Machine learning and Global Vegetation: Random Forests for Downscaling and Gapfilling

Barry van Jaarsveld[1], Sandra M. Hauswirth[1], and Niko Wanders[1]

[1]Utrecht University, Department of Physical Geography, Princetonlaan 8a, Utrecht, The Netherlands

Correspondence: Barry van Jaarsveld (a.s.vanjaarsveld@uu.nl)

Abstract. Drought is a devastating natural disaster, where water shortage often manifests itself in the health of vegetation. Unfortunately, it is difficult to obtain high-resolution vegetation drought impact, which is spatially and temporally consistent. While remotely sensed products can provide part of this information, they often suffer from data gaps and limitations in spatial or temporal resolutions. A persistent feature among remote sensing products are tradeoffs between spatial resolution and revisiting times, where high temporal resolution is met by coarse spatial resolution and *vice verse*. Machine learning methods have been successfully applied in a wide range of remote sensing and hydrological studies. However, global applications to resolve drought impacts on vegetation dynamics still need to be made available, while there is significant potential for such a product to aid improved drought impact monitoring. To this end, this study predicted global vegetation dynamics based on the Enhanced Vegetation Index (*evi*) and the popular Random Forest algorithm (RF) at 0.1°. We assessed the applicability of RF as a gap filling and downscaling tool to generate global *evi* estimates that are spatially and temporally consistent. To do this, we trained an RF regressor with 0.1° *evi* data using a host of features indicative of the water and energy balances experienced by vegetation and evaluated the performance of this new product. Next, to test whether the RF is robust in terms of spatial resolution, we downscale global *evi*, the model trained on 0.1° data is used to predict *evi* at 0.01° resolution. The results show that the RF can capture global *evi* dynamics at both the 0.1° (RMSE: 0.02 - 0.4) and at the finer 0.01° (RMSE: 0.04 - 0.6) resolution. Overall errors were higher in the down-scaled 0.01° compared to the 0.1° product. Yet, relative increases remained small, thus demonstrating that RF can be used to create downscaled and temporally consistent *evi* products. Additional error analysis reveals that errors vary spatiotemporally, with underrepresented land cover types and periods of extreme vegetation conditions having the highest errors. Finally, this model is used to produce global spatially continuous *evi* products at both the 0.1° and 0.01° spatial resolution for 2003-2013 at an 8-day frequency. Drought is a devastating natural disaster, where water shortage often manifests itself in the health of vegetation. Unfortunately, it is difficult to obtain high-resolution vegetation drought impact, which is spatially and temporally consistent. While remotely sensed products can provide part of this information, they often suffer from data gaps and limitations in spatial or temporal resolutions. A persistent feature among remote sensing products is tradeoffs between spatial resolution and revisiting times, where high temporal resolution is met by coarse spatial resolution and *vice verse*. Machine learning methods have been successfully applied in a wide range of remote sensing and hydrological studies. However, global applications to resolve drought impacts on vegetation dynamics still need to be made available, while there is significant potential for such a product to aid improved drought impact monitoring. To this end, this study predicted global vegetation dynamics based on the Enhanced Vegetation Index (*evi*) and the popular Random

Forest algorithm (RF) at 0.1°. We assessed the applicability of RF as a gap filling and downscaling tool to generate spatial and temporal consistent global *evi* estimates. To do this, we trained an RF regressor with 0.1° *evi* data using a host of features indicative of water and energy balances experienced by vegetation and we evaluated the performance of this new product. Next, to test whether the RF is robust in terms of spatial resolution, we downscale global *evi*, the model trained on 0.1° data is used to predict *evi* at 0.01° resolution. The results show that the RF can capture global *evi* dynamics at both the 0.1° (RMSE: 0.02 - 0.4) and at the finer 0.01° (RMSE: 0.04 - 0.6) resolution. Overall errors were higher in the down-scaled 0.01° compared to the 0.1° product. Yet, relative increases remained small, thus demonstrating that RF can be used to create downscaled and temporally consistent *evi* products. Additional error analysis reveals that errors vary spatiotemporally, with underrepresented land cover types and periods of extreme vegetation conditions having the highest errors. Finally, this model is used to produce global spatially continuous *evi* products at both the 0.1° and 0.01° spatial resolution for 2003-2013 at an 8-day frequency.

1 Introduction

The impacts of natural hazards are felt on a local scale, but creating impactful risk management strategies requires a global view on the driving processes and impacts (Ward et al., 2020). Given its complex and multivariate nature, a global perspective is especially necessary when considering drought hazards. Drought is one of the most disruptive natural hazards, causing negative repercussions on the environment, economy, and society, which can affect large areas and populations (Naumann et al., 2014; Vereinte Nationen, 2021). However, a universal definition of what constitutes a drought event remains elusive, and as a result, we lack a comprehensive understanding of the direct and indirect effects of drought on the environment and society (Blauhut et al., 2016; Vogt et al., 2018; Sutanto et al., 2019). Remotely sensed products that monitor earth system responses during drought periods are one promising tool that can enable a global perspective on drought hazards and their impacts (AghaKouchak et al., 2015; West et al., 2019). Yet, they suffer from trade-offs between spatial and temporal resolution, where we either have high-resolution low frequency products or vice versa. The production of high-resolution spatially continuous products can facilitate a more holistic view of drought responses and management by incorporating more relevant fine-scale processes (Chen et al., 2022; Schneider et al., 2017).

Vegetation is involved in numerous drought-impact pathways and using remote sensing to track vegetation responses has been widely used (Zhang et al., 2021c; AghaKouchak et al., 2015). Drought disrupts terrestrial water and carbon cycles, which can reduce the integrity of ecosystem dynamics and associated ecosystem services (Banerjee et al., 2013; Crausbay et al., 2017; Han et al., 2018; Smith et al., 2020). More subtlely, vegetation also affects the dynamics of drought propagation itself; under favourable antecedent conditions, vegetation overshoot may exacerbate and facilitate the onset of rapid and intense droughts (Zhang et al., 2021c). Vegetation is also expected to play a crucial role in shaping drought resistance under future climate change (Vereinte Nationen, 2021). In the absence of such resistance, interventions to alleviate the negative impacts of disrupted ecosystem services can cost up to a billion dollars per drought event (Banerjee et al., 2013; Cammalleri et al., 2020). It follows that formulating appropriate responses to drought and alleviating the negative effects of ecosystem disruption during these periods requires accurate predictions.

In recent decades, numerous satellite-based vegetation indices have been developed (Li et al., 2021a). For example, the Enhanced Vegetation Index (*evi*), have proven to be an indispensable tool for monitoring vegetation at multiple scales, from the fine scale, such as crop patches (Moussa Kourouma et al., 2021; Sharifi, 2021) to the global scale (Huang et al., 2021; Vicente-Serrano et al., 2010). However, a persistent feature among these products are trade-offs between spatial resolution and revisiting times, where high temporal resolution is met by coarse spatial resolution and *vice verse*. For example, the Moderate Resolution Imaging Spectroradiometer (MODIS) captures the entire Earth with a high temporal resolution every 1 to 2 days (Zhao and Duan, 2020) with a maximum resolution of 250 m. Landsat and Sentinel-2 data have a higher spatial resolution, 10 and 30 m, but longer revisiting times of approximately 10 and 5 days, respectively (Zhu, 2017; Li et al., 2021a). Revisiting times for Landsat and Sentinel-2 are further prolonged when sensors or retrievals are interrupted by cloud cover, pollution in the atmosphere, or even technical issues. In addition to temporal frequency, temporal coverage is another important consideration. Coarser scale products are associated with older satellites and have more extended temporal coverage than the newer ones; MODIS products reach as far back as 1999 whereas Sentinel-2 products only go back to 2017. The ideal product for monitoring vegetation dynamics would have global coverage, little to no data gaps, and high spatial and temporal resolution.

Machine learning (ML) methods have been used for downscaling and gap filling purposes in remote sensing products and can be seen as one tool that can lead to the production of high-quality remote sensing products and thus alleviate the limitations around resolution and coverage current products exhibit (Zhu et al., 2022; Zeng et al., 2013). ML methods have been successfully applied to a wide range of drought-related (Hauswirth et al., 2021; Shamshirband et al., 2020; Tufaner and Özbeyaz, 2020; Shen et al., 2019; Das et al., 2020; Hauswirth et al., 2022) and remotely detected vegetation studies (Roy, 2021; Li et al., 2021b; Reichstein et al., 2019). Compared to conventional statistical downscaling techniques, ML is considered the superior alternative; given that no strict statistical assumptions are required, complex and non-linear relationships are well captured and provide high precision (Ebrahimy et al., 2021).

One ML algorithm that has been widely applied for gap filling and downscaling in remote sensing data is the Random Forest Regressor (RF) (Zhang et al., 2021a; Fu et al., 2022; Liu et al., 2020; Wang et al., 2022). Gap-filling can achieved by training a RF on available data and then use the model to predict values where data is sparse or missing (Wang et al., 2022). Using RF to downscale data involves establishing an RF at a coarse scale and predicting targets at finer resolutions by feeding the algorithm with high-resolution auxiliary data (Liu et al., 2020). These studies have highlighted that ML methods can accurately predict the dynamics of vegetation (Roy, 2021; Gensheimer et al., 2022). However, studies applying ML methods to global vegetation dynamics and assessing their suitability to investigate drought responses are less prominent, and it remains to be seen whether this approach is applicable at the global scale (Li et al., 2021b; Zhang et al., 2021c; Chen et al., 2021).

This study aims to further our understanding on how well ML methods can be used create vegetation products that are useful for global drought impact applications. This will allow us to further quantify to what degree ML can facilitate continues drought monitoring by gap-filling and downscaling existing remote sensing products. We set out to establish whether ML methods can alleviate missing data and resolution limitations of remote sensing-based vegetation health products by linking vegetation condition (*evi*) with meteorological and hydrological data. This was done in three steps; first, we assess whether ML is an appropriate tool to predict the condition of vegetation on a global scale and act as a gap filling tool. Second, we determine whether ML can be used to downscale vegetation conditions and predict values at spatial scales finer than those provided during training. High degrees of transferability between scales could allow for further spatial up- or down-scaling of the vegetation status in future applications while still providing robust predictions. Last, to explore how these products can be applied to drought impact studies, we investigated how well the ML-based vegetation maps predict vegetation status during periods of drought.

2 Materials and Methods

The materials and methods are constructed so that each subsection corresponds to one of the objectives. We first provide an overview of the approach used to construct a RF, using a variety of input data, which is the used to assess how well a ML approach can be used as a gap filling and downscaling tool. We then detail how we trained the RF and which data was used, followed by how we tested the gap-filling and downscaling capabilities in two subsequent sections. Last, the gap-filled and down-scaled products are stress tested by investigating how well they can be used to derive insights into global vegetation dynamics, specifically under drought conditions.

The relative abundance of remotely sensed vegetation data provides an opportunity to effectively establish the suitability of ML based methods for gap-filling and downscaling. In this study we relied on already assimilated data products to test the applicability of RF as a downscaling and gap-filling tool. To do this we first set out to train an RF on a subset of the available *evi* data at 0.1°. As a test of its gap-filling abilities, the model was then used to predict *evi* values at locations not seen during training. To determine how viable the RF is for downscaling, we predicted *evi* at the 0.01° resolution by providing the model with high resolution auxiliary data that were available and regridding the data that were not available at a high resolution.

2.1 Random Forest Regressor

2.1.1 Data Sources

The data sources (Table 1) and further information in the following subsections were used to construct a 0.1° resolution dataset to train and test the ML model. The data set spans 10 years, from 2003 to 2013. The goal was to have all data at a 0.1° resolution, in cases where the resolution of the downloaded data was not, the relevant treatments are described below.

Vegetation Index - The reference data used in this study is the *evi* index. *evi* data provide the observational benchmark for the training and validation of the ML-based products created in this study. The *evi* can be used as an indicator of overall vegetation status and health, as it is sensitive to chlorophyll content and correlates with primary production, photosynthesis rates, and vegetation physiognomy (Box et al., 1989). Compared to the more widely used Normalized Difference Vegetation Index, *evi* is considered the superior index, as it is less sensitive to atmospheric conditions and saturation effects in areas of dense vegetation (Gao et al., 2000). These data arise from the Moderate Resolution Imaging Spectroradiometer aboard the Terra and Aqua satellites. Sensors aboard Terra and Aqua are identical, and the 16-day composite images from each sensor are released 8-days apart. In this study, Google Earth Engine's python Application Program Interface (Gorelick et al., 2017) through the geemap package (Wu, 2020) was used to access the terra (MOD13A2.006) and aqua (MYD13A2.006) *evi* data. These two products were combined to produce a quasi-eight-day time series (Didan, 2015, 2021). For the experimental setup used here, we required two sets of *evi* data, one at the 0.1° resolution, for training the RF and test its gap filling capability, and another at the 0.01° resolution to assess its downscaling abilities. To enable for the assessment of gap-filling and downscaling capabilities of the RF we downloaded one dataset at the 0.01° and another at the 0.1° resolution. The two different resolution datasets were acquired by relying on Google Earth Engines' Image Pyramiding Policy. This policy aggregates high resolution data to the required resolution using the mean for continuous variables (i.e., *evi*).

Table 1. Target variable (*evi*) and potential features with accompanying name, description, units, spatial resolution (Spat. Res.) and temporal resolution (Temp. Res.) and references.

Name	Description	Units	Spat. Res.	Temp. Res.	Reference
Target Variable					
evi	Enhanced Vegetation Index	-	0.01°	8 day	Gao et al. (2000)
Feature Variables					
lc	Land cover Types	-	500m	Yearly	Friedl, Mark and Sulla-Menashe, Damien (2019)
elv	Elevation	m	92m		
hnd	Height Above Nearest Drainage	m	92m	Static	Yamazaki et al. (2019)
aspect	Aspect	°	92m		
slope	Slope	°	92m		
tp	Total Precipitation	mm.day^{-1}	0.1°		
t2m	Two-meter Temperature	°C	0.1°		Muñoz-Sabater et al. (2021)
swvl1	Soil Water Volumetric Layer 1 (0-7cm)	-	0.1°	Hourly	
stl1	Soil Temperature Layer 1 (0-7cm)	°C	0.1°		
pet	Potential evapotranspiration	mm.day^{-1}	0.1°		Singer et al. (2021)
spi1,spi3,...spi24	Standerdized Precipitation Index	-	0.1°	Monthly	this study
spei1,spei3,...spei24	Standerdized Precipitation Evaporation Index	-	0.1°		

Highlighted rows indicate that features were dropped from further analysis after conducting feature selection prior to model fitting.

Feature Variables - Global vegetation dynamics are largely driven by terrestrial water and energy balances (Hawkins et al., 2003). Similarly, the responses of vegetation to drought are regulated, in part, by water and energy availability (Xu et al., 2010). Consequently, a suite of data indicative of terrestrial water and energy balances were selected as potential input variables. These variables are introduced below, and Table 1 provides an overview.

Meteorology - Hourly data for total precipitation (*tp*), two-meter temperature (*t2m*), volumetric soil moisture layer 1 (*swvl1*; 0-7cm), and soil temperature layer 1 (*stl1*; 0-7cm) were retrieved from the hourly ERA5-Land Reanalysis product by the European Centre for Medium-Range Weather Forecasts (Muñoz-Sabater et al., 2021). In addition, potential daily evaporation (*pet*) was acquired from Singer et al. (2021), *pet* is calculated following the Penman-Monteith formulation with ERA5-Land as the input data. *pet* was included as it is directly correlated to air temperature and radiation (Thornthwaite, 1948; Monteith, 1965; Priestley and Taylor, 1972) and the photosynthesis potential of plants and thus can account for a plural of other variables. All meteorological data were resampled to match the 8-day frequency of the *evi* data. *tp* was aggregated by taking the cumulative sum of the previous 8 days, whereas the remainder of the variables were averaged over a previous 8 day window.

Drought Indices - Aside from short-term changes in water availability, it is also key to understand the long-term dynamics to identify drought legacy effects on the current vegetation states (Schwalm et al., 2017). To this end, the Standardized Precipitation Index (*spi*) (McKee et al., 1993) and Standardized Precipitation Evapotranspiration Index (*spei*) (Vicente-Serrano et al., 2010) were used to characterise these legacy effects. The *spi* and *spei* were calculated at the 1, 3, 6, 9, 12 and 24-month aggregation lengths. The different lengths of aggregation are related to types of drought: precipitation, soil moisture, and hydrological droughts. Precipitation and soil moisture droughts mostly correlate short-term deficits in soil water (1-3 months), and are important for vegetation with shallow roots; hydrological drought (6-12 months) can be a good proxy for impacts on shrubs, bushes and trees that have deeper roots and are likely to rely on local groundwater for water (12-24 months). In addition, the inclusion of drought indices allows for the characterisation of past climate memory effects on current vegetation growth (Reichstein et al., 2019; Schwalm et al., 2017) associated with past climatic conditions. The equations and steps for calculating *spi* and *spei* are detailed in Appendix A2.

Land cover Types and Topography - Land cover type is an important predictor of vegetation abundance and health (Meza et al., 2020). Here, the Moderate Resolution Imaging Spectroradiometer Yearly Land cover Types (MCD12Q1.006) were retrieved from the Google Earth Engine. In this product, land cover types are classified according to the International Geosphere-Biosphere Programme classification scheme. Barren land, deserts, permanent snow and water bodies were masked in all further analyses. It is important to note that the RF was supplied with the remainder 15 unique land cover types; however, these were collapsed into eight broader classifications for brevity and clarity in the results, discussion and visualisations. Grasslands, wetlands, croplands, urban and mixed did not require grouping and represent the accompanying class in accordance with the Geosphere-Biosphere Programme classification scheme. Forests refers to the grouped class which contains evergreen needleleaf, evergreen broadleaf, deciduous needleleaf, deciduous broadleaf and mixed broadleaf forests. Shrubland refers to the grouped class containing closed and open shrubland; whereas savannas refer to the grouped class containing woody savannas and savannas. To capture the variations in water and energy availability attributable to topographic effects, elevation (*elv*) and height from the nearest drainage basin (*hand*) were accessed from MERIT Hydro, a high-resolution global hydrography map (Yamazaki et al., 2019), also through Google Earth Engine. To enable for the assessment of gap-filling and down-scaling capabilities of the RF we downloaded one dataset at the 0.1° and another at the 0.01° resolution using the Google Earth Engine's python Application Program Interface (Gorelick et al., 2017) through the geemap package (Wu, 2020).The two different resolution datasets were acquired by relying on Google Earth Engines' Image Pyramiding Policy. This policy aggregates high resolution data to the required resolution using the mode for land cover data and mean for continuous variables (i.e., *evi* and *hand*). Last, 0.1° and 0.01° *slope* and *aspect* was calculated from *elv* using the relevant functions in `xarray-spatial` (Hoyer and Hamman, 2017).

2.1.2 Random Forest Model

While an abundance of ML approaches has been used to predict vegetation status, here the Random Forests Regressor (RF) was selected to link meteorology, land cover, topography, and drought inputs to vegetation health. RF is an ensemble method that fits many decision trees on different subsets of data.

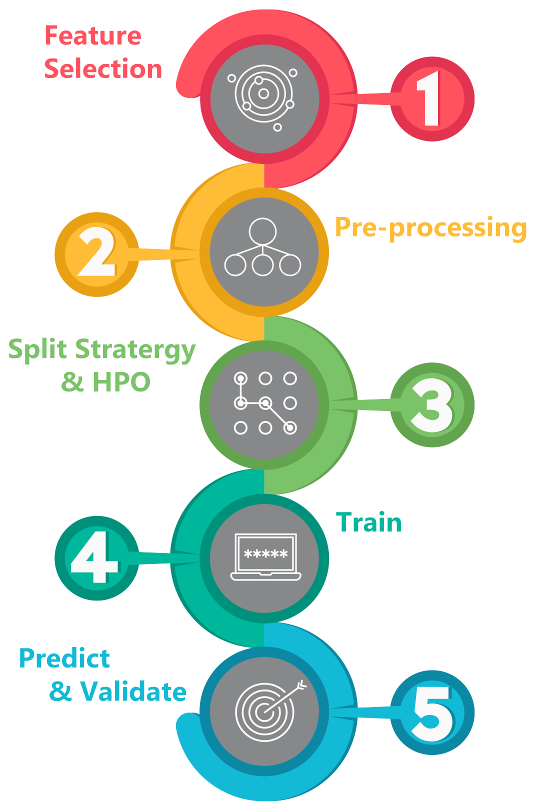

Figure 1. The five sequential steps followed during the RF fitting and evaluation.

RF is advantageous given its relatively straight-forwarded implementation, ability to incorporate categorical features, ability to easily identify causal links and limited risk of overfitting. The general pipeline used throughout consisted of five sequential steps (Fig. 1). Here, the RF was implemented in Python 3.9 (Rossum and Drake, 2010) under the `scikit_learn` framework (Pedregosa et al., 2011).

Feature Selection - In an attempt to include only relevant data in the ML model, the list of potential variables described in Section 2.1.1 and Table 1 was evaluated for their ability to provide meaningful information during model fitting. A pairwise Spearman rank correlation was calculated between all features to ensure that input data correlated with *evi*. Those variables that exhibited strong correlations were retained in further analysis, whereas variables that experienced weak correlations were excluded. *Aspect* did not exhibit strong correlations with *evi* (Fig. A1). Similarly, *spi* (at all aggregation times) did not correlate strongly with *evi*. In addition, *spi* and were closely correlated with *spei*, *spi* was excluded in favour of *spei* (Fig. A1). *spi* and *aspect* were excluded from further analyses; features that were excluded are highlighted in Table 1. Soil moisture and total precipitation exhibited some degree of cross-correlation in the global sense, yet these were retained to account for regions where soil moisture is independent of precipitation such as wetlands and groundwater dependent ecosystems.

Pre-processing - Given that the RF algorithm accepts 2-dimensional numeric arrays as input, the 3-dimensional data was processed so that each unique latitude and longitude was associated with a time series of each variable. The single categorical feature (*lc*) was converted to binary numeric. Each unique land cover type is assigned to a new feature, with 1 indicating presence and 0 indicating absence.

Split Strategy and Hyper-parameter optimisation - To refine the number of estimators and maximum depth, a 3-fold cross-validation approach using the `HalvingRandomSearchCV` was applied. This hyper-parameter optimisation provides the optimal configuration for the RF so that the critical vegetation dynamics are captured while simultaneously reducing the RF complexity and preventing over-fitting. The hyper-parameter optimisation focused on two parameter settings, namely, `Maximum_depth` and the `number_of_estimators`. The `Maximum_depth` determines the maximum depth of the decision tree and the `number_of_estimators` determines the number of decision trees used. The search space used for the `number_of_estimators` and `Maximum_depth` was 1-40 and 1-25, respectively. The upper bounds of the search space were largely determined by computational considerations, increasing the upper limits beyond 40 for `number_of_estimators` and 25 for `Maximum_depth` would result in impractical computation times. Nonetheless, even with this constraint, increasing the `Maximum_depth` and `number_of_estimators` past 12 and 13, respectively, yielded only marginal increases in test scores (Fig. 2a). Given that the risk of overfitting and computational time increases with increasing `Maximum_depth` and `number_of_estimators` combined with the fact that only marginal increases in test scores are observed past 12 and 13 these values were identified as optimal for `Maximum_depth` and `number_of_estimators` respectively.

After determining optimal parameter settings, the data were split into training and validation sets. However, three-dimensional data could conceivably be split along the temporal dimension where the model is trained on all locations with only a subset of the temporal availability (i.e., temporal splitting), or the data can be split according to location where only a subset of the grid pixels are selected for training but over the entire available period (i.e., spatial splitting). Given that previous research has highlighted that RF performance is sensitive to spatial vs temporal splitting, this is especially true for extreme events such as droughts (Hauswirth et al., 2021). We conducted a cursory analysis to determine whether a temporal or spatial splitting approach better balances trade-offs between computational complexity and learning rates. Learning curves for cursory RF models using each splitting approach were quantified and compared. Each model was supplied with increasing training sizes, and test scores were calculated and plotted to visualise learning curves. This cursory analysis revealed that spatial splitting yields faster learning curves than the temporal splitting approach (Fig. 2b); therefore, spatial splitting was identified as the preferred approach.

Train - For the final RF model, a spatial split with a (0.06:0.94) (train: predict) ratio was used to train the final model. A 0.06:0.94 split was chosen, and there was very little increase in performance past training sixes of 6% (Fig. 2b). `Maximum_depth` and `number_of_estimators` were set at 12 and 13, respectively. The parameters that were not subjected to hyper-parameter optimisation were set as follows: the `squared_error` criterion was used to measure the quality of the splits in branches, the maximum `number_of_features` considered in each split was set at `auto` which instructs the algorithm to consider all features when considering a split. The minimum `samples_per_leaf_node`, which determines the minimum number of samples required in a leaf node, was set at the default value of 1. The minimum `samples_per_split` was

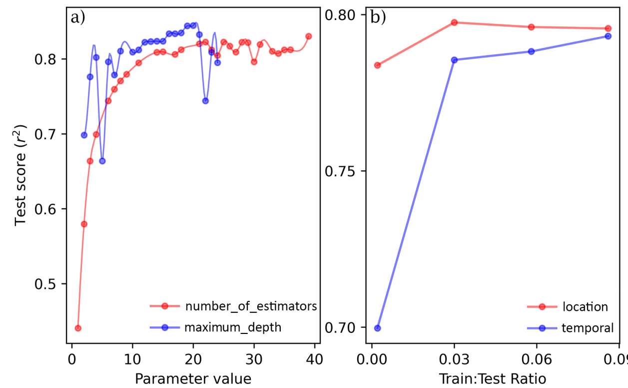

Figure 2. (a) Evolution of RF performance during `HalvingRandomSearchCV` hyper-parameter optimization of: `maximum_depth` and `number_of_estimators`. (b) RF performance following the incremental increase of train set size using a location based split approach compared to a temporal based split approach.

also set at the default value of 2, which means a split will only be considered if each branch left and right of an internal node has at least two samples in it.

2.2 Gapfilling *evi* using Random Forests

As a test of the RF gap-filling capabilities, we predicted *evi* for the 94% of the grid cells that were not used during training. The accuracy of these predictions was evaluated against the *evi* data obtained from MODIS. As a first-pass assessment of overall performance, the model was scored using default coefficient of determination (R^2) scorer in the RF implementation of scikit. The model predictions were further evaluated by calculating the root mean squared error (RMSE) and Pearson correlation coefficients. These were calculated independently for each grid cell to provide information on the spatial variation of errors. Last, to gain insight into which features were the most essential for predicting *evi*, global feature importance was calculated using Shapley Additive exPlanations' (SHAP) TreeExplainer (Lundberg et al., 2020).

2.3 Downscaling *evi* using Random Forests

In this section, the focus shifted toward whether RF can be used to downscale global *evi* values, that is, whether a model trained on 0.1° can accurately predict *evi* at a finer 0.01° scale. To this end, a 0.01° data set was compiled. In cases where data were not at the 0.01° resolution (i.e., meteorology and drought indices data) the nearest neighbour interpolation scheme from `xarray` (Hoyer and Hamman, 2017) was used to match the variables to the same spatial resolution. This data set was used as new input data to the already trained RF model to predict *evi* at the 0.01° scale. The evaluation approach for the downscaled values remained much the same, the overall model accuracy was assessed using (R^2) and (RMSE), and Pearson correlation coefficients were calculated for each grid cell.

2.4 Applicability of ML informed vegetation status products during periods of drought

One noticeable shortcoming of the RF is its relatively poor ability to predict extreme values depending on the training selection (Hauswirth et al., 2021). To determine to what extent this may influence the generality of the two products mentioned above, we further investigated the accuracy of the predicted *evi* under low growing conditions by calculating the anomaly correlation coefficient of *evi* ($eviACC$; Eq. 1), where $eviACC_{i,j}$ denotes *evi* anomaly for the month j in year i, $e\bar{v}i_{,j}$ denotes the average *evi* of month *j* over 2003-2013; σ stands for the standard deviation of *evi* during the period. We use this metric to assess the applicability of the RF based 0.1° and 0.01° *evi* predictions against remotely sensed *evi*. We consider $eviACC$ values greater than 0 as capable of capturing anomalies beyond the seasonal cycle and values exceeding 0.2 as good, given the strong seasonal cycle that is present in *evi* data.

$$eviACC_{i,j} = \frac{evi_{i,j} - e\bar{v}i_j}{\sigma} \tag{1}$$

3 Results

The results here are presented in three parts. First, the results of the model trained on the 0.1° data are presented; here, the focus is retained on the model's performance and ability to predict the status of the vegetation at the spatial resolution it is not trained at and in such acting as a gap-filling tool. We also touch on which features are most important in predicting the status of the vegetation. Subsequently, we present the model's performance when used to downscale *evi* and predict 0.01° data. We explore how this model can be used to gain insight into global vegetation dynamics by assessing the accuracy of both products under drought conditions.

3.1 Gapfilling *evi* using Random Forests

The model was able to reproduce global vegetation patterns by correctly predicting high vegetation density in tropical forests and low vegetation density in arid and urban regions of the world (Fig. 3). SHapley Additive exPlanations values provided an understanding of the relative importance of each feature in predicting *evi*. The most important features were those associated with meteorology, land cover type and elevation; drought indices and slope proved to be less important (Fig. 4).

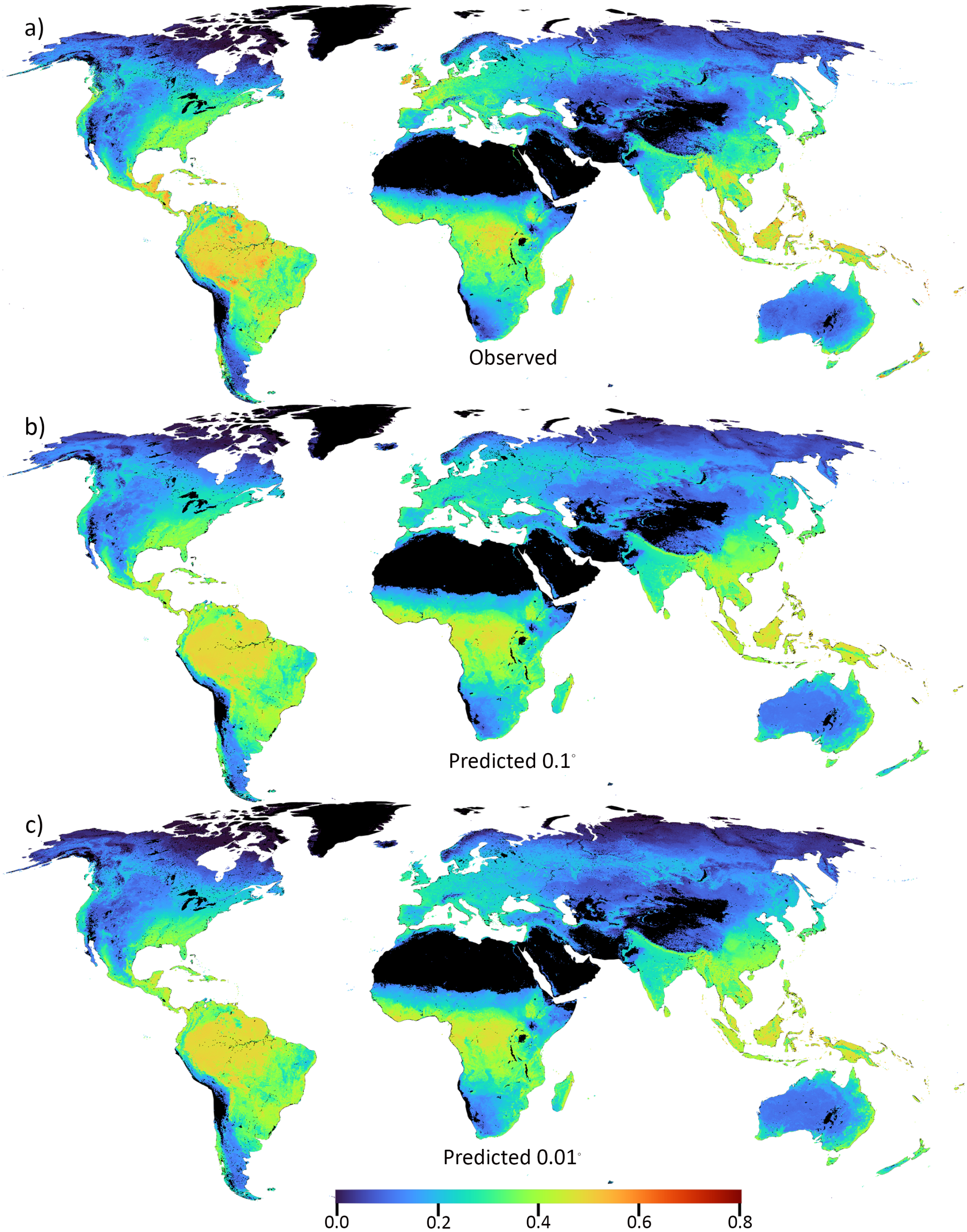

Figure 3. Mean *evi* (2003 - 2013) for the (a) observed 0.01°, (b) predicted 0.01° and (c) predicted 0.1° values by the RF. Barren land, deserts, permanent snow, and water bodies were masked and represented by black.

When trained on only 6% of the data (i.e., the point at which the use of additional data did not result in better predictions but increases the risk of overfitting), the RF was able to predict global *evi* accurately with a spatial resolution of 0.1° (R^2 = 0.86; Fig. 3, 5, 6 & 7a). Looking more closely at the distribution of errors, less than 1% of grid cells showed negative correlations and more than 80% showed correlations higher than 0.5 (Fig. 7c) and RMSE ranged between 0.02 and 0.4 (mean: 0.05 ± 0.03; Fig. 7d). However, it is important to note that the accuracy was neither spatially nor temporally uniform. Land cover types were an important feature in determining predictive ability. The predictions of *evi* in areas dominated by urban, mixed and crop land cover types showed the highest degree of error (Fig. 6a). On the contrary, the most natural types of land cover, such as forests and grasslands, were the most accurately represented by the model (Fig. 5a & 6a). For all types of land cover, the periods of maximum and minimum *evi* were less accurately predicted than the intermediate periods (Fig. 6a). Predicted *evi* was consistently overestimated by the model in urban land covers (Fig. 6a).

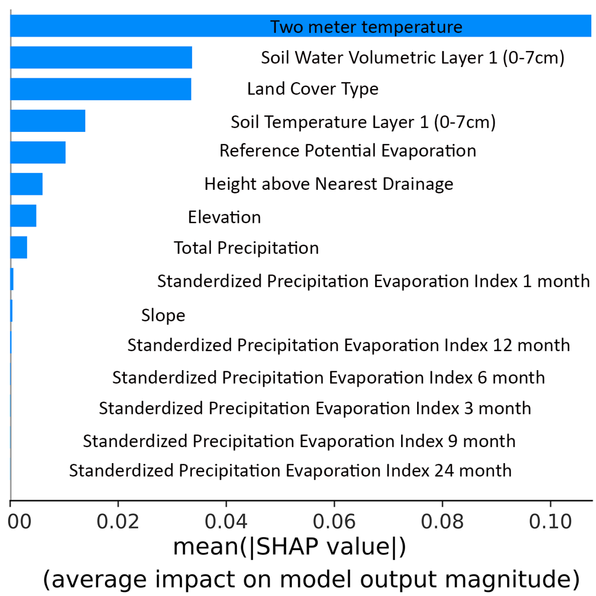

Figure 4. Feature importance for the RF based predicted *evi* at 0.1°. The features are ordered by level of importance, with higher mean SHAP values indicating higher importance.

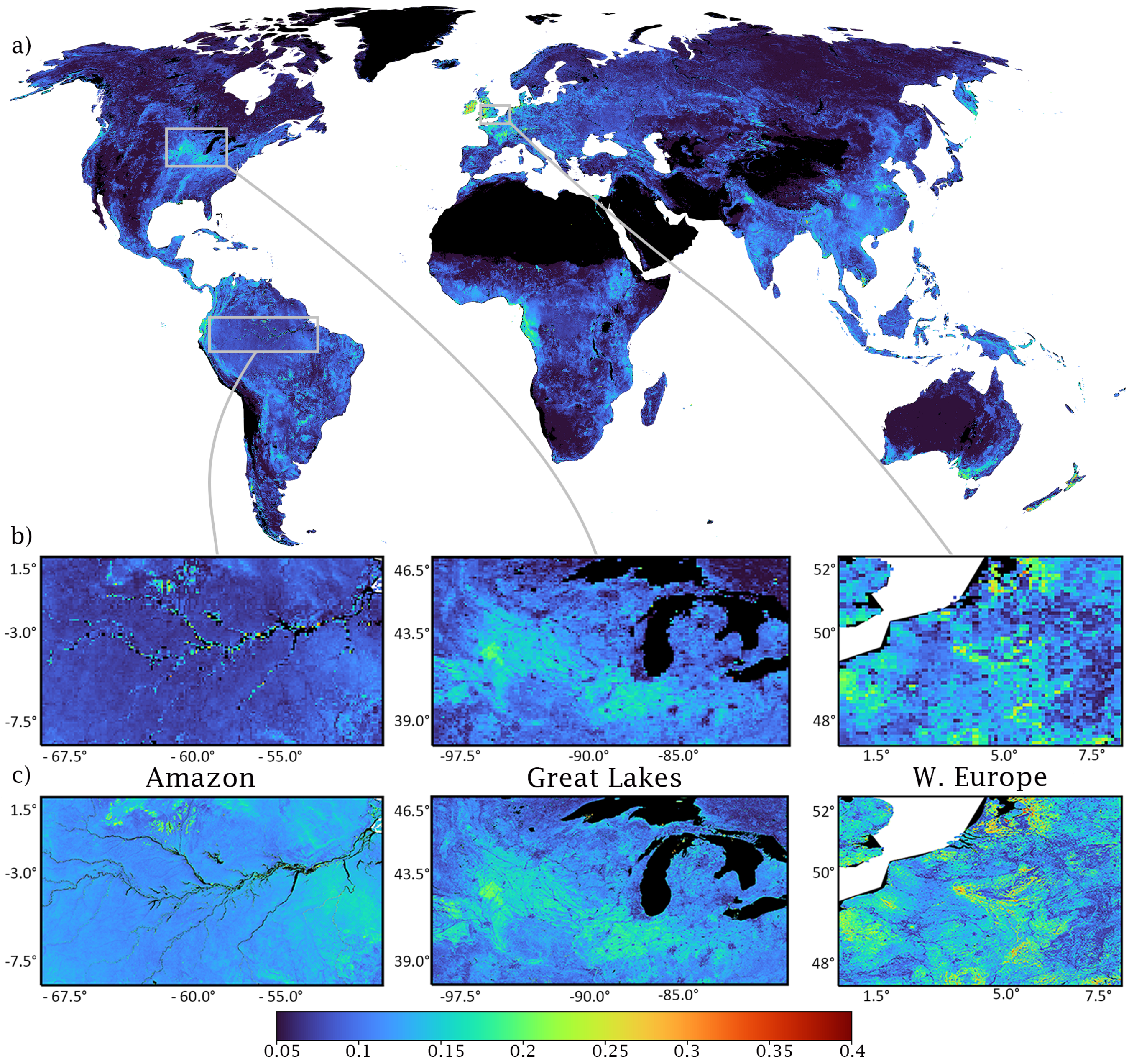

Figure 5. (a) RMSE for predicted *evi* using a RF model. Zoomed in inserts show Amazon Basin as representative of forest land cover, Great Lakes as a representative of croplands and Western Europe representative of urban land cover over at the (b) 0.1° and (c) 0.01°.

3.2 Downscaling *evi* using Random Forests

When the model trained on 0.1° data was used to predict *evi* at the 0.01° spatial resolution, there was a slight drop in accuracy but the model was still able to capture spatial and temporal vegetation dynamics when supplied with 0.01° data (Fig. 5 & 7b). The predictive capacity was still good but reduced compared to the 0.1° product, with a median R^2 of 0.75 (Fig. 7b). The errors also increased, 5% of grid cells displaying negative correlations (Fig. 7c) compared to less than 1% for the 0.1° product. RMSE ranged between 0.04 and 0.6 (mean: 0.09 ± 0.07; Fig. 7d), with the majority of the grid cells exhibiting RMSE around 0.05. Accuracy was again dependent on the land cover, with urban, mixed and crops performing the worst (Fig. 6b). Noticeably, for urban land cover types the model consistently overestimated *evi*.

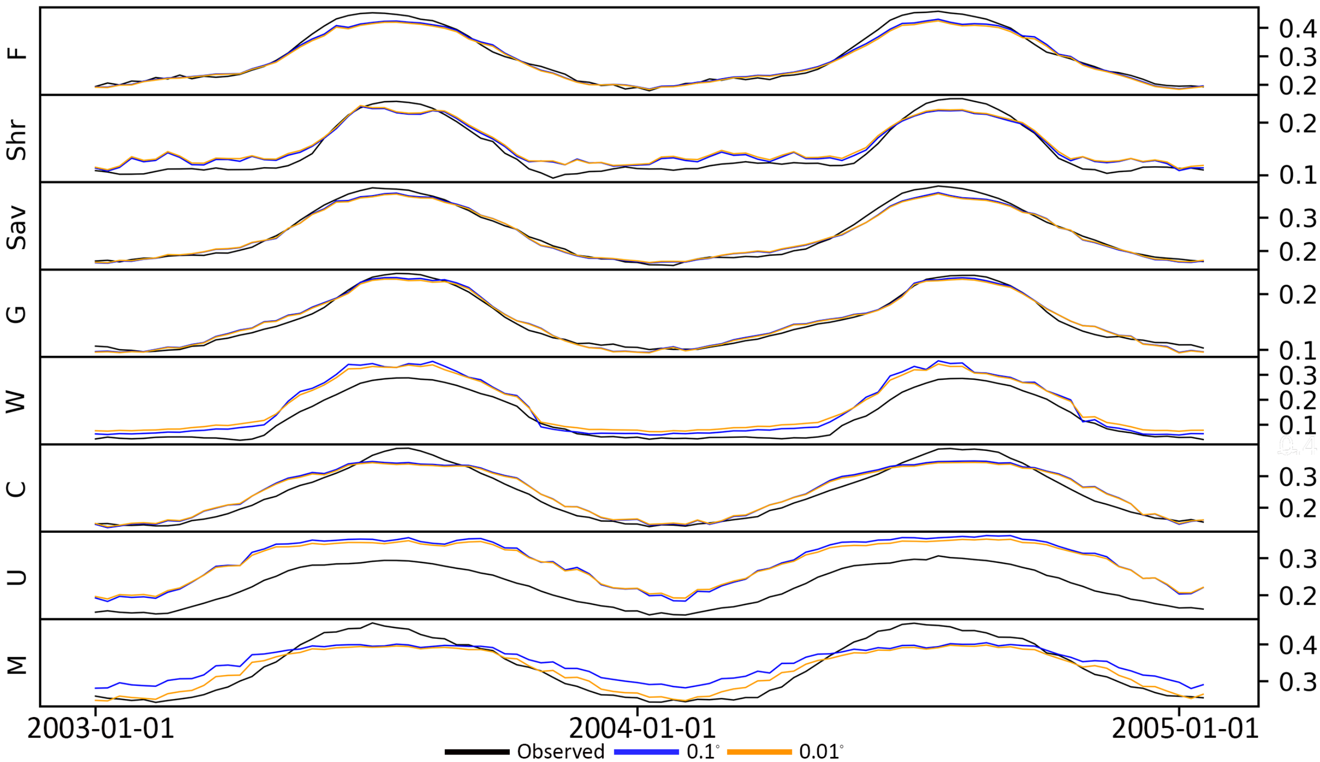

Figure 6. Time series of average and predicted *evi*, per major land cover type at 0.1° and 0.01°. F=Forest, Shr=Shrubland, Sav=Savanna, G=Grassland, W=Wetlands, C=Crops, U=Urban, M=Mixed.

3.3 Accuracy under drought conditions

The anomaly correlation analysis revealed that the RF was still able to capture *evi* anomalies (Fig. 8), but to a lesser extent compared to overall performance (Fig. 7c). The majority of grid cells showed positive correlations, with less than 10% displaying negative correlations; indicating that for that 90% of the locations where $eviACC$ was positive, the RF can reproduce

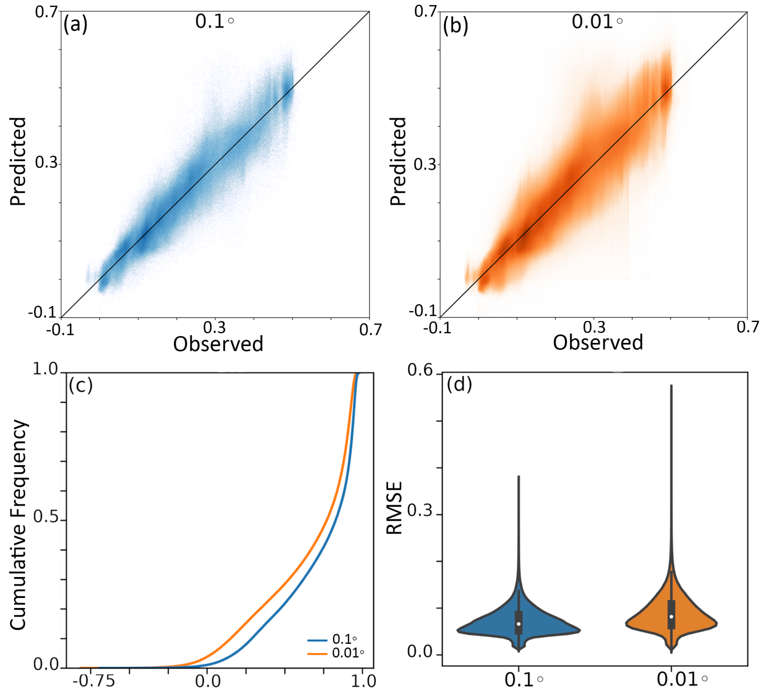

Figure 7. : (a) Scatter plot of observed and predicted evi at 0.1° and (b) 0.01°; Cumulative distribution function for (c) Pearson Correlation Coefficients for all grid points at 0.1° and 0.01°, (d) violin plot of RMSE for all grid points at 0.1° and 0.01°

anomalies from the average seasonal cycle and thus can be used to identify periods of negative or positive *evi* impacts resulting from droughts or more favourable growing conditions. More than 50% of grid cells exhibited an $eviACC$ of 0.25 for 0.1° compared to 45% when *evi* was predicted at 0.01° (Fig. 8).

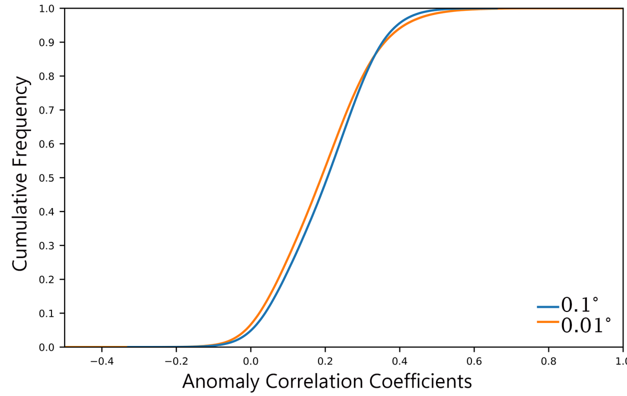

Figure 8. Cumulative distribution curves of anomaly correlation coefficients for *evi* predicted by a RF at 0.1° and 0.01°.

4 Discussion

4.1 Gapfilling *evi* using Random Forests

The results here show that RF can accurately predict *evi* at unseen geographic locations when trained on relatively few data; here training the RF on only 6% provides a representative sample of global distribution of *evi* values (see section 4.4 for further discussion on the influence of data representativity). Using RF as a gap-filling tool has previously been applied remotely sensed vegetation indices (Roy, 2021; Sarafanov et al., 2020; Sun et al., 2023; Wang et al., 2021; Moreno-Martínez et al., 2018) albeit mostly at the more local scale. Although challenging to directly compare the errors of a global product to other regional products; the errors and correlations reported here are comparable with the regional studies ($R^2 \approx 0.9$ and RMSE:0.02 - 0.4). Two previous studies have however applied ML techniques to predict *evi* at the global. These studies relied on Long Short-Term Memory (LSTM) networks, using only meteorological input data, to predict global 15 day and 8-day *evi* at the 0.5° (Chen et al., 2021) and 250m (Xiong et al., 2023) resolution, respectively. This study, using a more simple ML model, reports similar rates of error ($R^2 \approx 0.9$ and RMSE:0.02 - 0.4) compared to the more sophisticated methods in Chen et al. 2021 (RMSE=0.01) and Xiong et al. 2023 ($R^2 \approx 0.9$ and RMSE ≈ 0.07), which suggests that using multiple sources of input data is beneficial. The use of multiple sources of earth data in conjunction with RF has also been used for predicting global soil moisture (Zhang et al., 2021b). In addition to other ML based methods, this current work adds to the number of already available tools (reviewed in Peng et al. 2017) that can be used for gap-filling and the production of global and spatially continuous *evi* datasets.

4.2 Downscaling *evi* using Random Forests

The RF accurately predicted *evi* at finer spatial scales than was trained, successfully predicting *evi* at a scale of 0.01° using high-resolution auxiliary data. However, it should be noted that this resulted in a reduction in precision compared to the 0.1°product. This is an expected result, given that *evi* at the 0.01° resolution will exhibit greater variances and more extreme values during periods of high and low growth. Scale-dependent drivers of vegetation dynamics may be another phenomenon that contributes to decreased precision when predicting *evi* at the 0.01° using a model trained at a coarser resolution. Meteorology has been shown to be tightly coupled to vegetation at the ecosystem scale but less so at finer scales, where biotic processes, such as competition, herb ivory, disease, and fire, are more important (Franklin et al., 2020). When predicting *evi*, the relative increases in error remained small. Downscaling vegetation indices using ML methods have previously been applied to downscale other remotely sensed variables such as precipitation (Park et al., 2022), evapotranspiration (Hobeichi et al., 2023) and gross-primary productivity (Gensheimer et al., 2022).

4.3 Random Forests for predicting drought effects

The increase in error among extreme values is a known limitation of the RF (Hauswirth et al., 2021) and in accordance the RF was less capable of capturing extreme values of *evi* compared to the overall performance of the model. During RF training, an evaluation metric, in this case `squared_error`, is used to minimize the error for the model as a whole. In this scenario, optimal fits inevitably result in reduced errors for values close to the mean at the expense of inflated errors for the outliers (Ribeiro and Moniz, 2020). In the current study, this means that *evi* during normal growth periods is prioritized over periods of extremely low or high vegetation growth. Nonetheless, given that the majority of the grid cells exhibited positive anomaly correlations, the ability to predict vegetation status under drought is still a positive result in accordance with previous research (Prodhan et al., 2022; Hauswirth et al., 2021). Although, provided that more sophisticated machine learning models tend to predict extreme values more accurately than the RF used here (e.g., Kladny et al. (2024)) future studies should aim evaluate their feasibility and applicability to predict vegetation status under drought conditions at the global scale. Yet in comparison with RF, the more complex algorithms have larger computational requirements during training of the model and are less capable of capturing potential non-linearity's.

4.4 Importance of Land cover Types and Input Data

Varying error according to land cover type in the 0.1° and 0.01° is expected for at least two reasons. The first relates to the inherent features of the RF algorithm itself, and the second to the environmental process that affects the dynamics of *evi*. A limitation of the RF algorithm is that when data is imbalanced, underrepresented groups are less well explained by the algorithm. Accordingly, accuracy varied according to a proportional abundance of land cover types (Jung et al., 2020). Dominant land cover types, such as forests and grasslands, displayed the least amount of error; in contrast, minority land cover types regions that have undergone human modification (i.e., urban areas and croplands) were associated with the highest error. Second, the features used in this study may not incorporate processes critical to vegetation status equally among land cover types (Moussa Kourouma et al., 2021). Forests, grasslands, and other natural ecosystems are closely coupled to water availability determined by climatic variations processes. However, croplands and urban areas may be less influenced by weather and more influenced by anthropogenic manipulations of water and energy balances (Zhang et al., 2004; Hawkins et al., 2003; Tang et al., 2021). A potential solution to this problem is to rely on Extreme Gradient Boosted Decision Trees, which have been shown to provide more accurate predictions where data are imbalanced (Li et al., 2021b) or include information on human-water management to better represent drought responses (Wanders and Wada, 2015).

Land cover-specific variations in the model's ability to predict vegetation are an important outcome of this study. Apart from the statistical reasons detailed in the previous paragraph as potential mechanisms for this phenomenon, an additional, and most likely compounding explanation is that the data used to predict *evi* may be more relevant for some land cover types and levels of vegetation growth than others. For instance, vegetation status in urban areas and croplands shows weak correlation or high errors (Fig. 5 & 6). The meteorological data used here to predict *evi* may not be the only factor driving the vegetation dynamics in human-modified areas. It is possible that irrigation, harvesting, and water management influence vegetation. Indeed, vegetation in urban areas have been shown to grow more rapidly and have a longer growing season than rural counterparts; this is thought to be driven by higher temperatures, high concentration of airborne phosphorous and other aerosol pollutants (Sicard et al., 2018a, b; Pretzsch et al., 2017). In contrast, natural forests and grasslands show high levels of accuracy and correlations, thus suggesting that the data used here is appropriate for the machine learning models to capture vegetation dynamics. Similarly, poor accuracy in wetlands is not unexpected as wetland vegetation is primarily driven by water quality, salinity, and pH (Grieger et al., 2021). On the contrary, forests and grasslands show high accuracy when using meteorological variables, since these are important drivers of vegetation growth in these areas. Although not directly related to vegetation, Hauswirth et al. (2021) showed that by including water management practices in machine learning models, the predictions of groundwater head and stream flow were more accurately predicted. It is important to note that the relevancy of predictors in shaping *evi* does not only affect the accuracy between land cover types but also plays a role in determining the overall accuracy of the model. For instance, precipitation and soil moisture do not exhibit similar feature importance, whilst soil temperature and two meter temperature does. The amount of precipitation retained in soils is dependent on a number of factors, and these results suggest that soil water moisture is a more critical variable than precipitation in governing global *evi* dynamics; this aligns with the observation of the residence time of precipitated water in soils that are often much longer than the actual precipitation events (McColl et al., 2017). In addition, slope is known to be an important determinant of vegetation status at fine spatial resolutions (Chen et al., 2013). Yet, the relatively weak feature importance of slope suggests that the model could not find much meaningful information regarding vegetation status and slope at the $0.1°$ during training and subsequently would be unable to use this information when predicting *evi* at the $0.01°$ resolution.

One other possibility is that uncertainty in the input data prevents more accurate predictions by the model. The temperature of ERA5-Land is known to show weaker correlations with the observed data in the tropics compared to more northern and southern latitudes (Muñoz-Sabater et al., 2021). In accordance to that, the errors *evi* predicted using the RF model largely follow this pattern where errors are higher in the tropics compared to the temperate zones. The temperature from ERA5-Land show relatively higher errors along the Andes, the northern reaches of the African rainforest, and the Sichuan Basin in China and the errors in predicted *evi* mirror this uncertainty. Similarly, when comparing errors in soil water content from ERA5-Land; Gabon forest's, the Andes, Vietnam, New South Wales in Australia and the East African Rift Valley have relatively high errors (Lal et al., 2022). Again, the errors in predicted *evi* are also relatively high in these regions. When considering the quality of land cover data used here, some inconsistencies may affect the ability of the RF to accurately predict *evi*. For example, when croplands are smaller than the pixel size used in MODIS, these croplands are incorrectly assigned as natural vegetation. Furthermore, temperate evergreen needleleaf forests are misclassified as broadleaf evergreen forests, and some grassland areas are classified as savannas. The relatively poor predictive performance in mixed land cover types further reiterates the need to provide models with appropriate input data sources where string signals are present. In addition, data that is more relevant to vegetation dynamics could provide better results; for example the weak feature importance of slope and the various SPEI metrics at their various aggregation times suggests that these variables do not play a relatively important role predicting *evi* at the temporal and spatial scales here. For example at fine scales, slope and aspect are important for determining radiation intensity experienced by plants but training the model at 0.1° means that this effect becomes less important and is not learnt by the model. Perhaps a better result would be acquired if more scalable variables would be included, specifically for the downscaling component.

5 Conclusions

The results of this study reveal that the RF is an appropriate method for predicting *evi* on the global scale, at the 0.1° and downscaled 0.01° resolution. In general, RF was capable of predicting *evi* dynamics with high accuracy; global patterns of vegetation and temporal dynamics were well captured with land cover, and variables relating to energy and water balances experienced by plants, baring the most significance. The model was able to capture annual vegetation growth cycles and distinguish between the main global biomes with high accuracy. However, it is essential to note that higher errors were associated with under-represented land cover types and periods of extreme vegetation growth, such a drought periods. Lower accuracy for under-represented classes in unbalanced data sets and a hampered ability to predict extreme values is a common criticism of the RF. In accordance with this study, land cover types that account for a smaller fractional cover of the earth's surface, and periods of extreme vegetation growth, were associated with the highest error. Predicting *evi* at a finer resolution resulted in increased errors. This is attributed to higher variances in the 0.01° product compared to 0.1° and it is important to note that the relative increases remained small.

The results here also highlight the use of RF for efficiently and accurately predicting missing data and downscaling, which ultimately allows for the production of spatially continuous *evi* data at very high spatial and temporal resolutions. To this end, this study produces spatially continuous *evi* product at 0.1° and 0.01° resolution, and therefore this approach could be used to fill existing gaps in satellite observations or in conjunction with satellite data to improve the monitoring of drought impacts on vegetation. For example, Landsat and Sentinel-2 satellites can produce high-resolution vegetation products; however, retrievals are strongly affected by weather conditions, which results in data gaps. In addition, their relatively low orbiting altitude means that the spatial coverage for each pass over is low. Using this approach on such data could produce globally continuous vegetation products resolutions lower than 100m.

This study shows that ML can be used for drought monitoring at high spatial and temporal resolutions, however there are trade-offs when it comes to using machine learning for vegetation drought impact monitoring. ML based *evi* estimates can be used to assess the potential impact of droughts on vegetation, however this ML based estimates still require meteorological input dataset. The ML model also needs to be trained on actual remotely sensed *evi* observations to identify the relationship between these meteorological variables and vegetation drought impacts. This inherently makes the ML based estimates as good as the remotely sensed product, and as long as no reliable alternative exists it will be difficult to fully replace remotely sensed *evi* observations. However, there is an added benefit of having continues high resolution global coverage derived from a ML based *evi* estimate. Finally, the ML-based estimates also allow us to extrapolate the *evi* records to historical periods for which meteorological data exist but satellite remotely sensing was not yet available or use as post-processing in hydrological model simulations to directly estimate drought impacts.

This study adds to previous research efforts that have successfully applied the RF in predicting vegetation status. Here the RF was used to produce a global spatial and temporally continuous *evi* product at 0.1° and 0.01°, with a median R^2 of 0.86 & 0.75, respectively. The approach outlined in this study could be applied to Landsat and Sentinel-2, to produce continuous vegetation index data sets at the 30-10m spatial resolution. The RF algorithm is a powerful technique for predicting temporal and spatial vegetation dynamics from remote sensor data and can be used for gap-filling and downscaling. The novelty of this product, compared to previous studies, is that it has global coverage, high spatial and high temporal resolution.

Author contributions. **BvJ:** Data curation, Formal analysis, Writing – review & editing. **SH:** Data curation, Formal analysis, Writing – review & editing. **NW** Conceptualization, Formal analysis, Writing – review & editing.

Competing interests. NW is a member of the editorial board of journal Hydrology and Earth System Sciences. The peer-review process was guided by an independent editor, and the authors have also no other competing interests to declare.

Acknowledgements. Steven M. de Jong is thanked for his valuable input on previous versions of this manuscript. SH acknowledges funding from the Cooperate Innovation Program and the Department of Water, Transport and Environment at the Dutch National Water Authority, Rijkswaterstaat. NW acknowledges funding by the European Union (ERC-Stg, Multi-Dry, 101075354). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The five anonymous reviewers whose comments helped improve the manuscript are thanked.

Appendix: A1. Feature Selection

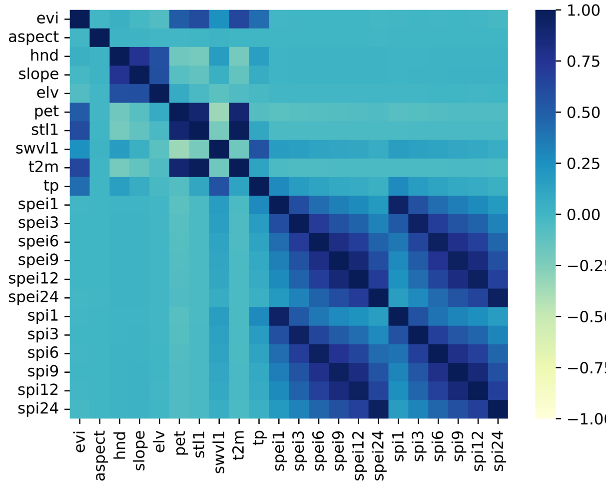

Figure A1. Correlation Matrix of pairwise Spearman rank correlation coefficients between all potential variables

Appendix: A2. Drought Indices Calculations

For the calculation of *spi*:

$$x = \sum_{i}^{m} tp_i \tag{A1}$$

where i is the month in question and $m = i - scale$.

For the calculation of *spei*:

$$x = \sum_{i}^{m} D_i \tag{A2}$$

where: $D_i = tp_i - pet_i$ and

$$x_{i,j}^{k} = \begin{cases} \sum_{l=13-k+j}^{12} tp_{i-j,l} + \sum_{l=1}^{j} tp_{i,l}, & \text{if } j < k \\ \sum_{l=j-k+1}^{j} tp_{i,l}, & \text{if } j \geq k \end{cases} \tag{A3}$$

This time series is then fitted to a gamma distribution taken the following steps:

First α and β fitting parameters as calculated as:

$$\hat{\alpha} = \frac{1}{4A}(1 + \sqrt{1 + \frac{4A}{3}}) \tag{A4}$$

Where $A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}$ with n being number of observations.

$$\hat{\beta} = \frac{\bar{x}}{\alpha} \tag{A5}$$

The gamma distribution probability density (Eq. A6) function with respect to *x* and including the calculated estimates for α and β can be inserted to produce an equation for the cumulative probability of a value for (Eq. A7).

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{x}{\beta}} \tag{A6}$$

where α is the shape parameter and β is the scale parameter and $\Gamma(a) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy$

$$G(x) = \frac{1}{\hat{\beta}^{\hat{\alpha}}\Gamma(\hat{\alpha})} \int_0^x x^{\hat{\alpha}} e^{\frac{-x}{\hat{\beta}}} dx \tag{A7}$$

then substituting t for $\frac{x}{\hat{\beta}}$ results in the incomplete gamma distribution (Eq. A8)

$$G(x) = \frac{1}{\Gamma(\hat{\alpha})} \int_0^x t^{\hat{\alpha}-1} e^{-1} dt \tag{A8}$$

Values of the incomplete gamma function can be computed using Eq. A9

$$H(x) = q + (1-q)G(x) \tag{A9}$$

Finally, values computed from Eq. A9 are transformed into the standard normal distribution to yield the *spi* and *spei* at the relevant time scales. These calculations were completed using the relevant algorithms in the `climate_indices` python package (Adams, 2021) using *tp*, *pet*, and *t2m* detailed in Section 2.1.2.

References

Adams, J.: Climate Indices in Python, https://github.com/monocongo/climate_indices, original-date: 2017-06-13T15:21:07Z, 2021.

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., and Hain, C. R.: Remote sensing of drought: Progress, challenges and opportunities: REMOTE SENSING OF DROUGHT, Reviews of Geophysics, 53, 452–480, https://doi.org/10.1002/2014RG000456, 2015.

Banerjee, O., Bark, R., Connor, J., and Crossman, N. D.: An ecosystem services approach to estimating economic losses associated with drought, Ecological Economics, 91, 19–27, https://doi.org/10.1016/j.ecolecon.2013.03.022, 2013.

Blauhut, V., Stahl, K., Stagge, J. H., Tallaksen, L. M., De Stefano, L., and Vogt, J.: Estimating drought risk across Europe from reported drought impacts, drought indices, and vulnerability factors, Hydrology and Earth System Sciences, 20, 2779–2800, https://doi.org/10.5194/hess-20-2779-2016, 2016.

Box, E. O., Holben, B. N., and Kalb, V.: Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO 2 flux, Vegetatio, 80, 71–89, publisher: Springer, 1989.

Cammalleri, C., Naumann, G., Mentaschi, L., Formetta, G., Forzieri, G., Gosling, S., Bisselink, B., De Roo, A., and Feyen, L.: Global warming and drought impacts in the EU: JRC PESETA IV project : Task 7., Publications Office, LU, https://data.europa.eu/doi/10.2760/597045, 2020.

Chen, J. M., Chen, X., and Ju, W.: Effects of vegetation heterogeneity and surface topography on spatial scaling of net primary productivity, Biogeosciences, 10, 4879–4896, https://doi.org/10.5194/bg-10-4879-2013, 2013.

Chen, Q., Timmermans, J., Wen, W., and van Bodegom, P. M.: A multi-metric assessment of drought vulnerability across different vegetation types using high resolution remote sensing, Science of The Total Environment, 832, 154 970, https://doi.org/10.1016/j.scitotenv.2022.154970, 2022.

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.

Crausbay, S. D., Ramirez, A. R., Carter, S. L., Cross, M. S., Hall, K. R., Bathke, D. J., Betancourt, J. L., Colt, S., Cravens, A. E., Dalton, M. S., Dunham, J. B., Hay, L. E., Hayes, M. J., McEvoy, J., McNutt, C. A., Moritz, M. A., Nislow, K. H., Raheem, N., and Sanford, T.: Defining Ecological Drought for the Twenty-First Century, Bulletin of the American Meteorological Society, 98, 2543–2550, https://doi.org/10.1175/BAMS-D-16-0292.1, 2017.

Das, P., Naganna, S. R., Deka, P. C., and Pushparaj, J.: Hybrid wavelet packet machine learning approaches for drought modeling, Environmental Earth Sciences, 79, 221, https://doi.org/10.1007/s12665-020-08971-y, 2020.

Didan, K.: MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC, 2015.

Didan, K.: MODIS/Aqua Vegetation Indices Monthly L3 Global 0.05 Deg CMG V061,[data set], NASA EOSDIS Land Processes DAAC, 2021.

Ebrahimy, H., Aghighi, H., Azadbakht, M., Amani, M., Mahdavi, S., and Matkan, A. A.: Downscaling MODIS Land Surface Temperature Product Using an Adaptive Random Forest Regression Method and Google Earth Engine for a 19-Years Spatiotemporal Trend Analysis Over Iran, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 2103–2112, https://doi.org/10.1109/JSTARS.2021.3051422, 2021.

Franklin, O., Harrison, S. P., Dewar, R., Farrior, C. E., Brännström, , Dieckmann, U., Pietsch, S., Falster, D., Cramer, W., Loreau, M., Wang, H., Mäkelä, A., Rebel, K. T., Meron, E., Schymanski, S. J., Rovenskaya, E., Stocker, B. D., Zaehle, S., Manzoni, S., van Oijen, M., Wright, I. J., Ciais, P., van Bodegom, P. M., Peñuelas, J., Hofhansl, F., Terrer, C., Soudzilovskaia, N. A., Midgley, G., and Prentice, I. C.: Organizing principles for vegetation dynamics, Nature Plants, 6, 444–453, https://doi.org/10.1038/s41477-020-0655-x, 2020.

Friedl, Mark and Sulla-Menashe, Damien: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006, https://doi.org/10.5067/MODIS/MCD12Q1.006, type: dataset, 2019.

Fu, R., Chen, R., Wang, C., Chen, X., Gu, H., Wang, C., Xu, B., Liu, G., and Yin, G.: Generating High-Resolution and Long-Term SPEI Dataset over Southwest China through Downscaling EEAD Product by Machine Learning, Remote Sensing, 14, 1662, https://doi.org/10.3390/rs14071662, 2022.

Gao, X., Huete, A. R., Ni, W., and Miura, T.: Optical–Biophysical Relationships of Vegetation Spectra without Background Contamination, Remote Sensing of Environment, 74, 609–620, https://doi.org/10.1016/S0034-4257(00)00150-4, 2000.

Gensheimer, J., Turner, A. J., Köhler, P., Frankenberg, C., and Chen, J.: A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet), Biogeosciences, 19, 1777–1793, https://doi.org/10.5194/bg-19-1777-2022, 2022.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, 202, 18–27, https://doi.org/https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Grieger, R., Capon, S. J., Hadwen, W. L., Mackey, B., Grieger, R., Capon, S. J., Hadwen, W. L., and Mackey, B.: Spatial variation and drivers of vegetation structure and composition in coastal freshwater wetlands of subtropical Australia, Marine and Freshwater Research, 72, 1746–1759, https://doi.org/10.1071/MF21023, 2021.

Han, D., Wang, G., Liu, T., Xue, B.-L., Kuczera, G., and Xu, X.: Hydroclimatic response of evapotranspiration partitioning to prolonged droughts in semiarid grassland, Journal of Hydrology, 563, 766–777, https://doi.org/10.1016/j.jhydrol.2018.06.048, 2018.

Hauswirth, S. M., Bierkens, M. F., Beijk, V., and Wanders, N.: The potential of data driven approaches for quantifying hydrological extremes, Advances in Water Resources, 155, 104 017, https://doi.org/10.1016/j.advwatres.2021.104017, 2021.

Hauswirth, S. M., Bierkens, M. F. P., Beijk, V., and Wanders, N.: The suitability of a hybrid framework including data driven approaches for hydrological forecasting, Hydrology and Earth System Sciences Discussions, pp. 1–20, https://doi.org/10.5194/hess-2022-89, 2022.

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J.-F., Kaufman, D. M., Kerr, J. T., Mittelbach, G. G., Oberdorff, T., O'Brien, E. M., Porter, E. E., and Turner, J. R. G.: Energy, water, and broad-scale geographic patterns of species richness, Ecology, 84, 3105–3117, https://doi.org/10.1890/03-8006, 2003.

Hobeichi, S., Nishant, N., Shao, Y., Abramowitz, G., Pitman, A., Sherwood, S., Bishop, C., and Green, S.: Using Machine Learning to Cut the Cost of Dynamical Downscaling, Earth's Future, 11, e2022EF003 291, https://doi.org/10.1029/2022EF003291, 2023.

Hoyer, S. and Hamman, J.: xarray: N-D labeled Arrays and Datasets in Python, Journal of Open Research Software, 5, 10, https://doi.org/10.5334/jors.148, 2017.

Huang, S., Tang, L., Hupy, J. P., Wang, Y., and Shao, G.: A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing, Journal of Forestry Research, 32, 1–6, https://doi.org/10.1007/s11676-020-01155-1, 2021.

Jung, M., Dahal, P. R., Butchart, S. H. M., Donald, P. F., De Lamo, X., Lesiv, M., Kapos, V., Rondinini, C., and Visconti, P.: A global map of terrestrial habitat types, Scientific Data, 7, 256, https://doi.org/10.1038/s41597-020-00599-8, bandiera_abtest: a Cc_license_type: cc_publicdomain Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Biodiversity;Biogeography;Environmental sciences;Macroecology Subject_term_id: biodiversity;biogeography;environmental-sciences;macroecology, 2020.

Kladny, K.-R., Milanta, M., Mraz, O., Hufkens, K., and Stocker, B. D.: Enhanced prediction of vegetation responses to extreme drought using deep learning and Earth observation data, Ecological Informatics, 80, 102 474, https://doi.org/10.1016/j.ecoinf.2024.102474, 2024.

Lal, P., Singh, G., Das, N. N., Colliander, A., and Entekhabi, D.: Assessment of ERA5-Land Volumetric Soil Water Layer Product Using In Situ and SMAP Soil Moisture Observations, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, https://doi.org/10.1109/LGRS.2022.3223985, 2022.

Li, S., Xu, L., Jing, Y., Yin, H., Li, X., and Guan, X.: High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques, International Journal of Applied Earth Observation and Geoinformation, 105, 102 640, https://doi.org/10.1016/j.jag.2021.102640, 2021a.

Li, X., Yuan, W., and Dong, W.: A Machine Learning Method for Predicting Vegetation Indices in China, Remote Sensing, 13, 1147, https://doi.org/10.3390/rs13061147, 2021b.

Liu, Y., Jing, W., Wang, Q., and Xia, X.: Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms, Advances in Water Resources, 141, 103 601, https://doi.org/10.1016/j.advwatres.2020.103601, 2020.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees, Nature Machine Intelligence, 2, 56–67, https://doi.org/10.1038/s42256-019-0138-9, 2020.

McColl, K. A., Alemohammad, S. H., Akbar, R., Konings, A. G., Yueh, S., and Entekhabi, D.: The global distribution and dynamics of surface soil moisture, Nature Geoscience, 10, 100–104, https://doi.org/10.1038/ngeo2868, 2017.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: Proceedings of the 8th Conference on Applied Climatology, vol. 17, pp. 179–183, Boston, issue: 22, 1993.

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.

Monteith, J. L.: Evaporation and environment, Symposia of the Society for Experimental Biology, 19, 205–234, https://repository.rothamsted.ac.uk/item/8v5v7/evaporation-and-environment, 1965.

Moreno-Martínez, , Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., Kramer, K., Cornelissen, J. H. C., Reich, P., Bahn, M., Niinemets, , Peñuelas, J., Craine, J. M., Cerabolini, B. E. L., Minden, V., Laughlin, D. C., Sack, L., Allred, B., Baraloto, C., Byun, C., Soudzilovskaia, N. A., and Running, S. W.: A methodology to derive global maps of leaf traits using remote sensing and climate data, Remote Sensing of Environment, 218, 69–88, https://doi.org/10.1016/j.rse.2018.09.006, 2018.

Moussa Kourouma, J., Eze, E., Negash, E., Phiri, D., Vinya, R., Girma, A., and Zenebe, A.: Assessing the spatio-temporal variability of NDVI and VCI as indices of crops productivity in Ethiopia: a remote sensing approach, Geomatics, Natural Hazards and Risk, 12, 2880–2903, https://doi.org/10.1080/19475705.2021.1976849, 2021.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, publisher: Copernicus GmbH, 2021.

Naumann, G., Barbosa, P., Garrote, L., Iglesias, A., and Vogt, J.: Exploring drought vulnerability in Africa: an indicator based analysis to be used in early warning systems, Hydrology and Earth System Sciences, 18, 1591–1604, https://doi.org/10.5194/hess-18-1591-2014, 2014.

Park, S., Singh, K., Nellikkattil, A., Zeller, E., Mai, T. D., and Cha, M.: Downscaling Earth System Models with Deep Learning, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3733–3742, ACM, Washington DC USA, https://doi.org/10.1145/3534678.3539031, 2022.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, : Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, http://jmlr.org/papers/v12/pedregosa11a.html, 2011.

Peng, J., Loew, A., Merlin, O., and Verhoest, N. E. C.: A review of spatial downscaling of satellite remotely sensed soil moisture, Reviews of Geophysics, 55, 341–366, https://doi.org/10.1002/2016RG000543, 2017.

Pretzsch, H., Biber, P., Uhl, E., Dahlhausen, J., Schütze, G., Perkins, D., Rötzer, T., Caldentey, J., Koike, T., Con, T. v., Chavanne, A., Toit, B. d., Foster, K., and Lefer, B.: Climate change accelerates growth of urban trees in metropolises worldwide, Scientific Reports, 7, 15 403, https://doi.org/10.1038/s41598-017-14831-w, 2017.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Monthly Weather Review, 100, 81–92, https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Prodhan, F. A., Zhang, J., Hasan, S. S., Pangali Sharma, T. P., and Mohana, H. P.: A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions, Environmental Modelling & Software, 149, 105 327, https://doi.org/10.1016/j.envsoft.2022.105327, 2022.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Ribeiro, R. P. and Moniz, N.: Imbalanced regression and extreme value prediction, Machine Learning, 109, 1803–1835, https://doi.org/10.1007/s10994-020-05900-9, 2020.

Rossum, G. v. and Drake, F. L.: The Python language reference, no. Pt. 2 in Python documentation manual / Guido van Rossum; Fred L. Drake [ed.], Python Software Foundation, Hampton, NH, release 3.0.1 [repr.] edn., 2010.

Roy, B.: Optimum machine learning algorithm selection for forecasting vegetation indices: MODIS NDVI & EVI, Remote Sensing Applications: Society and Environment, 23, 100 582, https://doi.org/10.1016/j.rsase.2021.100582, 2021.

Sarafanov, M., Kazakov, E., Nikitin, N. O., and Kalyuzhnaya, A. V.: A Machine Learning Approach for Remote Sensing Data Gap-Filling with Open-Source Implementation: An Example Regarding Land Surface Temperature, Surface Albedo and NDVI, Remote Sensing, 12, 3865, https://doi.org/10.3390/rs12233865, 2020.

Schneider, T., Lan, S., Stuart, A., and Teixeira, J.: Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations, Geophysical Research Letters, 44, https://doi.org/10.1002/2017GL076101, 2017.

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., Litvak, M., Ogle, K., Shaw, J. D., Wolf, A., Huntzinger, D. N., Schaefer, K., Cook, R., Wei, Y., Fang, Y., Hayes, D., Huang, M., Jain, A., and Tian, H.: Global patterns of drought recovery, Nature, 548, 202–205, https://doi.org/10.1038/nature23021, 2017.

Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., Kargar, K., Mosavi, A., Nabipour, N., and Chau, K.-W.: Predicting Standardized Streamflow index for hydrological drought using machine learning models, Engineering Applications of Computational Fluid Mechanics, 14, 339–350, https://doi.org/10.1080/19942060.2020.1715844, 2020.

Sharifi, A.: Yield prediction with machine learning algorithms and satellite images, Journal of the Science of Food and Agriculture, 101, 891–896, https://doi.org/10.1002/jsfa.10696, 2021.

Shen, R., Huang, A., Li, B., and Guo, J.: Construction of a drought monitoring model using deep learning based on multi-source remote sensing data, International Journal of Applied Earth Observation and Geoinformation, 79, 48–57, https://doi.org/10.1016/j.jag.2019.03.006, 2019.

Sicard, P., Agathokleous, E., Araminiene, V., Carrari, E., Hoshika, Y., De Marco, A., and Paoletti, E.: Should we see urban trees as effective solutions to reduce increasing ozone levels in cities?, Environmental Pollution, 243, 163–176, https://doi.org/10.1016/j.envpol.2018.08.049, 2018a.

Sicard, P., Agathokleous, E., Araminiene, V., Carrari, E., Hoshika, Y., De Marco, A., and Paoletti, E.: Should we see urban trees as effective solutions to reduce increasing ozone levels in cities?, Environmental Pollution, 243, 163–176, https://doi.org/10.1016/j.envpol.2018.08.049, 2018b.

Singer, M. B., Asfaw, D. T., Rosolem, R., Cuthbert, M. O., Miralles, D. G., MacLeod, D., Quichimbo, E. A., and Michaelides, K.: Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981-present, Scientific Data, 8, 224, https://doi.org/10.1038/s41597-021-01003-9, bandiera_abtest: a Cc_license_type: cc_publicdomain Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Hydrology Subject_term_id: hydrology, 2021.

Smith, N. E., Kooijmans, L. M. J., Koren, G., van Schaik, E., van der Woude, A. M., Wanders, N., Ramonet, M., Xueref-Remy, I., Siebicke, L., Manca, G., Brümmer, C., Baker, I. T., Haynes, K. D., Luijkx, I. T., and Peters, W.: Spring enhancement and summer reduction in carbon uptake during the 2018 drought in northwestern Europe, Philosophical Transactions of the Royal Society B: Biological Sciences, 375, 20190 509, https://doi.org/10.1098/rstb.2019.0509, 2020.

Sun, M., Gong, A., Zhao, X., Liu, N., Si, L., and Zhao, S.: Reconstruction of a Monthly 1 km NDVI Time Series Product in China Using Random Forest Methodology, Remote Sensing, 15, 3353, https://doi.org/10.3390/rs15133353, 2023.

Sutanto, S. J., van der Weert, M., Wanders, N., Blauhut, V., and Van Lanen, H. A. J.: Moving from drought hazard to impact forecasts, Nature Communications, 10, 4945, https://doi.org/10.1038/s41467-019-12840-z, 2019.

Tang, L., Chen, X., Cai, X., and Li, J.: Disentangling the roles of land-use-related drivers on vegetation greenness across China, Environmental Research Letters, 16, 124 033, https://doi.org/10.1088/1748-9326/ac37d2, 2021.

Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, Geographical Review, 38, 55–94, https://doi.org/10.2307/210739, 1948.

Tufaner, F. and Özbeyaz, A.: Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms, Environmental Monitoring and Assessment, 192, 576, https://doi.org/10.1007/s10661-020-08539-0, 2020.

Vereinte Nationen, ed.: Special report on drought 2021, no. 2021 in Global assessment report on disaster risk reduction, United Nations Office for Desaster Risk Reduction, Geneva, 2021.

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, 2010.

Vogt, J. V., Naumann, G., Masante, D., Spinoni, J., Cammalleri, C., Erian, W., Pischke, F., Pulwarty, R., and Barbosa, P.: Drought risk assessment and management: a conceptual framework., Publications Office, LU, https://data.europa.eu/doi/10.2760/057223, 2018.

Wanders, N. and Wada, Y.: Human and climate impacts on the 21st century hydrological drought, Journal of Hydrology, 526, 208–220, https://doi.org/10.1016/j.jhydrol.2014.10.047, 2015.

Wang, H., Seaborn, T., Wang, Z., Caudill, C. C., and Link, T. E.: Modeling tree canopy height using machine learning over mixed vegetation landscapes, International Journal of Applied Earth Observation and Geoinformation, 101, 102 353, https://doi.org/10.1016/j.jag.2021.102353, 2021.

Wang, Q., Wang, L., Zhu, X., Ge, Y., Tong, X., and Atkinson, P. M.: Remote sensing image gap filling based on spatial-spectral random forests, Science of Remote Sensing, 5, 100 048, https://doi.org/10.1016/j.srs.2022.100048, 2022.

Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., Emberson, R., Jenkins, S. F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G. A., Schäfer, A., Stanley, T., Veldkamp, T. I. E., and Winsemius, H. C.: Review article: Natural hazard risk assessments at the global scale, Natural Hazards and Earth System Sciences, 20, 1069–1096, https://doi.org/10.5194/nhess-20-1069-2020, 2020.

West, H., Quinn, N., and Horswell, M.: Remote sensing for drought monitoring & impact assessment: Progress, past challenges and future opportunities, Remote Sensing of Environment, 232, 111 291, https://doi.org/10.1016/j.rse.2019.111291, 2019.

Wu, Q.: geemap: A Python package for interactive mapping with Google Earth Engine, Journal of Open Source Software, 5, 2305, https://doi.org/10.21105/joss.02305, 2020.

Xiong, C., Ma, H., Liang, S., He, T., Zhang, Y., Zhang, G., and Xu, J.: Improved global 250 m 8-day NDVI and EVI products from 2000–2021 using the LSTM model, Scientific Data, 10, 800, https://doi.org/10.1038/s41597-023-02695-x, 2023.

Xu, Z., Zhou, G., and Shimizu, H.: Plant responses to drought and rewatering, Plant Signaling & Behavior, 5, 649–654, https://doi.org/10.4161/psb.5.6.11398, 2010.

Yamazaki, D., Ikeshima, Daiki, Sosa, Jeison, Allen, George H., Bates, Paul D., and Pavelsky, Tamlin M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resources Research, 55, 5053–5073, 2019.

Zeng, C., Shen, H., and Zhang, L.: Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis and a regularization method, Remote Sensing of Environment, 131, 182–194, https://doi.org/10.1016/j.rse.2012.12.012, 2013.

Zhang, J., Liu, K., and Wang, M.: Downscaling Groundwater Storage Data in China to a 1-km Resolution Using Machine Learning Methods, Remote Sensing, 13, 523, https://doi.org/10.3390/rs13030523, 2021a.

Zhang, L., Zeng, Y., Zhuang, R., Szabó, B., Manfreda, S., Han, Q., and Su, Z.: In Situ Observation-Constrained Global Surface Soil Moisture Using Random Forest Model, Remote Sensing, 13, 4893, https://doi.org/10.3390/rs13234893, 2021b.

Zhang, X., Friedl, M. A., Schaaf, C. B., and Strahler, A. H.: Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data: CLIMATE CONTROLS ON VEGETATION PHENOLOGICAL PATTERNS, Global Change Biology, 10, 1133–1145, https://doi.org/10.1111/j.1529-8817.2003.00784.x, 2004.

Zhang, Y., Keenan, T. F., and Zhou, S.: Exacerbated drought impacts on global ecosystems due to structural overshoot, Nature Ecology & Evolution, 5, 1490–1498, https://doi.org/10.1038/s41559-021-01551-8, 2021c.

Zhao, W. and Duan, S.-B.: Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data, Remote Sensing of Environment, 247, 111 931, https://doi.org/10.1016/j.rse.2020.111931, 2020.

Zhu, S., Clement, R., McCalmont, J., Davies, C. A., and Hill, T.: Stable gap-filling for longer eddy covariance data gaps: A globally validated machine-learning approach for carbon dioxide, water, and energy fluxes, Agricultural and Forest Meteorology, 314, 108 777, https://doi.org/10.1016/j.agrformet.2021.108777, 2022.

Zhu, Z.: Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications, ISPRS Journal of Photogrammetry and Remote Sensing, 130, 370–384, https://doi.org/10.1016/j.isprsjprs.2017.06.013, 2017.