{NDVI}_{min})}{(\text{NDVI}_{max} - \text{NDVI}_{min})}, \quad \text{with} \quad \text{VCI} \in [0, 100], \tag{2}$$

where NDVI is the weekly noise reduced NDVI, and NDVI_{min} and NDVI_{max} the multi-year weekly absolute minimum and maximum NDVI values, respectively. The thermal condition index is given by:

$$\text{TCI} = 100 \frac{(\text{BT}_{max} - \text{BT})}{(\text{BT}_{max} - \text{BT}_{min})}, \quad \text{with} \quad \text{TCI} \in [0, 100], \tag{3}$$

where BT is the weekly noise reduced BT, and BT_{min} and BT_{max} the multi-year weekly absolute minimum and maximum BT values, respectively. The vegetation health index is given by:

$$\text{VHI} = (\alpha)\text{VCI} + (1 - \alpha)\text{TCI}, \quad \text{with} \quad \text{VHI} \in [0, 100], \tag{4}$$

where α is a weighting coefficient. While VCI is a proxy for the moisture condition and its lower values reflect a water-related stress, TCI is a proxy for the thermal condition and its lower values indicate a temperature and wetness-related stress. The composite index VHI is a linear combination of the former two indices to approximate the vegetation health. VHI fluctuates annually between 0 (unfavourable condition) to 100 (favourable condition). The values of these indices above 100 and below 0 are clipped. The dataset is provided globally with $\sim 0.05°$ (~ 4 km) spatial resolution mapped into the Plate Carrée projection. NOAA VP have been broadly used for research and real-world applications. For a summary on the validation and studies that use this dataset for agricultural droughts monitoring, we refer to Yang et al. (2020).

---

[1]Center for Satellite Applications and Research (STAR) https://www.star.nesdis.noaa.gov/star/index.php

2.3 Preprocessing

In this section we describe the data preprocessing that is needed prior to apply DL. Overall the TSMP has 30 years of data (1989-2019). We reserved the years 1989-2009 (AVHRR era) and 2013-2016 (VIIRS era) for training, 2010-2011 (AVHRR era) and 2017 (VIIRS era) for validation, and 2012 (AVHRR era), 2018-2019 (VIIRS era) for testing. For TSMP, we excluded the lateral boundary relaxation zone by removing invalid grid points from the boundaries. This results in a final grid with 397×409 grid cells in the latitudinal and longitudinal direction, respectively. In order to connect local-related characteristics to climate conditions, we computed 3 additional static variables from the static variables described in Table A2. We computed slope (Horn, 1981) and roughness (Wilson et al., 2007) from orography and distance to water from the land/sea mask. Due to the fact that the remote sensing data were obtained from two different satellite systems, the data derived from VIIRS have to be first adjusted to insure continuity and consistency with the data derived from AVHRR. Yang et al. (2018, 2021b) showed that the discrepancy between sensors are mainly due to the differences in spectral response ranges and calibration parameters. Compared to BT/TCI, this has a greater impact on NDVI/VCI (Kogan et al., 2015). Considering this issue, we followed the same re-compositing approach as described in Yang et al. (2021b). The re-compositing approach can be used to generate cross-sensor vegetation products for the time period from 2013 to 2019. In fact NDVI/BT from different sensors can be decomposed into climatologies and VCI/TCI. The climatology provides information about the Ecosystem and it is sensor-specific. While VCI/TCI for the same ecosystem location are cross-sensor. Thus, using climatology from AVHRR and VCI from VIIRS, Eq. (2) can be reformulated to re-composite NDVI for AVHRR as following:

$$\text{NDVI}'_{(AVHRR)} = \left(\frac{\text{VCI}_{(VIIRS)}}{100} \right) (\text{NDVI}_{(max,AVHRR)} - \text{NDVI}_{(min,AVHRR)}) + \text{NDVI}_{(min,AVHRR)}, \tag{5}$$

where $\text{NDVI}'_{(AVHRR)}$ is the converted weekly noise reduced NDVI from VIIRS to AVHRR, $\text{VCI}_{(VIRRS)}$ is the Vegetation Condition Index derived from VIIRS, $\text{NDVI}_{(min,AVHRR)}$ and $\text{NDVI}_{(max,AVHRR)}$ are the multi-year weekly absolute minimum and maximum NDVI values (climatology) derived from AVHRR, respectively. Similarly from Eq. (3) we have:

$$\text{BT}'_{(AVHRR)} = \text{BT}_{(max,AVHRR)} - \left(\frac{\text{TCI}_{(VIIRS)}}{100} \right) (\text{BT}_{(max,AVHRR)} - \text{BT}_{(min,AVHRR)}), \tag{6}$$

where $\text{BT}'_{(AVHRR)}$ is the converted weekly noise reduced BT from VIIRS to AVHRR, $\text{TCI}_{(VIRRS)}$ is the Thermal Condition Index derived from VIIRS, $\text{BT}_{(min,AVHRR)}$ and $\text{BT}_{(max,AVHRR)}$ are the multi-year weekly absolute minimum and maximum BT values (climatology) derived from AVHRR, respectively. Please note that $\text{VCI}_{(VIIRS)}$ and $\text{TCI}_{(VIIRS)}$ were based on a pseudo long-term VIIRS climatology (for more details on this, please see Yang et al. (2018)). In addition, the TSMP simulation and target remote sensing data have to be spatially aligned in the same domain. After the continuity at NDVI and BT level has been realized, we mapped these two products into the TSMP rotated coordinate system over the EURO-CORDEX EUR-11 domain. For the mapping, we up-scaled the data from $0.05°$ to $0.11°$ resolution based on a first-order conservative mapping (Jones, 1999) using the package from Zhuang et al. (2020). For calculating the spatial mean, we excluded invalid, water, and coastal lines pixels. Afterwards, we computed VCI, TCI and VHI based on Eq. (2)-(4). We note that the weighted coefficient α in Eq. (4) can be empirically calibrated as a spatially variant factor (Zeng et al., 2022, 2023). Following previous works, we set

α to its standard value 0.5 in all experiments as in Yang et al. (2020). Furthermore, masks over desert and very cold areas were extracted from the quality assurance (QA) metadata provided with the data. Eventually, the preprocessed data are aggregated into data cubes ({variable, lat, lon}) on a weekly basis and stored as netCDF files. This remote sensing dataset can serve as a reference to train and evaluate the DL model performance. Overall, this includes 1263, 156, and 139 samples (weeks) for training, validation, and testing, respectively. To avoid overfitting or the domination of few input variables, we normalized the input of TSMP by subtracting the mean and dividing by the standard deviation corresponding to each input variable. These statistics were computed only from the years that are used for training. The invalid values of pixels were replaced with zero values as input to the DL model.

3 Methodology

Problem formulation. Given $\text{TSMP} \in \mathbb{R}^{V \times T \times W \times H}$ as a climate change simulation, where V is the number of output variables from the COSMO, CLM, and ParFlow models and the static forcing variables, T is the temporal dimension and W and H are the spatial extensions, our objective is to construct a mapping function to predict $\text{NDVI} \in \mathbb{R}^{I \times W \times H}$ and $\text{BT} \in \mathbb{R}^{I \times W \times H}$ on a weekly basis, where I is the number of weeks. To accomplish this, we propose to approximate this function as a function f using a DL model based on a U-Net (Ronneberger et al., 2015) with focal modulations (Yang et al., 2022) as building blocks:

$$f : (\text{TSMP}; \theta) \to (\text{NDVI}, \text{BT}), \tag{7}$$

where θ are the weights of the model. The input for DL is a data cube representing a specific week i of TSMP data and the output are NDVI and BT corresponding to the same week i. We denote the weekly averaged input data cube produced by TSMP as $\mathbf{X}^i \in \mathbb{R}^{V \times W \times H}$. Where we obtain $\mathbf{X}^i$ by taking the mean of the days corresponding to the week i. For simplicity, we will drop the notation i in the following sections. First, the network architecture is introduced in Sect. 3.1 and the focal modulation is then described in Sect. 3.2. Section 3.3 discusses the loss function and Sect. 3.4 outlines the baselines approaches. Implementation and technical details are given in Sect. 3.5. Finally, the evaluation metrics are described Sect. 3.6.

3.1 Model architectures

The recent applications of Vision Transformers (ViT) have covered many tasks in the field of computer vision. The network design of ViT along with the multi-head self-attention mechanism (Vaswani et al., 2017) allow ViT to stand as the state-of-the-art backbone in recent DL models. In contrast to CNNs, ViTs with self-attention modules can handle long-range interactions across tokens (pixels) more efficiently. In a nutshell, the self-attention module aims to transfer pixel representations of a given image into a new feature representation based on a weighted aggregation of interactions between every individual pixel and its surrounding. This mechanism allows the model to focus on more relevant regions of the input images. Despite this powerful transforming process, the computational requirement of a standard ViT has been a limitation when applying it to vision tasks. More recently, the Focal Modulation Network (Yang et al., 2022) has been introduced to substitute the self-attention mechanism with a lightweight focal module. In contrast to self-attention, focal modulation starts with contextual aggregation and ends with
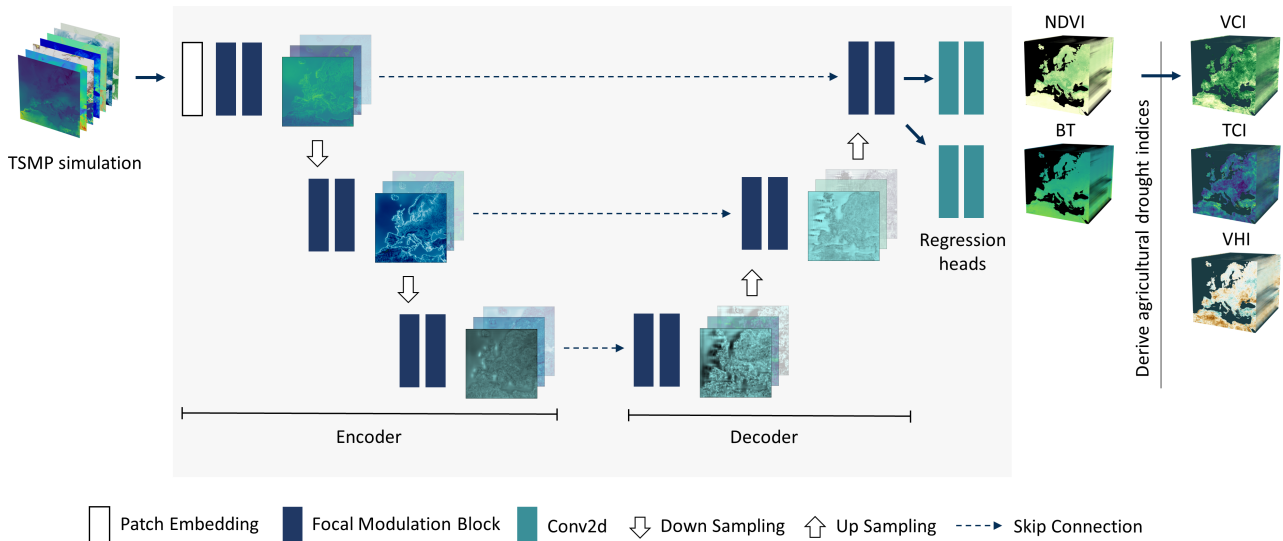
Figure 1. An overview of the proposed model to predict NDVI and BT from a TSMP climate simulation. The model follows the U-Net shape with encoder and decoder layers. We use focal modulation as the basic building block for the model. The input TSMP simulation is first encoded into a latent representation via encoder layers. In a subsequent step, the decoder constructs new features to be given as input to two separated regression heads that output NDVI and BT simultaneously. The predicted NDVI and BT can then be used to derive different agricultural drought indices such as VCI, TCI, and VHI.

interactions. Based on this recently introduced mechanism, DL models were developed for medical image segmentation (Naderi et al., 2022; Rasoulian et al., 2023), change detection for remote sensing data (Fazry et al., 2023), and video action recognition (Wasim et al., 2023). We build our model on focal modulation networks and extend their applications in Geoscience. Figure 1 provides an overview of the model architecture. The model design follows the U-Net shape with encoder and decoder layers connected via skip connections and followed by two regression heads. This allows the model to extract features in a hierarchical way and predict NDVI/BT with customized heads. In the following, we describe the main parts of the model:

Patch embedding. The patch embedding is implemented as a single 1D convolution, where one patch is equivalent to one pixel. The role of this embedding is to project the input $\mathbf{X}$ from V dimension into a channel dimension that matches the channel dimension $C_{(en,1)}$ of the first encoder block. In contrast to related works with transformers, we do not reduce the spatial resolution at this step. This is important to mitigate blurring effects for regression tasks. An analysis of the impact of the patch size for embedding is provided in Appendix E.

Encoder. The encoder consists of 3 encoding layers. Each layer has 2 consecutive focal modulation blocks that have the same number of channels. We use focal modulation to capture local to global dependencies in the domain (Sect. 3.2). We apply down-sampling on the output of the first two encoder layers to reduce the spatial resolution by a factor of 2 and double the number of channels. The down-sampling is implemented as a 2D convolution with 2×2 kernel size and stride of 2. We set $C_{(en,1)} = 96$ as the number of channels of the first encoder layer. Consequently, the encoder has the dimensionality

$\{C_{(en,1)} = 96, C_{(en,2)} = 192, C_{(en,3)} = 384\}$, where $C_{(en,2)}$ is the dimensionality for the second encoder layer and $C_{(en,3)}$ is the dimensionality for the third encoder layer. The encoder allows the network to extract low to high level features in a hierarchical way. Note that focal modulation allows an additional hierarchical feature extraction at each level (Sect. 3.2).

Skip connections. These connections copy outputs from each encoder layer into its corresponding decoder layer. The purpose of this is to enhance the gradient flow in the network and preventing vanishing gradient issues.

Decoder. The decoder has a similar design to the encoder. It consists of 3 decoder layers with 2 consecutive focal modulation blocks for each decoder layer. The input for the first decoder layer is the output of the last encoder layer copied via a skip connection. The input for the second and third decoder layers is a concatenation of the output from the previous decoder layer with the output of the corresponding encoder layer. The outputs of the first and second decoder layers are up-sampled to double the image size and reduce the dimensionality by a factor of 2. The up-sampling is implemented as a bilinear interpolation followed by a 2D convolution with 1×1 kernel size and stride of 1. The decoder layers has the dimensionality $\{C_{(de,1)} = C_{(en,3)} = 384, C_{(de,2)} = C_{(en,2)} + C_{(de,1)} = 384, C_{(de,3)} = C_{(en,1)} + C_{(de,2)} = 288\}$, where $C_{(de,1)}$, $C_{(de,2)}$, and $C_{(de,3)}$ are the dimensionality for the first, second, and third decoder layers, respectively. The purpose of the decoder is to gradually construct the input for the regression heads from the encoded features.

Regression heads. The ouput of the last decoder layer is then given as input to two separated regression heads to predict NDVI and BT. Each head has two 2D convolutions with 3×3 kernel size and stride of 1 with a LeakyReLU activation in between. The regression head reduces the dimensionality from $C_{(de,3)} = 288$ to 128, and then to 1.

3.2 Focal Modulations
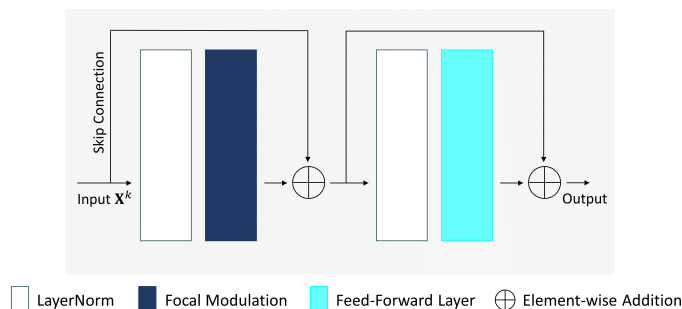
Figure 2. An illustration of the focal modulation block. It follows the typical transformer block with a focal modulation instead of self-attention. $\mathbf{X}^k$ represents the input to the k-th block.

We first describe how the block is implemented and then describe the main focal modulation module denoted as FM. Fig. 2 illustrates the architecture of the focal modulation block used in both the encoder and decoder layers. The design follows a typical transformer block. Let $\mathbf{X}^k \in \mathbb{R}^{N \times C^k \times W^k \times H^k}$ be the input at the k-th block, where N is the batch size (number of input tensors), C^k is the number of input channels, and W^k and H^k are the spatial resolution. First, the input is normalized across N via a layer normalization (Ba et al., 2016) denoted as LayerNorm. Using the indices $n \in \{1, \ldots, N\}, c^k \in \{1, \ldots, C^k\}$,

$w^k \in \{1, \ldots, W^k\}$, and $h^k \in \{1, \ldots, H^k\}$, the LayerNorm can be written as:

$$\text{LayerNorm}(\mathbf{X}^k; (\gamma_l^k, \beta_l^k)) = \left(\frac{\mathbf{X}_{n(c^k, w^k, h^k)}^k - \mu_n^k}{\sigma_n^k} \right) \cdot \gamma_{l\,(c^k)}^k + \beta_{l\,(c^k)}^k, \tag{8}$$

$$\mu_n^k = \frac{1}{C^k W^k H^k} \sum_{c^k=1}^{C^k} \sum_{w^k=1}^{W^k} \sum_{h^k=1}^{H^k} \mathbf{X}_{n(c^k, w^k, h^k)}^k, \tag{9}$$

$$\sigma_n^k = \sqrt{\frac{1}{C^k W^k H^k} \sum_{c^k=1}^{C^k} \sum_{w^k=1}^{W^k} \sum_{h^k=1}^{H^k} (\mathbf{X}_{n(c^k, w^k, h^k)}^k - \mu_n^k)^2}, \tag{10}$$

where $\mathbf{X}_{n(c^k, w^k, h^k)}^k$ is the input tensor of order n in the batch, μ_n^k and σ_l^k are the computed mean and standard deviation of the corresponding input $\mathbf{X}_{n(c^k, w^k, h^k)}^k$, and $\gamma_{l\,(c^k)}^k \in \mathbb{R}^{C^k}$ and $\beta_{l\,(c^k)}^k \in \mathbb{R}^{C^k}$ are per-element learnable parameters.

These learnable parameters are shared across input tensors. The output of LayerNorm is then passed into the function FM. After that, the output of the first part is normalized by a second LayerNorm and passed into a feed-forward layer (FFL). The FFL consists of one linear layer that maps the dimentionality to $r_{mlp} \times C^k$ followed by a GELU activation (Hendrycks and Gimpel, 2016) and a second linear layer to bring the dimensionality back to C^k, where r_{mlp} is the MLP ratio parameter. We set r_{mlp} to 4 for the encoder and decrease it to 2 for the decoder to reduce the number of model parameters. The output of each block can be formulated as follows:

$$\text{FocalModulationBlock}(\mathbf{X}^k) \triangleq \gamma_2^k \left(\text{FFL} \left(\text{LayerNorm}(\gamma_1^k \text{FM}(\text{LayerNorm}(\mathbf{X}^k)) + \mathbf{X}^k) \right) \right)$$
$$+ \left(\gamma_1^k \text{FM}(\text{LayerNorm}(\mathbf{X}^k)) + \mathbf{X}^k \right), \tag{11}$$

where $\gamma_1^k \in \mathbb{R}^{C^k}$ and $\gamma_2^k \in \mathbb{R}^{C^k}$ are learnable scaling parameters. The main component of each focal modulation block is the FM function. As seen in Fig. 3, it consists of three main steps: hierarchical contextualization, gated aggregation, and interactions.

Hierarchical contextualization. The objective of this part is to encode local to global range dependencies for every pixel. It is based on focal transformer (Yang et al., 2021a) and aims to extract features at 4 different levels. Let $\mathbf{X}^k$ be the input for FM and $L = 4$ be the number of levels. First, $\mathbf{X}^k$ is projected by a linear layer into a new representation $\mathbf{L}_0^k = \text{Linear}(\mathbf{X}^k) \in \mathbb{R}^{N \times C^k \times W^k \times H^k}$. Afterwards, the contexts are obtained in a recursive manner using a sequence of 3 depth-wise 2D convolutions (DWConv2D) with GeLU activation and with increased receptive fields. In DWConv2D, each output channel corresponds to a convolution on one input channel. We denote r_l as the kernel size at level l and start with $r_1 = 3$. Thereby, the kernel sizes at the focal levels have the values $r_1=3$, $r_2=5$, $r_3=7$. To obtain a global feature representation, a global average pooling (GAP) followed by a GeLU activation is applied at level $l = 4$. Using the index $l \in \{1, \ldots, L\}$, the hierarchical contextualization can be formulated as follows:

$$\mathbf{L}_l^k \triangleq \begin{cases} \text{GeLU}(\text{DWConv2D}(\mathbf{L}_{l-1}^k)), & \text{if } 1 \leq l < L, \\ \text{GeLU}(\text{GAP}(\mathbf{L}_{l-1}^k)), & \text{otherwise.} \end{cases} \tag{12}$$
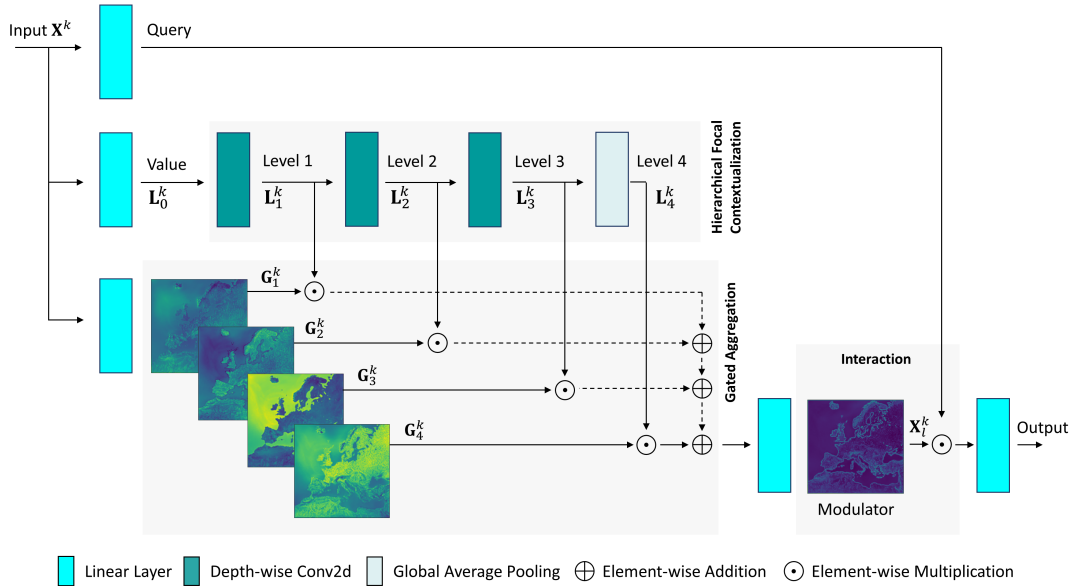
Figure 3. An illustration of the function FM at k-th block. It consists of 3 main parts: focal contextualization, gated aggregation, and interaction. First, the query, value and gates are obtained by projecting $\mathbf{X}^k$ with linear layers. Then, a stack of depth-wise 2D convolutions followed by a global pooling is used on the value to derive contextual features around pixels. Gates are used to adaptive aggregate contextual features into a modulator. Finally, the interaction between queried pixels and the modulator is performed and projected by a final linear layer to compute the output. The shown images are examples of learned gates along with the pixel-wise magnitude of corresponding modulator at the first block encoder. The bright colors (i.e., green to yellow) for specific regions represent higher values which correspond to higher attentions of the model to that regions.

Gated aggregation. The gated aggregation adaptively summarizes the extracted hierarchical contexts $\mathbf{L}_l^k$ into a modulator. First, $\mathbf{X}^k$ is projected by a linear layer into 4 gates $\mathbf{G}^k = \text{Linear}(\mathbf{X}^k) \in \mathbb{R}^{N \times L \times W^k \times H^k}$. As can be seen from the example in Fig. 3, the third gate focuses on the water area while other gates focus on different segmented regions. This allows each pixel to adaptively aggregate features from different semantic regions conditioned on its context. Pixels in a less dynamic environment may depend on more distant pixels while pixels in a more dynamic environment may depend more on the local context. The aggregation is performed over different focal levels and followed by a linear layer:

$$\mathbf{X}_L^k \triangleq \text{Linear}(\sum_{l=1}^{L} \mathbf{G}_l^k \odot \mathbf{L}_l^k), \tag{13}$$

where $\mathbf{X}_L^k \in \mathbb{R}^{N \times C^k \times W^k \times H^k}$ are the contextual aggregated features for each pixel called the modulator, $\mathbf{G}_l^k$ is the gate corresponding to level l, and $\odot$ is the Hadamard operator (element-wise multiplication).

Interaction. Finally, the interaction between the queried pixels and the modulator is given with the following formula:

$$\text{FM}(\mathbf{X}^k) \triangleq \mathbf{X}_L^k \odot \text{Linear}(\mathbf{X}^k) \in \mathbb{R}^{N \times C^k \times W^k \times H^k}. \tag{14}$$

3.3 Loss function

For training we use the Mean Absolute Error (MAE) as a loss function, since it is less sensitive to outliers than the Mean Squared Error (MSE):

$$\mathcal{L}_{MAE} = \frac{1}{NWH} \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{h=1}^{H} |Y_{(n,w,h)} - \hat{Y}_{(n,w,h)}|, \tag{15}$$

335   where N is the batch size, and $Y_{(n,w,h)}$ and $\hat{Y}_{(n,w,h)}$ are the predicted and observed images, respectively.

In addition, to increase local variability and balance the blurring effects from Eq. (15), we use a perceptual loss (Ledig et al., 2017; Johnson et al., 2016) based on a pre-trained VGG-19 network (Simonyan and Zisserman, 2014) on ImageNet (Deng et al., 2009). This additional loss constrains the generated images to have a similar structure and spatial variability to the observed images by comparing multi-level features extracted by a VGG classifier network from both the predicted and

340   observed images:

$$\mathcal{L}_{VGG} = 8\mathcal{L}_{VGG}^{1} + \sum_{j=2}^{J} \mathcal{L}_{VGG}^{j}, \tag{16}$$

$$\mathcal{L}_{VGG}^{j} = \frac{1}{NC^{j}W^{j}H^{j}} \sum_{n=1}^{N} \sum_{c=1}^{C^{j}} \sum_{w=1}^{W^{j}} \sum_{h=1}^{H^{j}} |\phi^{j}(Y_{(n,j,c,w,h,)}) - \phi^{j}(\hat{Y}_{(n,j,c,w,h,)})|, \tag{17}$$

where J is the number of levels from which the VGG features are extracted, W^{j} and H^{j} are the spatial extensions of the respective level within the VGG classifier, C^{j} is the number of channel dimension of the respective level, and $\phi^{j}(Y_{(n,j,c,w,h,)})$

345   and $\phi^{j}(\hat{Y}_{(n,j,c,w,h,)})$ are the extracted features at level j from the predicted and observed images, respectively. In contrast to classification problems where high level features play a more important role, we multiply the low level features by a weighting factor of 8 to preserve the local features and give them more importance since these are more relevant to our regression task. The VGG network was originally trained with RGB images and giving NDVI and BT as input is not directly possible. To solve this issue, we replicate NDVI and BT along the channel dimension and feed each of them separately to the VGG network.

350   The impact of using this perceptual loss is evaluated in Appendix D. The entire loss function to be minimized is thus given as follows:

$$\mathcal{L} = \mathcal{L}_{MAE}^{NDVI} + 0.1\mathcal{L}_{VGG}^{NDVI} + \mathcal{L}_{MAE}^{BT} + 0.1\mathcal{L}_{VGG}^{BT}, \tag{18}$$

where $\mathcal{L}_{MAE}^{NDVI}$ and $\mathcal{L}_{VGG}^{NDVI}$ are the MAE and VGG losses on NDVI and $\mathcal{L}_{MAE}^{BT}$ and $\mathcal{L}_{VGG}^{BT}$ are the MAE and VGG losses on BT, respectively. The weighting factor 0.1 is set to balance the losses. The model is trained with a stochastic gradient descent.

355   More technical details regarding the training are provided in Sect. 3.5.

3.4 Baseline approaches

We study the performance of recently developed vision transformers on our task. We achieve this by sharing the overall model architecture and implementing the main building block inside the encoder and decoder according to different algorithms. The implemented models are as follows:

360 **U-Net** (Ronneberger et al., 2015) serves as a baseline of typical U-Net models. We implemented this model based on a 2D CNN with residual convolutional blocks. The U-Net model does not use an attention mechanism.

Swin Transformer V1 (Liu et al., 2021) performs self-attention in shifted windows to reduce the computational complexity compared to the original ViT. Transformers based on this model have been commonly applied for variety of tasks in remote sensing and computer vision (Wang et al., 2022a; Gao et al., 2021; Wang et al., 2022b; Aleissaee et al., 2023).

365 **Swin Transformer V2** (Liu et al., 2022) is an improved model of Swin V1. The attention mechanism is replaced with a scaled cosine attention to measure pixel feature similarities. Swin V2 utilizes post normalization layers inside the main block thus making the optimization of large models more stable. In addition, it proposes to replace the positional encoding inside the windows with a log-spaced continuous one to ease downstream tasks with pre-trained models.

Wave-MLP (Tang et al., 2022) is a MLP-Mixer-based transformer model. The basic block is built on a stack of MLPs.
370 Wave-MLP represents each pixel as a wave function with amplitude features representing pixel contents and phase to measure the relations with other pixels.

Apart of these models, we report the results for two NDVI/BT climatology baselines. The climatology is based on multi-year mean values computed from remote sensing observations pixel-wise and on a weekly basis. The first is a **climatology-I** computed from the years 1981-1988 which represents a prescribed satellite phenology before the beginning of the simulation.
375 The second is a **climatology-II** computed from the training years 1989-2016 in an overlap with the simulation period. The later climatology represents a function that models the annual cycles and it can be used to check if the models generalize beyond the mean annual cycles of the predicted NDVI/BT.

3.5 Implementation details

We re-implemented all aforementioned DL models in our framework and trained them with 3 different random seeds, this
380 ensures a fair comparison and better estimation. All models have almost the same capacity with ~ 12 million parameters. The encoders for the transformer models were pre-trained on ImageNet-1K (Deng et al., 2009) while the weights in the decoders and regression heads were initialized randomly from a standard normal distribution. To increase generalization and robustness of the models, we use 4 augmentation techniques. This includes flipping and rotating of the input with a probability of 0.5 and randomly perturbing the input variables by adding noise from a normal distribution with zero mean and a standard
385 deviation of 0.02 with a probability of 0.5. In addition, to generate the input corresponding to week i during training, we randomly average two days corresponding to the week i as an additional augmentation technique. All models were trained with the $\mathcal{L}$ loss Eq. (18) using the Pytorch framework (Paszke et al., 2019) with a learning rate 0.0003 and a scheduler to decay the learning rate by a factor of 0.9 every 16 epochs. AdamW optimizer (Loshchilov and Hutter, 2019) was used for the gradient descent with ($\beta_1 = 0.9$, $\beta_2 = 0.999$) and a weight decay 0.05. We use dropout probability of 0.2 and a stochastic
390 depth rate of 0.3. We train with a batch size of $N = 2$ for 100 epochs. For Swin Transformers, we set the window size to 8 and use the following number of heads $\{3, 6, 12\}$ for the encoder and the same order for the decoder. The down sampling in the encoder followed the original implementation in Swin Transformer. Wave-MLP was trained with the dimensionality $\{C_{(en,1)} = 64, C_{(en,2)} = 128, C_{(en,3)} = 320\}$ and $r_{mlp} = 4$ for both the encoder and the decoder. Wave-MLP and Swin V2 use

a dropout probability of 0.1 and a stochastic depth rate of 0.2. In addition, we follow the official implementation of Wave-MLP
and use GroupNorm (Wu and He, 2018) with a group of 1 instead of LayerNorm. Finally, all models were trained on individual
NVIDIA RTX A6000 GPUs with 48 GB.

3.6 Evaluation metrics

To measure the model performance, we use the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), coefficient of
determination (R^2), Pearson Correlation Coefficient (R_p), and Spearman Correlation Coefficient (R_s). In addition, we compute
the Bias as (predicted $-$ observed $= Y_{(w,h)} - \hat{Y}_{(w,h)}$). We compute the metrics for each sample and then average the values to
obtain the final metrics. MAE is computed from Eq. (15). While RMSE can be calculated as follows:

$$\text{RMSE}(Y_{(w,h)}, \hat{Y}_{(w,h)}) = \sqrt{\frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \left(Y_{(w,h)} - \hat{Y}_{(w,h)}\right)^2}. \tag{19}$$

R^2 measures the variation of the perdition from the regression fitted line and it is calculated as follows:

$$R^2(Y_{(w,h)}, \hat{Y}_{(w,h)}) = 1 - \frac{\sum_{w=1}^{W} \sum_{h=1}^{H} \left(Y_{(w,h)} - \hat{Y}_{(w,h)}\right)^2}{\sum_{w=1}^{W} \sum_{h=1}^{H} \left(Y_{(w,h)} - \bar{\hat{Y}}_{(w,h)}\right)^2}, \tag{20}$$

where $\bar{\hat{Y}}_{(w,h)}$ is the overall mean observed value. The highest value for R^2 is 1 which represents a perfect fit. Please note that
R^2 measures the variability in $\hat{Y}_{(w,h)}$ predicted by the model thus it is by definition inversely proportional to the variance and
noise in the observations and should be interpreted carefully.

Pearson correlation (R_p) is a parametric correlation that measures the linear correlation between the predicted and observed
values:

$$R_p(Y_{(w,h)}, \hat{Y}_{(w,h)}) = \frac{\sum_{w=1}^{W} \sum_{h=1}^{H} (Y_{(w,h)} - \bar{Y}_{(w,h)})(\hat{Y}_{(w,h)} - \bar{\hat{Y}}_{(w,h)})}{\sqrt{\sum_{w=1}^{W} \sum_{h=1}^{H} \left(Y_{(w,h)} - \bar{Y}_{(w,h)}\right)^2} \sqrt{\sum_{w=1}^{W} \sum_{h=1}^{H} \left(\hat{Y}_{(w,h)} - \bar{\hat{Y}}_{(w,h)}\right)^2}}, \tag{21}$$

where $\bar{Y}_{(w,h)}$ is the mean predicted value. The best value for R_p is 1 which represents a perfect positive correlation.

Spearman correlation (R_s) is a non-parametric measure of relationship between predicted and observed values that can be
calculated as follows:

$$R_s(Y_{(w,h)}, \hat{Y}_{(w,h)}) = R_p(R(Y_{(w,h)}), R(\hat{Y}_{(w,h)})), \tag{22}$$

where $R(Y_{(w,h)})$ and $R(\hat{Y}_{(w,h)})$ are ranks obtained from the predicted and observed values, respectively. A perfect positive
correlation occurs when R_s is 1.

4 Experimental results and analysis

4.1 NDVI and BT prediction

The quantitative results of the models are shown in Tables 1 and 2. Pixels without a vegetation cover (i.e., pixels over desert)
were excluded from the results. Including these pixels will overestimate the model performance since they have small variations
throughout the years. For the masking, we use NOAA quality assurance (QA) metadata. As can be seen from Tables 1 and 2,
all DL models outperform the first climatology-I baseline with a huge margin. This is because the climatology was calculated
before the simulation run. This climatology can not capture the dynamic after 3 decades. The second climatology-II baseline is
stronger. It uses information from multiple years within the simulation run. All DL models still achieve better results indicating
that the models have learned the seasonal dynamic beyond climatology. In addition, these climatologies baselines can not be
used to derive drought indices (Sect. 4.2) since the inter-annual variability in NDVI/BT is neglected as average cycles are

Table 1. Comparing the performance of different DL models. The metrics are shown for the validation set.

Validation - Years (2010, 2011, 2017) - 156 weeks					
NDVI					
Algorithm	MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)
Climatology-I	0.0550	0.0680	0.5763	0.8939	0.8669
Climatology-II	0.0326	0.0416	0.8372	0.9353	0.9113
U-Net	0.0277 $\pm$0.0001	0.0365 $\pm$0.0002	0.8743 $\pm$0.0008	0.9406 $\pm$0.0005	0.9172 $\pm$0.0005
Wave-MLP	0.0272 $\pm$0.0003	**0.0358** $\pm$0.0003	0.8784 $\pm$0.0018	0.9422 $\pm$0.0018	0.9183 $\pm$0.0021
Swin Transformer V1	0.0273 $\pm$0.0003	0.0362 $\pm$0.0003	0.8759 $\pm$0.0022	0.9411 $\pm$0.0013	0.9161 $\pm$0.0023
Swin Transformer V2	0.0277 $\pm$0.0003	0.0369 $\pm$0.0003	0.8703 $\pm$0.0021	0.9415 $\pm$0.0010	0.9167 $\pm$0.0008
Focal Modulation	**0.0269** $\pm$0.0001	**0.0358** $\pm$0.0002	**0.8790** $\pm$0.0017	**0.9432** $\pm$0.0001	**0.9194** $\pm$0.0009
BT (K)					
Algorithm	MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)
Climatology-I	2.9130	3.7302	0.8454	0.9466	0.9408
Climatology-II	2.3017	3.0020	0.8963	0.9601	0.9539
U-Net	1.9377 $\pm$0.0093	2.6067 $\pm$0.0057	0.9243 $\pm$0.0014	0.9667 $\pm$0.0004	0.9603 $\pm$0.0007
Wave-MLP	1.9200 $\pm$0.0491	2.5834 $\pm$0.0486	0.9248 $\pm$0.0035	0.9668 $\pm$0.0006	0.9603 $\pm$0.0007
Swin Transformer V1	1.9642 $\pm$0.0246	2.6341 $\pm$0.0303	0.9221 $\pm$0.0012	0.9661 $\pm$0.0005	0.9590 $\pm$0.0006
Swin Transformer V2	1.9741 $\pm$0.0191	2.6420 $\pm$0.0258	0.9225 $\pm$0.0013	0.9659 $\pm$0.0011	0.9590 $\pm$0.0014
Focal Modulation	**1.9010** $\pm$0.0071	**2.5364** $\pm$0.0073	**0.9280** $\pm$0.0012	**0.9679** $\pm$0.0001	**0.9614** $\pm$0.0007

Table 2. Comparing the performance of different DL models. The metrics are shown for the test set.

Test - Years (2012, 2018, 2019) - 139 weeks					
NDVI					
Algorithm	MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)
Climatology-I	0.0567	0.0697	0.5529	0.8933	0.8704
Climatology-II	0.0314	0.0400	0.8507	0.9433	0.9254
U-Net	0.0274 $\pm$0.0004	0.0359 $\pm$0.0005	0.8772 $\pm$0.0006	0.9435 $\pm$0.0006	0.9237 $\pm$0.0009
Wave-MLP	**0.0261** $\pm$0.0006	**0.0343** $\pm$0.0008	**0.8861** $\pm$0.0043	**0.9467** $\pm$0.0024	0.9252 $\pm$0.0011
Swin Transformer V1	0.0269 $\pm$0.0003	0.0355 $\pm$0.0004	0.8795 $\pm$0.0029	0.9442 $\pm$0.0010	0.9239 $\pm$0.0014
Swin Transformer V2	0.0270 $\pm$0.0005	0.0359 $\pm$0.0005	0.8766 $\pm$0.0038	0.9447 $\pm$0.0012	0.9251 $\pm$0.0020
Focal Modulation	0.0266 $\pm$0.0003	0.0350 $\pm$0.0004	0.8808 $\pm$0.0014	0.9454 $\pm$0.0009	**0.9253** $\pm$0.0016
BT (K)					
Algorithm	MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)
Climatology-I	2.8806	3.6864	0.8447	0.9485	0.9470
Climatology-II	2.2024	2.8880	0.9036	0.9623	0.9606
U-Net	1.9920 $\pm$0.0148	2.6652 $\pm$0.0262	0.9164 $\pm$0.0021	0.9644 $\pm$0.0009	0.9616 $\pm$0.0005
Wave-MLP	1.9376 $\pm$0.0184	2.6221 $\pm$0.0177	0.9172 $\pm$0.0005	0.9647 $\pm$0.0005	0.9619 $\pm$0.0008
Swin Transformer V1	1.9563 $\pm$0.0329	2.6381 $\pm$0.0397	0.9169 $\pm$0.0038	0.9649 $\pm$0.0009	0.9627 $\pm$0.0008
Swin Transformer V2	1.9516 $\pm$0.0639	2.6277 $\pm$0.0874	0.9183 $\pm$0.0060	0.9641 $\pm$0.0025	0.9619 $\pm$0.0020
Focal Modulation	**1.9179** $\pm$0.0458	**2.5745** $\pm$0.0470	**0.9204** $\pm$0.0030	**0.9664** $\pm$0.0007	**0.9636** $\pm$0.0006

used. Furthermore, comparing the correlation and results of BT with NDVI, we can observe that all models achieve higher correlation metrics (R^2, R_p, and R_s) on BT than NDVI. This can be explained by the fact that NDVI is a composition of two bands while BT is only derived from the infrared band, thus it is harder for the models to estimate NDVI than BT. In general, all DL models provide close results and are considered suitable for the task. Focal Modulation clearly outperformed other DL models on the validation set for both NDVI and BT predictions. For the test set on NDVI, it comes slightly after the Wave-MLP model. However, Focal Modulation can generalize better for BT thus providing a balanced prediction between NDVI and BT and consequently it is capable to generate an overall better prediction.

In Table 3, we report the estimated inference time for the DL models. For the Focal Modulation model, the estimated inference time to generate one sample for NDVI and BT containing $397 \times 409 \times 2$ grid points is 0.24 ± 0.01 seconds on one NVIDIA GeForce RTX 3090 GPU and 12 ± 0.1 seconds on one AMD Ryzen 9 3900X 12-Core CPU. U-Net with 2D CNN dose not include operations for the attention mechanism thus it is the fastest but the performance is lower.

Table 3. Inference time in seconds for different DL models.

Algorithm	GPU[1]	CPU[2]
U-Net	0.09 ±0.02	5 ±0.2
Wave-MLP	0.28 ±0.00	10 ±0.3
Swin Transformer V1	0.18 ±0.00	11 ±0.2
Swin Transformer V2	0.19 ±0.00	11 ±0.2
Focal Modulation	0.24 ±0.01	12 ±0.1

[1]NVIDIA GeForce RTX 3090 GPU

[2]AMD Ryzen 9 3900X 12-Core CPU

Qualitative results for the model prediction with Focal Modulation are shown in Figs. 4 and 5. We take weeks from different seasons through the years and remove pixels over desert for the calculations of bias distribution and regression line. Positive bias values mean that the model overestimates NDVI (BT) while negative ones indicate that the model underestimates NDVI (BT). As shown in Figs. 4 and 5, the biases vary across the weeks and locations. For week 7 in 2012, the biases for both NDVI and BT are relatively high. Week 26 in 2019 exhibits similar high biases in both NDVI and BT over high latitudes regions. The respective distribution of biases is also shown in Figs. 4 and 5. Overall, the results show that the dynamics over the years are well captured. The biases for both NDVI and BT are closely centered around zero with a shift for the center of bias distribution from zeros. This shift is however in the same direction for both NDVI and BT. We can also observe that the model fits the regression lines better for weeks 14, 26, and 39 than for week 7 in winter 2012. The comparison between the distributions of predicted and observed NDVI/BT confirms also the observation that the model captured the dynamic throughout the years.

While this provides examples of the performance for individual samples, in Fig. 6 we provide an additional experiment where we analyze biases of model predictions within different seasons of the year and over PRUDENCE regions (see Appendix Fig. C1 for the definition of PRUDENCE regions). This allows us to assess the model weaknesses and strengths with different seasonality and spatial variability. The mean biases were computed pixel-wise from both the validation and test years time-series where we computed the biases for each pixels from the weeks that belong to a specific season and averaged the results to obtain the last metric. In addition, we computed Pearson Correlation R_p pixel-wise in a similar way. As seen in Fig. 6, there are clusters of positive/negative biases that vary with seasons over specific regions. For instance, for NDVI prediction, the eastern part of British Isles exhibits positive biases for all seasons while Iceland and north Africa show constant negative biases. For BT, Southeastern Europe has persistent positive biases with larger errors during winter. Pixels over desert, i.e., north Africa, show less variability in NDVI where only little seasonality is shown as in Fig. 4. Thus, such regions are easier to predict with relatively small biases. However any fluctuation in NDVI prediction over these pixels will lead to lower correlation compared to other regions since the time series primarily represent small variations around the mean NDVI value. In comparison to other seasons, the winter season has relatively poor predictions especially in the high latitudes regions. One possible explanation for these
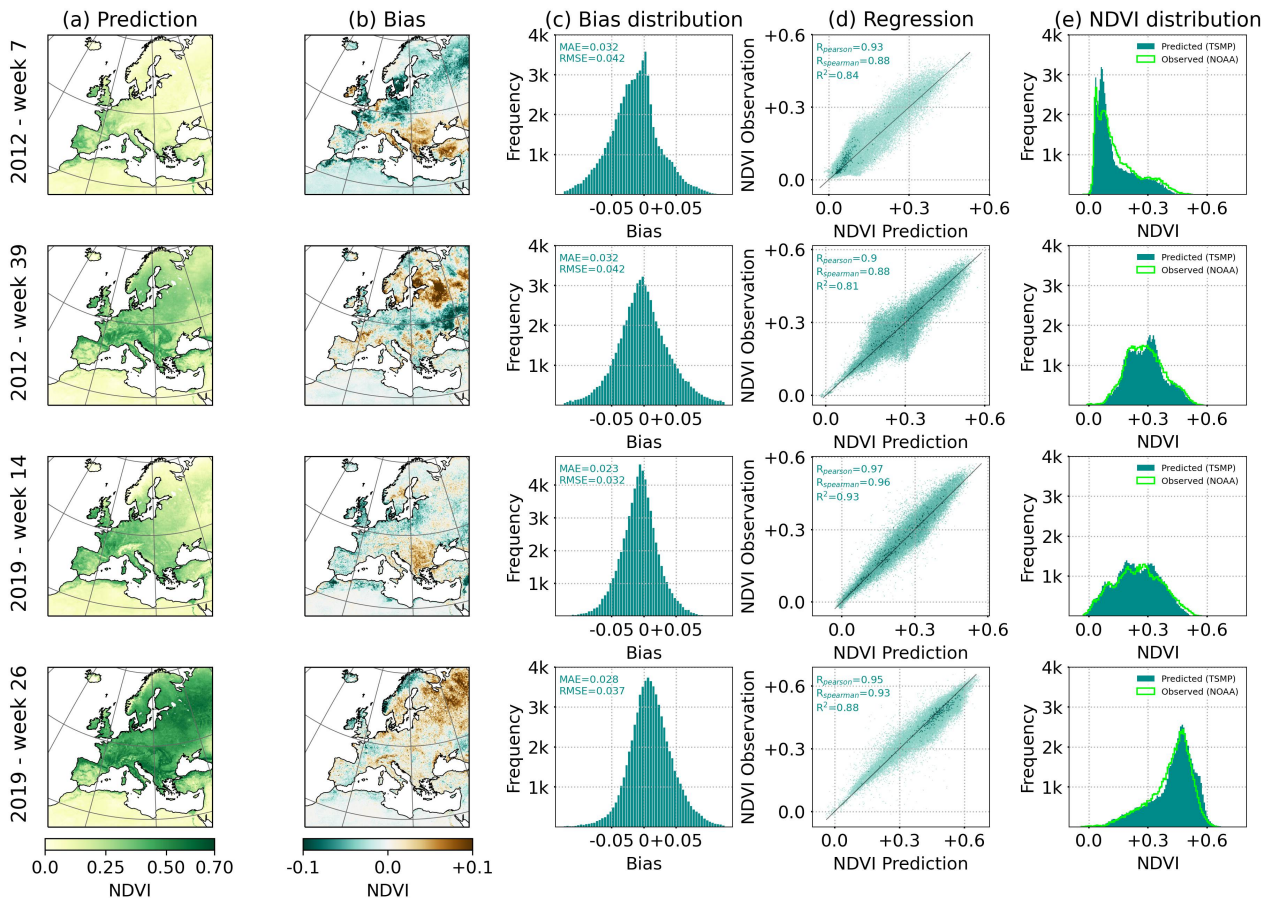
Figure 4. Examples predictions for weekly NDVI from the test set. (a) Predicted NDVI. (b) Bias computed as prediction minus observed. (c) Distribution of biases. (d) Regression results as predicted versus observed. (e) Distribution of NDVI values for NOAA observation and model prediction. The metrics are computed over all pixels with vegetation cover.

errors is the lack of accurate training data in Scandinavian regions during winter. For instance, previous studies on ParFlow-CLM models showed that hydrological modeling performs worse in northeastern Europe due to errors in snow dynamics and regional forces (Naz et al., 2023; Furusho-Percot et al., 2019). It was also shown by Yang et al. (2020) and Eisfelder et al. (2023) that high latitude regions are less reliable to derive vegetation products due to snow cover and its effects on the albedo and larger sensor zenith angles. Another source of model errors is that NOAA vegetation products depend on temporal compositing to handle high frequency and atmosphere transmittance (Yang et al., 2020). The absence of a generalized physical-based model to enhance accuracy over various surfaces and for all conditions generates difficulties for satellite products (Kogan, 1995b). Nagol et al. (2009) assessed the uncertainty of NDVI in this regards. These issues add some uncertainties to the model training and evaluation. Using more recent atmospheric correction methods such as in (Moravec et al., 2021) could also enhance the results. Furthermore, as mentioned in Sect. 2.1, the TSMP simulation was performed in a free mode and had no modelling
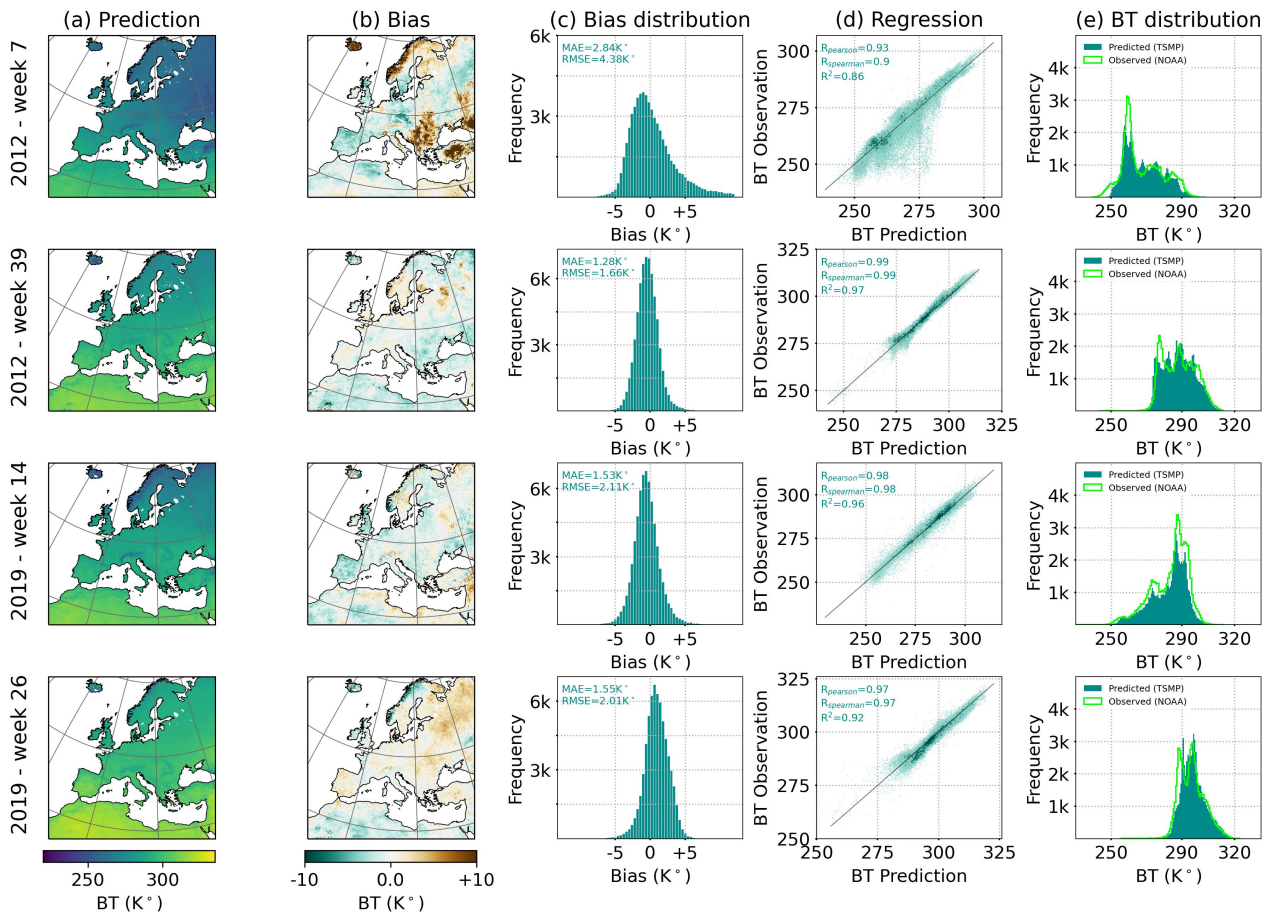
Figure 5. Examples predictions for weekly BT from the test set. (a) Predicted BT. (b) Bias computed as prediction minus observed. (c) Distribution of biases. (d) Regression results as predicted versus observed. (e) Distribution of BT values for NOAA observation and model prediction. The metrics are computed over all pixels with vegetation cover.

of anthropogenic-related influences. Given that agricultural systems and human activities which are interlinked with drought events could change and follow adaptation strategies (Van Loon et al., 2016), this certainly contributes to the error budget of the model. Developing realistic land use and water management scenarios within a probabilistic TSMP could reduce these errors. In addition, the uncertainty in TSMP is highly linked to potential errors in the driving forces and spin-up initialization. While these errors are common limitations of simulations and remote sensing data, it should be noted that the prediction of a DL model has its own uncertainty. Therefore, more efforts are needed to recognize the sources of uncertainty in model prediction (Sect. 4.2).

In Fig. 7, we visualize the computations over each PRUDENCE region separately. For Figs. 7a and 7b, we fit a normal distribution over the normalized histogram of biases for each season and over all PRUDENCE regions. For instance, positive
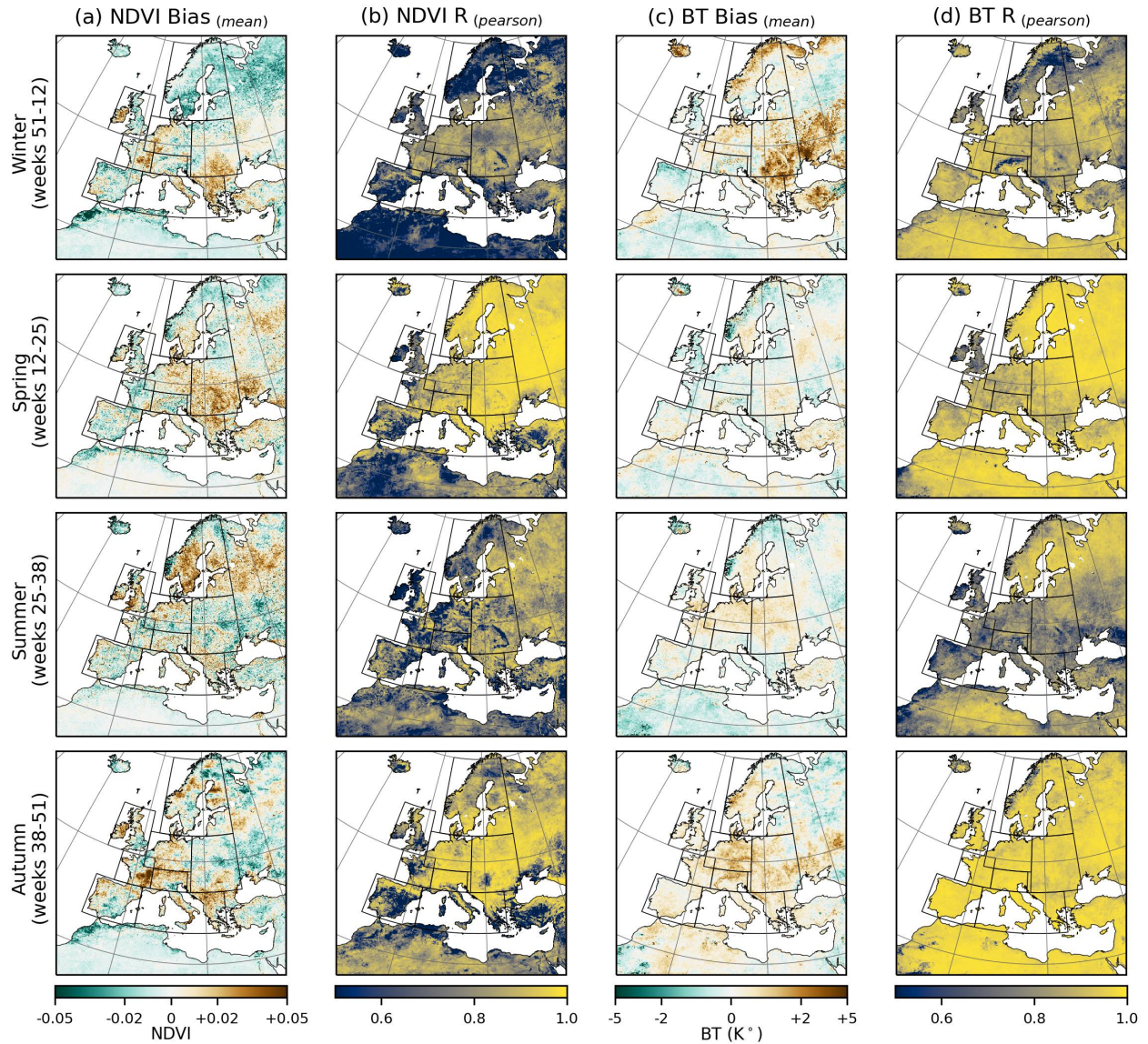
Figure 6. An analysis of uncertainty and model generalization for different times of the year. The analysis was performed on the validation and test sets as one set. (a) NDVI mean bias. (b) NDVI mean Pearson Correlation. (c) BT mean bias. (d) BT mean Pearson Correlation.

480 shifts of the estimated means are shown in NDVI for both FR and AL regions during autumn. The same pattern is shown for SC and BI during summer. As can also be seen in Fig. 7b, a positive shift for BT is shown for all regions during autumn. Furthermore, the shape of the distribution gives an overview of the prediction homogeneity within the region, i.e., the prediction is highly uncertain over EA during winter and consequently has a relatively high standard deviation. The mean values in Figs. 7c and 7d represent the expected MAE for all seasons combined. Fig. 7c indicates that in general the model predictions for NDVI
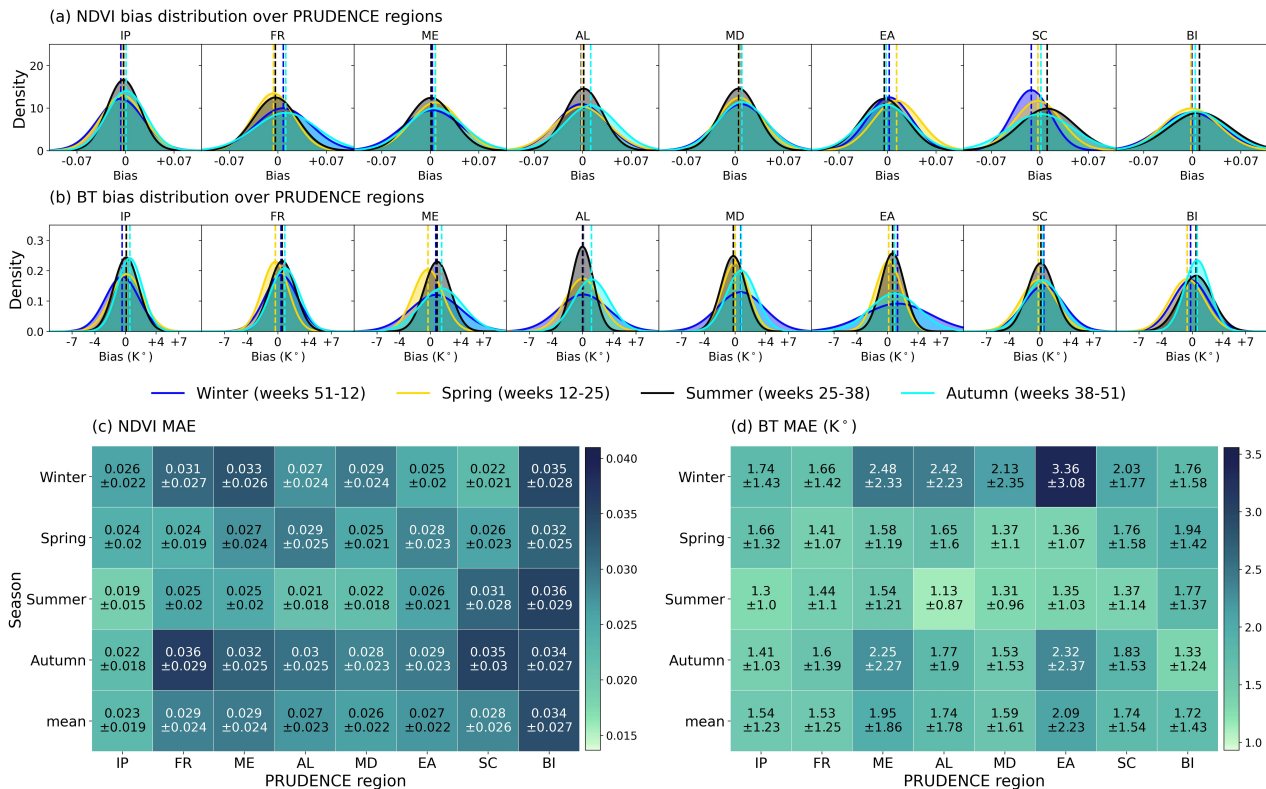
Figure 7. An analysis of uncertainty and model generalization for different time of the year over each PRUDENCE region. The analysis was performed on the validation and test sets as one set. (a) NDVI bias distribution. (b) BT bias distribution. Shown are the probability density functions. (c) NDVI MAE. (d) BT MAE.

are less certain during autumn in comparison to other periods and over BI within the PRUDENCE regions. For BT, it can be seen from Fig. 7d that the prediction is less certain during winter and over ME and EA regions.

4.2 Agricultural drought assessment

In this section, we assess the model capability to predict different agricultural droughts indices on a high temporal resolution (weekly basis). More specifically, we use the predicted NDVI and BT along with their multi-year climatology to derive NDVI anomaly, BT anomaly, VCI, TCI, and VHI drought indices. NDVI and BT anomalies were computed by subtracting the mean value of the respective pixel and week from the predictions (observations). VCI, TCI, and VHI were computed from Eq. (2)-(4). Figs. 8 and 9 compare the predicted agricultural drought indices VCI, TCI and VHI by the focal modulation model with the observed ones from NOAA remote sensing data for the years 2010-2012 (Fig. 8) and 2017-2019 (Fig. 9). We spatially average the values inside each PRUDENCE region and plot their respective time-series on a weekly basis. Generally, values below 40 are identified as abnormally dry conditions (Kogan et al., 2015; Yang et al., 2020). Overall, the prediction resembles
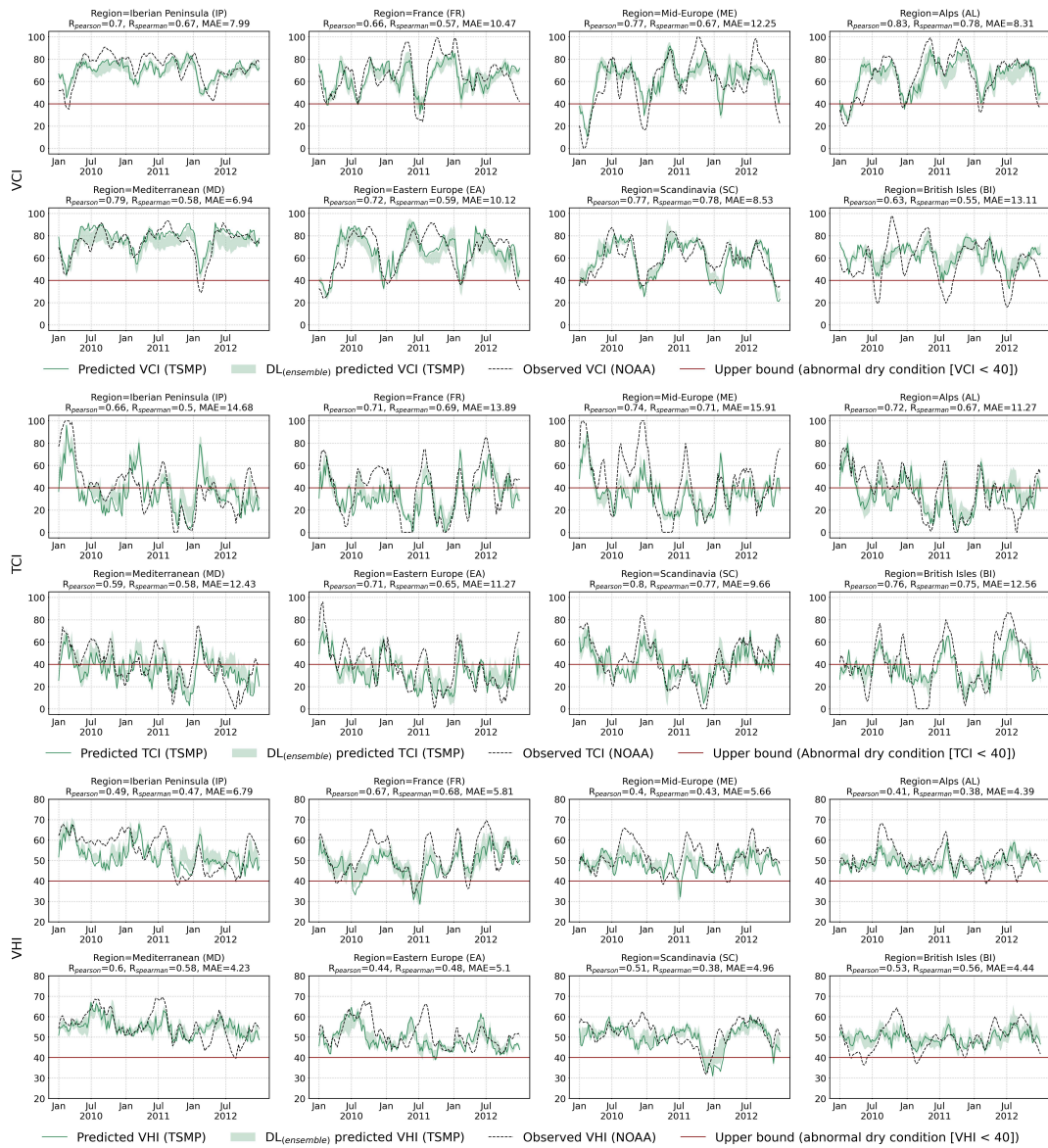
Figure 8. Comparison of spatially averaged weekly agricultural drought indices between the model prediction and NOAA observation over each PRUDENCE region. Drought indices were computed from the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with the Focal Modulation Network. The ensemble model is the result of all DL models described in Sect. 3.4. NDVI and BT anomalies are provided in Appendix Fig. F1.

the seasonal wetness and dryness at the regional scale. The agreements between predictions and observations vary across regions and time with satisfactory R_p values ranging from 0.50 to 0.77, 0.38 to 0.70, and 0.50 to 0.75 for VCI, TCI, and VHI, respectively. MAE values fluctuate in the range 9.99-6.81, 13.88-10.24, and 5.80-2.69 for VCI, TCI, and VHI, respectively.

Figure 9. Comparison of spatially averaged weekly agricultural drought indices between the model prediction and NOAA observations over each PRUDENCE region. Drought indices were computed based on the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with the Focal Modulation Network. The ensemble model is the result of all DL models described in Sect. 3.4. NDVI and BT anomalies are provided in Appendix Fig. F2.

While there is a satisfactory agreement with observations, there are some obvious discrepancies, i.e., in TCI over the Iberian Peninsula (IP) during summer 2018. More interestingly, we show the bounded results of an ensemble of DL models. This ensemble is the results of all DL models. As can be seen, all DL models which are based on different algorithms yield close

500
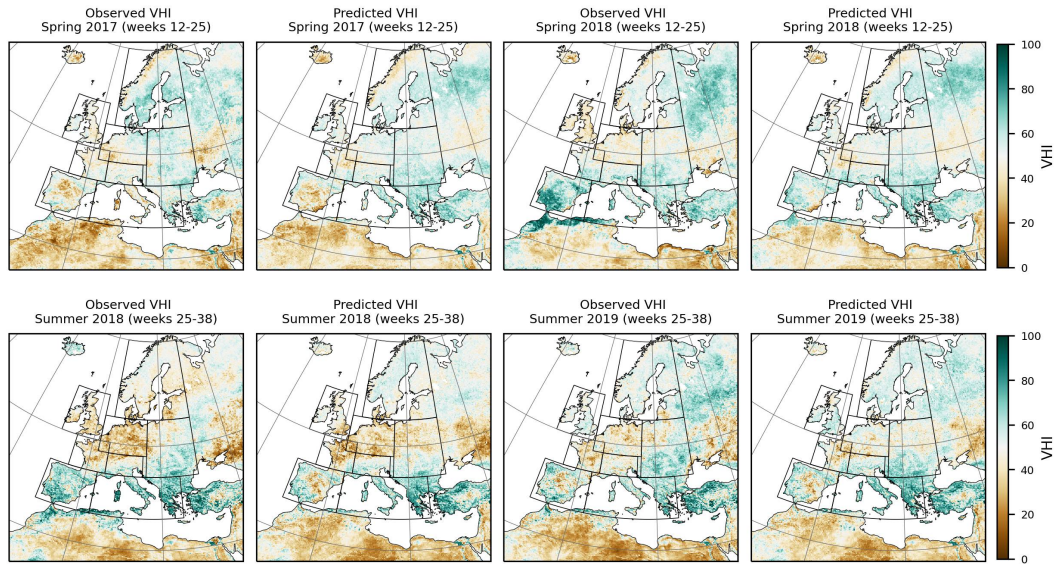
Figure 10. Comparison between the seasonal predicted Vegetation Health Index (VHI) and NOAA observations over Pan-Europe domain.

predictions with small standard deviations. This supports that errors in model prediction are probably to be more attributed to biases in the TSMP model and remote sensing reference data. In this respect, Yang et al. (2021b) showed that vegetation products over regions with extreme little seasonality, i.e., desert and high mountains have higher errors. This can be seen from Eq. (2)-(4), where small differences between maximum and minimum values could lead to higher deviation in the vegetation indices.
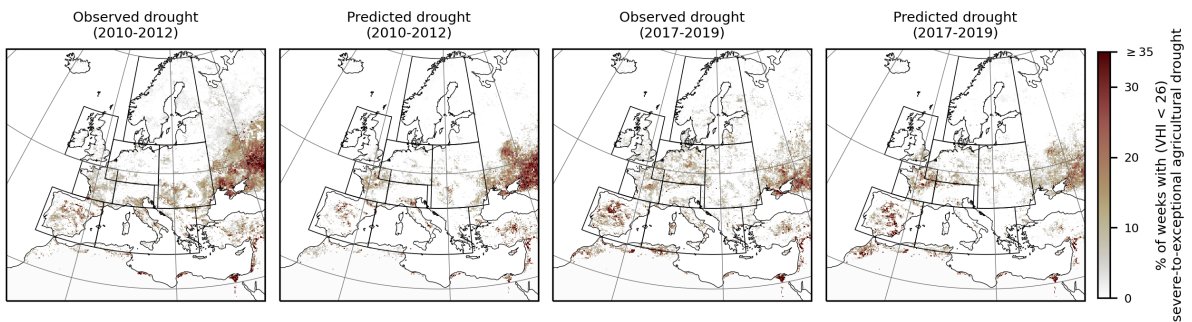
Figure 11. Comparison between the predicted drought frequency and NOAA observations over Pan-Europe domain. Frequency represents the percent of weeks with Severe-to-Exceptional drought events (VHI < 26).

Finally, as observed from the plots, the thermal surface condition represented by TCI contributes more to the agricultural drought events over Europe than the deficiency in vegetation moisture condition approximated as VCI. This is in agreement
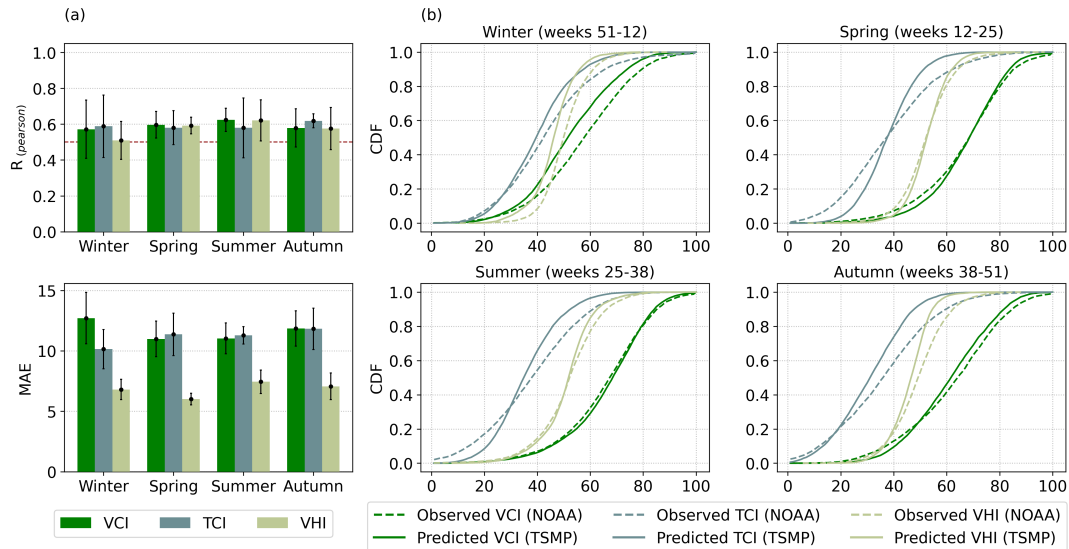
Figure 12. An evaluation of seasonally predicted agricultural drought indices with ground truth NOAA observations at the resolution $0.88°$. (a) Bottom is mean absolute errors (MAE) and top is Pearson Correlations (R_p) for different seasons. (b) Comparison of the cumulative distribution functions between prediction and observations.

with (Zeng et al., 2023), who showed that drought affecting vegetation is more likely to be associated with high abnormal temperatures in Europe. This is critical for studies that rely on NDVI as the solely vegetation product to identify drought events over Europe (Sect. 1). In the Appendix, we show the time-series for NDVI and BT anomalies in Figs. F1 and F2. We also show vegetation health maps for different seasons from the validation and test years. These predicted maps are depicted in Fig. 10. As shown, the model predicts an increase of agricultural droughts in the summer of 2018 in the Mid Europe and France regions. Xoplaki et al. (2023) associated this extremely dry summer with compound extreme events.

Furthermore, in Fig. 11, we provide an analysis about the frequency of extreme droughts for the two periods 2010-2012 and 2017-2019. Frequency represents the percent of weeks with Severe-to-Exceptional drought events where VHI < 26 (Kogan et al., 2020). While Figs. 8 and 9 provide overviews of the averaged values over the regions, the analysis in Fig. 11 provides a spatial comparison between the model prediction and observations. The major hotspots for the highest extremes are found outside the Prudence regions (North of the Black Sea, Northwest Africa, Egypt and Northwest of the Middle East). In comparison to the Prudence regions, the Iberian Peninsula and France exhibit more extreme droughts. The model predicts more extreme droughts in those regions and agrees with observations. For the period 2010-2012, the model predicts less extreme droughts in the Mediterranean and Eastern Europe. While for the 2017-2019 period, the model underestimates the frequency of extremes in the Mid Europe region.

Moreover, Fig. 12 evaluates the model capability to capture seasonal dynamic in drought indices. As seen in Fig.12a, the mean R_p values are greater than 0.5 and around 0.6 for all seasons. MAE values show the highest error in VCI for the winter season. One notable observation is that the error bars have relatively large values indicating a variation in prediction accuracy

across the years within the same seasons. This can be attributed to the seasonality shift in the long-term trends. Klimavičius et al. (2023) showed that meteorological forces like air temperature have strong impact on growing seasons and phenological trends of NDVI (VCI). The cumulative distribution functions (CDF) in Fig. 12b expresses the main difference in CDF for VCI during winter. While the model prediction overestimates TCI over the seasons.

4.3 Variable importance

To analyze the impact of each TSMP model components on the model prediction, we present in Table 4 the prediction results obtained with COSMO, CLM, and ParFlow. For this experiment, we train 3 models based on focal modulation with the dimensionality $\{C_{(en,1)} = 64, C_{(en,2)} = 128, C_{(en,3)} = 256\}$. As seen in Table 4, compared to CLM and ParFlow, COSMO achieves the best results for the validation set while CLM outperforms both for the test set. COSMO has important variables related to water contents and clouds along with other variables related to the atmospheric effects on the reflected signal on the ground. CLM has complementary variables related to heat fluxes and evapotranspiration. ParFlow can approximate the hydrology and serve as a proxy for the soil conditions. The results show that all model components are useful and the best result is obtained when all these models are used.

Table 4. Impact of TSMP model components on the model performance. The metrics are shown for the validation and test sets. All models were trained with Focal Modulation Network.

Validation - Years (2010, 2011, 2017) - 156 weeks										
	NDVI					BT (K)				
Model	MAE($\downarrow$)	RMSE($\downarrow$)	$R^2(\uparrow)$	$R_p(\uparrow)$	$R_s(\uparrow)$	MAE($\downarrow$)	RMSE($\downarrow$)	$R^2(\uparrow)$	$R_p(\uparrow)$	$R_s(\uparrow)$
COSMO	0.0281	0.0372	0.8696	0.9403	0.9160	1.9975	2.6389	0.9227	0.9667	0.9615
CLM	0.0289	0.0382	0.8586	0.9369	0.9115	2.0187	2.7080	0.9160	0.9653	0.9600
ParFlow	0.0303	0.0396	0.8500	0.9314	0.9042	2.2029	2.9254	0.9052	0.9617	0.9545
COSMO + CLM + ParFlow	**0.0270**	**0.0359**	**0.8781**	**0.9433**	**0.9184**	**1.8981**	**2.5433**	**0.9266**	**0.9679**	**0.9613**
Test - Years (2012, 2018, 2019) - 139 weeks										
	NDVI					BT (K)				
Model	MAE($\downarrow$)	RMSE($\downarrow$)	$R^2(\uparrow)$	$R_p(\uparrow)$	$R_s(\uparrow)$	MAE($\downarrow$)	RMSE($\downarrow$)	$R^2(\uparrow)$	$R_p(\uparrow)$	$R_s(\uparrow)$
COSMO	0.0285	0.0372	0.8619	0.9437	0.9238	2.0847	2.7549	0.9060	0.9633	0.9612
CLM	0.0269	0.0355	0.8782	0.9443	0.9238	1.9362	2.6303	0.9185	0.9650	0.9637
ParFlow	0.0291	0.0379	0.8648	0.9396	0.9175	2.2663	2.9481	0.8962	0.9635	0.9604
COSMO + CLM + ParFlow	**0.0268**	**0.0353**	**0.8795**	**0.9452**	**0.9243**	**1.8730**	**2.5277**	**0.9227**	**0.9672**	**0.9642**

Figure 13. Channel attention for TSMP input variables. The activations are shown for both NDVI (top) and BT (bottom) with respect to all weeks in the validation and test sets.

540 While Table 4 provides an overview on the importance of model components, apriori choice of proper input variables from each of these model components to predict NDVI and BT requires substantive efforts and assumptions. Especially, when the underlying physical process to construct albedo/emissivity from TSMP and tracing the atmospheric effects with satellite and solar geometry is very complex. Channel attention (Woo et al., 2018; Hu et al., 2018) was commonly used in the field of computer vision and remote sensing to enhance feature representations inside DL models. A channel attention module aims

545 to calibrate the input variables/channels by learning an input-dependent scale for each channel. Thus, it can model the inter-correlation across variables adaptively. In this work, we propose to use channel attention to determine the relative importance of TSMP input variables. Implementation details about the module are provided in Sect. B and Fig. B1. We used channel attention directly before the patch embedding for the U-Net model. To disentangle the correlation between NDVI and BT, we trained two separated models. One to predict NDVI and another one to predict BT. Note that we only used channel attention

550 for this experiment. Fig. 13 provides example attentions induced for each input variable from COSMO, CLM, and ParFlow with respect to all weeks in the test and validation sets. The attention value is the mean value and it represents the variable importance to predict NDVI (BT). Error bars show how the attention changes across the weeks and input samples. We observe that the distributions of attention values for NDVI and BT is close. This indicates that the importance of highly relevant input variables are probably shared for both NDVI and BT. In addition, the standard deviations (error bars) suggest that the choice

555 of prior explanatory variables is not trivial since the relative importance can change with time and input samples.

Overall, not all variables are relevant for the model. For COSMO, atmosphere water divergence (hudiv), humidity-related variables (hus, hur), precipitation variables (pr, prc, prg), surface air pressure (ps), drag coefficient of heat (tch) and geopotential height (zg200) receive the highest attention from the DL model. For CLM, all variables are considered important with snowfall flux (prsn) and precipitation on ground (prso) being less important. Regarding ParFlow variables, it can be seen that the model considers most underground water-related variables as relatively important. This is intuitive since water and the amount of underground water storage are important factors for the vegetation growth. The availability of groundwater supply can reduce vulnerability to agricultural drought (Meza et al., 2020; Ma et al., 2021). Some previous studies showed that precipitation and temperature are strong predictors of NDVI (Miao et al., 2015; Wu et al., 2020; Gao et al., 2023). In addition, the climatology of long-term NDVI is highly correlated with precipitation and the biome classification (Yang et al., 2021b). The relatively high value for zg200 in BT prediction can be explained as the decrease in zg200 increases the likelihood of heatwave occurrence (Miralles et al., 2019). The attention values for COSMO can be interpreted as Nagol et al. (2009) showed that scattering and absorption in the atmosphere affect the visible and near infrared radiance considerably. Shi et al. (2018) and Geiss et al. (2021) analyzed the influence of clouds related parametrization on visible and infrared satellite images and found that the accuracy is closely related to the cloud representation. A further study about the impact of surface and air pressure and water and ice clouds on visible and near-infrared bands can be found in Baur et al. (2023). It needs to be emphasized that the correlations shown in Fig. 13 must not be interpreted as a causal reasoning. One main reason is that data in Earth science are subject to complicated interactions and inherently inter-dependent. There may be hidden confounding variables that influence the explanatory variables as well as the evolution of the climate and vegetation variability. It is also worth noting that the learned variable importance by machine learning models is dependent on how the variables are represented in the training data (Betancourt et al., 2022). Furthermore, some variables have larger biases than others since TSMP was run in a free mode simulation. This may drive the model to rely less on such variables even if they are considered important in scientific literature. The same thing applies to highly correlated variables where changing the model architecture may alter dependencies as well (Betancourt et al., 2022).

5 Conclusions and outlook

In this paper, we presented a new deep learning based approach for vegetation health prediction from a regional climate simulation. The developed model enabled the prediction of variables which are not part of the input simulation. In particular, we developed a vision transformer model with focal modulation to predict NDVI and BT images from a long-term TSMP (Ground to Atmosphere G2A) simulation at $0.11°$ resolution and on a weekly basis. We further validated the approach with NOAA remote sensing satellite observations and identified regions of uncertainty in the model predictions. As part of this, agricultural drought assessment was performed based on vegetation health products, namely VCI, TCI and VHI, which were derived from the predicted NDVI and BT, as well as long-term climatology. In this regard, the applicability of the model was spatially and temporally analyzed at a continental scale. Additionally, we extended the commonly used explanatory variables by using plenty of TSMP variables and analyzed their relative importance for the task with channel attention as an explainable

AI method. The evaluation confirms that a DL model that was trained on observations has the capacity to predict NDVI and BT from a TSMP climate simulation with a sufficiently good agreement with real-world satellite observations.

Although our model is trained to predict vegetation products as they would be observed from the AVHRR platform, it would be possible to predict target variables from different platforms or following different atmospheric corrections. This could be done as future work by training multiple DL models. Moreover, our work can be extended to predict other vegetation products from different satellite platforms depending on requirements. The proposed approach can be used to predict future trends in the vegetation dynamic based on climate scenarios. Providing this information, the model can help to recognize regions that are expected to be more vulnerable to agricultural drought risks. The predicted satellite-based indices can be combined with different meteorological drought indices to provide more comprehensive drought assessments under future climate change. We believe that our approach could also be useful to combine deep learning with data assimilation, i.e., to simulate remote sensing products from down-scaled simulations and to be used as a supportive evaluation framework to further investigate the predictive capability of the simulation to reproduce drought events and consequently to improve the TSMP model development.

Code and data availability. The source code and the pretrained models to reproduce the results are published at https://zenodo.org/records/10015049 (Shams Eddin and Gall, 2023a). The source code is also available on GitHub at https://github.com/HakamShams/Focal_TSMP. The pre-processed data used in this study are available at https://doi.org/10.5281/zenodo.10008815 (Shams Eddin and Gall, 2023b). The original TSMP data are stored at Jülich Research Centre at https://datapub.fz-juelich.de/slts/cordex/index.html (Furusho-Percot et al., 2019a), as well as at PANGAEA at https://doi.pangaea.de/10.1594/PANGAEA.901823 (Furusho-Percot et al., 2019b). The raw vegetation health products can be downloaded from the National Oceanic and Atmospheric Administration (NOAA), Center for Satellite Applications and Research (STAR) at https://www.star.nesdis.noaa.gov/star/index.php (Yang et al., 2020).

Appendix A: Datasets

Table A1. Technical details on the output variables in the TSMP EUR-11 simulation. For more information on the data, we refer to Furusho-Percot et al. (2019).

Model	Variable name	Long name	Unit	Level
COSMO	awt	Atmosphere total water content	kg m^{-2}	1
	capec	Specific convectively available potential energy	j kg^{-1}	1
	capeml	Cape of mean surface layer parcel	j kg^{-1}	1
	ceiling	Cloud ceiling height (above mean sea level)	m	1
	cli	Vertical integrated cloud ice	kg m^{-2}	1
	clt	Total cloud fraction	1	1
	clw	Vertical integrated cloud water	kg m^{-2}	1
	hudiv	Atmosphere water divergence	kg m^{-2}	1
	hur2	2m relative humidity	%	1
	hur(200, 500, 850)	Relative humidity (at 200, 500 and 850 hpa)	%	3
	hus2	2m specific humidity	1	1
	hur(200, 500, 850)	Relative humidity (at 200, 500 and 850 hpa)	1	3
	incml	Convective inhibition of mean surface layer parcel	j kg^{-1}	1
	pr	Precipitation	kg m^{-2}	1
	prc	Convective precipitation	kg m^{-2}	1
	prg	Large scale precipitation	kg m^{-2}	1
	prt	Total rain water content vertically integrated	kg m^{-2}	1
	ps	Surface air pressure	pa	1
	psl	Sea level pressure	pa	1
	snt	Total snow content vertically integrated	kg m^{-2}	1
	ta(200, 500, 850)	Air temperature (at 200, 500 and 850 hpa)	K	3
	tch	Drag coefficient of heat	1	1
	td2	2m dew point temperature	K	1
	ua(200, 500, 850)	Eastward wind (at 200, 500 and 850 hpa)	m s^{-1}	3
	uas	Eastward near-surface wind velocity	m s^{-1}	1
	va(200, 500, 850)	Northward wind (at 200, 500 and 850 hpa)	m s^{-1}	3
	vas	Northward near-surface wind velocity	m s^{-1}	1
	zg(200, 500, 850)	Geopotential height (at 200, 500 and 850 hpa)	m	3
	zmla	Height of boundary layer	m	1
CLM	evspsbl	Evapotranspiration	mm s^{-1}	1
	hfls	Surface upward sensible heat flux	w m^{-2}	1
	hfss	Surface upward sensible heat flux	w m^{-2}	1
	prsn	Snowfall flux	$\text{kg m}^{-2}\,\text{s}^{-2}$	1
	prso	Precipitation on ground	$\text{kg m}^{-2}\,\text{s}^{-2}$	1
	rlds	Incoming shortwave radiation	w m^{-2}	1
	tas	Near-surface air temperature	K	1
	trspsbl	Transpiration	w m^{-2}	1
ParFlow	sgw	Groundwater saturation	1	15
	wtd	Water table depth	m	1

Table A2. Technical details on the static variables from CLM in the TSMP EUR-11 simulation and the computed static variables.

Model	Variable name	Long name	Unit	Level
CLM	orog	Surface height or digital elevation model (DEM)	m	1
	sftlf	Land-sea fraction	%	1
	zbot	Atmospheric reference height (from COSMO to CLM)	m	1
Computed from Land-sea fraction	-	Distance to water	km	1
Computed from Orography	-	Roughness	1	1
Computed from Orography	-	Slope	○	1

Table A3. Technical details on the spectral channel characteristics for Advanced Very High Resolution Radiometer (AVHRR) and Visible Infrared Imaging Radiometer Suite (VIIRS).

Satellite system	Spectral band	Spectral range (μm)
AVHRR	ρ_R	0.58 - 0.68
	ρ_{NIR}	0.725 - 1.1
	ρ_{IR}	10.3 - 11.3
VIIRS	ρ_R	0.600 - 0.680
	ρ_{NIR}	0.846 - 0.885
	ρ_{IR}	10.500 - 12.400

Appendix B: Channel Attention

610 Channel attention aims to condense the input channels into a lower dimensionality and then construct channel scales with a sigmoid activation function ($\text{Sigmoid}(x) = \frac{1}{1+\mathrm{e}^{-x}} \in [0,1]$). In this manner, the neural network learns to calibrate the input channels with the learned scaling depending on the input channels. Given $\mathbf{X} \in \mathbb{R}^{V \times W \times H}$ as input TSMP simulation, where V is the number of output variables from COSMO, CLM, and ParFlow models, and W and H are the spatial extensions, the channel attention is computed as follows:

615
$$\text{ChannelAttention}(\mathbf{X}) \triangleq \text{Sigmoid}\Big(\text{MLP}\big(\text{GAP}(\mathbf{X})\big) + \text{MLP}\big(\text{GSD}(\mathbf{X})\big)\Big) \in \mathbb{R}^{V \times 1 \times 1}, \tag{B1}$$

where Sigmoid is the sigmoid function, MLP consists of two linear layers with a ReLU activation in between. The first decreases the dimension to $\frac{V}{r_{att}}$ and the subsequent layer maps it back to V. GAP is global average pooling, and GSD is the global standard deviation. For the experiments in Sect. 4.3, we trained two separated models for NDVI and BT independently with ($r_{att} = 3$, $r_{att} = 5$) and with the dimensionality $\{C_{(en,1)} = 64, C_{(en,2)} = 128, C_{(en,3)} = 256\}$, and averaged the results.
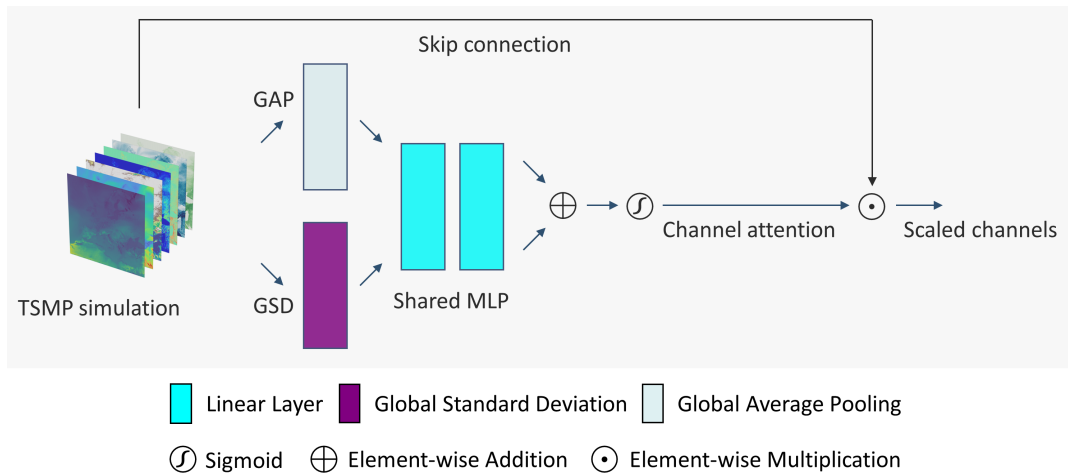
Figure B1. Illustration of the channel attention implementation. The output of channel attention is multiplied with the input TSMP to scale the channels from COSMO, CLM, and ParFlow according to their activation values.

Appendix C: PRUDENCE scientific regions

620
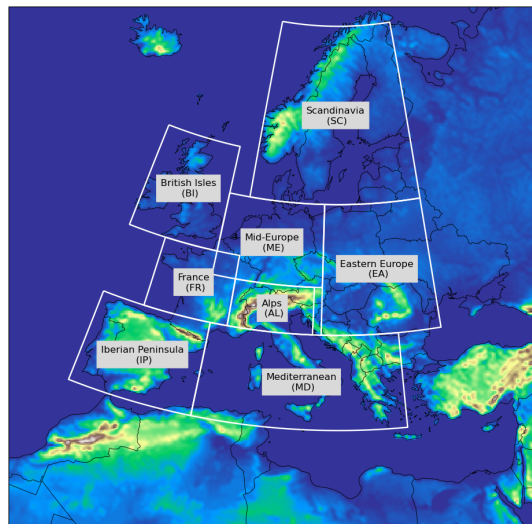
Figure C1. Orography over the EURO-CORDEX domain. The white boundaries with the labeled names inside define the PRUDENCE regions. The time series for validating and testing agricultural drought indices were computed over these regions.

Appendix D: Ablation Study

In Table D1, we provide an additional analysis about the impact of the perceptual VGG loss described in Eq. (16). When adding a perceptual loss for training, we observe a consistent improvement for all metrics while residuals are slightly bigger for the test set. As shown in Figs. D1 and D2, adding the loss $\mathcal{L}_{VGG}$ reduces the blurring effect and increases variability.

Table D1. Ablation study on the perceptual VGG loss described in Eq. (16). The metrics are shown for the validation and test sets as one set. The used model is a U-Net based on focal modulation.

	Loss function	NDVI					BT (K)				
		MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)	MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)
Val	$\mathcal{L}_{MAE}$	0.0274	0.0364	0.8744	0.9400	0.9139	1.9562	2.5945	0.9255	0.9664	0.9597
Val	$\mathcal{L}_{MAE} + \mathcal{L}_{VGG}$	**0.0270**	**0.0359**	**0.8781**	**0.9433**	**0.9184**	**1.8981**	**2.5433**	**0.9266**	**0.9679**	**0.9613**
Test	$\mathcal{L}_{MAE}$	**0.0266**	**0.0350**	**0.8819**	0.9443	0.9219	1.9642	2.6329	0.9181	0.9639	0.9610
Test	$\mathcal{L}_{MAE} + \mathcal{L}_{VGG}$	0.0268	0.0353	0.8795	**0.9452**	**0.9243**	**1.8730**	**2.5277**	**0.9227**	**0.9672**	**0.9642**
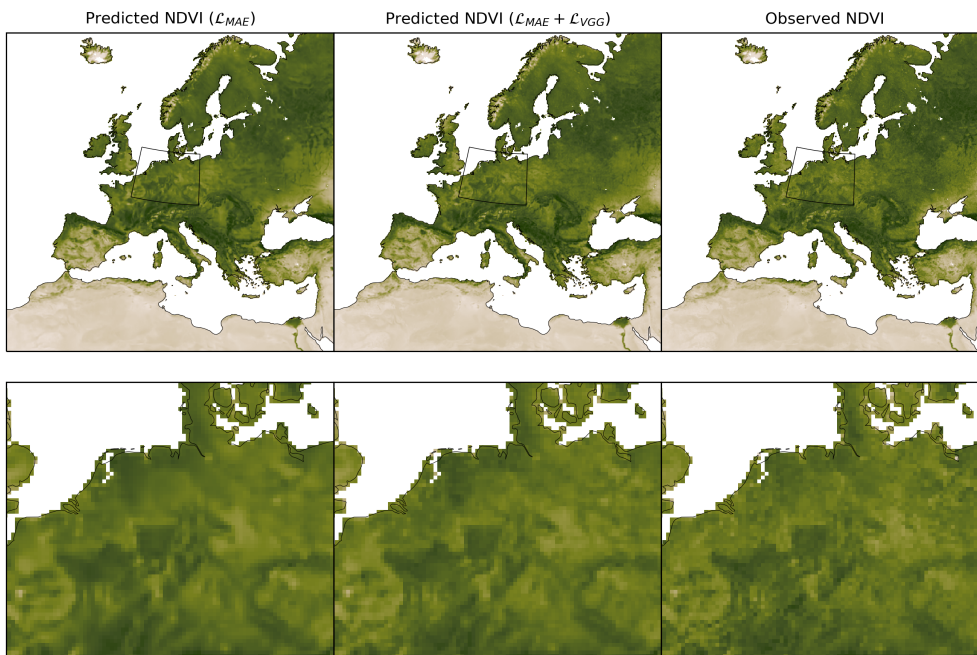
Figure D1. Impact of the perceptual VGG loss on NDVI predictions and image sharpness. The shown example is for the week 30 in the year 2018. Best seen in digital formats with colors.
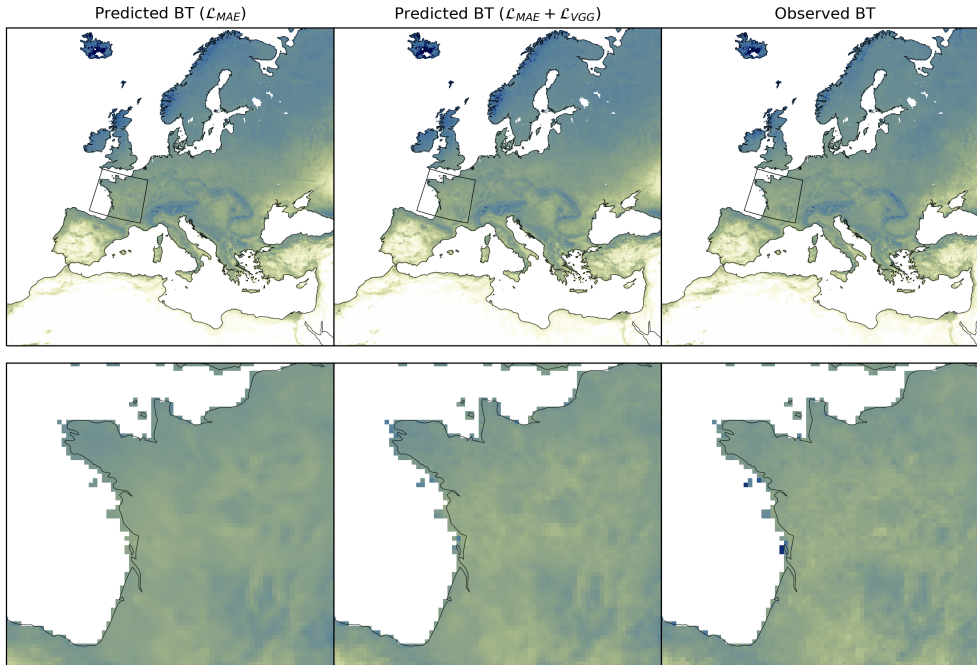
Figure D2. Impact of the perceptual VGG loss on BT predictions and image sharpness. The shown example is for the week 30 in the year 2018. Best seen in digital formats with colors.

Appendix E: Patch embedding

Patch embedding with a patch size > 1 is commonly used in vision transformer architectures. The main aim of this embedding is to increase the channel dimension and reduce the computational demands of the self-attention modules. This can be done by merging and embedding neighborhood pixels/tokens thus reducing the spatial or temporal resolution. In Table E1, we show that decreasing the spatial dimension of the raw input for the encoder has negative effects on our image-to-image regression task in both quantitative and qualitative terms. This can be understood as the information was lost and the model struggles to output the original resolution. Note that for all experiments we keep using the down- and up-sampling with a factor of 2 in both encoder and decoder while we only change the patch size before the first encoder layer. To match the original spatial resolution, we used an additional bilinear up-sampling after the last decoder layer.

Table E1. Impact of patch size for patch embedding before the first encoder layer. The metrics are shown for the validation and test sets. The used model is a U-Net based on focal modulation model.

	Patch size	NDVI					BT (K)				
		MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)	MAE($\downarrow$)	RMSE($\downarrow$)	R^2($\uparrow$)	R_p($\uparrow$)	R_s($\uparrow$)
Val	1×1	**0.0270**	**0.0359**	**0.8781**	**0.9433**	**0.9184**	**1.8981**	**2.5433**	**0.9266**	**0.9679**	**0.9613**
	2×2	0.0280	0.0369	0.8707	0.9374	0.9116	1.9372	2.6108	0.9243	0.9664	0.9604
	4×4	0.0291	0.0383	0.8625	0.9345	0.9075	2.0033	2.6957	0.9184	0.9633	0.9570
Test	1×1	**0.0268**	**0.0353**	**0.8795**	**0.9452**	**0.9243**	**1.8730**	**2.5277**	**0.9227**	**0.9672**	**0.9642**
	2×2	0.0271	0.0355	0.8786	0.9422	0.9185	1.9638	2.6669	0.9157	0.9645	0.9618
	4×4	0.0286	0.0375	0.8644	0.9363	0.9141	2.1741	2.9132	0.8977	0.9594	0.9580

Appendix F: Supplementary Results

635 *Author contributions.* This research was coordinated and supervised by JG. M.H S.E developed the software, performed the experiments, developed the method, and wrote the initial manuscript. JG reviewed and edited the manuscript. The authors read and approved the final manuscript.

Competing interests. The authors declared that they have no competing interests.

Acknowledgements. We thank Jülich Research Centre for providing the TSMP dataset to the community and Dr. Klaus Goergen for the
640 technical discussion related to the TSMP simulation. We would also like to thank Dr. Leonhard Scheck for the helpful discussions on radiative transfer models and Dr. Petra Friederichs for the thoughtful discussions on detection and attribution of weather and climate extremes.

Financial support. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1502/1–2022 - Project-number: 450058266 within the Collaborative Research Centre (CRC) for the project Regional Climate Change: Disentangling the Role of Land Use and Water Management (DETECT).
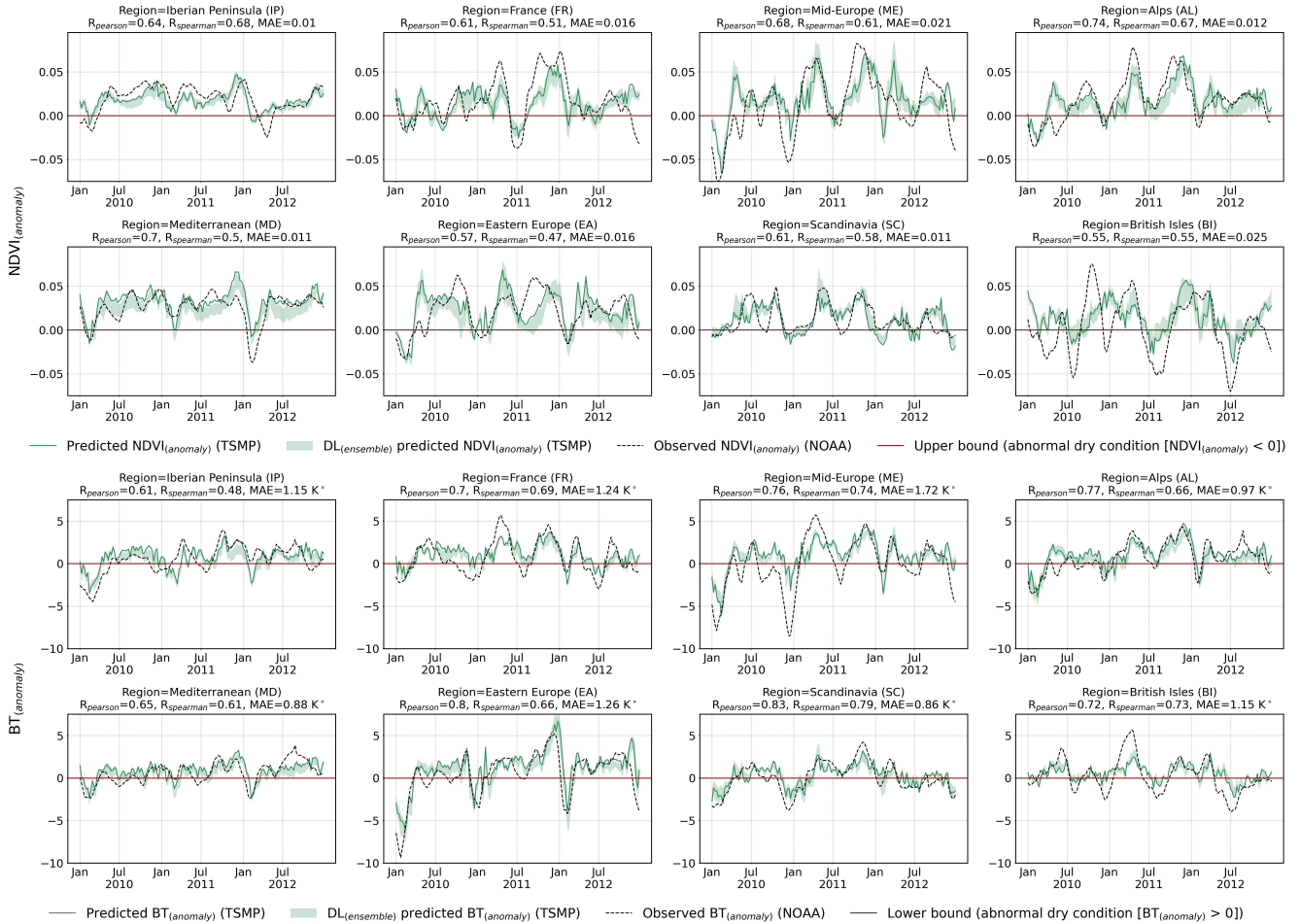
Figure F1. Supplementary results to Fig. 8. Comparison of spatially averaged weekly NDVI anomalies between the model prediction and NOAA observation over each PRUDENCE region. Anomaly was computed by subtracting the mean values from predictions (observations). The mean values were computed from the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with a DL model based on Focal Modulation Network. The ensemble model is the result of all DL models described in Sect. 5 and Table 1.
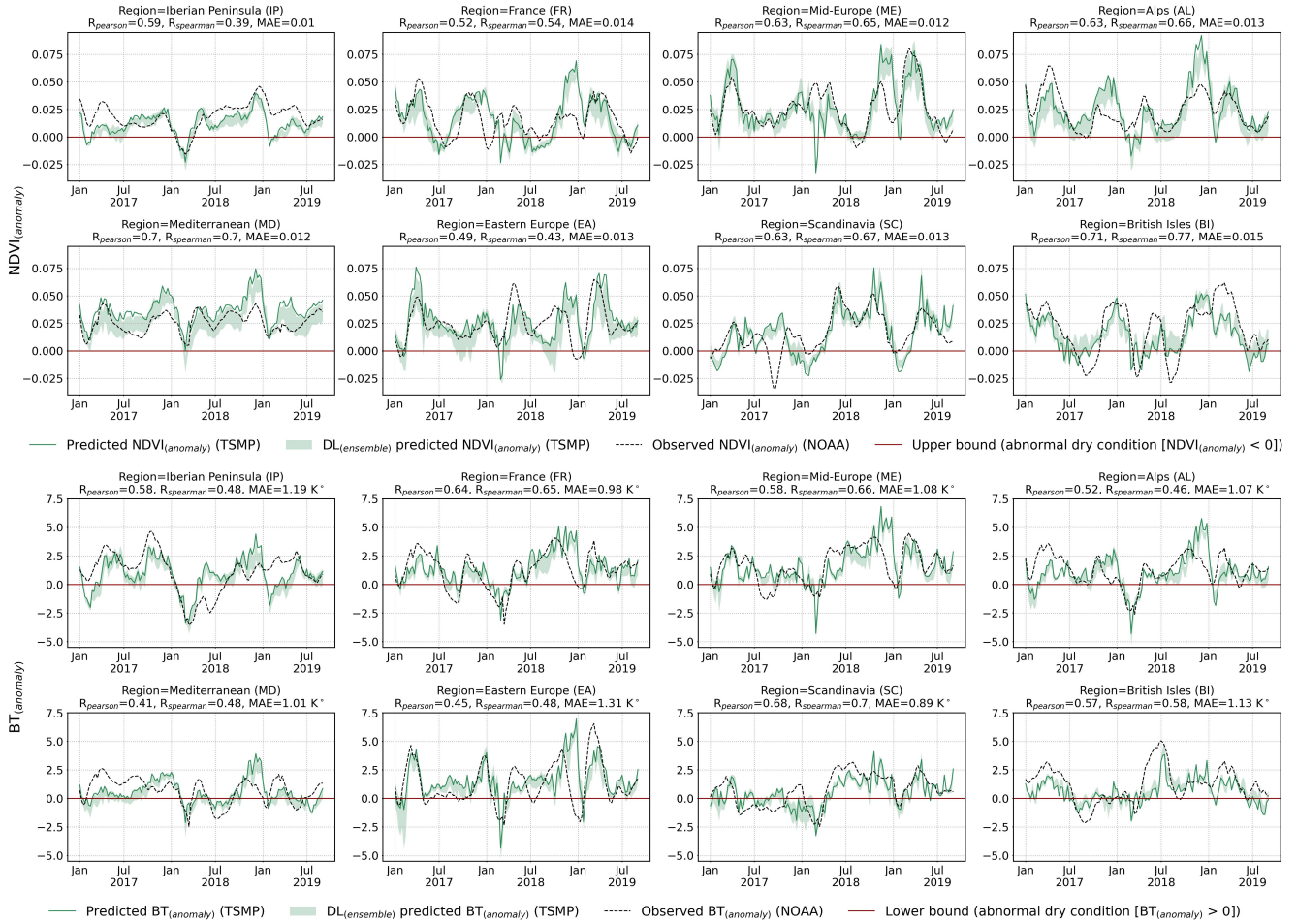
Figure F2. Supplementary results to Fig. 8. Comparison of spatially averaged weekly BT anomalies between the model prediction and NOAA observation over each PRUDENCE region. Anomaly was computed by subtracting the mean values from predictions (observations). The mean values were computed from the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with a DL model based on Focal Modulation Network. The ensemble model is the result of all DL models described in Sect. 5 and Table 1.

References

Adede, C., Oboko, R., Wagacha, P. W., and Atzberger, C.: A Mixed Model Approach to Vegetation Condition Prediction Using Artificial Neural Networks (ANN): Case of Kenya's Operational Drought Monitoring, Remote Sensing, 11, https://doi.org/10.3390/rs11091099, 2019.

Aleissaee, A. A., Kumar, A., Anwer, R. M., Khan, S., Cholakkal, H., Xia, G.-S., and Khan, F. S.: Transformers in Remote Sensing: A Survey, Remote Sensing, 15, https://doi.org/10.3390/rs15071860, 2023.

Ba, J. L., Kiros, J. R., and Hinton, G. E.: Layer normalization, arXiv preprint arXiv:1607.06450, 2016.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Monthly Weather Review, 139, 3887 – 3905, https://doi.org/10.1175/MWR-D-10-05013.1, 2011.

Baur, F., Scheck, L., Stumpf, C., Köpken-Watts, C., and Potthast, R.: A neural network-based method for generating synthetic $1.6\,\mu$m near-infrared satellite images, EGUsphere, 2023, 1–31, https://doi.org/10.5194/egusphere-2023-353, 2023.

Ben-Bouallegue, Z., Clare, M. C. A., Magnusson, L., Gascon, E., Maier-Gerber, M., Janousek, M., Rodwell, M., Pinault, F., Dramsch, J. S., Lang, S. T. K., Raoult, B., Rabier, F., Chevallier, M., Sandu, I., Dueben, P., Chantry, M., and Pappenberger, F.: The rise of data-driven weather forecasting, arXiv preprint arXiv:2307.10128, 2023.

Benson, V., Requena-Mesa, C., Robin, C., Alonso, L., Cortés, J., Gao, Z., Linscheid, N., Weynants, M., and Reichstein, M.: Forecasting localized weather impacts on vegetation as seen from space with meteo-guided video prediction, arXiv preprint arXiv:2303.16198, 2023.

Bergen, K. J., Johnson, P. A., de Hoop, M. V., and Beroza, G. C.: Machine learning for data-driven discovery in solid Earth geoscience, Science, 363, eaau0323, https://doi.org/10.1126/science.aau0323, 2019.

Betancourt, C., Stomberg, T. T., Edrich, A.-K., Patnala, A., Schultz, M. G., Roscher, R., Kowalski, J., and Stadtler, S.: Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties, Geoscientific Model Development, 15, 4331–4354, https://doi.org/10.5194/gmd-15-4331-2022, 2022.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

Blanchard, A., Parashar, N., Dodov, B., Lessig, C., and Sapsis, T.: A Multi-Scale Deep Learning Framework for Projecting Weather Extremes, in: NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning, https://www.climatechange.ai/papers/neurips2022/65, 2022.

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., Chen, X., Ma, L., Zhang, T., Su, R., et al.: FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead, arXiv preprint arXiv:2304.02948, 2023.

Chen, Z., Liu, H., Xu, C., Wu, X., Liang, B., Cao, J., and Chen, D.: Modeling vegetation greenness and its climate sensitivity with deep-learning technology, Ecology and Evolution, 11, 7335–7345, https://doi.org/10.1002/ece3.7564, 2021.

Christian, J. I., Basara, J. B., Hunt, E. D., Otkin, J. A., Furtado, J. C., Mishra, V., Xiao, X., and Randall, R. M.: Global distribution, trends, and drivers of flash drought occurrence, Nature communications, 12, 6330, https://doi.org/10.1038/s41467-021-26692-z, 2021.

Christian, J. I., Martin, E. R., Basara, J. B., Furtado, J. C., Otkin, J. A., Lowman, L. E., Hunt, E. D., Mishra, V., and Xiao, X.: Global projections of flash drought show increased risk in a warming climate, Communications Earth & Environment, 4, 165, https://doi.org/10.1038/s43247-023-00826-1, 2023.

Das, M. and Ghosh, S. K.: Deep-STEP: A Deep Learning Approach for Spatiotemporal Prediction of Remote Sensing Data, IEEE Geoscience and Remote Sensing Letters, 13, 1984–1988, https://doi.org/10.1109/LGRS.2016.2619984, 2016.

de Burgh-Day, C. O. and Leeuwenburg, T.: Machine learning for numerical weather and climate modelling: a review, Geoscientific Model Development, 16, 6433–6477, https://doi.org/10.5194/gmd-16-6433-2023, 2023.

685 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597,
690 https://doi.org/10.1002/qj.828, 2011.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848, 2009.

Diaconu, C.-A., Saha, S., Günnemann, S., and Xiang Zhu, X.: Understanding the Role of Weather Data for Earth Surface Forecasting using a ConvLSTM-based Model, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.
695 1361–1370, https://doi.org/10.1109/CVPRW56347.2022.00142, 2022.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, in: International Conference on Learning Representations, https://openreview.net/forum?id=YicbFdNTTy, 2021.

Düben, P., Modigliani, U., Geer, A., Siemen, S., Pappenberger, F., Bauer, P., Brown, A., Palkovic, M., Raoult, B., Wedi, N., and Baousis, V.:
700 Machine learning at ECMWF: A roadmap for the next 10 years, https://doi.org/10.21957/ge7ckgm, 2021.

Eisfelder, C., Asam, S., Hirner, A., Reiners, P., Holzwarth, S., Bachmann, M., Gessner, U., Dietz, A., Huth, J., Bachofer, F., and Kuenzer, C.: Seasonal Vegetation Trends for Europe over 30 Years from a Novel Normalised Difference Vegetation Index (NDVI) Time-Series-The TIMELINE NDVI Product, Remote Sensing, 15, https://doi.org/10.3390/rs15143616, 2023.

Essa, Y. H., Hirschi, M., Thiery, W., El-Kenawy, A. M., and Yang, C.: Drought characteristics in Mediterranean under future climate change,
705 npj Climate and Atmospheric Science, 6, 133, https://doi.org/10.1038/s41612-023-00458-4, 2023.

Fazry, L., Ramadhan, M. M. L., and Jatmiko, W.: Change Detection of High-Resolution Remote Sensing Images Through Adaptive Focal Modulation on Hierarchical Feature Maps, IEEE Access, 11, 69 072–69 090, https://doi.org/10.1109/ACCESS.2023.3292531, 2023.

Ferchichi, A., Abbes, A. B., Barra, V., and Farah, I. R.: Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review, Ecological Informatics, 68, 101 552,
710 https://doi.org/10.1016/j.ecoinf.2022.101552, 2022.

Furusho-Percot, C., Goergen, K., Hartick, C., Kulkarni, K., Keune, J., and Kollet, S.: Pan-European groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation, Scientific Data, 6, 320, https://doi.org/10.1038/s41597-019-0328-7, 2019.

Furusho-Percot, C., Goergen, K., Keune, J., Kulkarni, K., and Kollet, S.: Pan-european, physically consistent simulations from groundwater to the atmosphere with the Terrestrial Systems Modeling Platform, TerrSysMP (1989-2018 daily time-series),
715 https://doi.org/10.17616/R31NJMGR, 2019a.

Furusho-Percot, C., Goergen, K., Keune, J., Kulkarni, K., and Kollet, S.: Pan-european, physically consistent simulations from groundwater to the atmosphere with the Terrestrial Systems Modeling Platform, TerrSysMP (1989-2018 daily time-series), https://doi.org/10.1594/PANGAEA.901823, 2019b.

Furusho-Percot, C., Goergen, K., Hartick, C., Poshyvailo-Strube, L., and Kollet, S.: Groundwater Model Impacts Multiannual Simulations of Heat Waves, Geophysical Research Letters, 49, e2021GL096 781, https://doi.org/10.1029/2021GL096781, e2021GL096781 2021GL096781, 2022.

Gao, L., Liu, H., Yang, M., Chen, L., Wan, Y., Xiao, Z., and Qian, Y.: STransFuse: Fusing Swin Transformer and Convolutional Neural Network for Remote Sensing Image Semantic Segmentation, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 10 990–11 003, https://doi.org/10.1109/JSTARS.2021.3119654, 2021.

Gao, P., Du, W., Lei, Q., Li, J., Zhang, S., and Li, N.: NDVI Forecasting Model Based on the Combination of Time Series Decomposition and CNN – LSTM, Water Resources Management, 37, 1481–1497, https://doi.org/10.1007/s11269-022-03419-3, 2023.

Gasper, F., Goergen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M., and Kollet, S.: Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP v1.0) in a massively parallel supercomputing environment – a case study on JUQUEEN (IBM Blue Gene/Q), Geoscientific Model Development, 7, 2531–2543, https://doi.org/10.5194/gmd-7-2531-2014, 2014.

Geiss, S., Scheck, L., de Lozar, A., and Weissmann, M.: Understanding the model representation of clouds based on visible and infrared satellite observations, Atmospheric Chemistry and Physics, 21, 12 273–12 290, https://doi.org/10.5194/acp-21-12273-2021, 2021.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework., Bulletin - World Meteorological Organization, 58, 175–183, 2009.

Gutowski Jr., W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., and Tangang, F.: WCRP COordinated Regional Downscaling EXperiment (CORDEX): a diagnostic MIP for CMIP6, Geoscientific Model Development, 9, 4087–4095, https://doi.org/10.5194/gmd-9-4087-2016, 2016.

Hammad, A. T. and Falchetta, G.: Probabilistic forecasting of remotely sensed cropland vegetation health and its relevance for food security, Science of The Total Environment, 838, 156 157, https://doi.org/10.1016/j.scitotenv.2022.156157, 2022.

Hendrycks, D. and Gimpel, K.: Gaussian error linear units (gelus), arXiv preprint arXiv:1606.08415, 2016.

Horn, B.: Hill shading and the reflectance map, Proceedings of the IEEE, 69, 14–47, https://doi.org/10.1109/PROC.1981.11918, 1981.

Hu, J., Shen, L., and Sun, G.: Squeeze-and-Excitation Networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141, https://doi.org/10.1109/CVPR.2018.00745, 2018.

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., Benestad, R., Boberg, F., Buonomo, E., Cardoso, R. M., et al.: Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community, Regional Environmental Change, 20, 51, https://doi.org/10.1007/s10113-020-01606-9, 2020.

Jefferson, J. L. and Maxwell, R. M.: Evaluation of simple to complex parameterizations of bare ground evaporation, Journal of Advances in Modeling Earth Systems, 7, 1075–1092, https://doi.org/10.1002/2014MS000398, 2015.

Jin, H., Vicente-Serrano, S. M., Tian, F., Cai, Z., Conradt, T., Boincean, B., Murphy, C., Farizo, B. A., Grainger, S., López-Moreno, J. I., et al.: Higher vegetation sensitivity to meteorological drought in autumn than spring across European biomes, Communications Earth & Environment, 4, 299, https://doi.org/10.1038/s43247-023-00960-w, 2023.

Johnson, J., Alahi, A., and Fei-Fei, L.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution, in: Computer Vision – ECCV 2016, edited by Leibe, B., Matas, J., Sebe, N., and Welling, M., pp. 694–711, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-46475-6_43, 2016.

Jones, J. E. and Woodward, C. S.: Newton–Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems, Advances in Water Resources, 24, 763–774, https://doi.org/10.1016/S0309-1708(00)00075-0, 2001.

Jones, P. W.: First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates, Monthly Weather Review, 127, 2204 – 2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, 1999.

Kew, S. F., Philip, S. Y., Hauser, M., Hobbins, M., Wanders, N., van Oldenborgh, G. J., van der Wiel, K., Veldkamp, T. I. E., Kimutai, J., Funk, C., and Otto, F. E. L.: Impact of precipitation and increasing temperatures on drought trends in eastern Africa, Earth System Dynamics, 12, 17–35, https://doi.org/10.5194/esd-12-17-2021, 2021.

Kladny, K.-R., Milanta, M., Mraz, O., Hufkens, K., and Stocker, B. D.: Deep learning for satellite image forecasting of vegetation greenness, bioRxiv, https://doi.org/10.1101/2022.08.16.504173, 2022.

Klimavičius, L., Rimkus, E., Stonevičius, E., and Mačiulytė, V.: Seasonality and long-term trends of NDVI values in different land use types in the eastern part of the Baltic Sea basin, Oceanologia, 65, 171–181, https://doi.org/10.1016/j.oceano.2022.02.007, 2023.

Kogan, F.: Application of vegetation index and brightness temperature for drought detection, Advances in Space Research, 15, 91–100, https://doi.org/10.1016/0273-1177(95)00079-T, natural Hazards: Monitoring and Assessment Using Remote Sensing Technique, 1995a.

Kogan, F., Guo, W., and Jelenak, A.: Global Vegetation Health: Long-Term Data Records, in: Use of Satellite and In-Situ Data to Improve Sustainability, edited by Kogan, F., Powell, A., and Fedorov, O., pp. 247–255, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-90-481-9618-0_28, 2011.

Kogan, F., Goldberg, M., Schott, T., and Guo, W.: Suomi NPP/VIIRS: improving drought watch, crop loss prediction, and food security, International Journal of Remote Sensing, 36, 5373–5383, https://doi.org/10.1080/01431161.2015.1095370, 2015.

Kogan, F., Guo, W., and Yang, W.: Near 40-year drought trend during 1981-2019 earth warming and food security, Geomatics, Natural Hazards and Risk, 11, 469–490, https://doi.org/10.1080/19475705.2020.1730452, 2020.

Kogan, F. N.: Remote sensing of weather impacts on vegetation in non-homogeneous areas, International Journal of Remote Sensing, 11, 1405–1419, https://doi.org/10.1080/01431169008955102, 1990.

Kogan, F. N.: Droughts of the Late 1980s in the United States as Derived from NOAA Polar-Orbiting Satellite Data, Bulletin of the American Meteorological Society, 76, 655 – 668, https://doi.org/10.1175/1520-0477(1995)076<0655:DOTLIT>2.0.CO;2, 1995b.

Kollet, S. J. and Maxwell, R. M.: Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, Advances in Water Resources, 29, 945–958, https://doi.org/10.1016/j.advwatres.2005.08.006, 2006.

Kraft, B., Jung, M., Körner, M., Requena Mesa, C., Cortés, J., and Reichstein, M.: Identifying Dynamic Memory Effects on Vegetation State Using Recurrent Neural Networks, Frontiers in Big Data, 2, https://doi.org/10.3389/fdata.2019.00031, 2019.

Kuffour, B. N. O., Engdahl, N. B., Woodward, C. S., Condon, L. E., Kollet, S., and Maxwell, R. M.: Simulating coupled surface–subsurface flows with ParFlow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model, Geoscientific Model Development, 13, 1373–1397, https://doi.org/10.5194/gmd-13-1373-2020, 2020.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., Ravuri, S., Ewalds, T., Alet, F., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Stott, J., Vinyals, O., Mohamed, S., and Battaglia, P.: GraphCast: Learning skillful medium-range global weather forecasting, arXiv preprint arXiv:2212.12794, 2022.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description

of New Features, Benchmarking, and Impact of Forcing Uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245–4287,
https://doi.org/10.1029/2018MS001583, 2019.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W.: Photo-
Realistic Single Image Super-Resolution Using a Generative Adversarial Network, in: 2017 IEEE Conference on Computer Vision and
Pattern Recognition (CVPR), pp. 105–114, https://doi.org/10.1109/CVPR.2017.19, 2017.

Lees, T., Tseng, G., Atzberger, C., Reece, S., and Dadson, S.: Deep Learning for Vegetation Health Forecasting: A Case Study in Kenya,
Remote Sensing, 14, https://doi.org/10.3390/rs14030698, 2022.

Lessig, C., Luise, I., Gong, B., Langguth, M., Stadler, S., and Schultz, M.: AtmoRep: A stochastic model of atmosphere dynamics using
large scale representation learning, arXiv preprint arXiv:2308.13280, 2023.

Li, J., Geer, A. J., Okamoto, K., Otkin, J. A., Liu, Z., Han, W., and Wang, P.: Satellite all-sky infrared radiance assimilation: Recent progress
and future perspectives, Advances in Atmospheric Sciences, 39, 9–21, https://doi.org/10.1007/s00376-021-1088-9, 2022.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B.: Swin Transformer: Hierarchical Vision Trans-
former using Shifted Windows, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9992–10 002,
https://doi.org/10.1109/ICCV48922.2021.00986, 2021.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., and Guo, B.: Swin Transformer V2: Scaling
Up Capacity and Resolution, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11 999–12 009,
https://doi.org/10.1109/CVPR52688.2022.01170, 2022.

Loshchilov, I. and Hutter, F.: Decoupled Weight Decay Regularization, in: International Conference on Learning Representations, https:
//openreview.net/forum?id=Bkg6RiCqY7, 2019.

Ma, Y., Montzka, C., Bayat, B., and Kollet, S.: An Indirect Approach Based on Long Short-Term Memory Networks to Esti-
mate Groundwater Table Depth Anomalies Across Europe With an Application for Drought Analysis, Frontiers in Water, 3,
https://doi.org/10.3389/frwa.2021.723548, 2021.

Marj, A. F. and Meijerink, A. M. J.: Agricultural drought forecasting using satellite images, climate indices and artificial neural network,
International Journal of Remote Sensing, 32, 9707–9719, https://doi.org/10.1080/01431161.2011.575896, 2011.

Martínez-Fernández, J., González-Zamora, A., Sánchez, N., and Gumuzzio, A.: A soil water based index as a suitable agricultural drought
indicator, Journal of Hydrology, 522, 265–273, https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.12.051, 2015.

Maxwell, R. M., Condon, L. E., and Kollet, S. J.: A high-resolution simulation of groundwater and surface water over most of the continental
US with the integrated hydrologic model ParFlow v3, Geoscientific Model Development, 8, 923–937, https://doi.org/10.5194/gmd-8-923-
2015, 2015.

McKee, T. B.: Drought monitoring with multiple time scales, in: Proceedings of 9th Conference on Applied Climatology, Boston, 1995,
1995.

Meza, I., Siebert, S., Döll, P., Kusche, J., Herbert, C., Eyshi Rezaei, E., Nouri, H., Gerdener, H., Popat, E., Frischen, J., Naumann, G., Vogt,
J. V., Walz, Y., Sebesvari, Z., and Hagenlocher, M.: Global-scale drought risk assessment for agricultural systems, Natural Hazards and
Earth System Sciences, 20, 695–712, https://doi.org/10.5194/nhess-20-695-2020, 2020.

Miao, L., Ye, P., He, B., Chen, L., and Cui, X.: Future Climate Impact on the Desertification in the Dry Land Asia Using AVHRR GIMMS
NDVI3g Data, Remote Sensing, 7, 3863–3877, https://doi.org/10.3390/rs70403863, 2015.

Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J.: Land–atmospheric feedbacks during droughts and heatwaves: state of the
science and current challenges, Annals of the New York Academy of Sciences, 1436, 19–35, https://doi.org/10.1111/nyas.13912, 2019.

Moravec, D., Komárek, J., López-Cuervo Medina, S., and Molina, I.: Effect of Atmospheric Corrections on NDVI: Intercomparability of Landsat 8, Sentinel-2, and UAV Sensors, Remote Sensing, 13, https://doi.org/10.3390/rs13183550, 2021.

Naderi, M., Givkashi, M., Piri, F., Karimi, N., and Samavi, S.: Focal-UNet: UNet-like Focal Modulation for Medical Image Segmentation, arXiv preprint arXiv:2212.09263, 2022.

Nagol, J. R., Vermote, E. F., and Prince, S. D.: Effects of atmospheric variation on AVHRR NDVI data, Remote Sensing of Environment, 113, 392–397, https://doi.org/10.1016/j.rse.2008.10.007, 2009.

Nay, J., Burchfield, E., and Gilligan, J.: A machine-learning approach to forecasting remotely sensed vegetation health, International Journal of Remote Sensing, 39, 1800–1816, https://doi.org/10.1080/01431161.2017.1410296, 2018.

Naz, B. S., Sharples, W., Ma, Y., Goergen, K., and Kollet, S.: Continental-scale evaluation of a fully distributed coupled land surface and groundwater model, ParFlow-CLM (v3.6.0), over Europe, Geoscientific Model Development, 16, 1617–1639, https://doi.org/10.5194/gmd-16-1617-2023, 2023.

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., and Grover, A.: ClimaX: A foundation model for weather and climate, arXiv preprint arXiv:2301.10343, 2023.

Oleson, K., YJ, D., Bosilovich, M., Dickinson, R., Dirmeyer, P., Hoffman, F., Houser, P., Levis, S., Niu, G.-Y., Thornton, P., Vertenstein, M., Yang, Z.-L., and XB, Z.: Technical Description of the Community Land Model (CLM), p. 174 pp., https://doi.org/10.5065/D6N877R0, 2004.

Oleson, K. W., Niu, G.-Y., Yang, Z.-L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., Stöckli, R., Dickinson, R. E., Bonan, G. B., Levis, S., Dai, A., and Qian, T.: Improvements to the Community Land Model and their impact on the hydrological cycle, Journal of Geophysical Research: Biogeosciences, 113, https://doi.org/10.1029/2007JG000563, 2008.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems 32, pp. 8024–8035, Curran Associates, Inc., http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf, 2019.

Patacca, M., Lindner, M., Lucas-Borja, M. E., Cordonnier, T., Fidej, G., Gardiner, B., Hauf, Y., Jasinevičius, G., Labonne, S., Linkevičius, E., Mahnken, M., Milanovic, S., Nabuurs, G.-J., Nagel, T. A., Nikinmaa, L., Panyatov, M., Bercak, R., Seidl, R., Ostrogović Sever, M. Z., Socha, J., Thom, D., Vuletic, D., Zudin, S., and Schelhaas, M.-J.: Significant increase in natural disturbance impacts on European forests since 1950, Global Change Biology, 29, 1359–1376, https://doi.org/10.1111/gcb.16531, 2023.

Patakchi Yousefi, K. and Kollet, S.: Deep learning of model- and reanalysis-based precipitation and pressure mismatches over Europe, Frontiers in Water, 5, https://doi.org/10.3389/frwa.2023.1178114, 2023.

Patil, S. D., Gu, Y., Dias, F. S. A., Stieglitz, M., and Turk, G.: Predicting the spectral information of future land cover using machine learning, International Journal of Remote Sensing, 38, 5592–5607, https://doi.org/10.1080/01431161.2017.1343512, 2017.

Pirret, J., Fung, F., Lowe, J., McInnes, R., Mitchell, J., and Murphy, J.: UKCP Factsheet: Soil Moisture, https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/data/factsheets, 2020.

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J. E. S., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M. G., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A. G., Liu, J., Papadimitriou, L., Schewe, J., Schmied, H. M., Stacke, T. I., Telteu, C.-E., Thiery, W., Veldkamp, T. I. E., Zhao, F., and Wada, Y.: Global terrestrial water storage and drought severity under climate change, Nature Climate Change, 11, 226 – 233, https://doi.org/10.1038/s41558-020-00972-w, 2021.

Poshyvailo-Strube, L., Wagner, N., Goergen, K., Furusho-Percot, C., Hartick, C., and Kollet, S.: Groundwater in terrestrial systems modelling: a new climatology of extreme heat events in Europe, Earth System Dynamics Discussions, 2022, 1–25, https://doi.org/10.5194/esd-2022-53, 2022.

Prodhan, F. A., Zhang, J., Yao, F., Shi, L., Pangali Sharma, T. P., Zhang, D., Cao, D., Zheng, M., Ahmed, N., and Mohana, H. P.: Deep Learning for Monitoring Agricultural Drought in South Asia Using Remote Sensing Data, Remote Sensing, 13, https://doi.org/10.3390/rs13091715, 2021.

Qin, Q., Wu, Z., Zhang, T., Sagan, V., Zhang, Z., Zhang, Y., Zhang, C., Ren, H., Sun, Y., Xu, W., and Zhao, C.: Optical and Thermal Remote Sensing for Monitoring Agricultural Drought, Remote Sensing, 13, https://doi.org/10.3390/rs13245092, 2021.

Rasoulian, A., Salari, S., and Xiao, Y.: Weakly supervised segmentation of intracranial aneurysms using a 3D focal modulation UNet, arXiv preprint arXiv:2308.03001, 2023.

Requena-Mesa, C., Benson, V., Reichstein, M., Runge, J., and Denzler, J.: EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1132–1142, https://doi.org/10.1109/CVPRW53098.2021.00124, 2021.

Reyniers, N., Osborn, T. J., Addor, N., and Darch, G.: Projected changes in droughts and extreme droughts in Great Britain strongly influenced by the choice of drought index, Hydrology and Earth System Sciences, 27, 1151–1171, https://doi.org/10.5194/hess-27-1151-2023, 2023.

Robin, C., Requena-Mesa, C., Benson, V., Poehls, J., Alonzo, L., Carvalhais, N., and Reichstein, M.: Learning to forecast vegetation greenness at fine resolution over Africa with ConvLSTMs, in: NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning, https://www.climatechange.ai/papers/neurips2022/88, 2022.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241, Springer, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Salakpi, E. E., Hurley, P. D., Muthoka, J. M., Barrett, A. B., Bowell, A., Oliver, S., and Rowhani, P.: Forecasting vegetation condition with a Bayesian auto-regressive distributed lags (BARDL) model, Natural Hazards and Earth System Sciences, 22, 2703–2723, https://doi.org/10.5194/nhess-22-2703-2022, 2022a.

Salakpi, E. E., Hurley, P. D., Muthoka, J. M., Bowell, A., Oliver, S., and Rowhani, P.: A dynamic hierarchical Bayesian approach for forecasting vegetation condition, Natural Hazards and Earth System Sciences, 22, 2725–2749, https://doi.org/10.5194/nhess-22-2725-2022, 2022b.

Satoh, Y., Shiogama, H., Hanasaki, N., Pokhrel, Y., Boulange, J. E. S., Burek, P., Gosling, S. N., Grillakis, M., Koutroulis, A., Schmied, H. M., Thiery, W., and Yokohata, T.: A quantitative evaluation of the issue of drought definition: a source of disagreement in future drought assessments, Environmental Research Letters, 16, 104 001, https://doi.org/10.1088/1748-9326/ac2348, 2021.

Scheck, L., Frèrebeau, P., Buras-Schnell, R., and Mayer, B.: A fast radiative transfer method for the simulation of visible satellite imagery, Journal of Quantitative Spectroscopy and Radiative Transfer, 175, 54–67, https://doi.org/10.1016/j.jqsrt.2016.02.008, 2016.

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., and Stadtler, S.: Can deep learning beat numerical weather prediction?, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20200 097, https://doi.org/10.1098/rsta.2020.0097, 2021.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, p. 1513–1766, Cambridge University Press, https://doi.org/10.1017/9781009157896.013, 2021.

Shams Eddin, M. H. and Gall, J.: Focal-TSMP: Deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation, https://doi.org/10.5281/zenodo.10015048, 2023a.

Shams Eddin, M. H. and Gall, J.: Focal-TSMP: Deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation, https://doi.org/10.5281/zenodo.10008814, 2023b.

Sheffield, J., Wood, E. F., and Roderick, M. L.: Little change in global drought over the past 60 years, Nature, 491, 435–438, https://api.semanticscholar.org/CorpusID:205231154, 2012.

Shi, X., Li, Y., Liu, J., Xiang, X., and Liu, L.: Simulation of FY-2D infrared brightness temperature and sensitivity analysis to the errors of WRF simulated cloud variables, Science China Earth Sciences, 61, 1–16, https://doi.org/10.1007/s11430-017-9150-0, 2018.

Shrestha, P., Sulis, M., Masbou, M., Kollet, S., and Simmer, C.: A Scale-Consistent Terrestrial Systems Modeling Platform Based on COSMO, CLM, and ParFlow, Monthly Weather Review, 142, 3466 – 3483, https://doi.org/10.1175/MWR-D-14-00029.1, 2014.

Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.

Tang, Y., Han, K., Guo, J., Xu, C., Li, Y., Xu, C., and Wang, Y.: An Image Patch is a Wave: Phase-Aware Vision MLP, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10 925–10 934, https://doi.org/10.1109/CVPR52688.2022.01066, 2022.

Tucker, C. J.: Red and photographic infrared linear combinations for monitoring vegetation, Remote Sensing of Environment, 8, 127–150, https://doi.org/10.1016/0034-4257(79)90013-0, 1979.

Tuia, D., Schindler, K., Demir, B., Camps-Valls, G., Zhu, X. X., Kochupillai, M., Džeroski, S., van Rijn, J. N., Hoos, H. H., Del Frate, F., et al.: Artificial intelligence to advance Earth observation: a perspective, arXiv preprint arXiv:2305.08413, 2023.

Valcke, S.: The OASIS3 coupler: a European climate modelling community software, Geoscientific Model Development, 6, 373–388, https://doi.org/10.5194/gmd-6-373-2013, 2013.

Valmassoi, A., Keller, J. D., Kleist, D. T., English, S., Ahrens, B., Ďurán, I. B., Bauernschubert, E., Bosilovich, M. G., Fujiwara, M., Hersbach, H., Lei, L., Löhnert, U., Mamnun, N., Martin, C. R., Moore, A., Niermann, D., Ruiz, J. J., and Scheck, L.: Current challenges and future directions in data assimilation and reanalysis, Bulletin of the American Meteorological Society, https://doi.org/10.1175/BAMS-D-21-0331.1, 2022.

Van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., Gleeson, T., Van Dijk, A. I. J. M., Tallaksen, L. M., Hannaford, J., Uijlenhoet, R., Teuling, A. J., Hannah, D. M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., and Van Lanen, H. A. J.: Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches, Hydrology and Earth System Sciences, 20, 3631–3650, https://doi.org/10.5194/hess-20-3631-2016, 2016.

Van Oldenborgh, G. J., van Der Wiel, K., Kew, S., Philip, S., Otto, F., Vautard, R., King, A., Lott, F., Arrighi, J., Singh, R., et al.: Pathways and pitfalls in extreme event attribution, Climatic Change, 166, 13, https://doi.org/10.1007/s10584-021-03071-7, 2021.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Advances in neural information processing systems, 30, 2017.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, Journal of Climate, 23, 1696 – 1718, https://doi.org/10.1175/2009JCLI2909.1, 2010.

Vicente-Serrano, S. M., Peña-Angulo, D., Beguería, S., Domínguez-Castro, F., Tomás-Burguera, M., Noguera, I., Gimeno-Sotelo, L., and El Kenawy, A.: Global drought trends and future projections, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 380, 20210 285, https://doi.org/10.1098/rsta.2021.0285, 2022.

Vo, T. Q., Kim, S.-H., Nguyen, D. H., and hyo Bae, D.: LSTM-CM: a hybrid approach for natural drought prediction based on deep learning and climate models, Stochastic Environmental Research and Risk Assessment, 37, 2035–2051, https://doi.org/10.1007/s00477-022-02378-w, 2023.

Vreugdenhil, M., Greimeister-Pfeil, I., Preimesberger, W., Camici, S., Dorigo, W., Enenkel, M., van der Schalie, R., Steele-Dunne, S., and Wagner, W.: Microwave remote sensing for agricultural drought monitoring: Recent developments and challenges, Frontiers in Water, 4, https://doi.org/10.3389/frwa.2022.1045451, 2022.

Wang, L., Li, R., Duan, C., Zhang, C., Meng, X., and Fang, S.: A Novel Transformer Based Semantic Segmentation Scheme for Fine-Resolution Remote Sensing Images, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, https://doi.org/10.1109/LGRS.2022.3143368, 2022a.

Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., and Atkinson, P. M.: UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 190, 196–214, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2022.06.008, 2022b.

Wasim, S. T., Khattak, M. U., Naseer, M., Khan, S., Shah, M., and Khan, F. S.: Video-FocalNets: Spatio-Temporal Focal Modulation for Video Action Recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 13 778–13 789, 2023.

Wei, J., Liu, X., and Zhou, B.: Sensitivity of Vegetation to Climate in Mid-to-High Latitudes of Asia and Future Vegetation Projections, Remote Sensing, 15, https://doi.org/10.3390/rs15102648, 2023.

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., and Grehan, A. J.: Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope, Marine Geodesy, 30, 3–35, https://doi.org/10.1080/01490410701295962, 2007.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), pp. 3–19, https://doi.org/10.1007/978-3-030-01234-2_1, 2018.

Wu, T., Feng, F., Lin, Q., and Bai, H.: A spatio-temporal prediction of NDVI based on precipitation: an application for grazing management in the arid and semi-arid grasslands, International Journal of Remote Sensing, 41, 2359–2373, https://doi.org/10.1080/01431161.2019.1688418, 2020.

Wu, Y. and He, K.: Group Normalization, in: Computer Vision – ECCV 2018, edited by Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., pp. 3–19, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-01261-8_1, 2018.

Xoplaki, E., Ellsäßer, F., Grieger, J., Nissen, K. M., Pinto, J., Augenstein, M., Chen, T.-C., Feldmann, H., Friederichs, P., Gliksman, D., Goulier, L., Haustein, K., Heinke, J., Jach, L., Knutzen, F., Kollet, S., Luterbacher, J., Luther, N., Mohr, S., Mudersbach, C., Müller, C., Rousi, E., Simon, F., Suarez-Gutierrez, L., Szemkus, S., Vallejo-Bernal, S. M., Vlachopoulos, O., and Wolf, F.: Compound events in Germany in 2018: drivers and case studies, EGUsphere, 2023, 1–43, https://doi.org/10.5194/egusphere-2023-1460, 2023.

Yang, J., Li, C., Zhang, P., Dai, X., Xiao, B., Yuan, L., and Gao, J.: Focal Attention for Long-Range Interactions in Vision Transformers, in: Advances in Neural Information Processing Systems, edited by Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., vol. 34, pp. 30 008–30 022, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2021/file/fc1a36821b02abbd2503fd949bfc9131-Paper.pdf, 2021a.

Yang, J., Li, C., Dai, X., and Gao, J.: Focal modulation networks, Advances in Neural Information Processing Systems, 35, 4203–4217, https://proceedings.neurips.cc/paper_files/paper/2022/file/1b08f585b0171b74d1401a5195e986f1-Paper-Conference.pdf, 2022.

Yang, W., Guo, W., and Kogan, F.: VIIRS-based high resolution spectral vegetation indices for quantitative assessment of vegetation health: second version, International Journal of Remote Sensing, 39, 7417–7436, https://doi.org/10.1080/01431161.2018.1470701, 2018.

Yang, W., Kogan, F., and Guo, W.: An Ongoing Blended Long-Term Vegetation Health Product for Monitoring Global Food Security, Agronomy, 10, https://doi.org/10.3390/agronomy10121936, 2020.

Yang, W., Kogan, F., Guo, W., and Chen, Y.: A novel re-compositing approach to create continuous and consistent cross-sensor/cross-production global NDVI datasets, International Journal of Remote Sensing, 42, 6023–6047, https://doi.org/10.1080/01431161.2021.1934597, 2021b.

Yu, S., Hannah, W. M., Peng, L., Bhouri, M. A., Gupta, R., Lin, J., Lütjens, B., Will, J. C., Beucler, T., Harrop, B. E., et al.: ClimSim: An open large-scale dataset for training high-resolution physics emulators in hybrid multi-scale climate simulators, arXiv preprint arXiv:2306.08754, 2023.

Yu, W., Li, J., Liu, Q., Zhao, J., Dong, Y., Wang, C., Lin, S., Zhu, X., and Zhang, H.: Spatial–Temporal Prediction of Vegetation Index With Deep Recurrent Neural Networks, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, https://doi.org/10.1109/LGRS.2021.3064814, 2022.

Yuan, X., Wang, Y., Ji, P., Wu, P., Sheffield, J., and Otkin, J. A.: A global transition to flash droughts under climate change, Science, 380, 187–191, https://doi.org/10.1126/science.abn6301, 2023.

Zeng, J., Zhang, R., Qu, Y., Bento, V. A., Zhou, T., Lin, Y., Wu, X., Qi, J., Shui, W., and Wang, Q.: Improving the drought monitoring capability of VHI at the global scale via ensemble indices for various vegetation types from 2001 to 2018, Weather and Climate Extremes, 35, 100 412, https://doi.org/10.1016/j.wace.2022.100412, 2022.

Zeng, J., Zhou, T., Qu, Y., Bento, V., Qi, J., Xu, Y., Li, Y., and Wang, Q.: An improved global vegetation health index dataset in detecting vegetation drought, Scientific Data, 10, 338, https://doi.org/10.1038/s41597-023-02255-3, 2023.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., et al.: Greening of the Earth and its drivers, Nature climate change, 6, 791–795, https://doi.org/10.1038/nclimate3004, 2016.

Zhuang, J., raphael dussin, Jüling, A., and Rasp, S.: JiaweiZhuang/xESMF: v0.3.0 Adding ESMF.LocStream capabilities, https://doi.org/10.5281/zenodo.3700105, 2020.