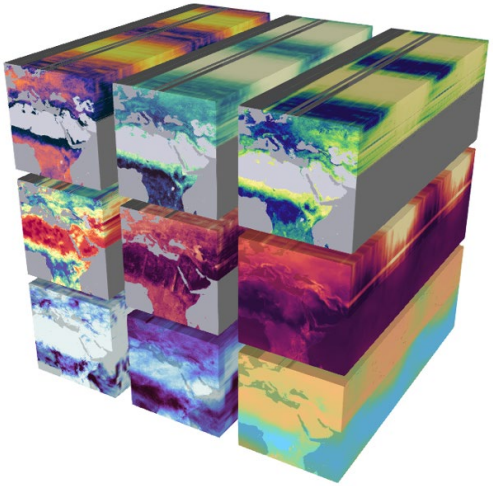


The Earth System Data Lab

Idea, examples, and way forward

Miguel Mahecha (1), Gunnar Brandt (2), Fabian Gans (3), Norman Fomferra (2), Alicja, Balfanz (2), Guido Kraemer (1), Maximilian Söchting (1), David Montero (1), Markus Reichstein (3), Carsten Brockmann (2), **early adopters & other contributors!**

(1) Remote Sensing Center for Earth System Research, Leipzig University, Germany
(2) Max Planck Institute for Biogeochemistry, Germany
(3) Brockmann Consult GmbH, Germany

@EarthSysDataLab

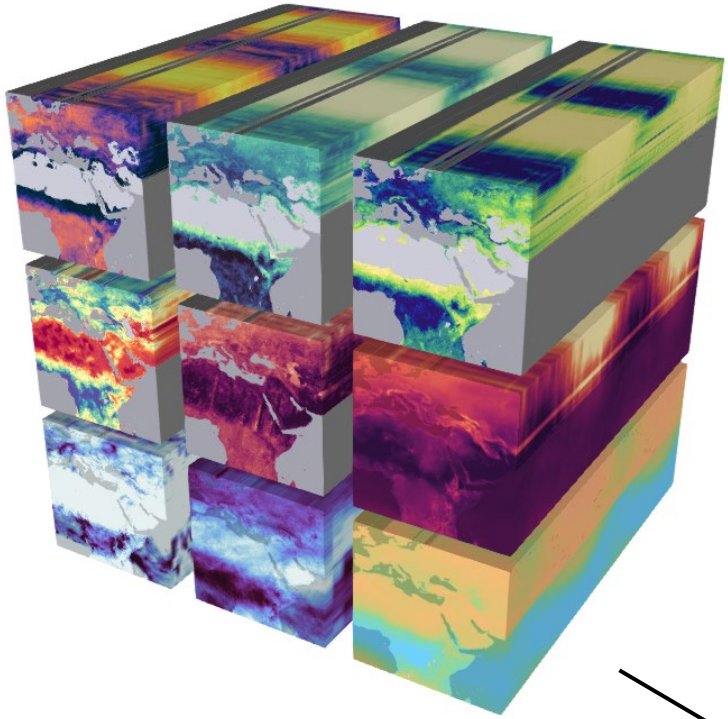
The original Idea

Solve all interoperability issues of downstream data streams

- *Make all downstream Earth data products truly interoperable*
- *Scientists should focus on data exploration only!*
- *Overcome scalability issues etc.*

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

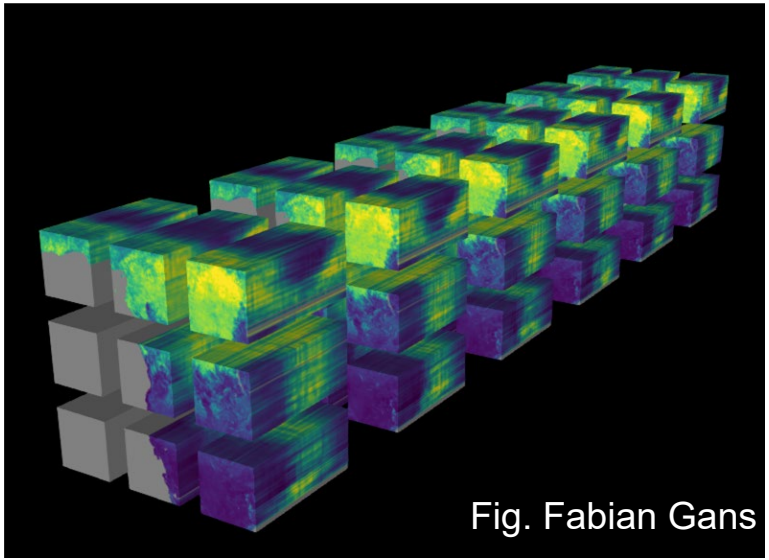
Every dimension is of equal relevance and accessibility



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

Order $\|L\|$	Set of data cubes $\mathcal{C}(L)$	Description of $\mathcal{C}(L)$
0	$\mathcal{C}(\{\})$	Scalar value where no dimension is defined.
1	$\mathcal{C}(\{lat\})$	Univariate latitudinal profile.
1	$\mathcal{C}(\{lon\})$	Univariate longitudinal profile.
1	$\mathcal{C}(\{time\})$	Univariate time series.
1	$\mathcal{C}(\{var\})$	Single multivariate observation.
2	$\mathcal{C}(\{lat, lon\})$	Univariate static geographical map.
2	$\mathcal{C}(\{lat, time\})$	Univariate Hovmöller diagram: zonal pattern over time.
2	$\mathcal{C}(\{lat, var\})$	Multivariate latitudinal profile.
2	$\mathcal{C}(\{lon, time\})$	Univariate Hovmöller diagram: meridional pattern over time.
2	$\mathcal{C}(\{lon, var\})$	Multivariate longitudinal profile.
2	$\mathcal{C}(\{time, var\})$	Multivariate time series.
2	$\mathcal{C}(\{time, freq\})$	Univariate time frequency plane.
3	$\mathcal{C}(\{lat, lon, time\})$	Univariate data cube.
3	$\mathcal{C}(\{lat, lon, var\})$	Multivariate map, e.g. a global map of different soil properties.
3	$\mathcal{C}(\{lat, time, var\})$	Multivariate latitudinal Hovmöller diagram.
3	$\mathcal{C}(\{lon, time, var\})$	Multivariate longitudinal Hovmöller diagram.
3	$\mathcal{C}(\{time, freq, var\})$	Multivariate spectrally decomposed time series.
4	$\mathcal{C}(\{lat, lon, time, var\})$	Multivariate spatiotemporal cube.
4	$\mathcal{C}(\{lat, lon, time, freq\})$	Univariate spectrally decomposed data cube.
5	$\mathcal{C}(\{lat, lon, time, var, ens\})$	Multivariate ensemble of model simulations.

Data cubes should live in the cloud

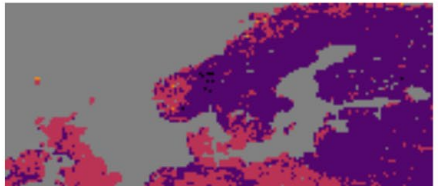


Accessible for the user via jupyter labs e.g. in Julia via

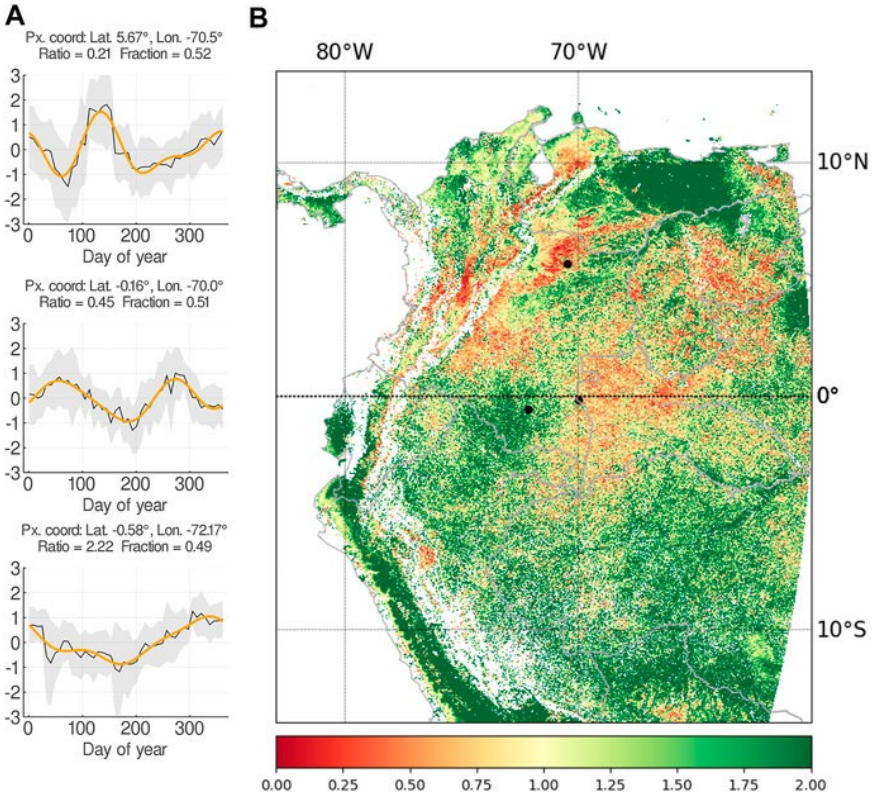
- DiskArrays.jl *(operations on any array data …)*
- YAXArrays.jl *(out of core operations on arrays)*
- EarthDataLab.jl *(specific functionality for Earth)*

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

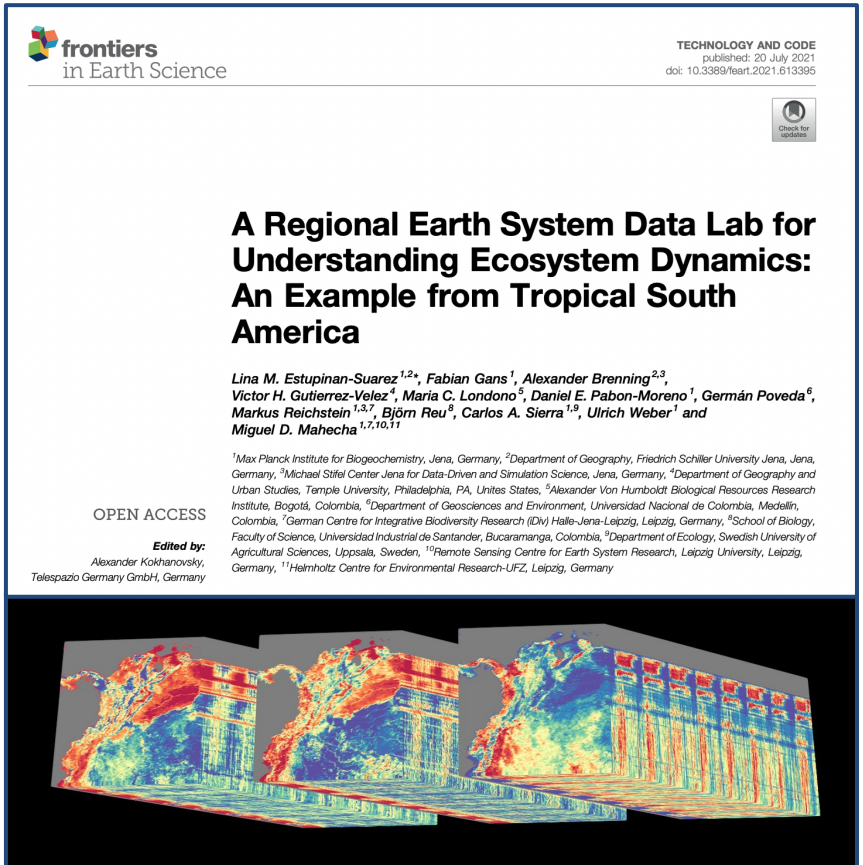
```
[ ]: function sufficient_dimensions(xin::AbstractArray, expl_var::Float64 = 0.95)

       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

[ ]: cube_int_dim = mapslices(sufficient_dimensions, cube_fill, dims = ("Time","Variable"))

[36]: plotMAP(cube_int_dim)

[36]:
```



Developing regional data labs



Colour values of >1 → Dominant annual cycle in comparison to the semiannual mode.



frontiers in Earth Science

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

Check for updates

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, Unites 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

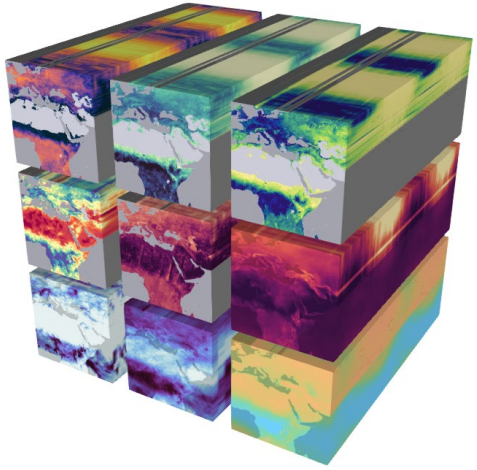
OPEN ACCESS

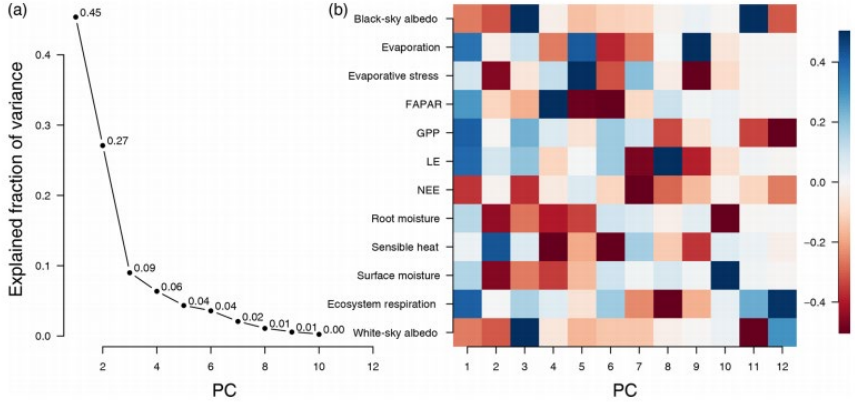
Edited by:
Alexander Kokhanovsky, Telespazio Germany GmbH, Germany

Scientific examples

Low-dimensional trajectories in the land-surface

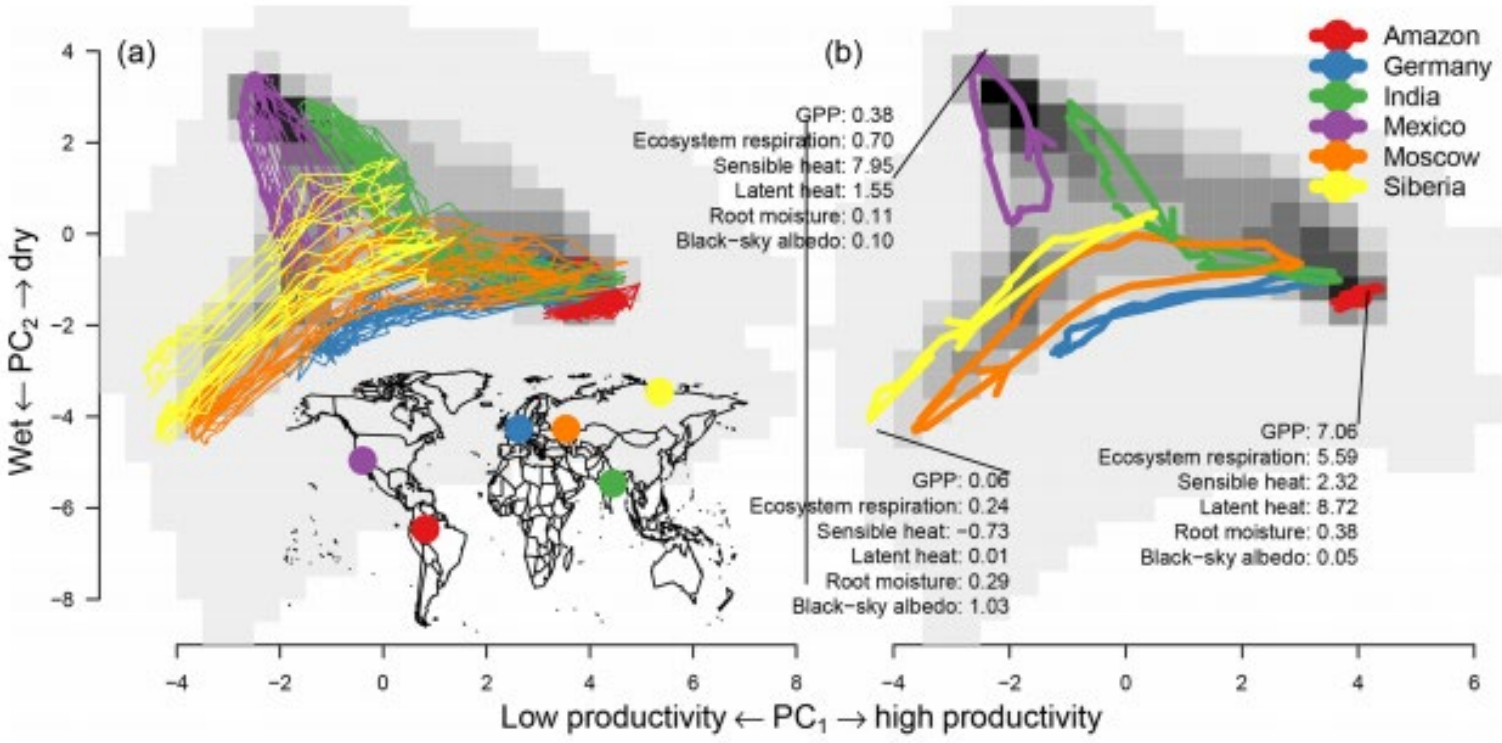
- What are the intrinsic dimensions of change”?
- What is the meaning of such indicators?
- What is their temporal dynamics?





Kraemer, Camps-Valls, Reichstein, Mahecha (2020)
Biogeosciences, **17**, 2397–2424

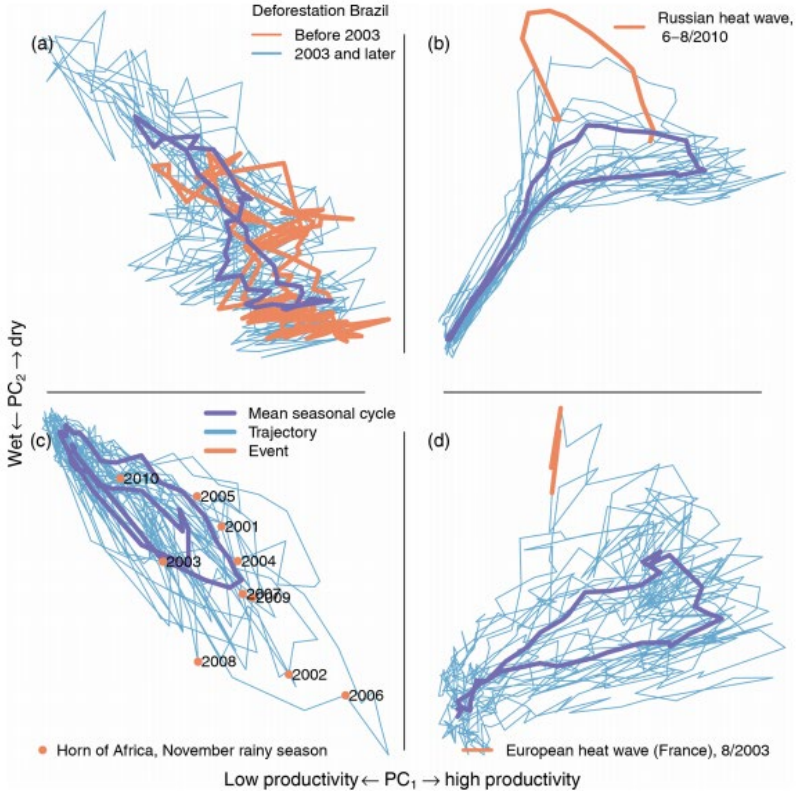
Low-dimensional trajectories in the land-surface



Kraemer, Camps-Valls, Reichstein, Mahecha (2020)
Biogeosciences, 17, 2397–2424

Low-dimensional trajectories in the land-surface

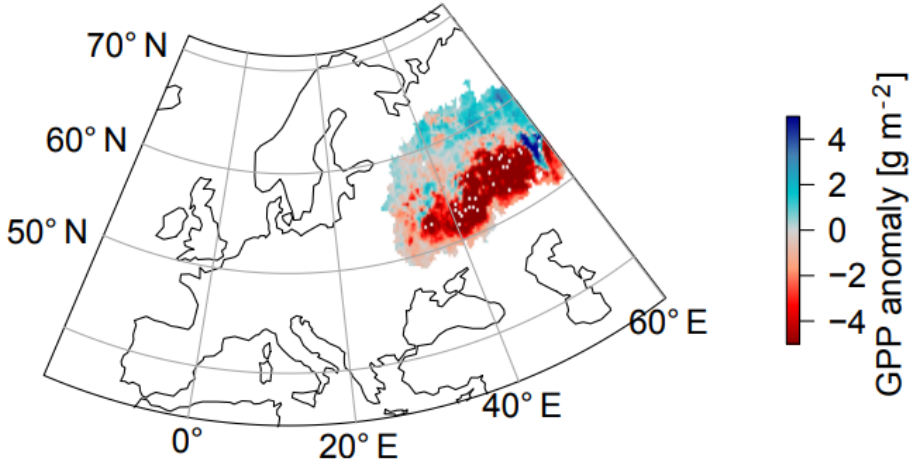
- Reflect deforestation
- Extreme events
- Long-term trends
- Modulations in amplitudes

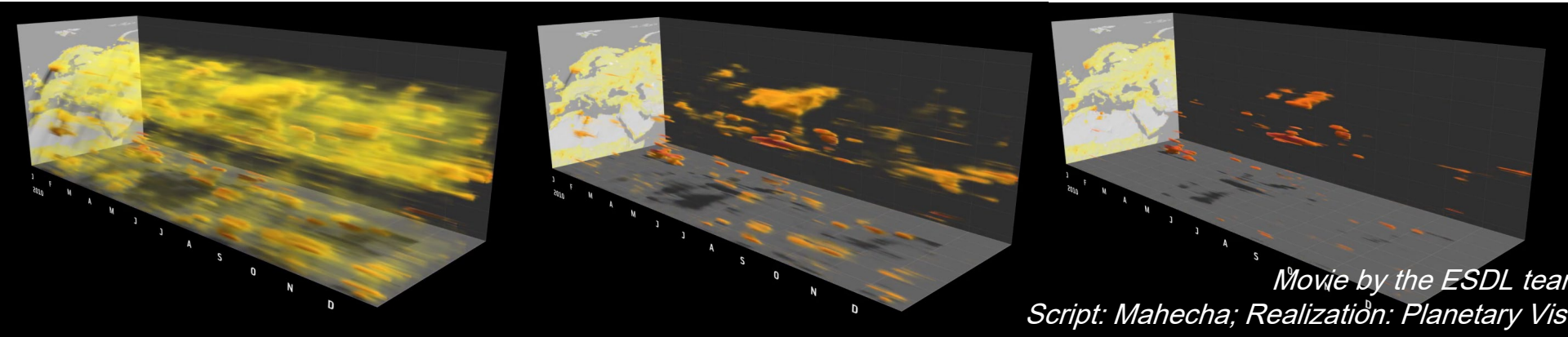


Kraemer, Camps-Valls, Reichstein, Mahecha (2020)
Biogeosciences, **17**, 2397–2424

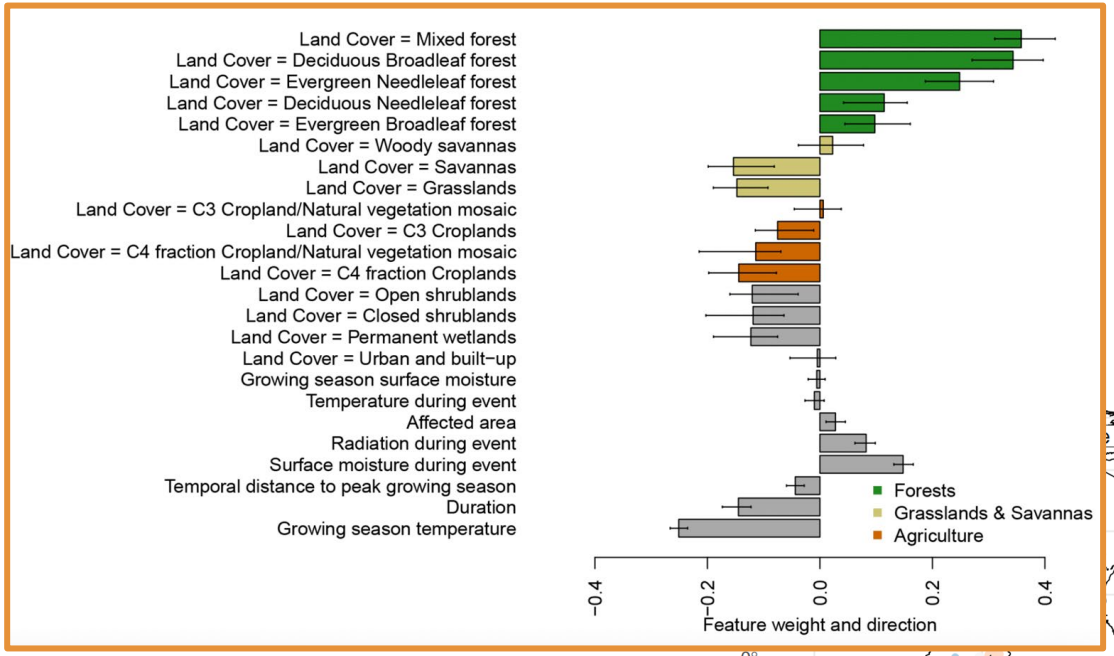
Extreme events in the ESDL

Flach et al. (2018) Biogeosciences, 15, 6067-6085



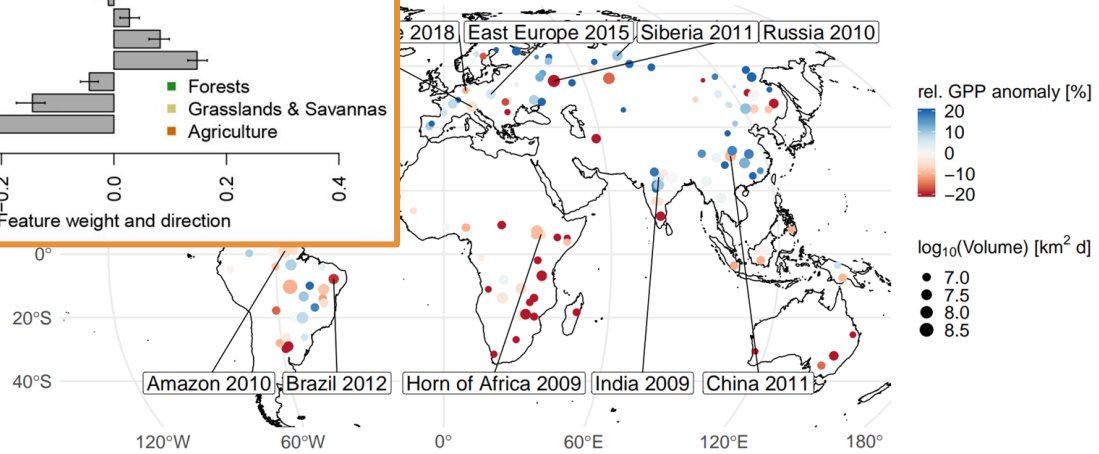


Extreme events in the ESDL - Global study



Explaining with ML what determines the direction of impacts

*Flachet al. (2021)*Biogeosciences https://doi.org/10.5194/bg-2020 -80,



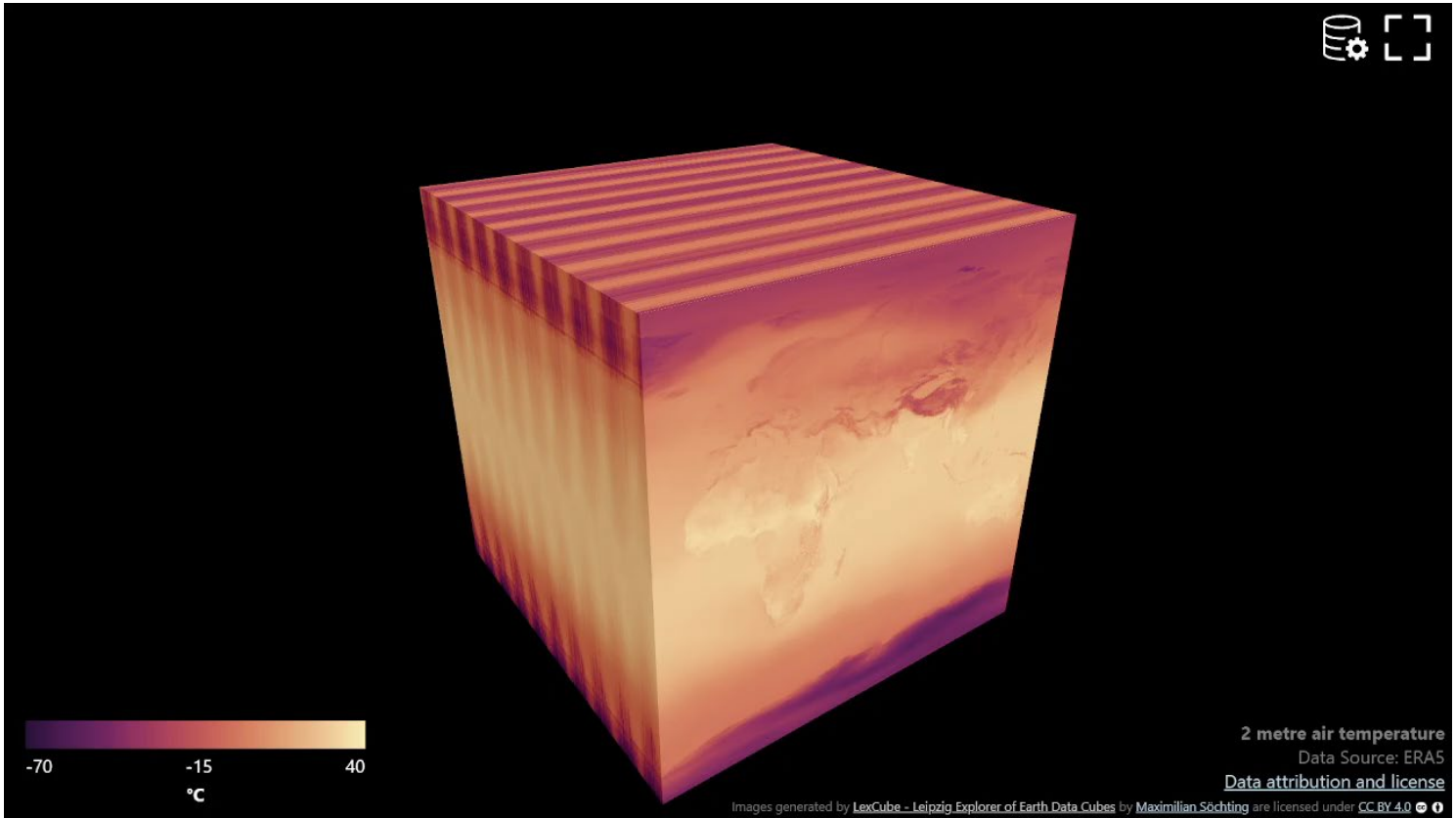
Emerging wishlist

A very very long wish list emerged …

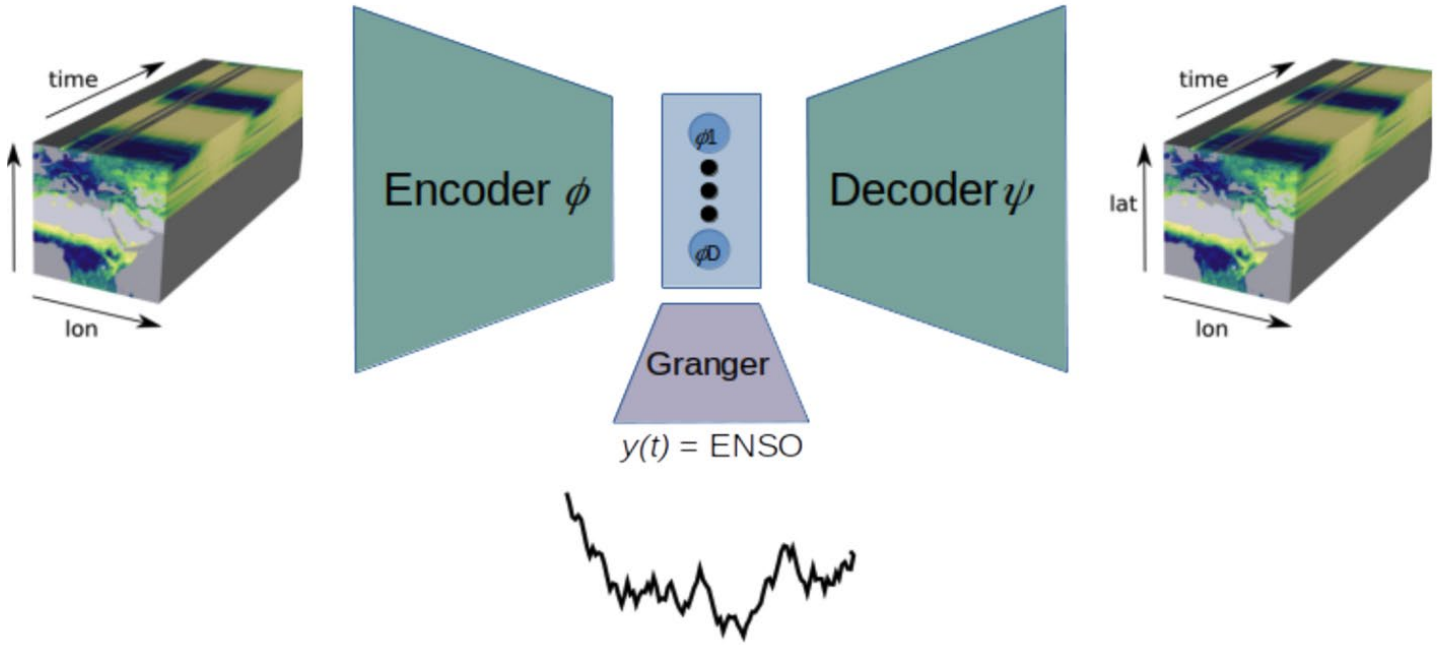
- On-demand data cubes for any application
- High-res global data cubes
- ML adjusted data cube
- On-the fly visualization …
- *And many many more*

Interactive visualization, see: *lexcube.org*

A PhD project by Maximilian Söchting – and supported by the German Science Foundation via the NFDI4Earth



Machine learning readiness



Varando, G., Fernández-Torres, M. A., & Camps-Valls, G. (2021) https://www.climatechange.ai/papers/icml2021/34

Future

Aim of the follow-up project: DeepESDL

- ESDL to become an **integrator of scientific data products** from different activities in a single infrastructure.
- 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** from ESA, particularly from the Science Clusters or from the scientific community worldwide.
- Enable implementation of **AI workflows on ESA data suite** for understanding Earth System Dynamics.

… and study in Leipzig! ☺

UNIVERSITÄT LEIPZIG

M. Sc. Earth System Data Science and Remote Sensing

International Master Program
Starting Oct 2022

Remote sensing
- Multispectral
- Hyperspectral
- Spectroscopy
- Radar & Lidar
- Ground-truthing

Data science
- Big data
- Machine learning
- Timeseries analysis
- Geostatistics
- Data management

Earth system data science

Specialization
- Earth sciences
- Physical geography
- Biodiversity science
- Climatology
- ...

2 years
120 ECTS
Teaching in English

'Facing environmental problems with the tools of our time'

A study program of the Remote Sensing Centre for Earth System Research
More information: http://rsc4earth.de/