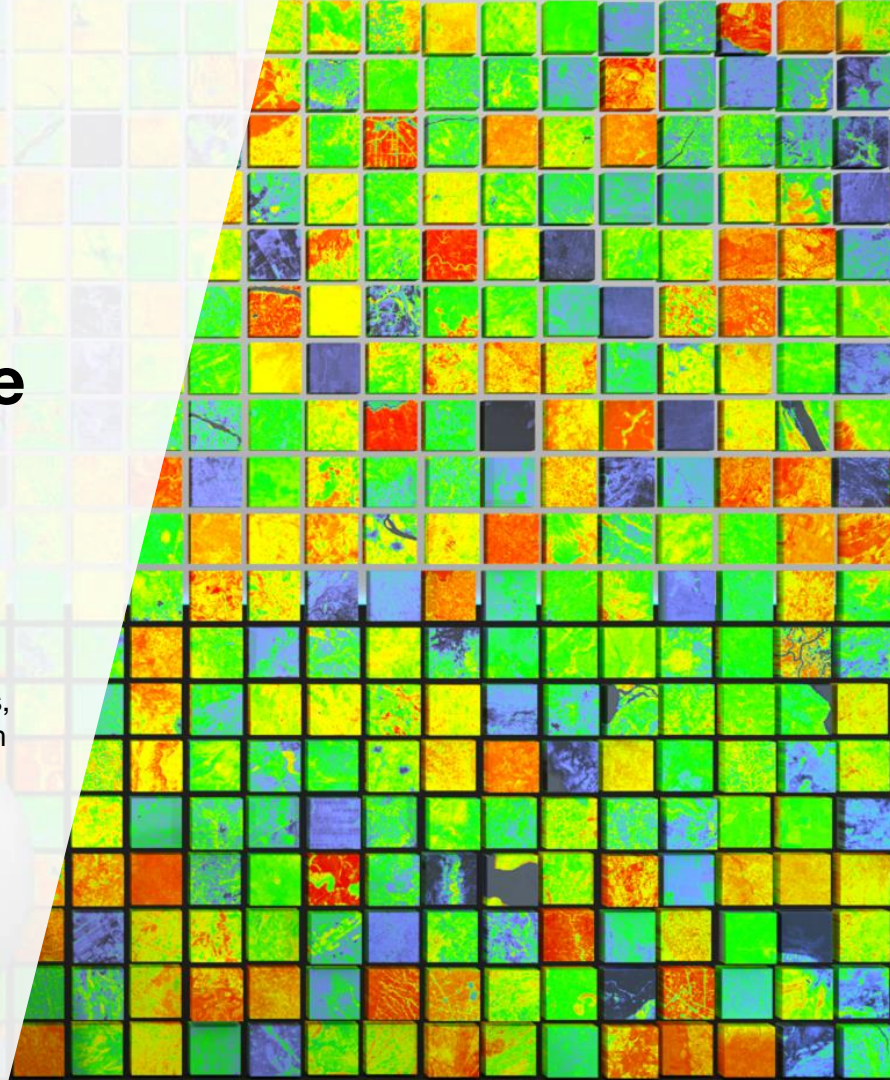
EarthNet: Bringing Machine Learning and Biogeoscience Together

Claire Robin
Max Planck Institute for Biogeochemistry, ELLIS Unit Jena, and Wageningen University

EarthNet Initiative: Vitus Benson, Lazaro Alonso, Mélanie Weynants, Markus Zehner, Marc Rußwurm, Nuno Carvalhais, Markus Reichstein

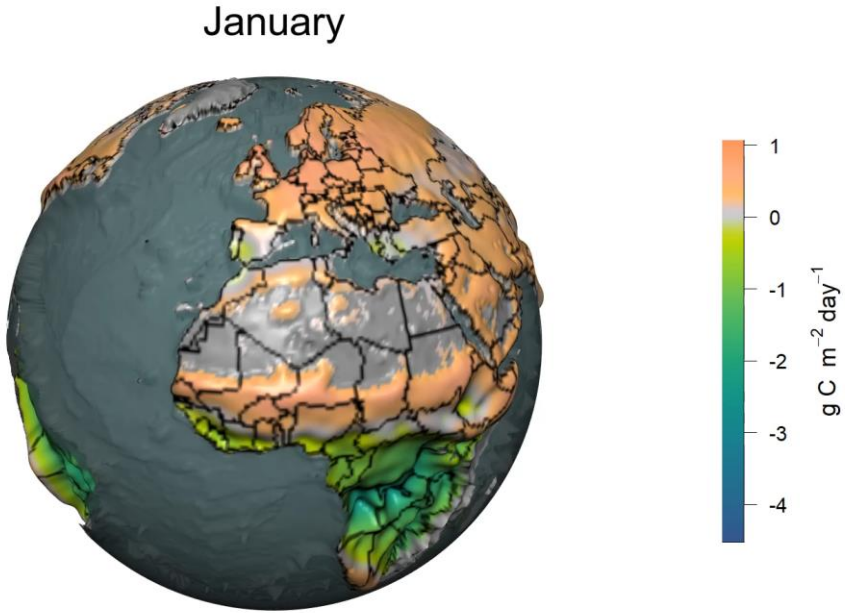
ELLIS Doctoral Symposium 2024 | 28.08.2024

crobin@bgc-jena.mpg.de



Biogeoscience

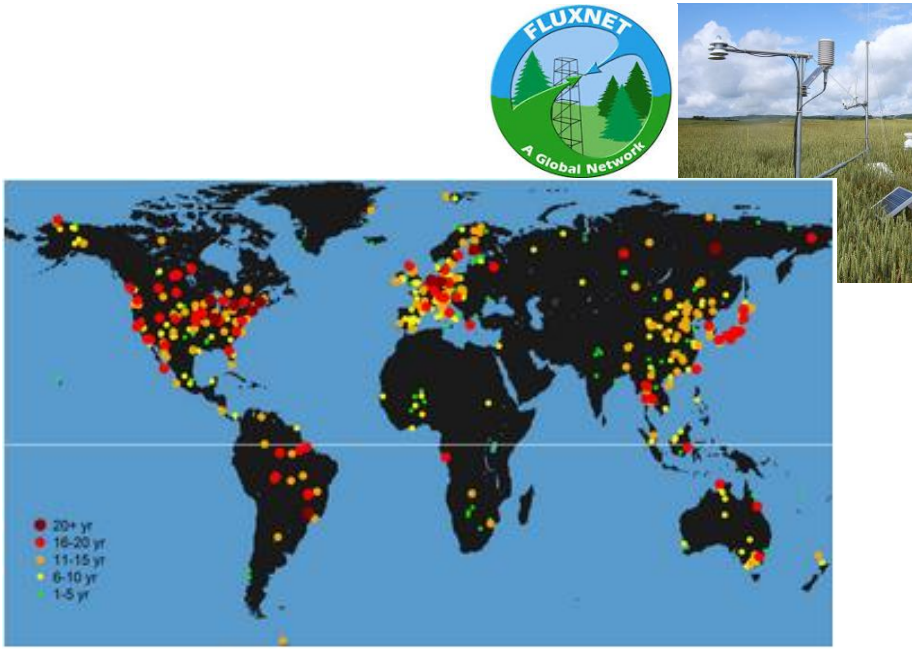
Study of global biogeochemical cycles and their long-term interactions with the biosphere, the atmosphere, the geosphere and the entire climate system.

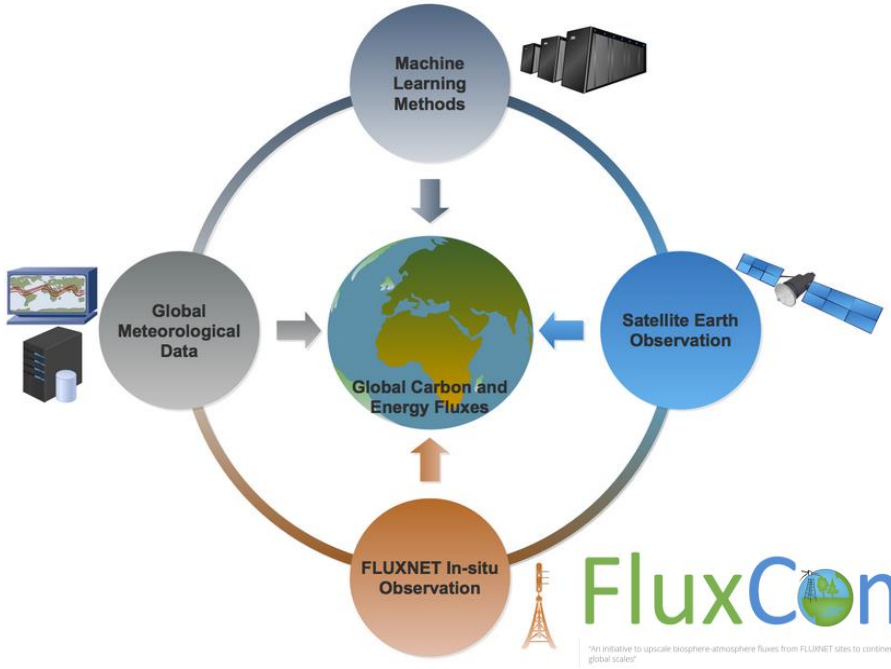


“Breathing Nature” – CO2 uptake and release by ecosystems
Data-driven, AI-based estimates

Net Ecosystem Exchange estimate (MPI-BGC)

Spatial upscaling with Machine Learning



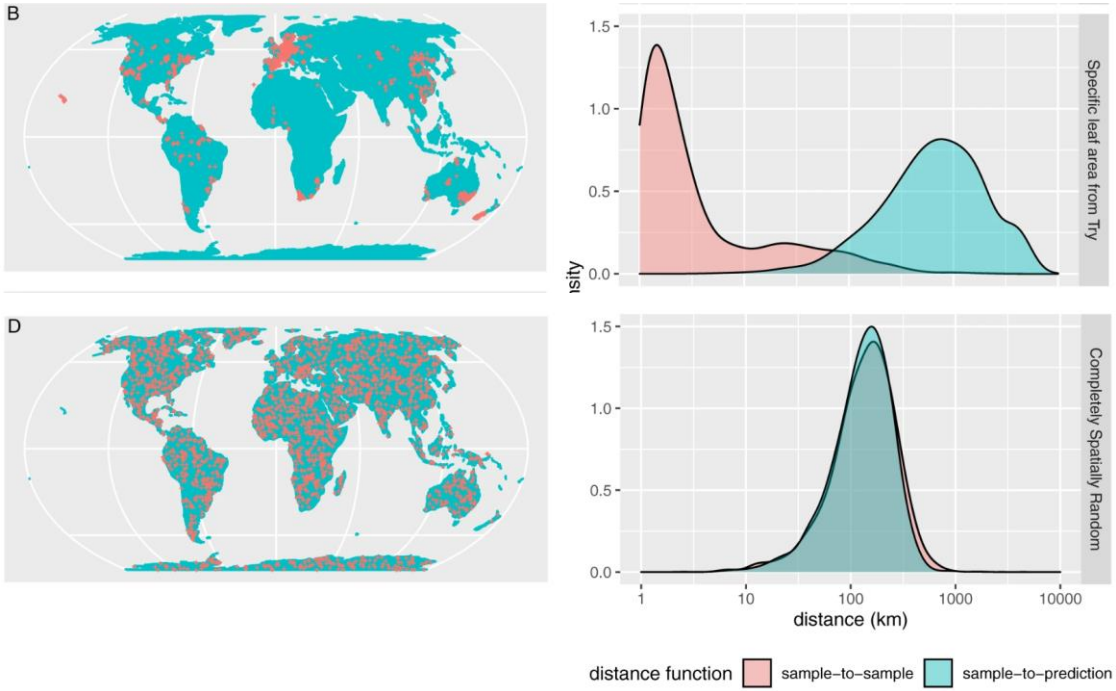


Jung, M., M. Reichstein, and Alberte Bondeau. "Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model." Biogeosciences 6.10 (2009): 2001-2013.

Jung, Martin, et al. "Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations." Journal of Geophysical Research: Biogeosciences 116.G3 (2011).

Jung, Martin, et al. "Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach." *Biogeosciences* 17.5 (2020): 1343-1365.

The Earth is not identically independently distributed (IID)



specific leaf area from the Try database

simulated completely spatially random sample

[Meyer, 2022]

[Meyer, 2022] Meyer, Hanna, and Edzer Pebesma. "Machine learning-based global maps of ecological variables and the challenge of assessing them." Nature Communications 13.1 (2022): 2208.

Sweet, Lily-belle, et al. "Cross-validation strategy impacts the performance and interpretation of machine learning models." *Artificial Intelligence for the Earth Systems* 2.4 (2023): e230026.

Rolf, Esther. "Evaluation challenges for geospatial ML." *arXiv preprint arXiv:2303.18087* (2023).

Effects of the 2018/2019 drought from space

https://www.esa.int/ESA_Multimedia/Videos/2018/07from_green_to_brown_in_a_month

Sentinel 2 Satellite Mission from ESA

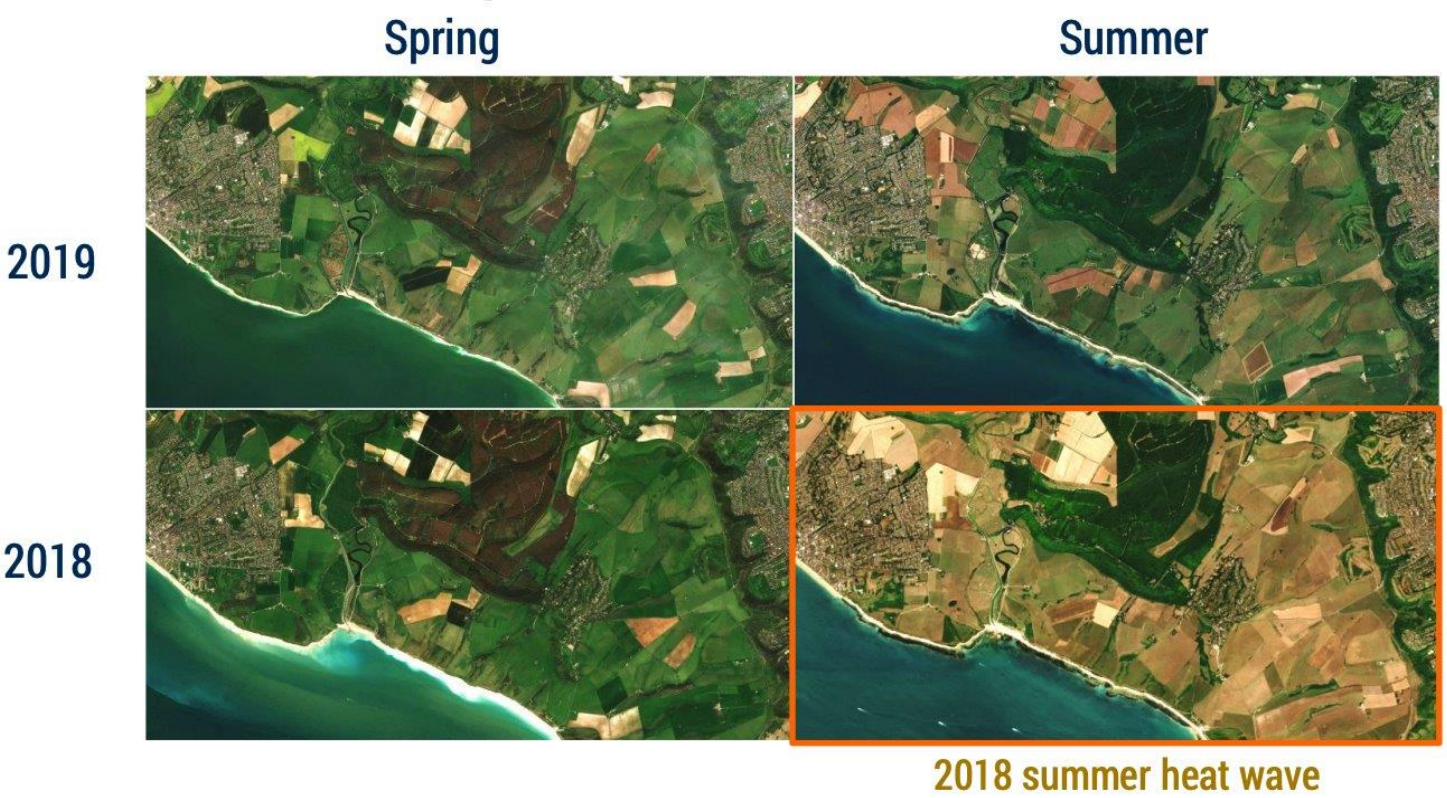
10m resolution

5-day revisit

Since 2017
7 full years

13 Bands
(Visible Light + SWIR)

Impact of droughts visible from space

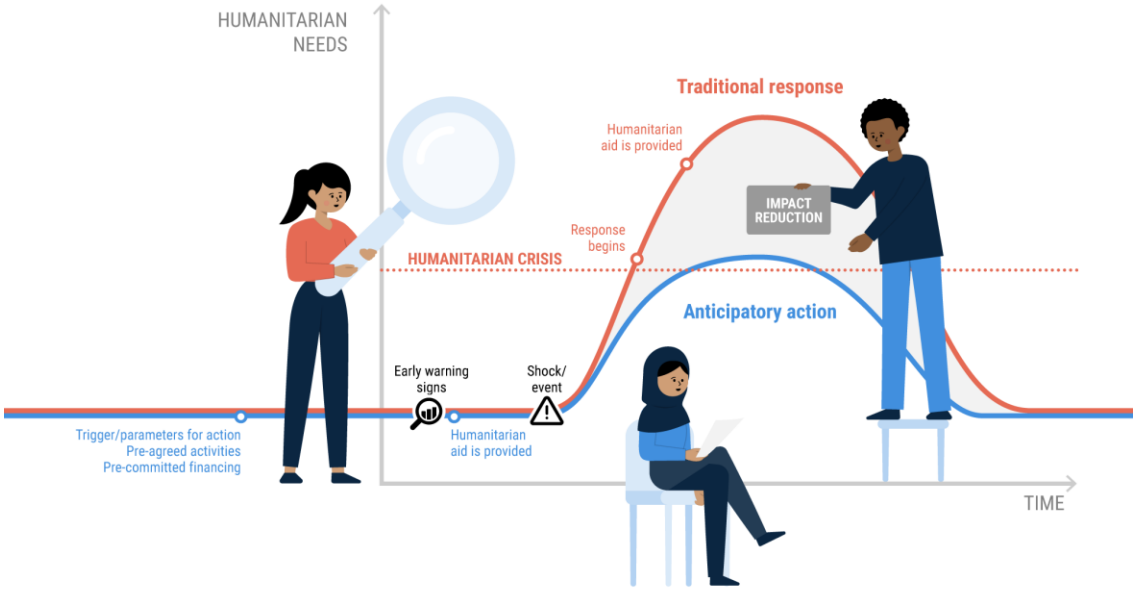


Predicting Climatic Extremes impact to support Anticipatory Action

Seasonal Weather Forecast Ensemble

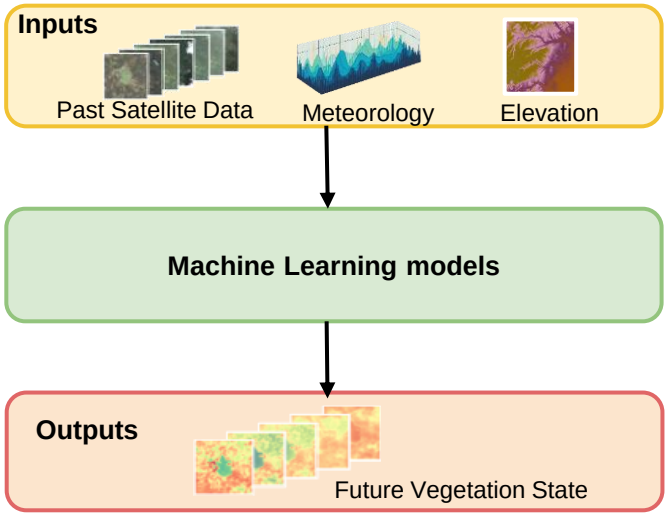
Vegetation Impact Scenarios

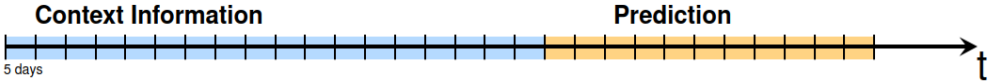
Trigger:
Community-scale NDVI < threshold (?)

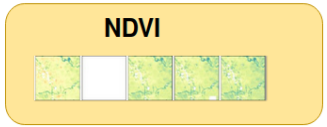


The EarthNet Modeling Framework

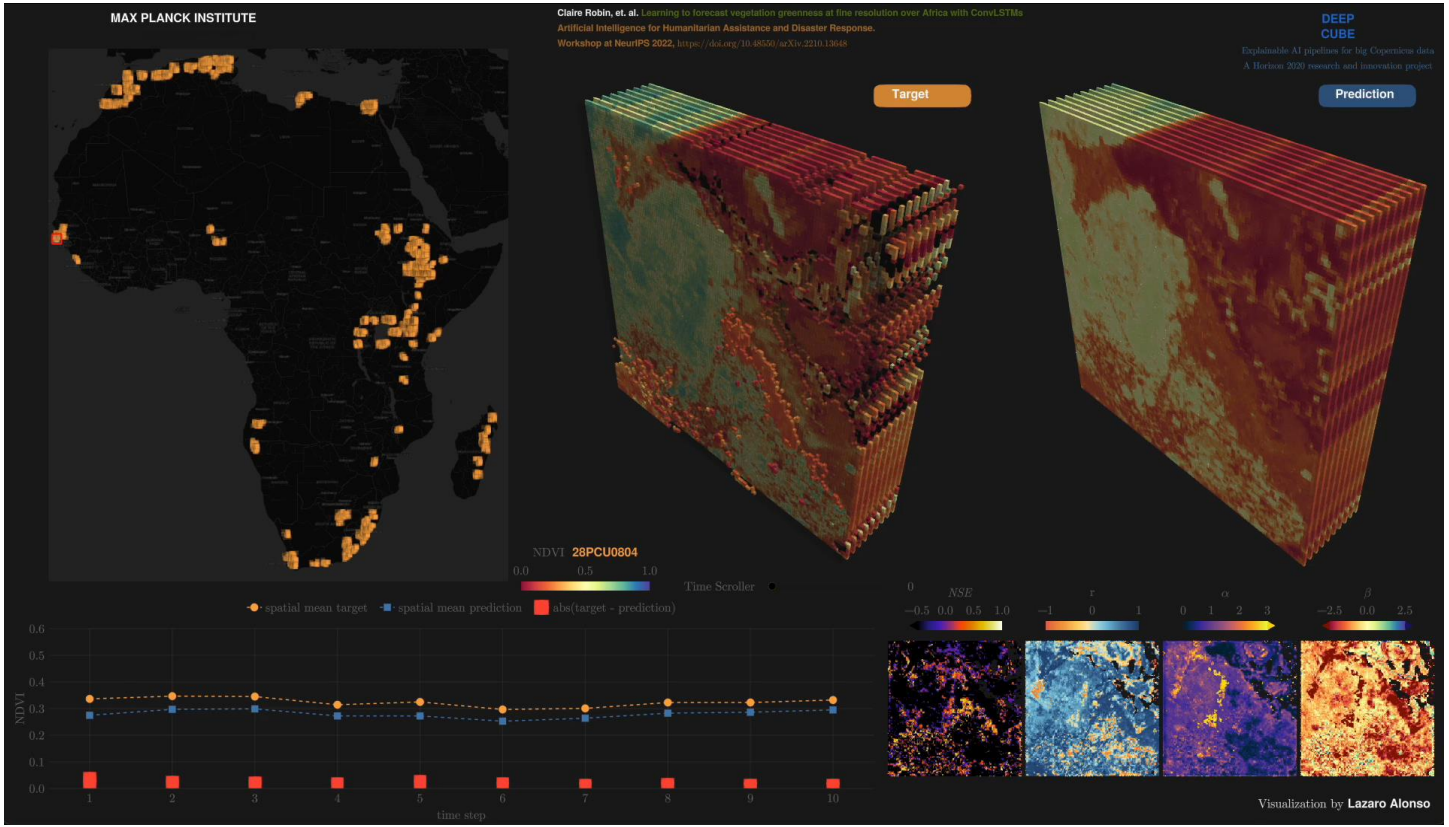
Objective: Predict the impact of climate extremes on vegetation using ML method.







Vegetation Forecasting Prediction with a ConvLSTM



EarthNet - Datasets

EarthNet2021

- First Dataset
- 25k Minicubes
- Europe
- Naïve Baselines

[Requena, 2021]

GreenEarthNet

- EarthNet2021 Version 2
- Improved Cloud Mask
- More Test sets
- Climatology Baseline

[Benson, 2024]

EarthNet2022

- 50k Minicubes
- Africa
- Focus on Droughts
- ConvLSTM Model (SOTA)

[Robin, 2022]

"DeepExtremeCubes"

- 50k Minicubes
- Global
- Compound Events: Droughts + Heatwaves

[Ji, 2024]

[Requena, 2021] Requena-Mesa, Christian et al. (2021). EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1132-1142).

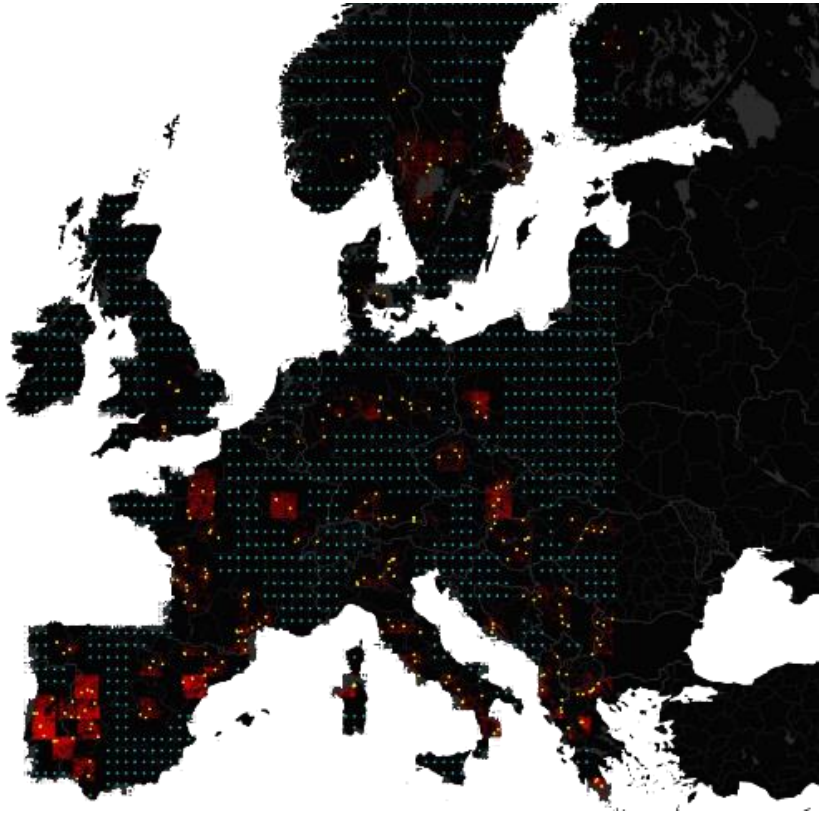
[Robin, 2022] Robin, Claire et al. "Learning to forecast vegetation greenness at fine resolution over Africa with ConvLSTMs." *arXiv preprint arXiv:2210.13648* (2022).

[Benson, 2024] Benson, Vitus et al. (2024). Multi-modal learning for geospatial vegetation forecasting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 27788-27799).

[Ji, 2024] Ji, Chaonan, et al. "DeepExtremeCubes: Integrating Earth system spatio-temporal data for impact assessment of climate extremes." Under review.

ContextFormer

New dataset GreenEarthNet tailored for geospatial vegetation forecasting



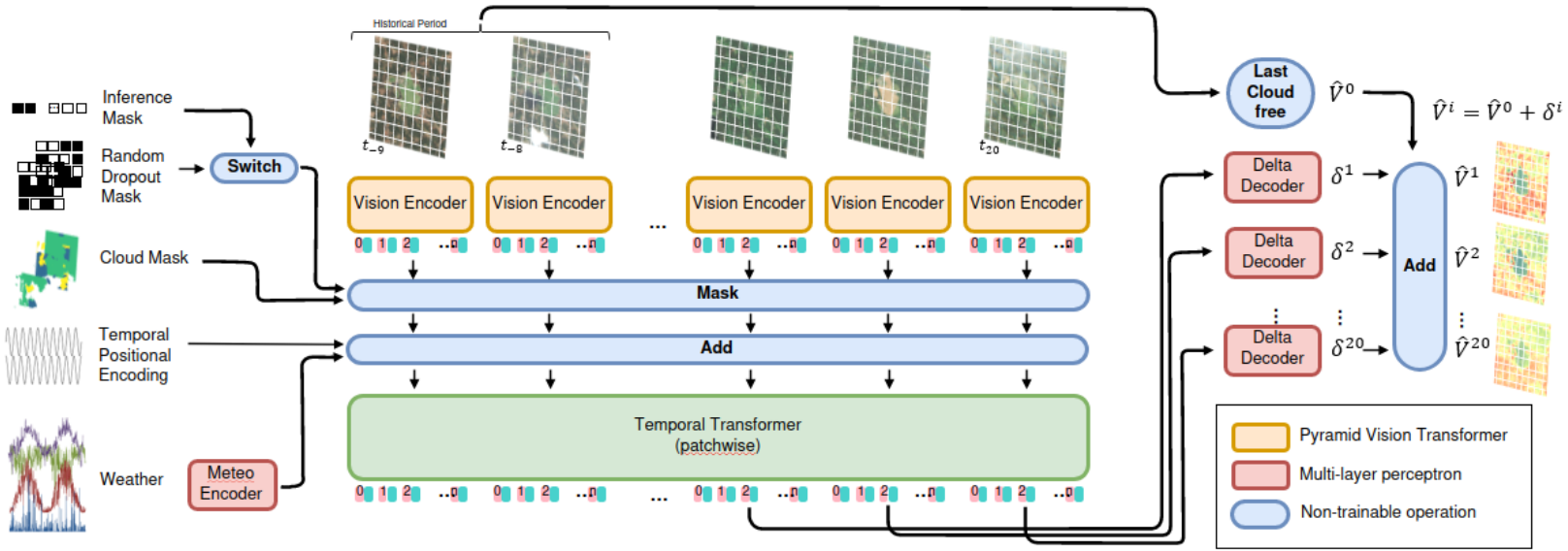
[Benson, 2024] Benson, Vitus et al. (2024). Multi-modal learning for geospatial vegetation forecasting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 27788-27799).

@vitusbenson
vbenson@bgc-jena.mpg.de

ContextFormer

New model Contextformer: **separation of spatial & temporal Transformers**

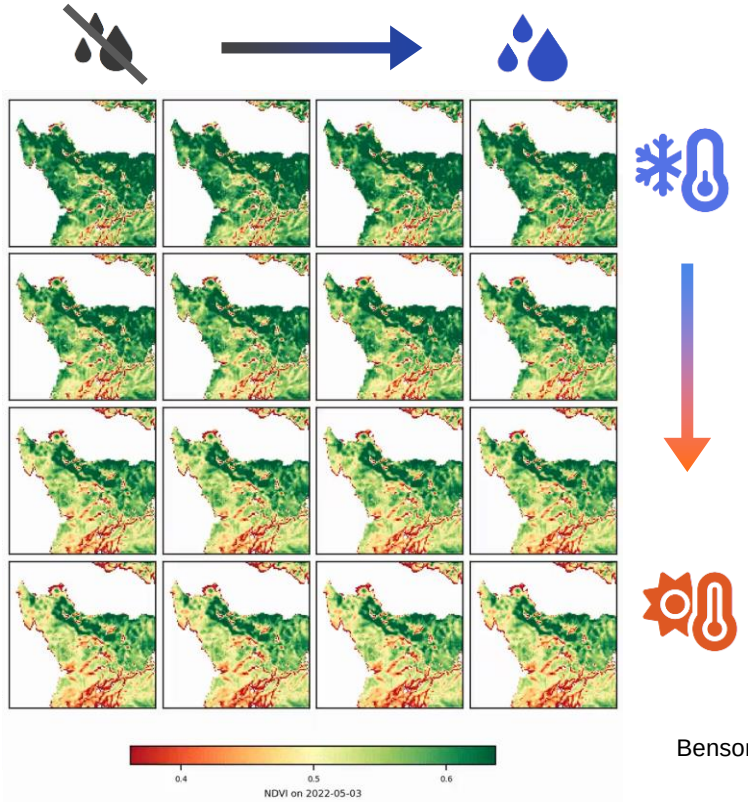
Contextformer outperforms diverse set of baseline models (incl. previous SOTA)



[Benson, 2024] Benson, Vitus et al. (2024). Multi-modal learning for geospatial vegetation forecasting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 27788-27799).

@vitusbenson
vbenson@bgc-jena.mpg.de

Worse Case Weather Scenario

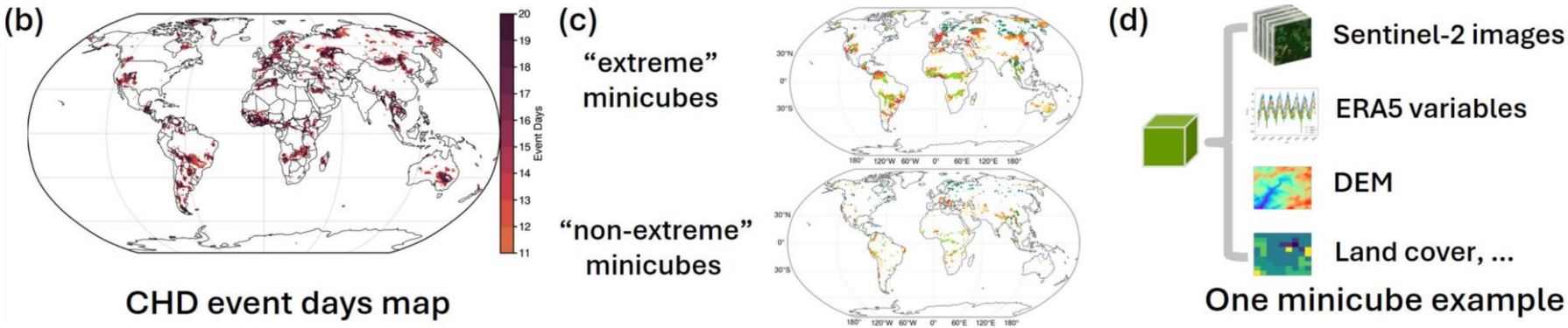


Benson, et. al. (2024)

: EarthNet Under Extremes

- Unbalanced problem
- Vegetation responses to climates extremes vary in **delay, intensity, duration**

Defining Extremes samples using Climate extremes

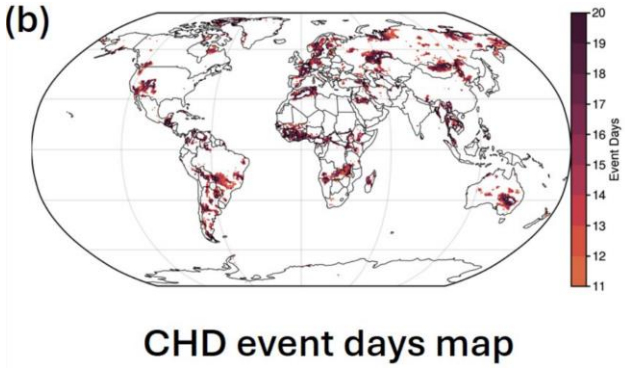


DeepExtremeCubes: Extremes minicubes are **defined using climate extremes** [Ji, 2024].

[Ji, 2024] Ji, Chaonan, et al. "DeepExtremeCubes: Integrating Earth system spatio-temporal data for impact assessment of climate extremes." Under review.

www.earthnet.tech

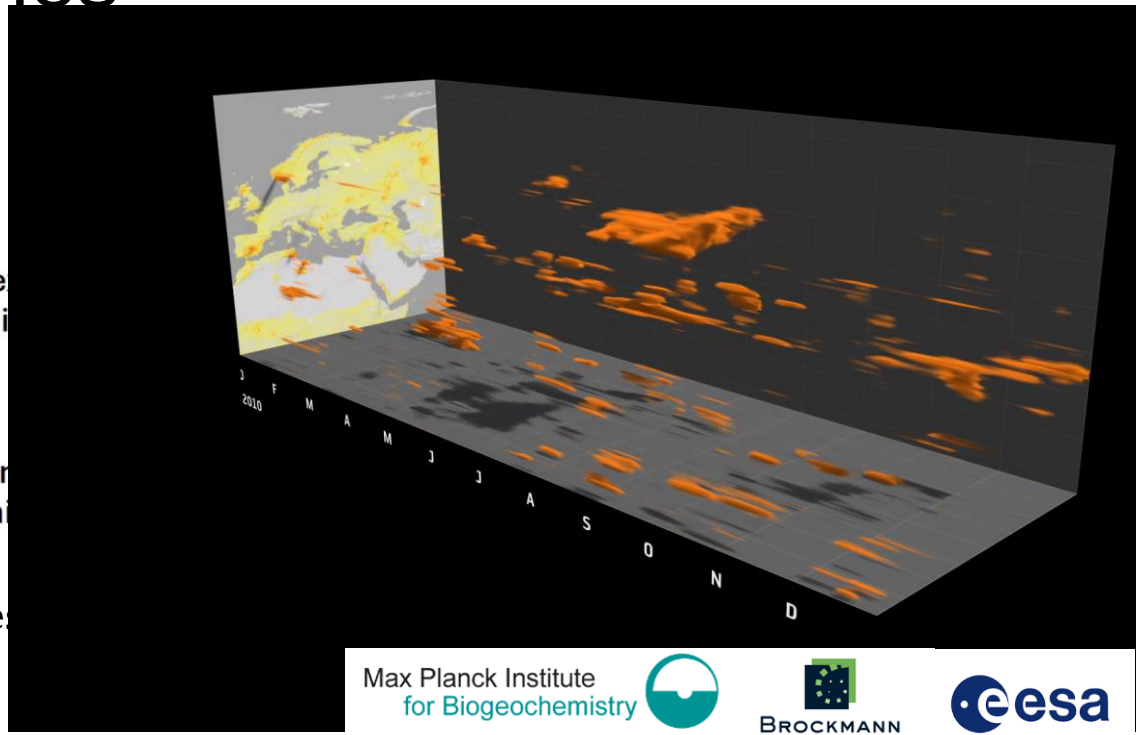
Defining Extremes samples using Climate extremes



(c)

“e
mi

“nor
mi



DeepExtremeCubes: Extremes minicube

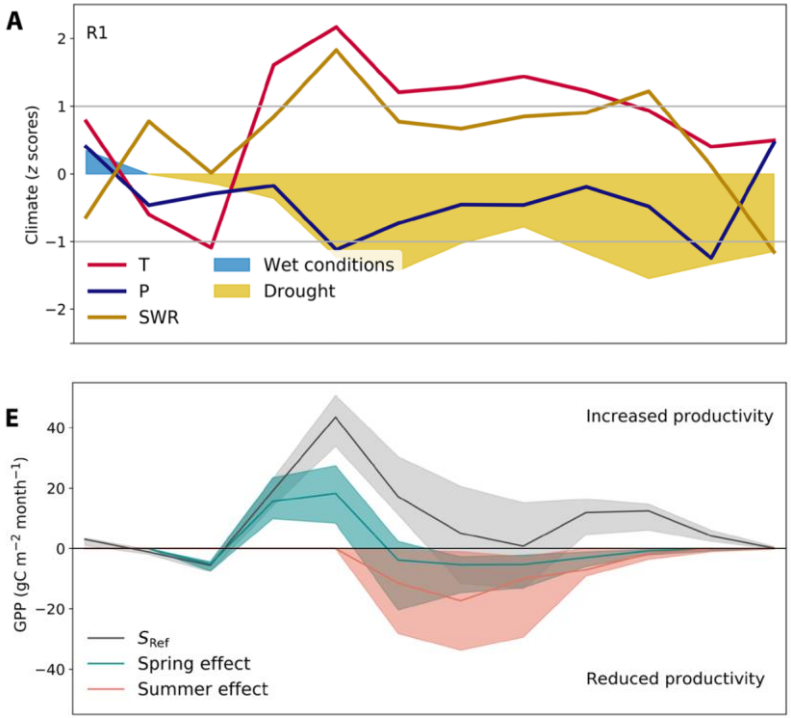
Max Planck Institute for Biogeochemistry

Brockmann Consult GmbH

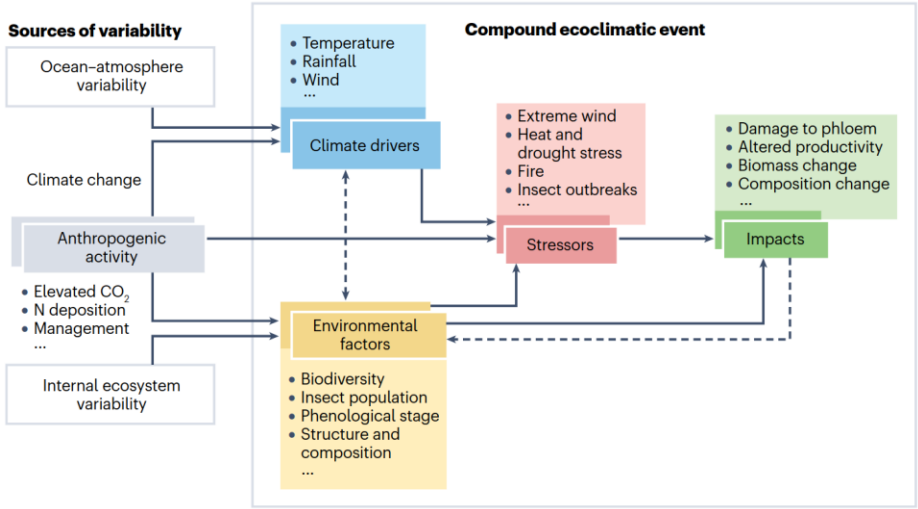
esa

[Ji, 2024] Ji, Chaonan, et al. "DeepExtremeCubes: Integrating Earth system spatio-temporal data for impact assessment of climate extremes." Under review.

Taking into account the Ecological responses



[Bastos, 2020]



[Bastos, 2023]

[Bastos, 2020] Bastos, Ana, et al. "Direct and seasonal legacy effects of the 2018 heat wave and drought on European ecosystem productivity." Science advances 6.24 (2020).[Bastos, 2023] Bastos, Ana, et al. "A joint framework for studying compound ecoclimatic events." Nature Reviews Earth & Environment 4.5 (2023): 333-350.

Sentinel 2 Satellite Mission from ESA

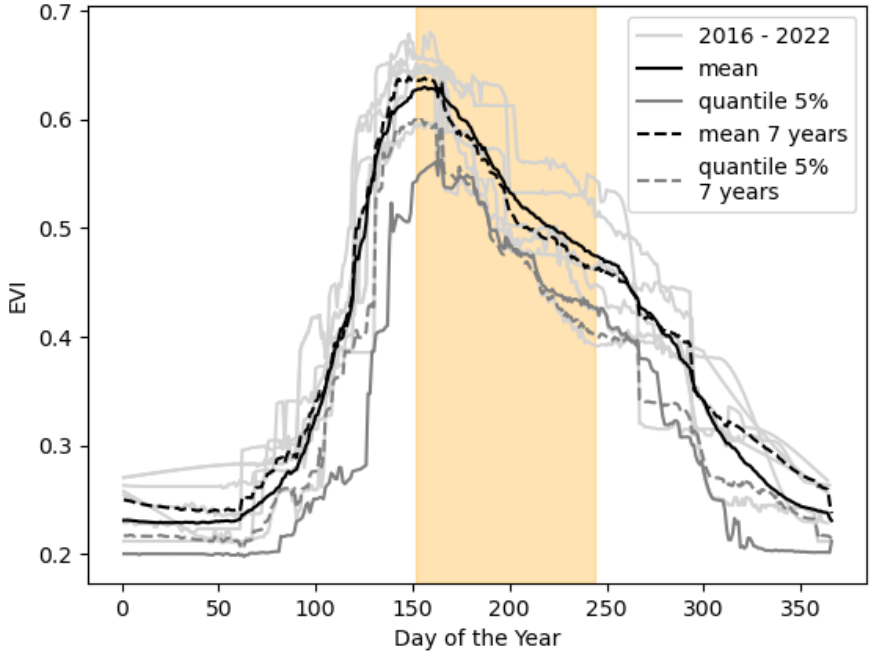
10m resolution

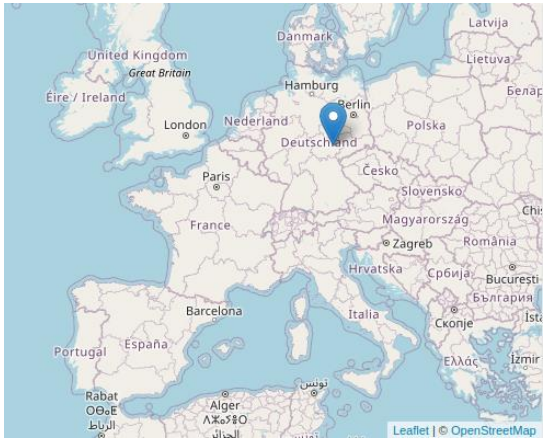
5-day revisit

Since 2017
7 full years

13 Bands
(Visible Light + SWIR)

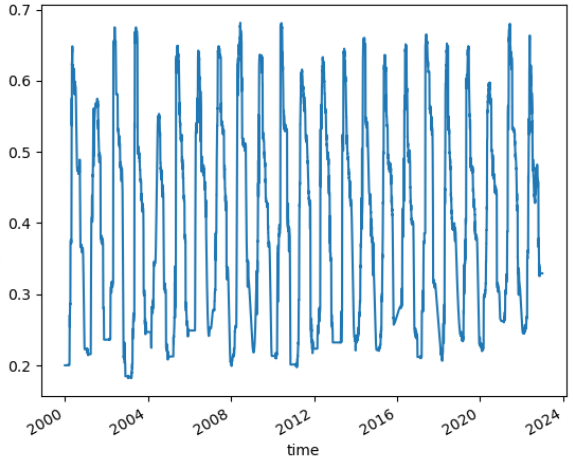
Defining Extremes with limited time series

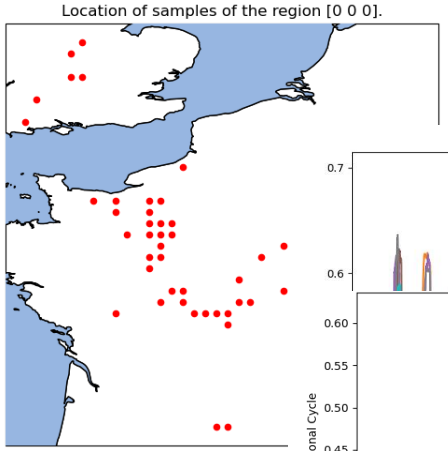


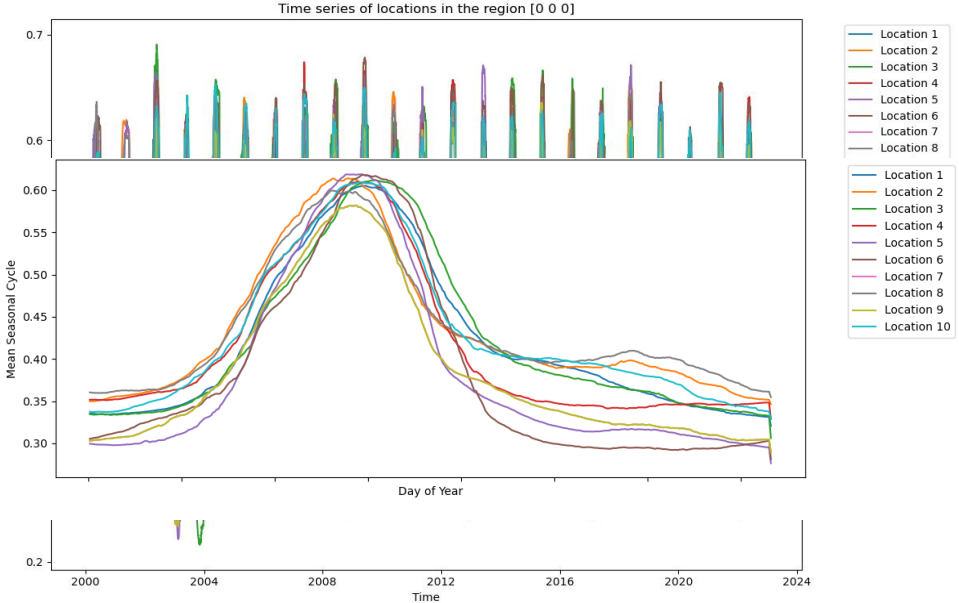


Example of mean and quantile shift with only 7 years of data.

Use Space for Time to define Extreme across a Spatio-Temporal Area (In Progress)

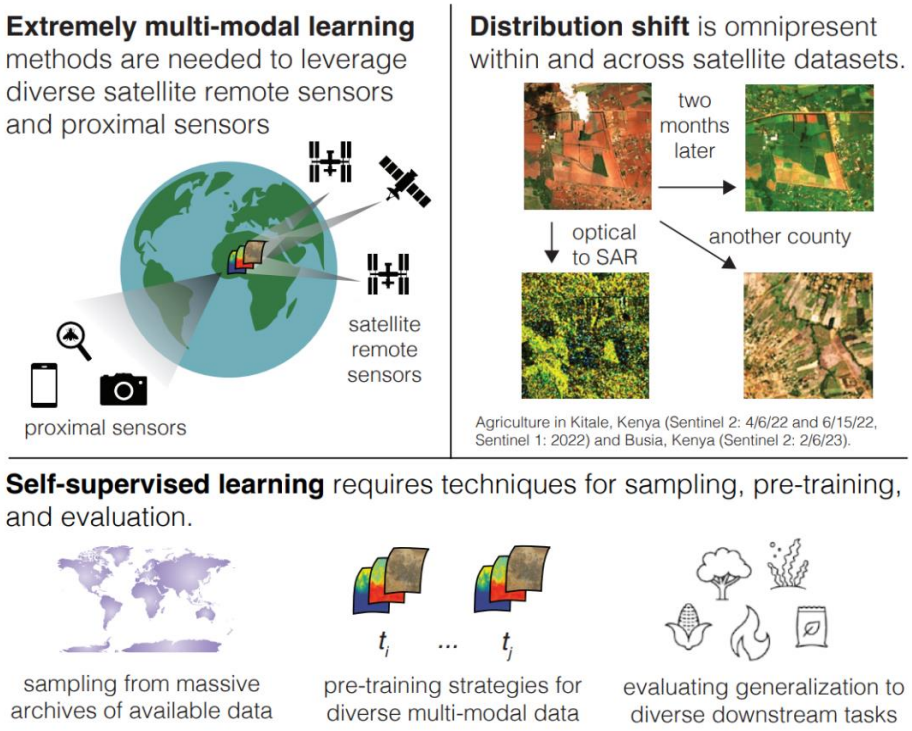






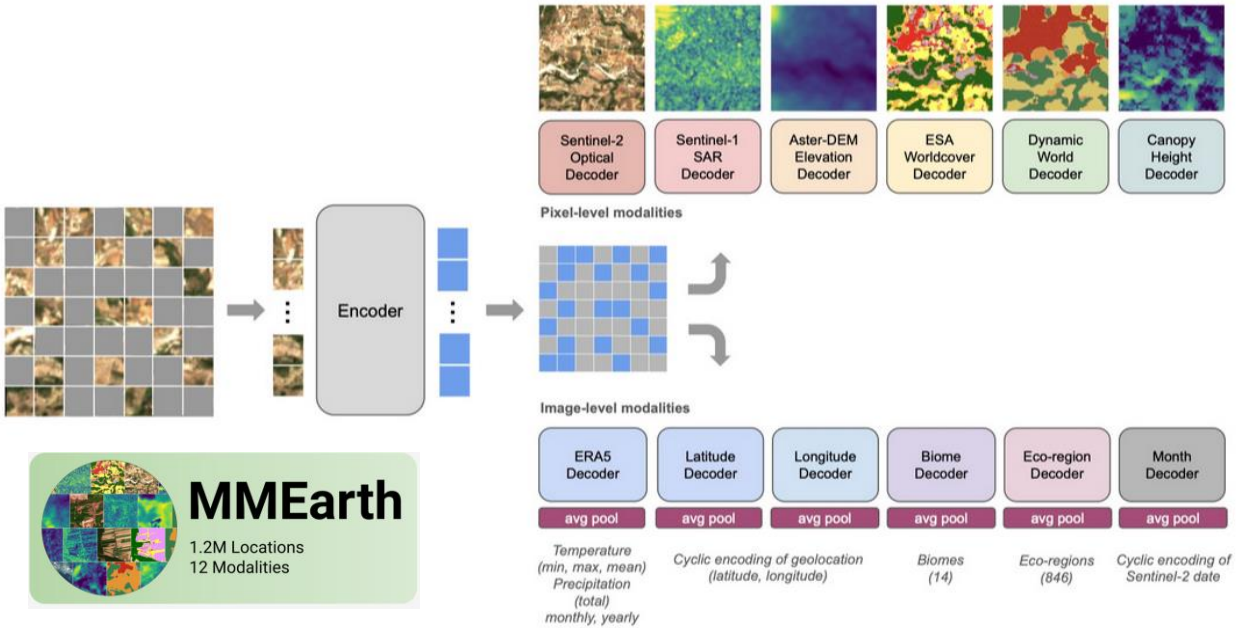
Miguel D Mahecha, Fabian Gans, Sebastian Sippel, Jonathan F Donges, Thomas Kaminski, Stefan Metzger, Mirco Migliavacca, Dario Papale, Anja Rammig, and Jakob Zscheischler. Detecting impacts of extreme events with ecological in situ monitoring networks. Biogeosciences, 14(18):4255–4277, 2017

Satellite Data is a Distinct Modality in Machine Learning



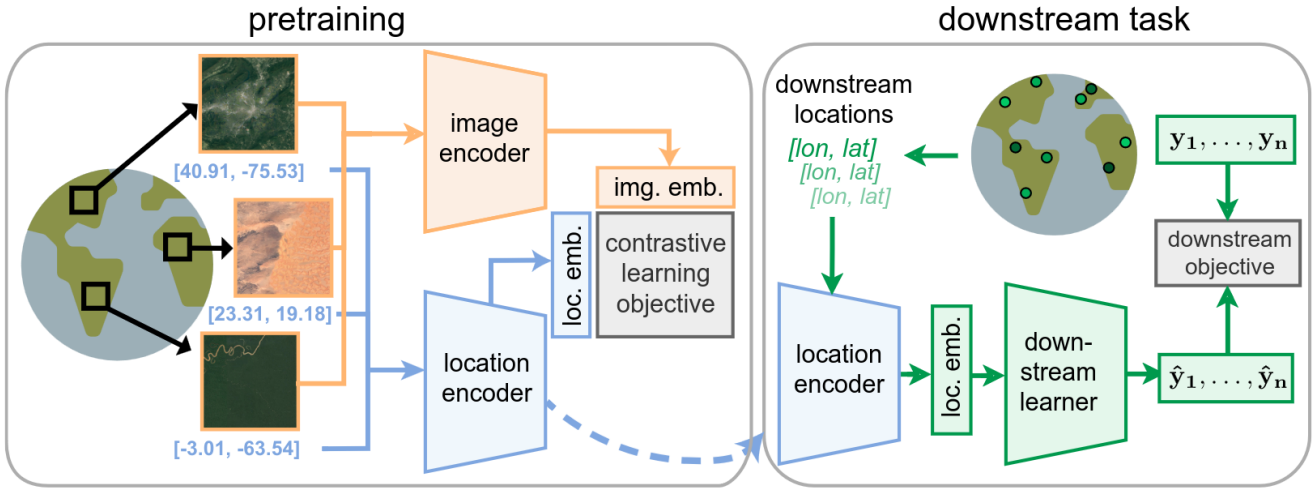
Rolf, Esther, et al. "Mission Critical--Satellite Data is a Distinct Modality in Machine Learning." *arXiv preprint arXiv:2402.01444* (2024).

Towards the development of a general model with a meaningful representation for biogeoscience



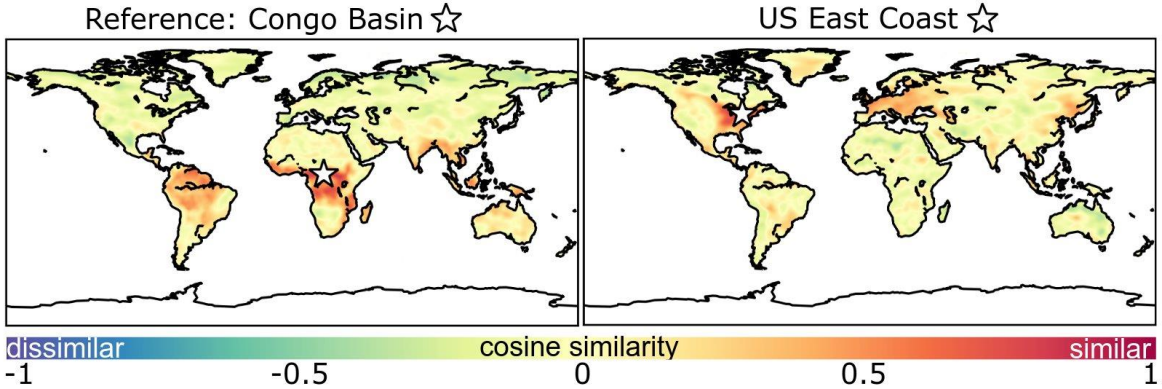
Nedungadi, Vishal, et al. "MMEarth: Exploring multi-modal pretext tasks for geospatial representation learning." *arXiv preprint arXiv:2405.02771* (2024).

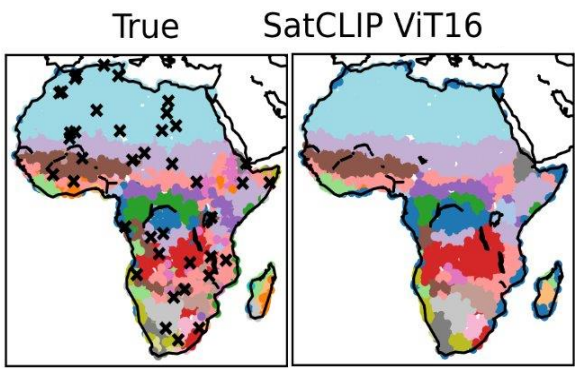
SatCLIP: Global, General-Purpose Location Embeddings with Satellite Imagery



Klemmer, K., Rolf, E., Robinson, C., Mackey, L., & Rußwurm, M. (2023). Satclip: Global, general-purpose location embeddings with satellite imagery. arXiv preprint arXiv:2311.17179.

SatCLIP: Global, General-Purpose Location Embeddings with Satellite Imagery





Klemmer, K., Rolf, E., Robinson, C., Mackey, L., & Rußwurm, M. (2023). Satclip: Global, general-purpose location embeddings with satellite imagery. arXiv preprint arXiv:2311.17179.

ellis | UNIT JENA

ELIAS

www.earthnet.tech

Take away

- Machine learning is promising to tackle most pressing problems
- Earthnet: ... Including predict local impact of extremes events such as drought and heatwaves and support early warning system

crobin@bgc-jena.mpg.de

WAGENINGEN UNIVERSITY & RESEARCH

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

ellis | UNIT JENA

ELIAS

www.earthnet.tech

Take away

- Machine learning is promising to tackle most pressing problems
- Earthnet: ... Including predict local impact of extremes events such as drought and heatwaves and support early warning system
- Interdisciplinary problems require deep understanding of both disciplines
- ... and create new exciting machine learning challenges!

crobin@bgc-jena.mpg.de

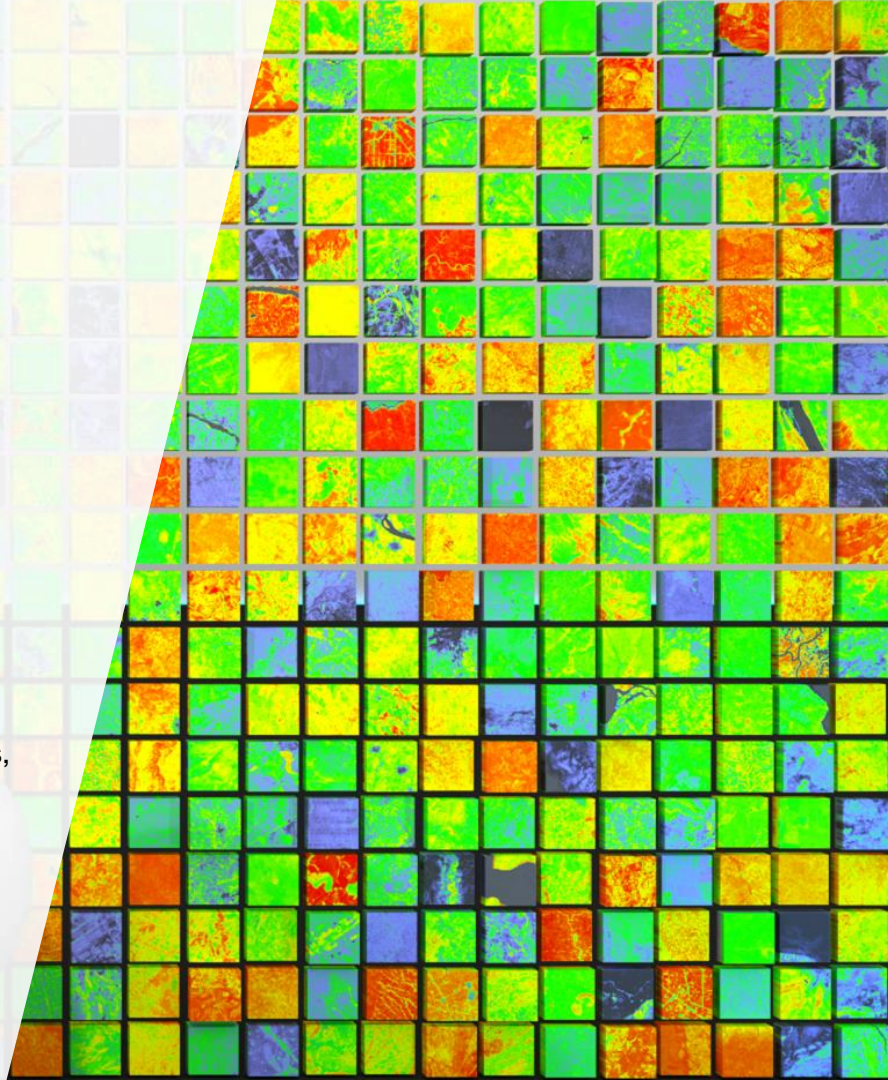
WAGENINGEN UNIVERSITY & RESEARCH

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

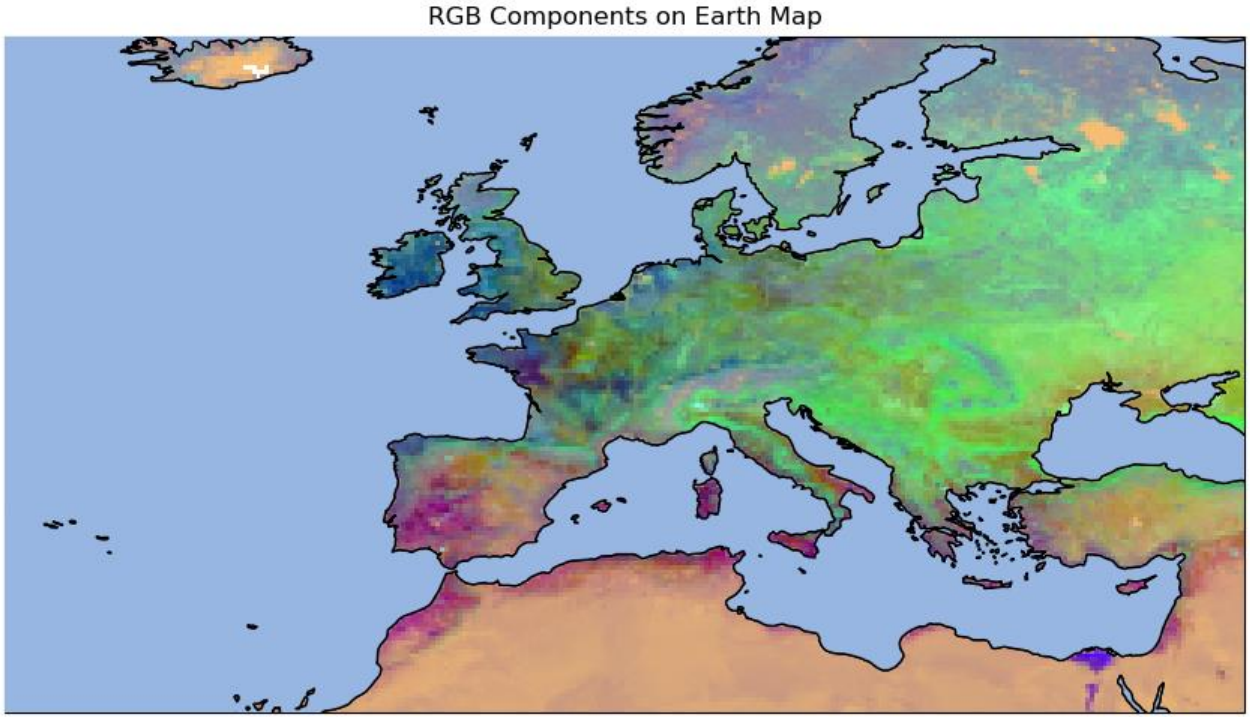
Thank you !

Research is a team effort: Vitus Benson, Lazaro Alonso, Mélanie Weynants, Markus Zehner, Marc Rußwurm, Nuno Carvalhais, Markus Reichstein

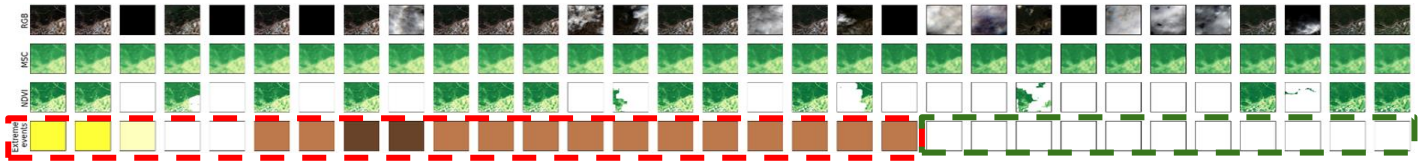
crobin@bgc-jena.mpg.de



Results in progress

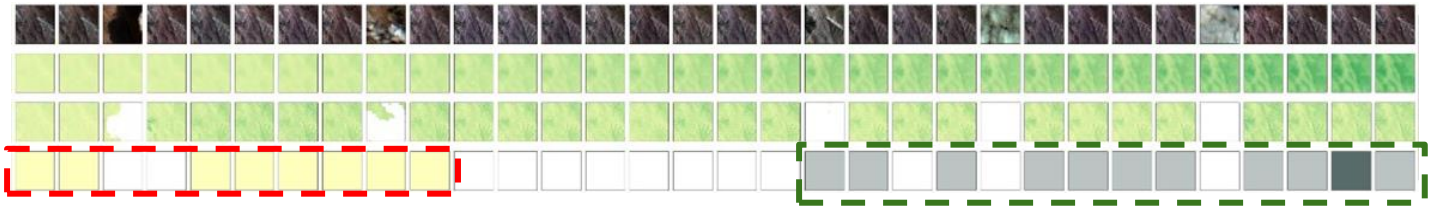


Climatic extremes do not necessary lead to a strong vegetation response



Climatic extremes (1% threshold droughts anomaly)

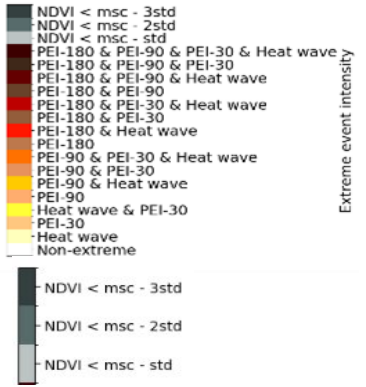
Vegetation under normal condition



Climatic extremes (1% threshold Heat waves)

Vegetation response

RGB
MSC
NDVI
Extremes labels



Climatic extremes such as drought elicit different vegetation responses

Regional Extremes

- **Climatic extremes are defined over a long time period** (30 years) using local percentage threshold.

- **Environmental variables are limited in time**:
    - 24 years for Modis.
    - **8 years of Sentinel-2**.
- **Regional Extremes define extremes using regions of similar phenology**
    - Avoid uniform distribution of extremes.

2- Process

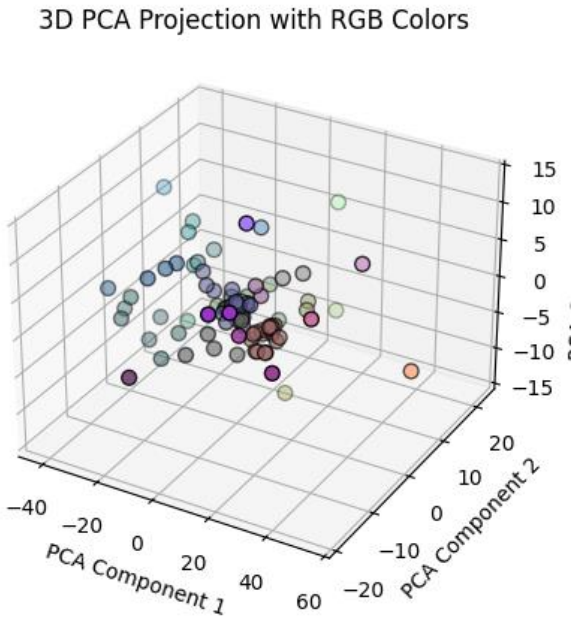
Data:

PEI

- Removed ocean and polar area.
- Between 1970 - 2022.
- Compute the mean seasonal cycle.
- Normalize the data

Model:

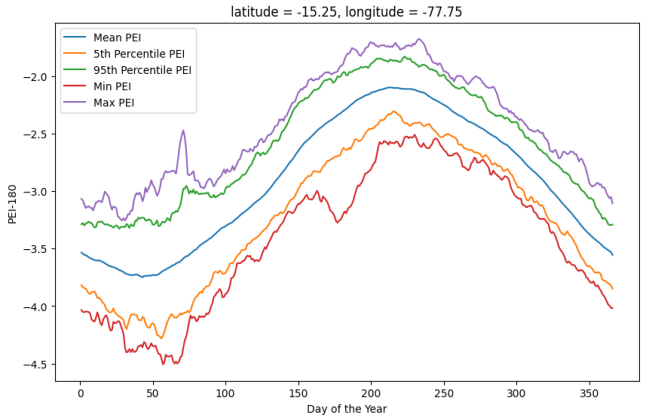
- PCA with 3 Components
- Train on a subset of 10000 samples (on lands).
- Applied to the full dataset.
- Compute the limits of the bins.
- Attribute a bin for each sample.



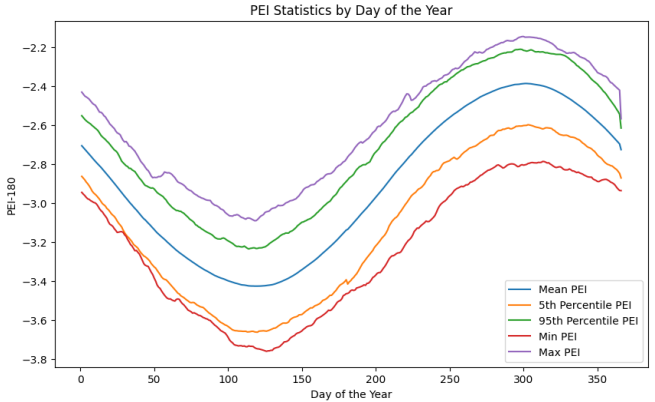
Climatic Extremes: PEI - Mean Seasonal Cycle

PEI: daily differences of Precipitation and Evapotranspiration averaged over 30, 90, and 180 days (PEI30, PEI90, PEI180).

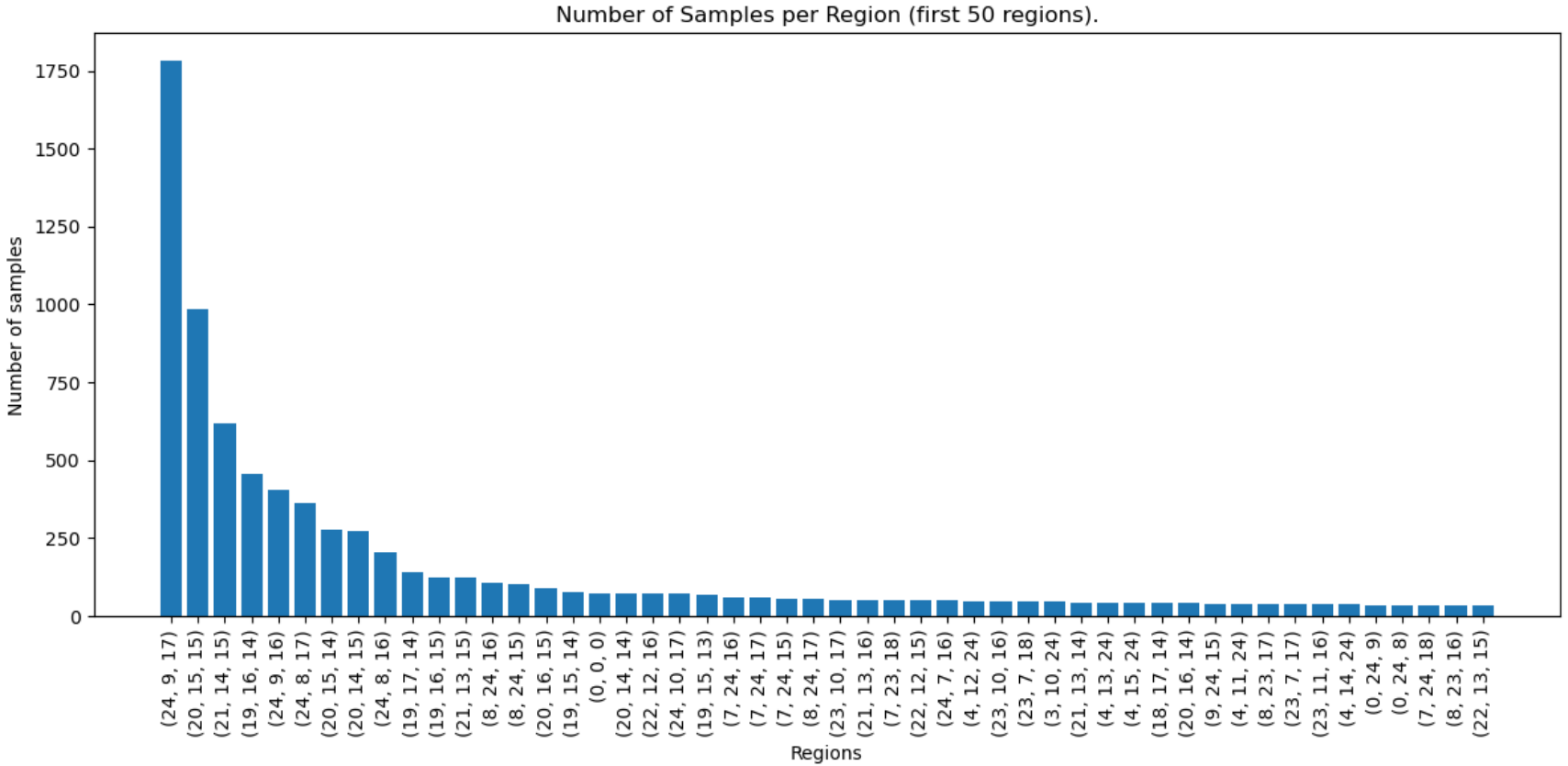
PEI 30



PEI 180



[Weynants, 2024] Mélanie Weynants, Chaonan Ji, Nora Linscheid, Ulrich Weber, and Fabian Gans. A global database of hot and dry extreme events from 1950 to 2022. In Prep.


Number of Samples per Region (first 50 regions).
Number of samples
0
250
500
750
1000
1250
1500
1750
Regions
(24, 9, 17)
(20, 15, 15)
(21, 14, 15)
(19, 16, 14)
(24, 9, 16)
(24, 8, 17)
(20, 15, 14)
(20, 14, 15)
(24, 8, 16)
(19, 17, 14)
(19, 16, 15)
(21, 13, 15)
(8, 24, 16)
(8, 24, 15)
(20, 16, 15)
(19, 15, 14)
(0, 0, 0)
(20, 14, 14)
(22, 12, 16)
(24, 10, 17)
(19, 15, 13)
(7, 24, 16)
(7, 24, 17)
(7, 24, 15)
(8, 24, 17)
(23, 10, 17)
(21, 13, 16)
(7, 23, 18)
(22, 12, 15)
(24, 7, 16)
(4, 12, 24)
(23, 10, 16)
(23, 7, 18)
(3, 10, 24)
(21, 13, 14)
(4, 13, 24)
(4, 15, 24)
(18, 17, 14)
(20, 16, 14)
(9, 24, 15)
(4, 11, 24)
(8, 23, 17)
(23, 7, 17)
(23, 11, 16)
(4, 14, 24)
(0, 24, 9)
(0, 24, 8)
(7, 24, 18)
(8, 23, 16)
(22, 13, 15)