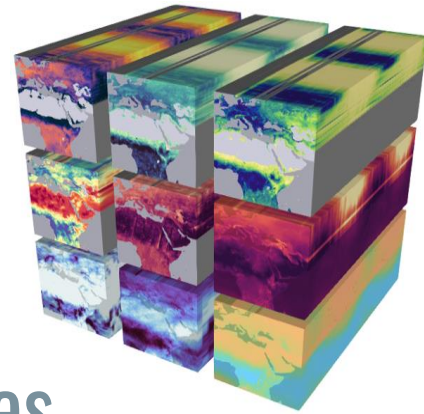# Earth System Data Lab & DeepExtremes

ESA projects relevant to study C cycle extremes

Miguel Mahecha (1), Gunnar Brandt (2) **& many other contributors!**

(1) Remote Sensing Center for Earth System Research, Leipzig University, Germany
(2) Brockmann Consult GmbH, Germany

Twitter: @EarthSysDataLab, @DeepExtremes

UNIVERSITÄT LEIPZIG

BROCKMANN CONSULT GMBH

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

EOX
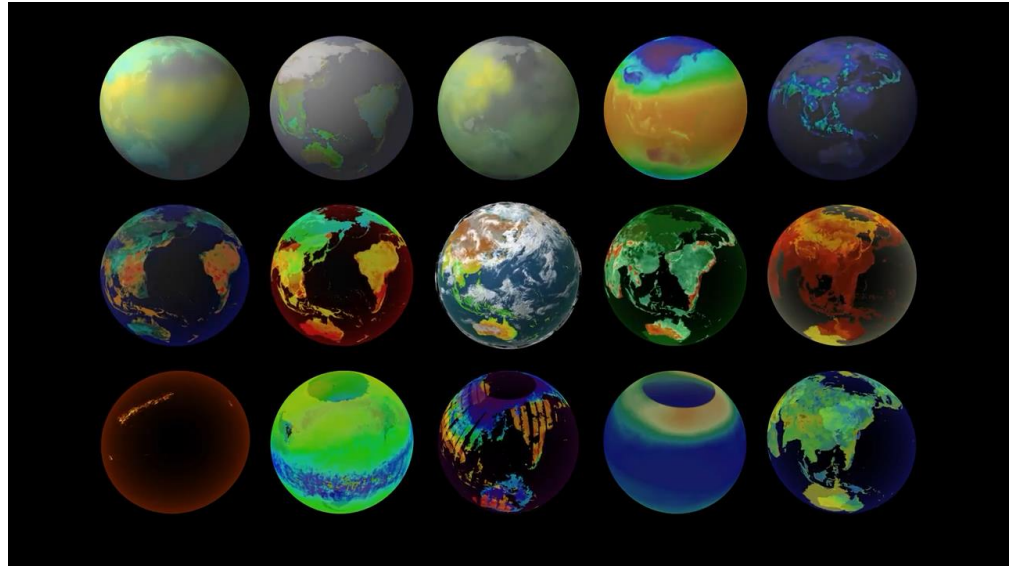
earthwave

VNIVERSITAT DE VALÈNCIA

Deep Earth System Data Lab - *idea*

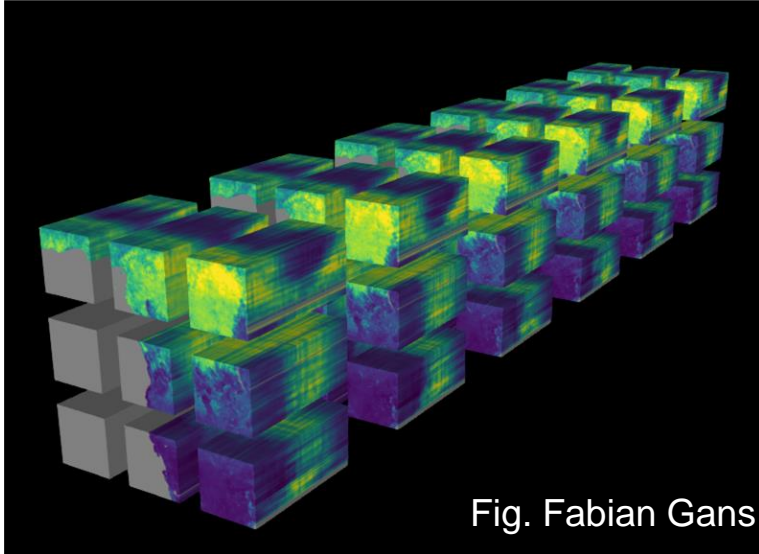# ESDL should solve interoperability issues of data streams

- *Make all downstream Earth data products truly interoperable*

- *Scientists should focus on data exploration only!*

- *Overcome scalability issues etc.*

Mahecha, Gans et al. (2020)
Earth System Dynamics, **11**, 201-234.

Movie concept by the ESDL team - for ESA by Planetary Visions

Data cubes should live in the cloud

Fig. Fabian Gans

Accessible for the user via jupyter labs

```julia
function sufficient_dimensions(xin::AbstractArray, expl_var::Float64 = 0.95)

    any(ismissing,xin) && return NaN
    npoint, nvar = size(xin)
    means = mean(xin, dims = 1)
    stds  = std(xin, dims = 1)
    xin   = broadcast((y,m,s) -> s>0.0 ? (y-m)/s : one(y), xin, means, stds)
    pca = fit(PCA, xin', pratio = 0.999, method = :svd)
    return findfirst(cumsum(principalvars(pca)) / tprincipalvar(pca) .> expl_var)
end
```
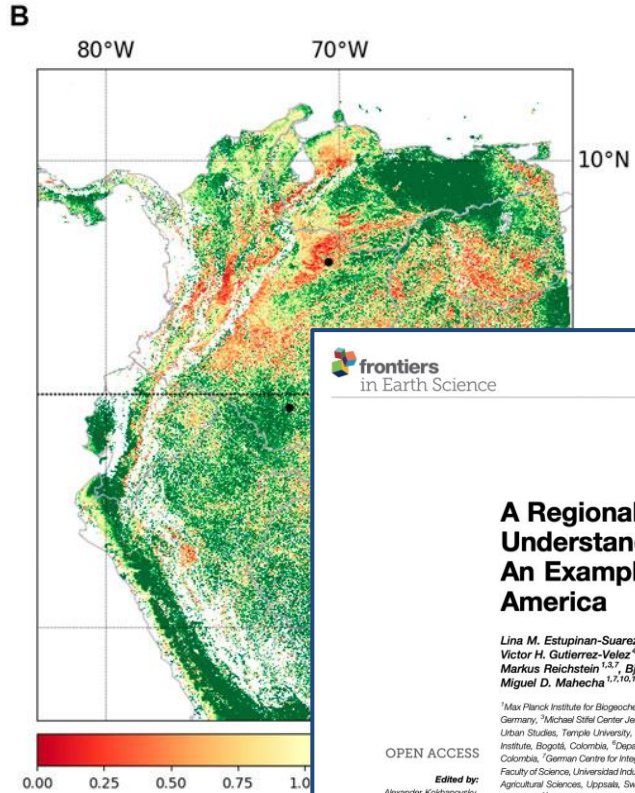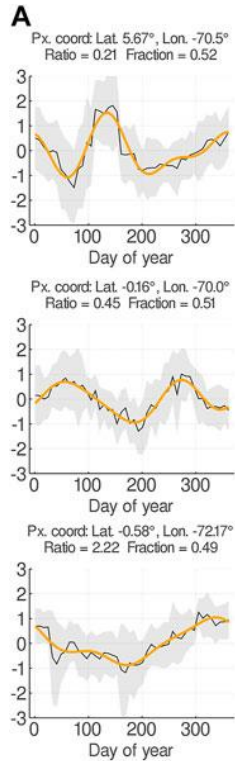
```julia
cube_int_dim = mapslices(sufficient_dimensions, cube_fill, dims = ("Time","Variable"))
```

```julia
plotMAP(cube_int_dim)
```


Cloud-optimized data formats:
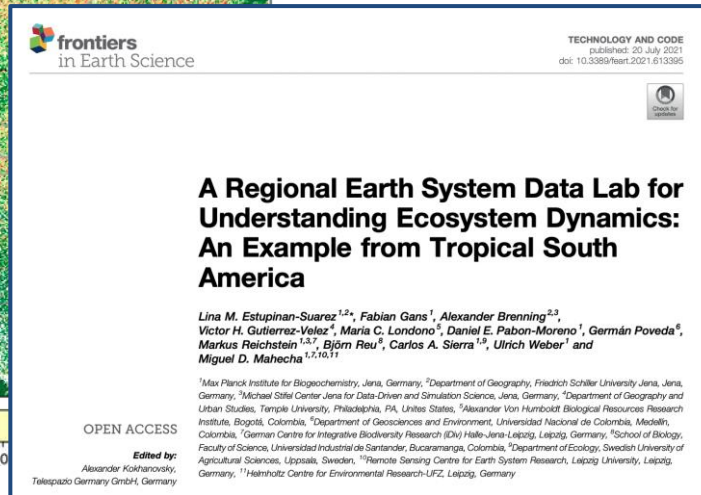Zarr - compatible with https://pangeo.io/

Now developing regional "on demand" data cubes

A

Px. coord: Lat. 5.67°, Lon. -70.5°
Ratio = 0.21 Fraction = 0.52

Px. coord: Lat. -0.16°, Lon. -70.0°
Ratio = 0.45 Fraction = 0.51

Px. coord: Lat. -0.58°, Lon. -72.17°
Ratio = 2.22 Fraction = 0.49

B

Under development
- Baltic sea data cube
- Polar data cube
- Etc…

frontiers
in Earth Science

TECHNOLOGY AND CODE
published: 20 July 2021
doi: 10.3389/feart.2021.613395

A Regional Earth System Data Lab for Understanding Ecosystem Dynamics: An Example from Tropical South America

Lina M. Estupinan-Suarez[1,2]*, Fabian Gans[1], Alexander Brenning[2,3], Victor H. Gutierrez-Velez[4], Maria C. Londono[5], Daniel E. Pabon-Moreno[1], Germán Poveda[6], Markus Reichstein[1,3,7], Björn Reu[8], Carlos A. Sierra[1,9], Ulrich Weber[1] and Miguel D. Mahecha[1,7,10,11]

[1]Max Planck Institute for Biogeochemistry, Jena, Germany, [2]Department of Geography, Friedrich Schiller University Jena, Jena, Germany, [3]Michael Stifel Center Jena for Data-Driven and Simulation Science, Jena, Germany, [4]Department of Geography and Urban Studies, Temple University, Philadelphia, PA, United States, [5]Alexander Von Humboldt Biological Resources Research Institute, Bogotá, Colombia, [6]Department of Geosciences and Environment, Universidad Nacional de Colombia, Medellín, Colombia, [7]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, [8]School of Biology, Faculty of Science, Universidad Industrial de Santander, Bucaramanga, Colombia, [9]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden, [10]Remote Sensing Centre for Earth System Research, Leipzig University, Leipzig, Germany, [11]Helmholtz Centre for Environmental Research-UFZ, Leipzig, Germany

Colour values of >1 → Dominant annual cycle in comparison to the semiannual mode.
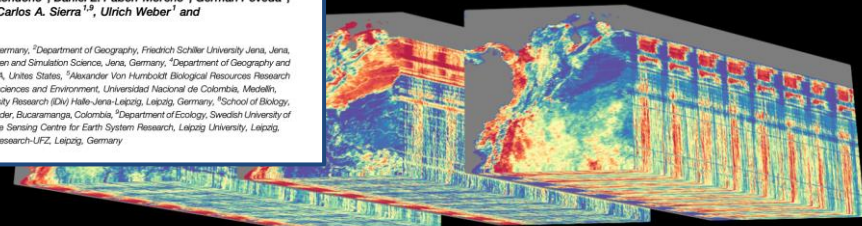
# Bring all cubes closer to machine learning

- Couple ARDCs to state-of-the-art deep learning libraries and other ML methods (xcube + ML libraries)

- Loading and transfer ARDCs (memory) / efficient scheme for generation of training sets

- Adapt existing data loading and transformation mechanisms from Keras and PyTorch (DataGenerator, DataLoader)

- Implement re-partitioning and re-sampling strategies on large cloud-hosted data stets

- Development of use case specific Jupyter Notebooks for user friendly application and guidance

Relevance for C cycle extremes

Low-dimensional trajectories in the land-surface

- What are the intrinsic dimensions of change"?
- What is the meaning of such indicators?
- What is their temporal dynamics?
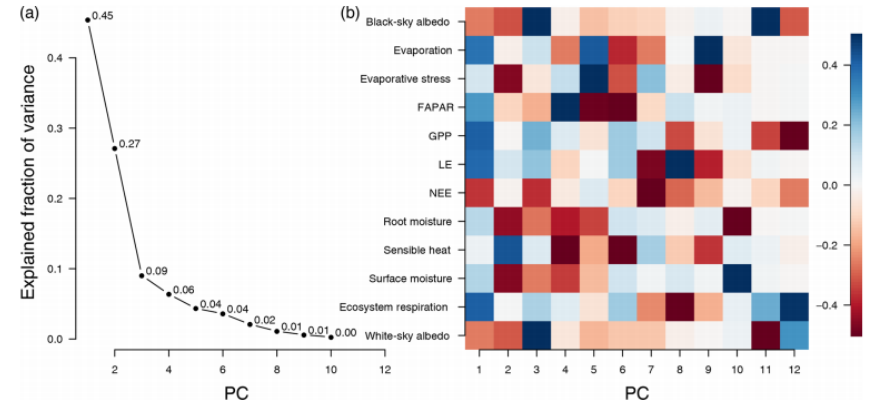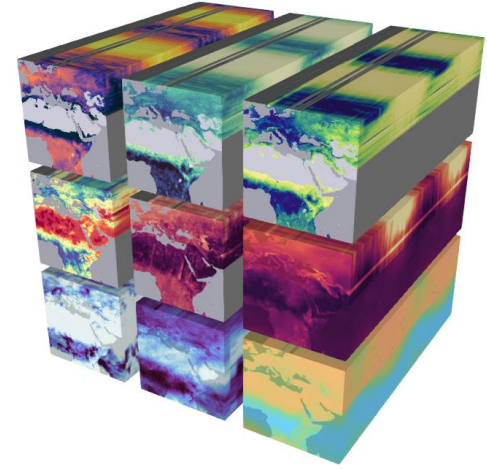
Kraemer, Camps-Valls, Reichstein, Mahecha (2020)
Biogeosciences, **17**, 2397–2424
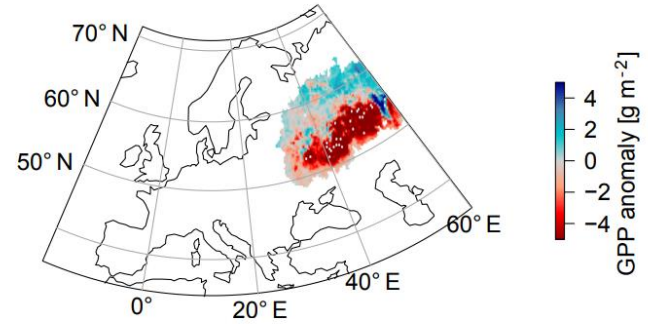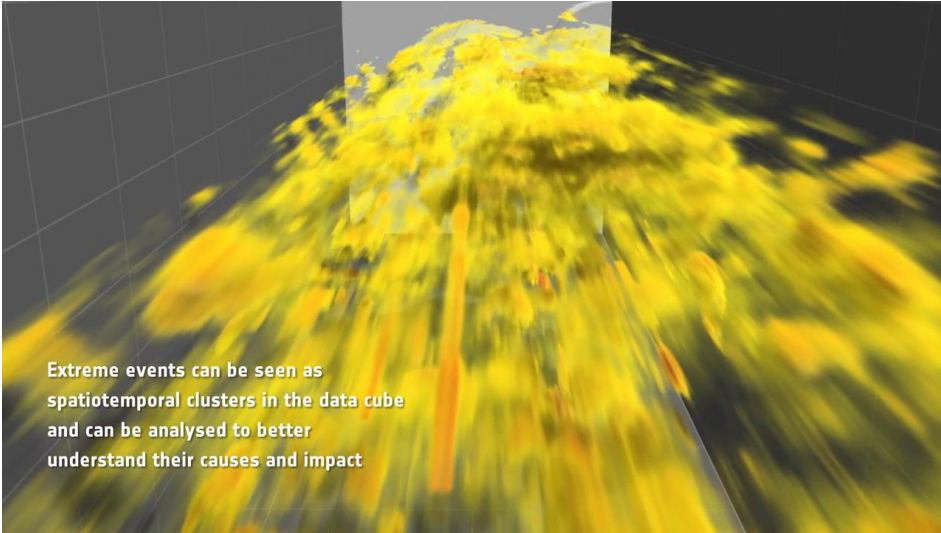
Low-dimensional trajectories in the land-surface

Guido Kraemer, *Camps-Valls, Reichstein, Mahecha* (2020) Biogeosciences, **17**, 2397–2424
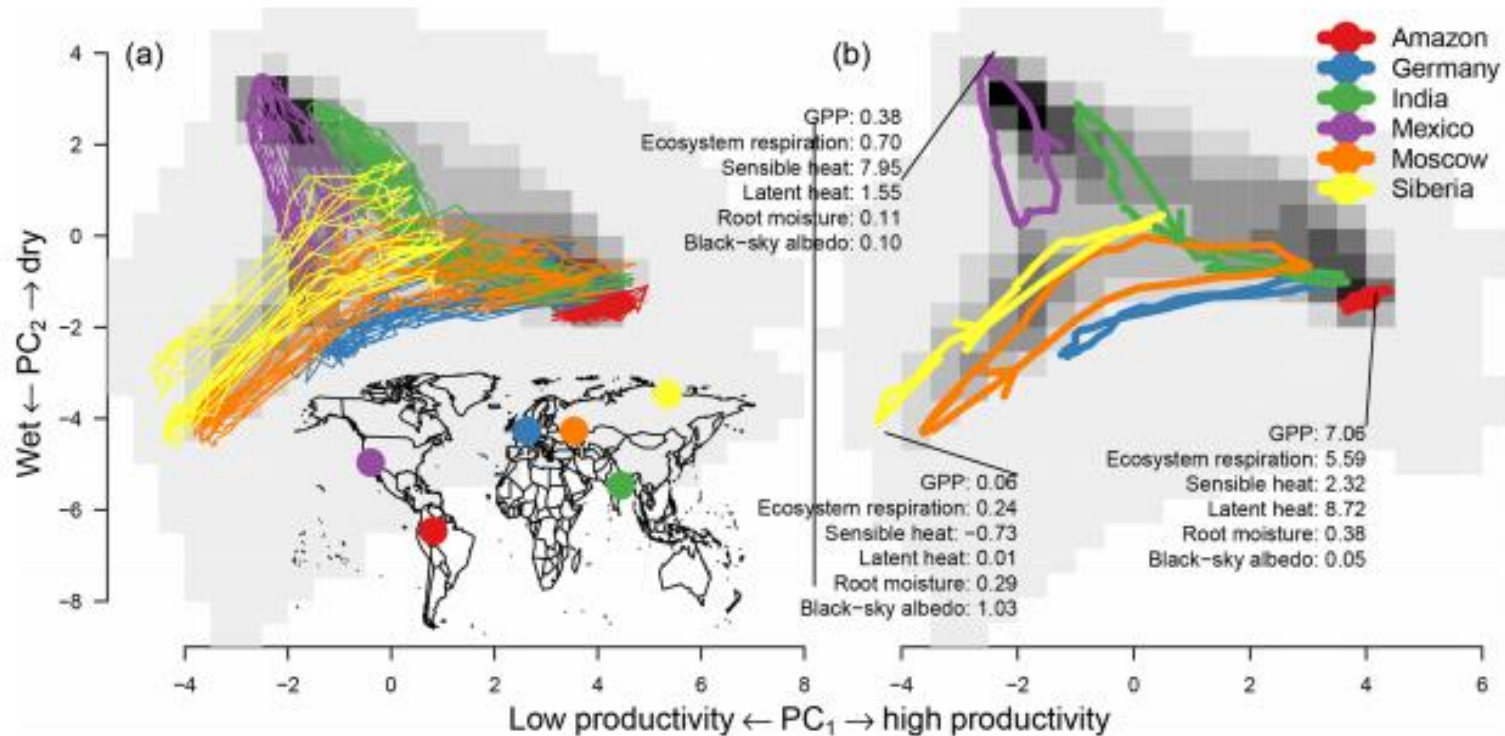
Low-dimensional trajectories in the land-surface

- Reflect deforestation
- Extreme events
- Long-term trends
- Modulations in amplitudes
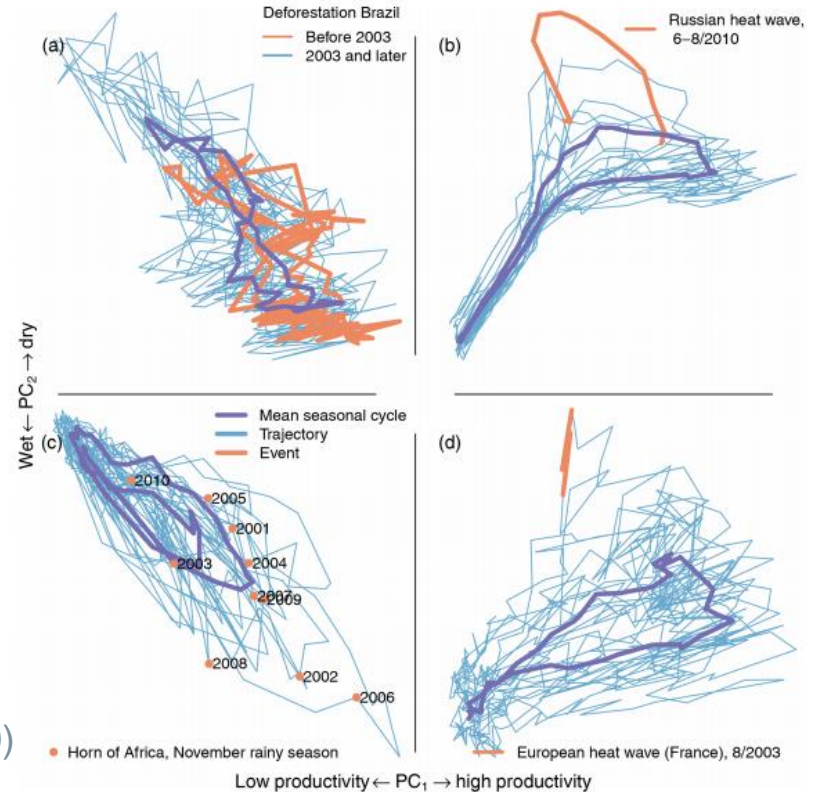
Guido Kraemer, Camps-Valls, Reichstein, Mahecha (2020)
Biogeosciences, **17**, 2397–2424

Extreme events in the ESDL

Extreme events can be seen as spatiotemporal clusters in the data cube and can be analysed to better understand their causes and impact

Milan Flach et al. (2018) Biogeosciences, 15, 6067-608
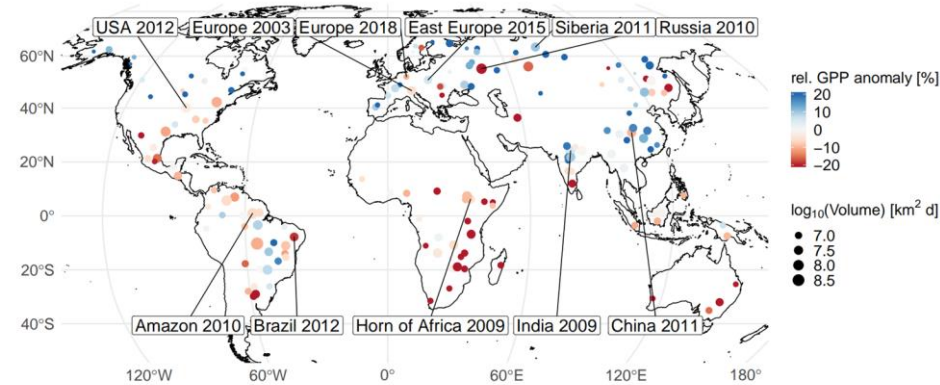
MiaIn Flach et al. (2021) Biogeosciences, 18, 39-53

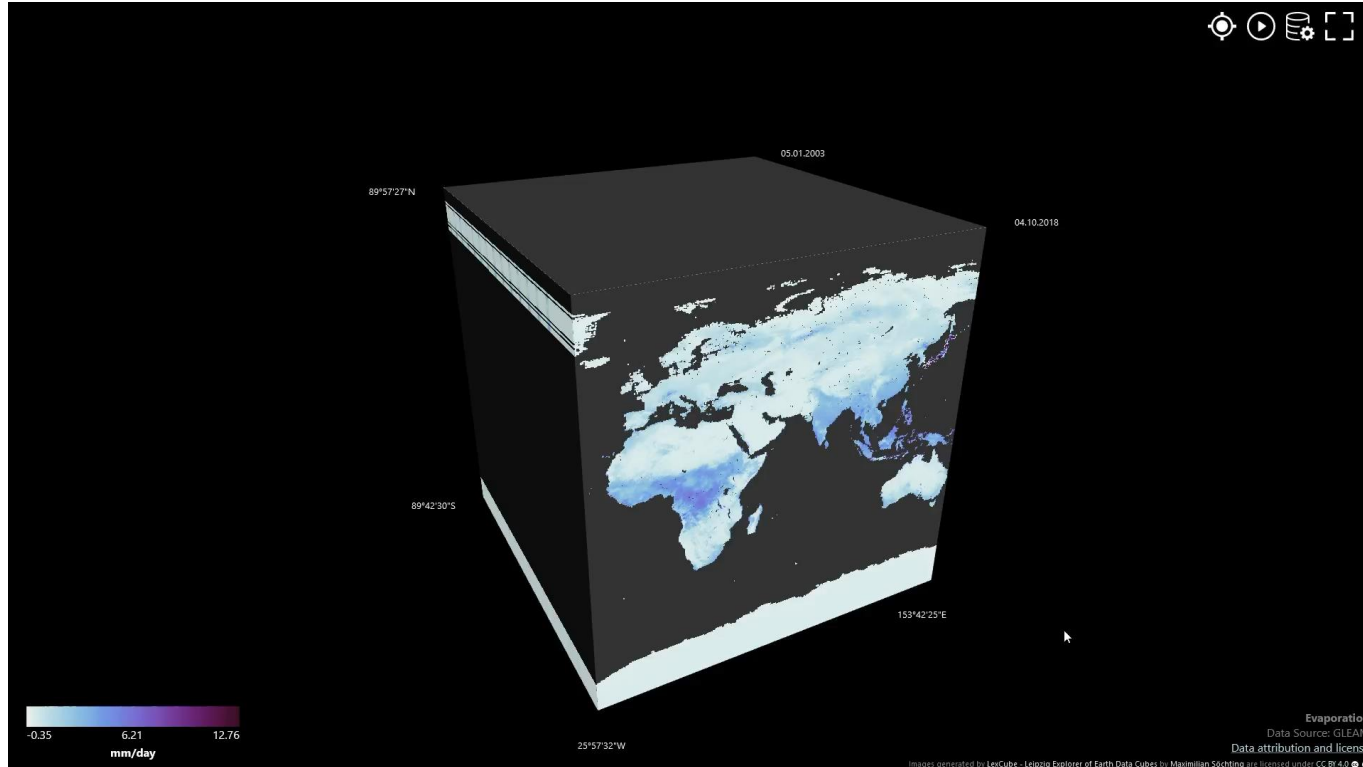# Extreme events in the ESDL - Global study

Milan Flach et al. (2021) Biogeosciences, 18, 39-53
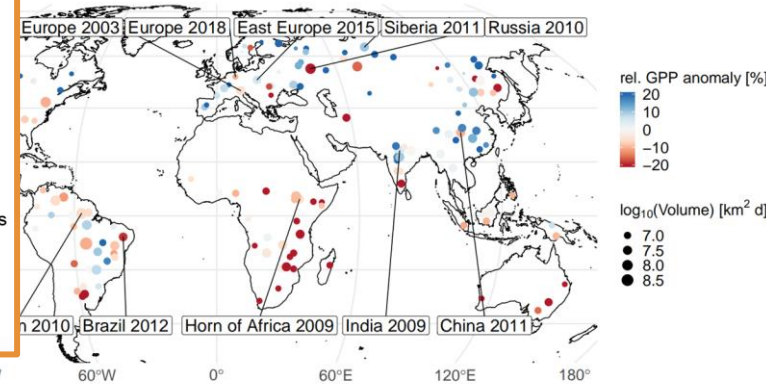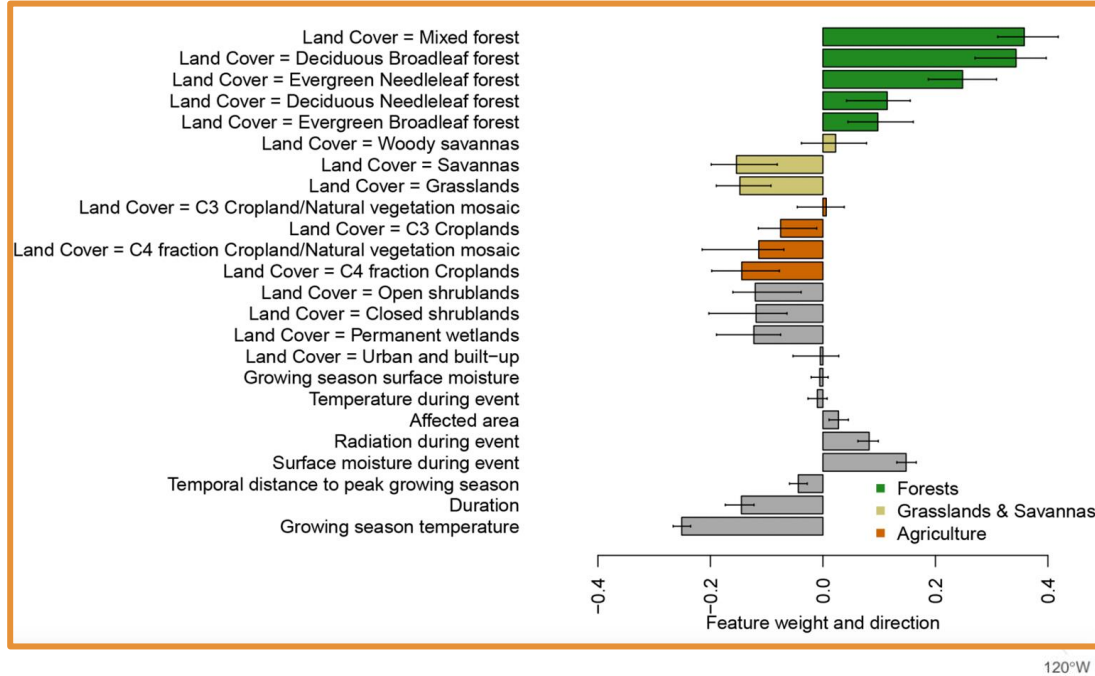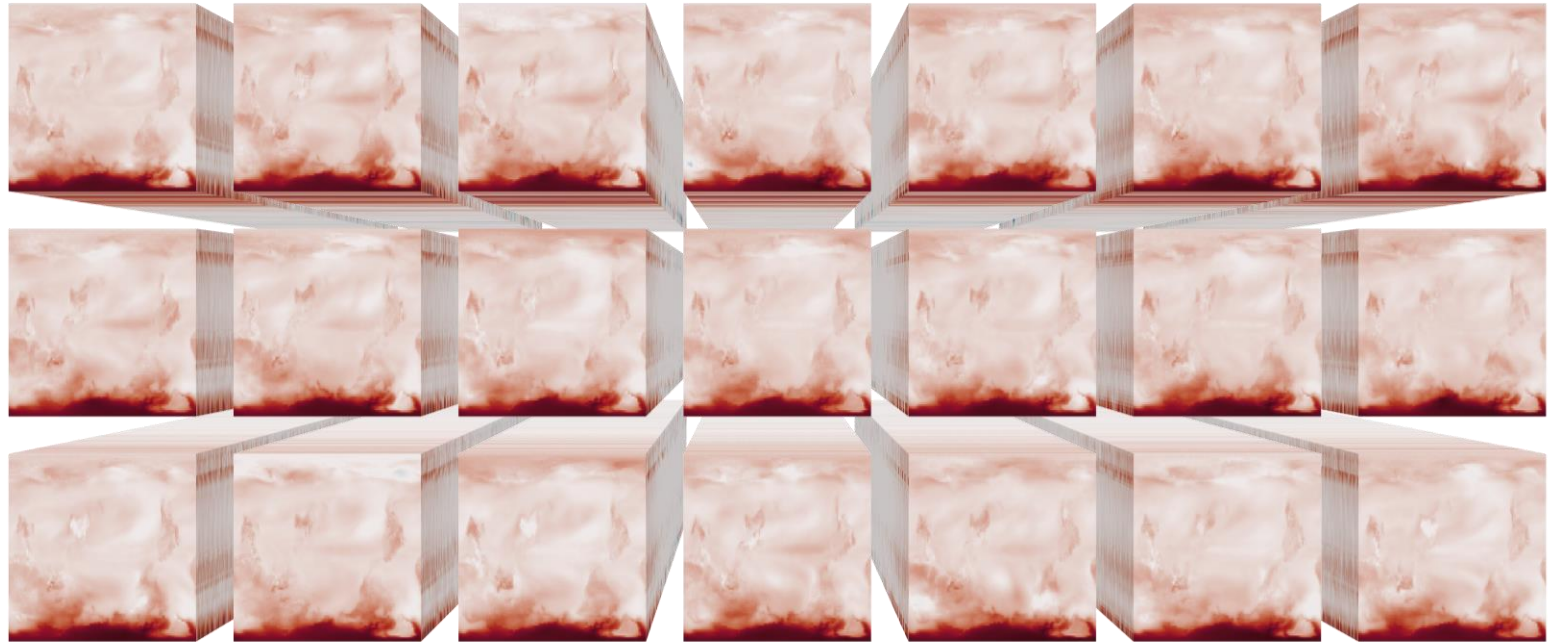
Visual analytics

Lexcube.org Interactive visualization of data cubes

__Maximilan Söchting__ et al. (submitted) – here using GLEAM

Lexcube.org Interactive visualization of data cubes

Maximilan Söchting, Sebastian Sippel et al. (in prep) Visualization of a fully coupled ocean-atmosphere climate model ensemble (from Fischer et al. 2021).

# Lexcube.org Interactive visualization of data cubes

Figure 1: A user with the Lexcube application on her smartphone, visualizing a data set of the protected area "Auwald forest", Leipzig, Germany, while walking through it. The spatiotemporal data visualization shows longitude by latitude on the cube front face, and latitude/longitude by time on the side faces. This data cube view enables visual comprehension of temporal patterns and trends. Here, the kNDVI spectral index (Kernel Normalized Difference Vegetation Index) based on satellite imagery from a commercial data provider is on display.

Maximilan Söchting et al. (submitted)

Lexcube.org Interactive visualization of data cubes

Maximilan Söchting et al. (submitted)
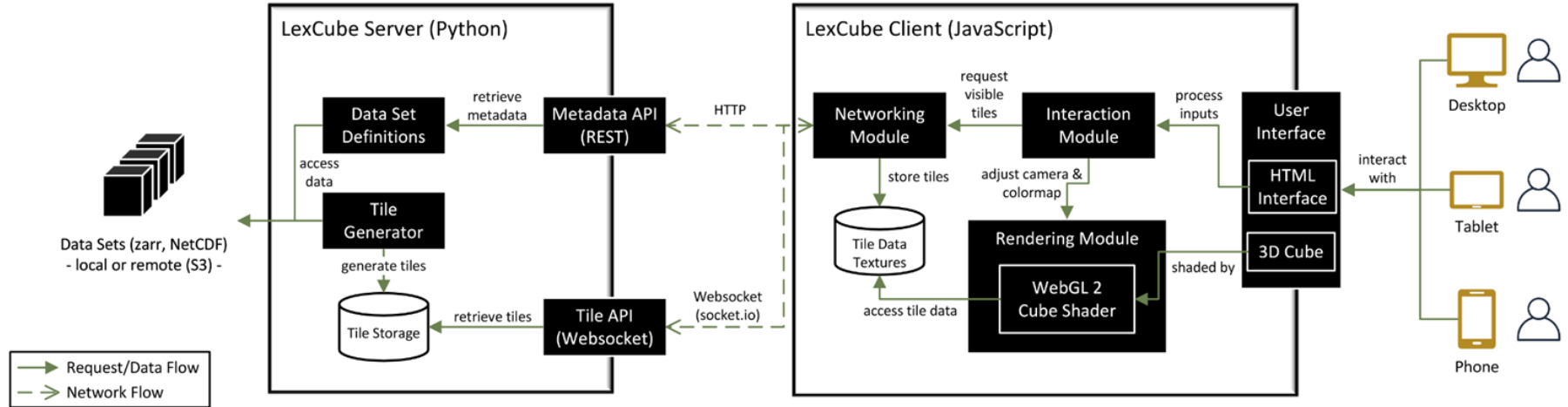
Dataset	Data Provider	Extent				Size	
		Parameters	*Time*	*Latitude*	*Longitude*	*Compressed*	*Original*
EAC4 CAMS Global Reanalysis	ECMWF	33	222	241	480	2.4 GB	3.2 GB
Spectral Indices in National Park Hainich	Sentinel-2, ESA	97	195	512	512	31 GB	36.9 GB
Earth System Data Cube	Various	73	1840	2160	4320	301 GB	5.1 TB
Global Daily Spectral Indices	MODIS, NASA	15	8236	3600	7200	2.2 TB	9.3 TB

DeepExtremes - *idea*

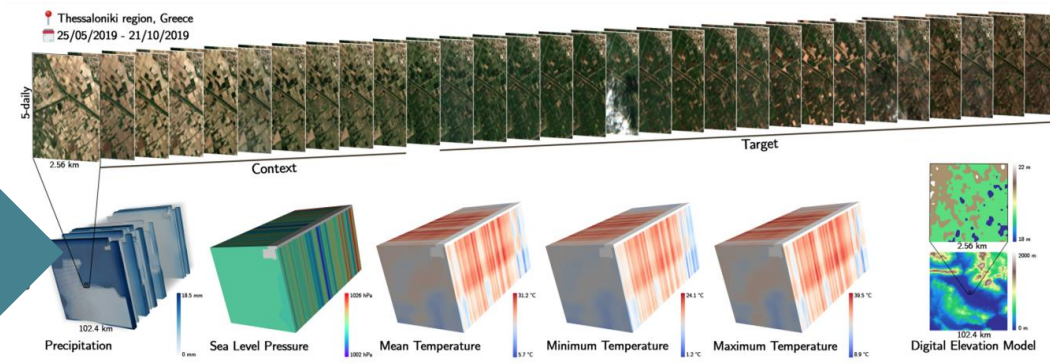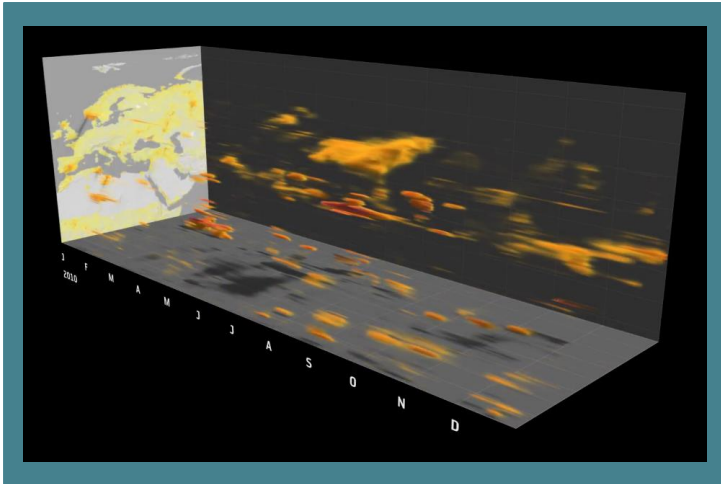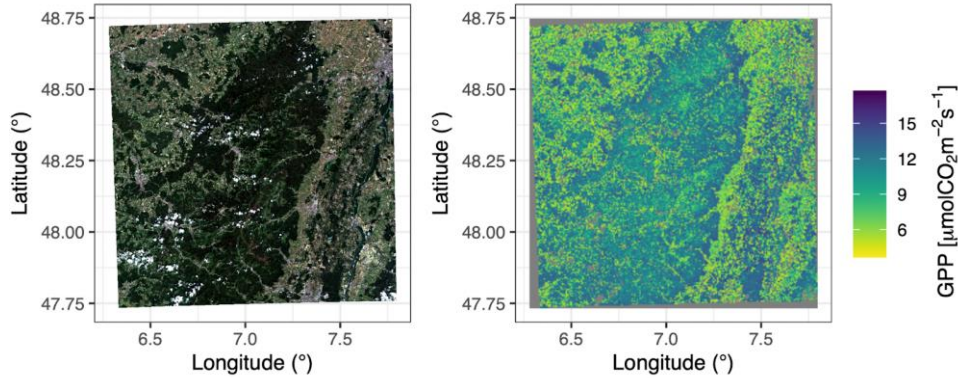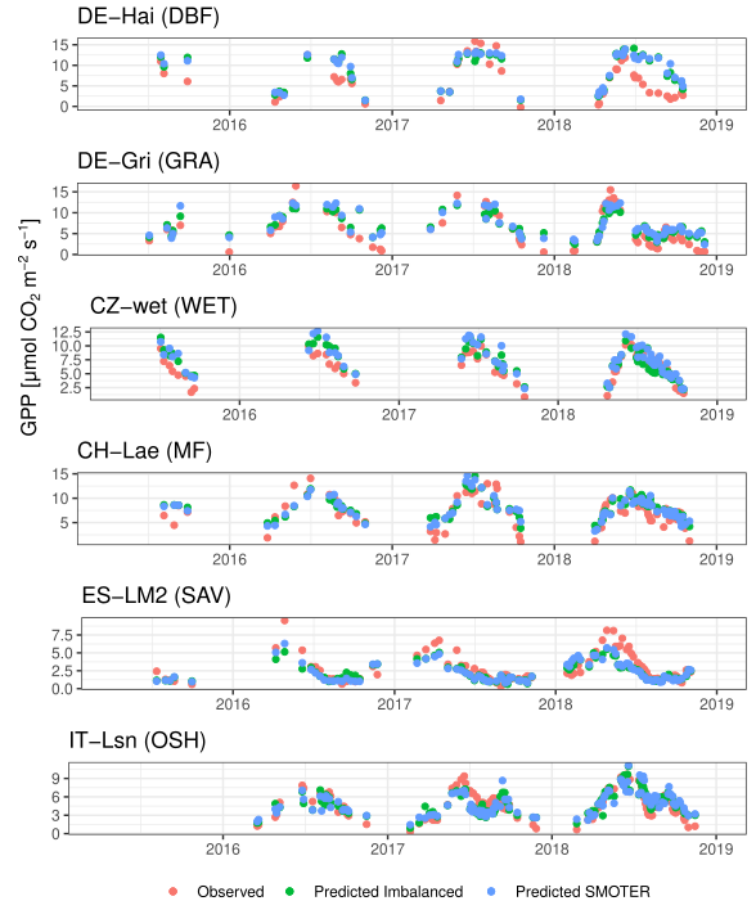# DeepExtremes: A competition of predicting extremes

Fig. Requena-Mesa et al. (2021) "EarthNet"

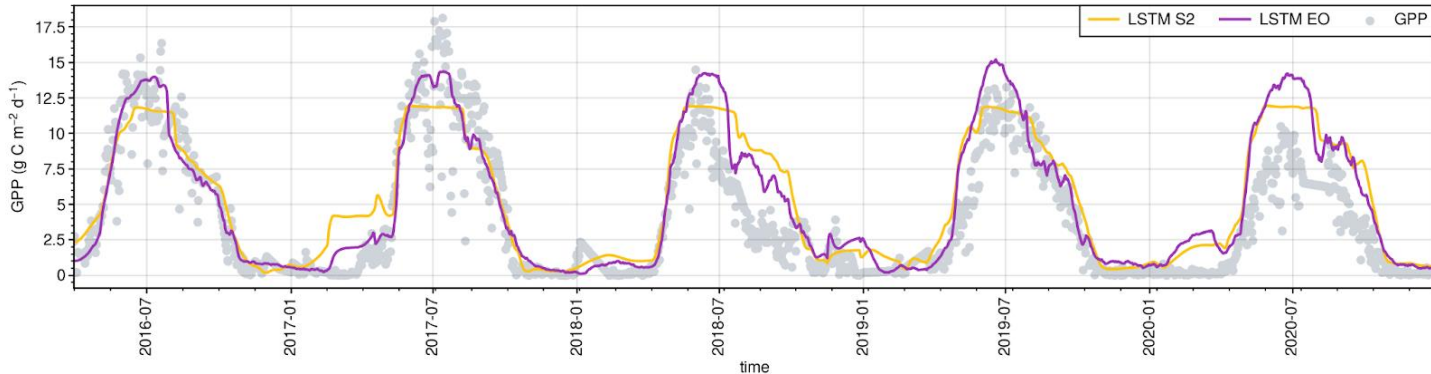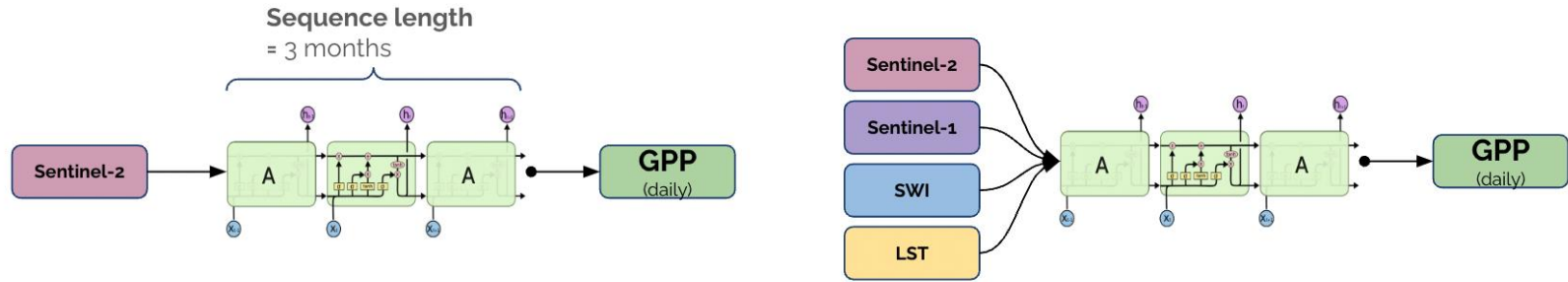# DeepExtremes: A competition of predicting extremes

Daniel Pabon, Migliavacca, M., Reichstein, M. & Mahecha, M. D. (2022). *IEEE Transactions on Geoscience and Remote Sensing*. 10.1109/TGRS.2022.3152272
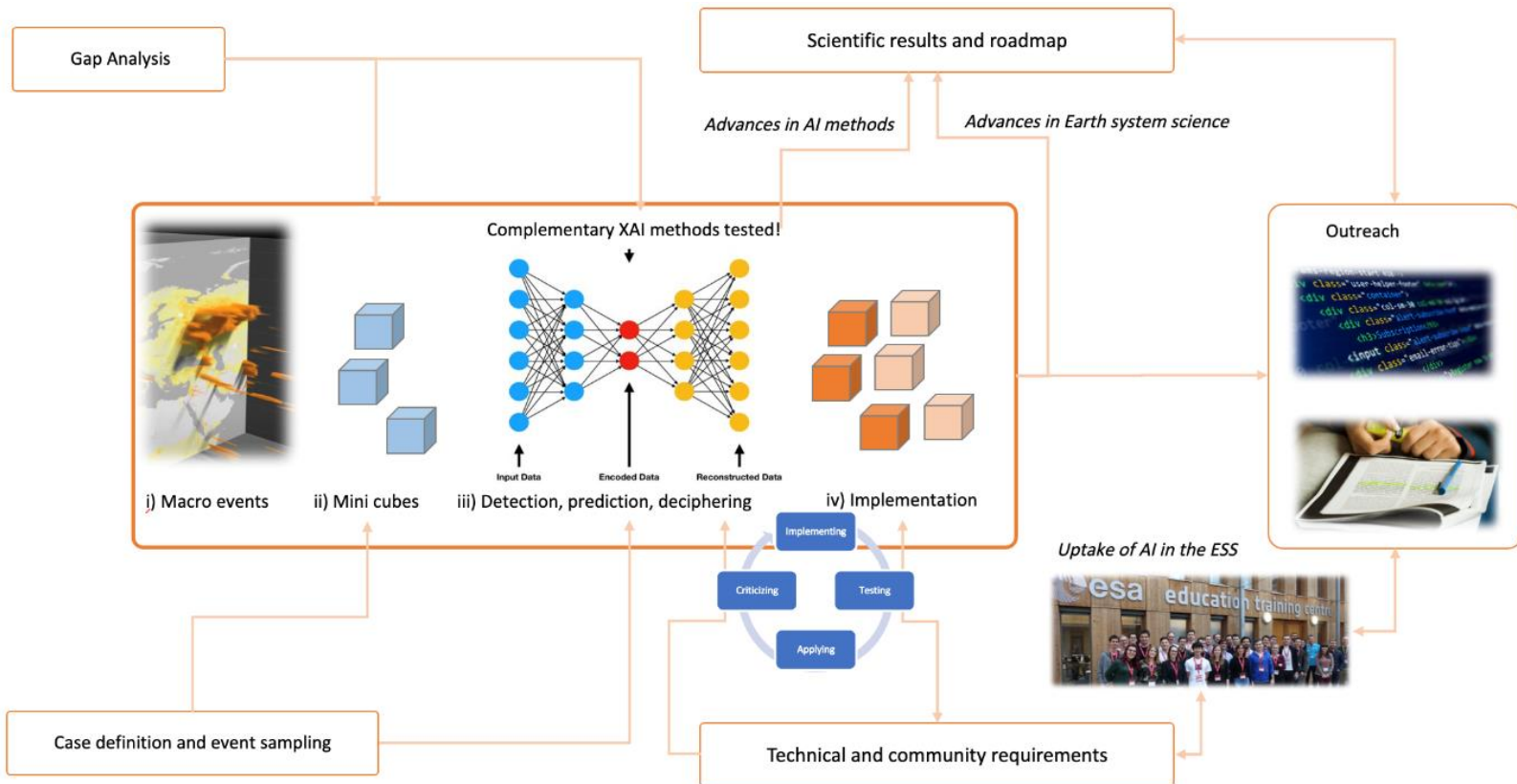
DeepExtremes: A competition of predicting extremes

David Montero, Sebastian Wieneke et al. (in prep.)

DeepExtremes: A competition of predicting extremes

DeepESDL ...

- An **integrator of scientific data products** from different activities in a single infrastructure.

- ESDL is a **platform for collaborative research** allowing different scientists and teams to work together in a collective scientific effort sharing data, tools and expertise.

- Support for the **execution of individual projects** e.g. via dedicated Jupyter hubs

DeepExtremes ...

- A **scientific project to advance the capacity to predict extremes** from Sentinel-2 data plus ancillary information

- Explore **the limits of predictability** allowing different scientists and teams to work together in a collective scientific effort sharing data, tools and expertise.

Complementary XAI methods tested!

Macro events   ii) Mini cubes   iii) Detection, prediction, deciphering   iv) Implementation

Input Data   Encoded Data   Reconstructed Data