EarthNet2021: A novel large-scale dataset and challenge for forecasting localized climate impacts.
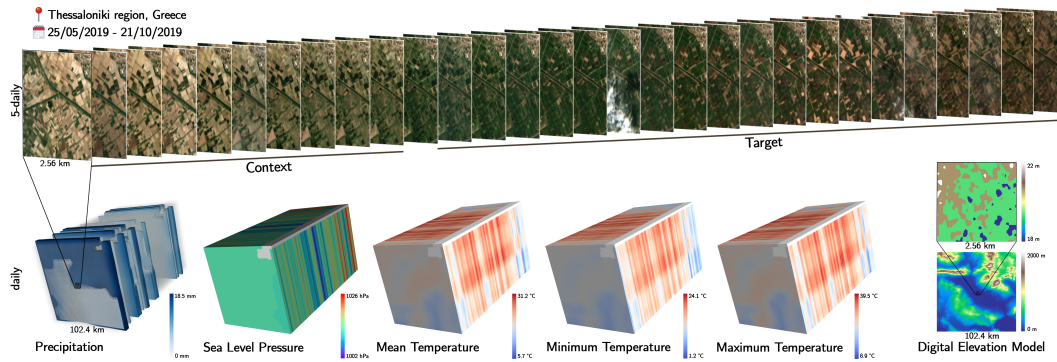

Figure 1: The EarthNet2021 dataset, an overview visualization of a single sample.

Christian Requena-Mesa[1,2,3,*]   **Vitus Benson**[1,*]

Joachim Denzler[2,4]   **Jakob Runge**[3]   **Markus Reichstein**[1,4]

[1] Department Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena
[2] Computer Vision Group, Computer Science, FSU Jena
[3] Institute of Data Science, German Aerospace Center (DLR), Jena
[4] Michael-Stifel-Center Jena for Data-driven and Simulation Science
[*] Joint first authors. E-Mail: {crequ, vbenson}@bgc-jena.mpg.de

Abstract

Climate change is global, yet its concrete impacts can strongly vary between different locations in the same region. Regional climate projections and seasonal weather forecasts currently operate at the mesoscale (> 1 km). For more targeted mitigation and adaptation, modelling impacts to < 100 m is needed. Yet, the relationship between driving variables and Earth's surface at such local scales remains unresolved by current physical models and is partly unknown; hence, it is a source of considerable uncertainty. Large Earth observation datasets now enable us to create machine learning models capable of translating coarse weather information into high-resolution Earth surface forecasts encompassing localized climate impacts.Here, we define high-resolution Earth surface forecasting as video prediction of satellite imagery conditional on mesoscale weather forecasts. Video prediction has been tackled with deep learning models. Developing such models requires analysis-ready datasets. We introduce **EarthNet2021**, a new, curated dataset containing target spatio-temporal Sentinel 2 satellite imagery at 20 m resolution, matched with high-resolution topography and mesoscale (1.28 km) weather variables. With over 32000 samples it is suitable for training deep neural networks. Comparing multiple Earth surface forecasts is not trivial. Hence, we define the EarthNetScore, a novel ranking criterion for models forecasting Earth surface reflectance. For model intercomparison we frame EarthNet2021 as a challenge with four tracks based on different test sets. These allow evaluation of model validity and robustness as well as model applicability to extreme events and the complete annual vegetation cycle. In addition to forecasting directly observable weather impacts through satellite-derived vegetation indices, capable Earth surface models will enable downstream applications such as crop yield prediction, forest health assessments, coastline management, or biodiversity monitoring. Find data, code, and how to participate at `www.earthnet.tech`.
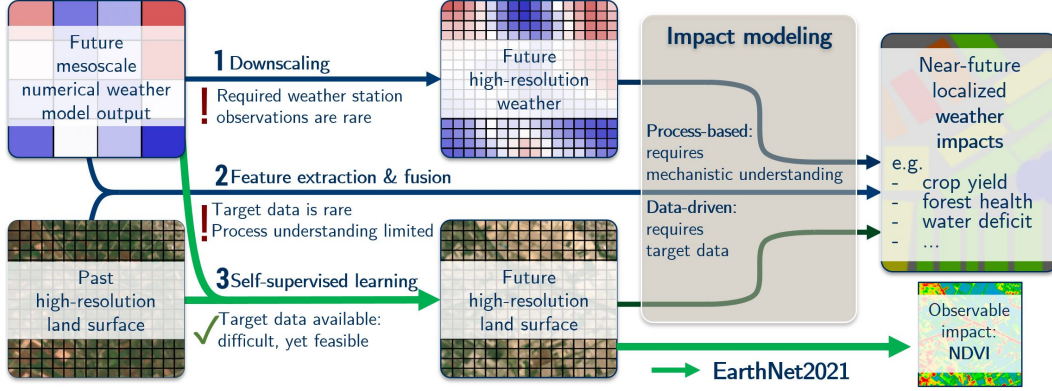
Figure 2: Predicting localized climate impacts can be done in a variety of ways. EarthNet2021 places Earth surface prediction as an intermediate task (green path). This gives directly obtainable impacts (NDVI) and enables further analyses to extract impacts from future Earth surface reflectance (lower dark green path in gray box). Compared to directly modeling any given impacts, where target labels are scarce, large amounts of satellite imagery are available. Hence, localized prediction of climate impacts becomes feasible since self-supervised deep learning can be leveraged.

1 Motivation

The terrestrial surface of Earth is home to most of the planet's species and houses human economical and societal systems. As already noticed by Alexander von Humboldt, climate is a key factor shaping vegetation cover and soils on Earth. Yet, importantly, the impact of the climatic drivers onto the surface is highly modulated by the fine-grained local conditions, such as geomorphology, geological substrate and vegetation and animals themselves. In particular extreme events can have very heterogeneous impacts at the local scale depending on the conditions (Kogan, 1990). For example, ecosystems next to a river might survive droughts better than those on south-facing slopes. However, the resolution of seasonal weather predictions is not fine enough to deploy effective prevention and mitigation strategies. Machine learning (Reichstein et al., 2019; Rolnick et al., 2019) can step in to increase resolution.

Predicting localized climate impacts can be tackled in three main ways (see Fig. 2). All approaches make use of seasonal weather forecasts (2 – 6 months ahead; Cantelaube & Terres, 2005). The first approach (Fig. 2, path 1), aims to reconstruct hyper-resolution weather forecasts for particular geolocations using statistical (Boé et al., 2006; Vrac et al., 2007) or dynamical (Lo et al., 2008) downscaling, that is, correlating the past observed weather with past mesoscale model outputs and using the estimated relationship. The downscaled weather can then be used in mechanistic models (e.g. of river discharge) for impact extraction. However, weather downscaling is a difficult task because it requires ground observations from weather stations, which are sparse.

A more direct way (Fig. 2, path 2) is to correlate a desired future impact variable with past data. Doing this for tangible impacts such as on crop yield (Peng et al., 2018) again suffers from a lack of data. An alternative is to forecast impacts obtainable from remote sensing, such as the normalized differenced vegetation index (NDVI), yet this has only been done at coarse resolution (Tadesse et al., 2010; Asoka & Mishra, 2015; Kraft et al., 2019; Foley et al., 2020).

Instead, we propose *Earth surface forecasting* as video prediction of satellite imagery with guidance of mesoscale weather projections for forecasting localized weather impacts (Fig. 2, path 3). From satellite imagery, we can directly observe NDVI and thus climate impacts on vegetation. Additionally, it can also be used to extract further processed weather impact data products, such as the biodiversity state (Fauvel et al., 2020), crop yields (Schwalbert et al., 2020), soil moisture (Efremova et al., 2019), or ground biomass (Ploton et al., 2017). We believe Earth surface forecasting is feasible since numerous studies suggest predicting satellite imagery works under a range of specific conditions (Zhu et al., 2015; Das & Ghosh, 2016; Hong et al., 2017; Requena-Mesa et al., 2019; Lee et al., 2019).

EarthNet2021 aims at providing analysis-ready data and a benchmark challenge for model intercomparison on the broad area of Europe to accelerate model development.
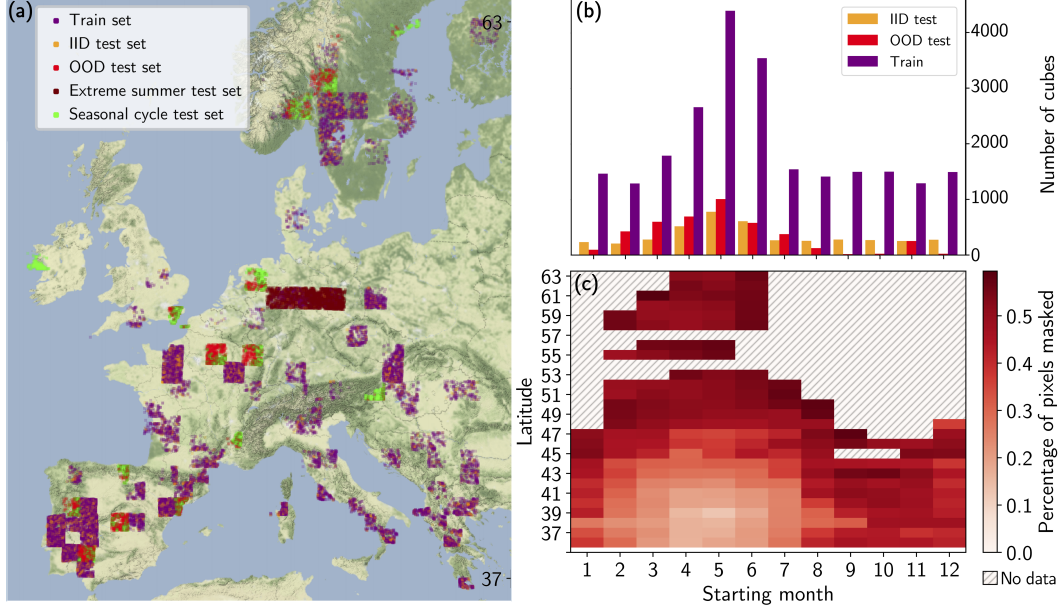
Figure 3: Panel (a) shows the spatial distribution of the multicubes in EarthNet2021. Panel (b) shows the monthly number of multicubes and panel (c) shows the data quality measured by the percentage of masked (mainly cloudy) pixels over both, months and latitude.

2 Data

The EarthNet2021 dataset combines three publicly available EU-funded data sources into spatio-temporal data samples for training deep learning models. Each sample, which we call a data multicube, is based on a timeseries of Sentinel 2 level 2A imagery (Louis et al., 2016) combined with a timeseries of daily climatic conditions from E-OBS (Cornes et al., 2018) and the EU-DEM digital surface model (Bashfield & Keim, 2011). Training deep learning models with raw geospatial data is usually not possible and there is need for analysis-ready datasets. An overview of the dataset generation pipeline, turning the raw geospatial data into the analysis-ready EarthNet2021 dataset, is shown in Fig. 4.
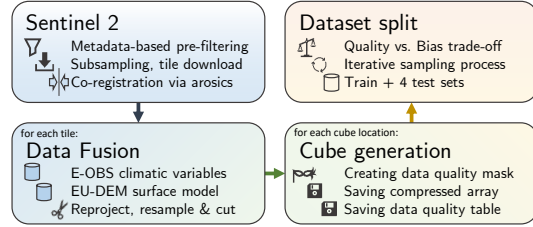

Figure 4: This dataset generation scheme roughly describes how we obtain the analysis-ready EarthNet2021 dataset from raw, public geospatial datasets.

After data processing, EarthNet2021 contains over 30000 samples, which we call *data multicubes*. A single multicube is visualized in Fig. 1. It contains 30 5-daily frames (128×128 pixel or 2.56×2.56 km) of four channels (blue, green, red, near-infrared) of satellite imagery with binary quality masks at high-resolution (20 m), 150 daily frames (80×80 pixel or 102.4×102.4 km) of five dynamic climatic variables (precipitation, sea level pressure, mean, minimum and maximum temperature) at mesoscale resolution (1.28 km) and a static digital elevation model at both high- and mesoscale resolution.

The entirety of the multicubes have been split across the training set and various test sets, which are related to different tracks in the EarthNet2021 challenge (sec. 3). EarthNet2021 is an imbalanced dataset, as during the data generation there is a direct trade-off between high data quality and low selection bias. For example, high-quality (cloud-free) samples are mostly found during summer on the Iberian Peninsula, whereas there are few consecutive weeks without clouds on the British Islands. In Fig. 3 we try to make some of the selection bias visible by showing the spatial and temporal distribution of the multicubes in EarthNet2021 across the different sets.

3

3 Challenge

The EarthNet2021 challenge aims at Earth surface forecasting model intercomparison. Due to its novelty, there is not yet a commonly used criterion for Earth surface predictions.

EarthNetScore. Specifically for Earth surface forecasting, we define the *EarthNetScore* as a ranking criterion balancing multiple goals in a harmonic mean as follows:

$$\text{ENS} = \frac{4}{\left(\frac{1}{\text{MAD}} + \frac{1}{\text{OLS}} + \frac{1}{\text{EMD}} + \frac{1}{\text{SSIM}}\right)}. \quad (1)$$

The four components of ENS are the median absolute deviation MAD; the difference of ordinary least squares linear regression slopes of pixelwise NDVI timeseries OLS; the Earth mover distance EMD between pixelwise NDVI time series and the structural similarity index SSIM. All component scores are modified to work properly in the presence of a data quality mask, normalized to lie between 0 (worst) and 1 (best) and rescaled to match difficulty. Since Earth surface forecasting is a stochastic task models may predict multiple future trajectories. Over a full test set, we aggregate these by only considering the best predicted trajectory per multicube (in line with video prediction common practice). Then we average the component scores of these and calculate the ENS by feeding the averages to eq. 1. Thus, the ENS ranges from 0 (bad) to 1 (perfect).

Figure 5: Evaluation pipeline for models on Earth-Net2021. For predicting the t_T target frames, a model can use satellite images from the t_C context frames, the static DEM and mesoscale climatic variables including those from the target time steps.

Test set	ENS	MAD	OLS	EMD	SSIM
IID	0.26	0.23	0.32	0.21	0.33
OOD	0.25	0.22	0.32	0.21	0.31
Extreme	0.19	0.22	0.28	0.16	0.16
Seasonal	0.27	0.23	0.38	0.20	0.32

Table 1: Persistence baseline performance.

Tracks. Multiple models are compared within the EarthNet2021 challenge by measuring their EarthNetScores on various tracks (see Fig. 5). The *main (IID) track* checks model validity. Models get 10 context frames of high resolution 5-daily multispectral satellite imagery (time [t-45, t]), mesoscale dynamic climate conditions for in total 150 past and future days (time [t-50, t+100]) and static topography at both resolutions. Models shall output 20 frames of sentinel 2 bands red, green, blue and near-infrared for the next 100 days (time [t+5, t+100]). These predictions are evaluated with the EarthNetScore on unmasked (cloud-free) pixels from the ground truth. The *Robustness (OOD) track* checks model performance on an out-of-domain (OOD) test set, since even on the same satellite data, deep learning models might generalize poorly across geolocations (Benson & Ecker, 2020). Furthermore, EarthNet2021 contains two tracks focused on Earth system science hot topics, which should both be understood as more experimental. The *extreme summer track* contains cubes from the extreme summer 2018 in northern Germany (Bastos et al., 2020), with 4 months of context (20 frames) starting from February and 6 months (40 frames) starting from June to evaluate predictions. The *seasonal cycle track* contains multicubes with 1 year (70 frames) of context frames and 2 years (140 frames) to evaluate predictions, thus checking models applicability to the vegetation cycle.

EarthNet2021 Framework. To facilitate research we provide the EarthNet2021 framework. It contains 1) the packaged evaluation pipeline as the EarthNet2021 toolkit which leverages multiprocessing for fast inference, 2) the model intercomparison suite, which gives one entry point for running a wide range of models and 3) a naive baseline (cloud-free mean, see table 1) and templates for PyTorch and Tensorflow. Further information can be found on `www.earthnet.tech`.

4 Outlook

Forecasting impacts of climate and weather on the Earth surface is a simultaneously societally important and scientifically challenging task. With the EarthNet2021 dataset, first models for Europe can be designed and the EarthNet2021 challenge offers a model intercomparison framework for identifying their strengths and limitations. We expect deep learning based video prediction models to be great starting points for solutions, in perspective allowing for high-resolution prediction of localized climate impacts.
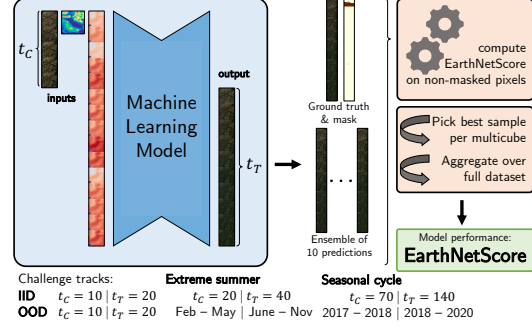
References

Akarsh Asoka and Vimal Mishra. Prediction of vegetation anomalies to improve food security and water management in india. *Geophysical Research Letters*, 42(13):5290–5298, 2015. URL `https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015GL063991`.

A Bashfield and A Keim. Continent-wide dem creation for the european union. In *34th International Symposium on Remote Sensing of Environment. The GEOSS Era: Towards Operational Environmental Monitoring. Sydney, Australia*, pp. 10–15, 2011. URL `https://www.isprs.org/proceedings/2011/ISRSE-34/211104015Final00143.pdf`.

Ana Bastos, P Ciais, P Friedlingstein, S Sitch, Julia Pongratz, L Fan, JP Wigneron, Ulrich Weber, Markus Reichstein, Z Fu, et al. Direct and seasonal legacy effects of the 2018 heat wave and drought on european ecosystem productivity. *Science advances*, 6(24):eaba2724, 2020. URL `https://advances.sciencemag.org/content/6/24/eaba2724.abstract`.

Vitus Benson and Alexander Ecker. Assessing out-of-domain generalization for robust building damage detection. *AI for Humanitarian Assistance and Disaster Response workshop (NeurIPS 2020)*, 2020. URL `https://arxiv.org/abs/2011.10328`.

J Boé, L Terray, F Habets, and E Martin. A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling. *Journal of Geophysical Research: Atmospheres*, 111(D23), 2006. URL `https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2005JD006889`.

Pierre Cantelaube and Jean-Michel Terres. Seasonal weather forecasts for crop yield modelling in europe. *Tellus A: Dynamic Meteorology and Oceanography*, 57(3):476–487, 2005. URL `https://www.tandfonline.com/doi/abs/10.3402/tellusa.v57i3.14669`.

Richard C Cornes, Gerard van der Schrier, Else JM van den Besselaar, and Philip D Jones. An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409, 2018. URL `https://doi.org/10.1029/2017JD028200`.

Monidipa Das and Soumya K Ghosh. Deep-step: A deep learning approach for spatiotemporal prediction of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 13(12):1984–1988, 2016. URL `https://ieeexplore.ieee.org/abstract/document/7752890`.

Natalia Efremova, Dmitry Zausaev, and Gleb Antipov. Prediction of soil moisture content based on satellite data and sequence-to-sequence networks. *NeurIPS 2018 Women in Machine Learning workshop*, 2019. URL `https://arxiv.org/abs/1907.03697`.

Mathieu Fauvel, Mailys Lopes, Titouan Dubo, Justine Rivers-Moore, Pierre-Louis Frison, Nicolas Gross, and Annie Ouin. Prediction of plant diversity in grasslands using sentinel-1 and-2 satellite image time series. *Remote Sensing of Environment*, 237:111536, 2020. URL `https://www.sciencedirect.com/science/article/pii/S0034425719305553`.

Conrad James Foley, Sagar Vaze, Mohamed El Amine Seddiq, Alexey Unagaev, and Natalia Efremova. Smartcast: Predicting soil moisture interpolations into the future using earth observation data in a deep learning framework. *Tackling Climate Change with Machine Learning workshop at ICLR 2020*, 2020. URL `https://www.climatechange.ai/papers/iclr2020/13/paper.pdf`.

Seungkyun Hong, Seongchan Kim, Minsu Joh, and Sa-Kwang Song. Psique: Next sequence prediction of satellite images using a convolutional sequence-to-sequence network. *Workshop on Deep Learning for Physical Sciences (NeurIPS 2017)*, 2017. URL `https://arxiv.org/abs/1711.10644`.

Felix N Kogan. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of remote sensing*, 11(8):1405–1419, 1990. URL `https://www.tandfonline.com/doi/abs/10.1080/01431169008955102`.

Basil Kraft, Martin Jung, Marco Körner, Christian Requena Mesa, José Cortés, and Markus Reichstein. Identifying dynamic memory effects on vegetation state using recurrent neural networks. *Frontiers in Big Data*, 2, 2019. URL `https://www.frontiersin.org/articles/10.3389/fdata.2019.00031`.

Jae-Hyeok Lee, Sangmin S Lee, Hak Gu Kim, Sa-Kwang Song, Seongchan Kim, and Yong Man Ro. Mcsip net: Multichannel satellite image prediction via deep neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3):2212–2224, 2019. URL `https://ieeexplore.ieee.org/abstract/document/8933126`.

Jeff Chun-Fung Lo, Zong-Liang Yang, and Roger A Pielke Sr. Assessment of three dynamical climate downscaling methods using the weather research and forecasting (wrf) model. *Journal of Geophysical Research: Atmospheres*, 113(D9), 2008. URL `https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2007JD009216`.

Jérôme Louis, Vincent Debaecker, Bringfried Pflug, Magdalena Main-Knorn, Jakub Bieniarz, Uwe Mueller-Wilm, Enrico Cadau, and Ferran Gascon. Sentinel-2 sen2cor: L2a processor for users. In *Proceedings Living Planet Symposium 2016*, pp. 1–8. Spacebooks Online, 2016. URL `https://elib.dlr.de/107381/`.

Bin Peng, Kaiyu Guan, Ming Pan, and Yan Li. Benefits of seasonal climate prediction and satellite data for forecasting us maize yield. *Geophysical Research Letters*, 45(18):9662–9671, 2018. URL `https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018GL079291`.

Pierre Ploton, Nicolas Barbier, Pierre Couteron, CM Antin, Narayanan Ayyappan, N Balachandran, N Barathan, J-F Bastin, G Chuyong, Gilles Dauby, et al. Toward a general tropical forest biomass prediction model from very high resolution optical satellite images. *Remote sensing of environment*, 200:140–153, 2017. URL `https://www.sciencedirect.com/science/article/pii/S0034425717303553`.

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019. URL `https://www.nature.com/articles/s41586-019-0912-1`.

Christian Requena-Mesa, Markus Reichstein, Miguel Mahecha, Basil Kraft, and Joachim Denzler. Predicting landscapes from environmental conditions using generative networks. In *German Conference on Pattern Recognition*, pp. 203–217. Springer, 2019. URL `https://link.springer.com/chapter/10.1007/978-3-030-33676-9_14`.

David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019. URL `https://arxiv.org/abs/1906.05433`.

Raí A Schwalbert, Telmo Amado, Geomar Corassa, Luan Pierre Pott, PV Vara Prasad, and Ignacio A Ciampitti. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern brazil. *Agricultural and Forest Meteorology*, 284:107886, 2020. URL `https://www.sciencedirect.com/science/article/pii/S0168192319305027`.

Tsegaye Tadesse, Brian D Wardlow, Michael J Hayes, Mark D Svoboda, and Jesslyn F Brown. The vegetation outlook (vegout): A new method for predicting vegetation seasonal greenness. *GIScience & Remote Sensing*, 47(1):25–52, 2010. URL `https://www.tandfonline.com/doi/abs/10.2747/1548-1603.47.1.25`.

Mathieu Vrac, Michael Stein, and Katharine Hayhoe. Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Climate Research*, 34(3):169–184, 2007. URL `https://www.int-res.com/abstracts/cr/v34/n3/p169-184/`.

Zhe Zhu, Curtis E Woodcock, Christopher Holden, and Zhiqiang Yang. Generating synthetic landsat images based on all available landsat data: Predicting landsat surface reflectance at any given time. *Remote Sensing of Environment*, 162:67–83, 2015. URL `https://www.sciencedirect.com/science/article/pii/S0034425715000590`.