The ESA Earth System Lab:

A light-weight data cube approach

Miguel Mahecha[1], Gunnar Brandt[2], Fabian Gans[1], Norman Fomferra[2],
Carsten Brockmann[2], Markus Reichstein[1]
et al.

[1]Max Planck Institute for Biogeochemistry
[2]Brockmann Consult GmbH

November 15, 2018

Idea and Concept

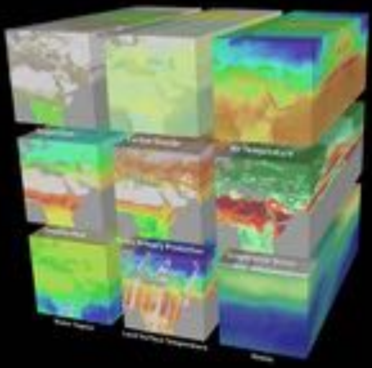
Implementation

Scientific Applications

Participation

Outlook

Idea & concept



Our study object

Intertwined Earth System:

- *How does the metabolism of the Earth system work?*
- *What do we need to know about the couplings across all subsystems?*
- *What trajectory do we follow these days?*

Figure: Composite by the MPI-BGC

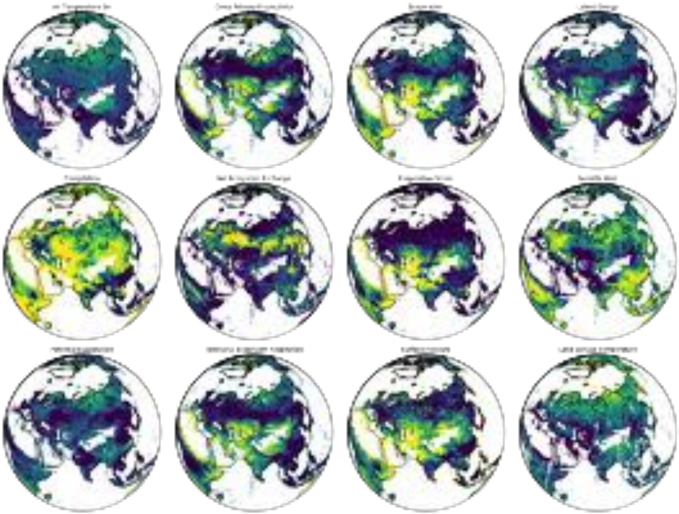
Our study object

Intertwined Earth System:

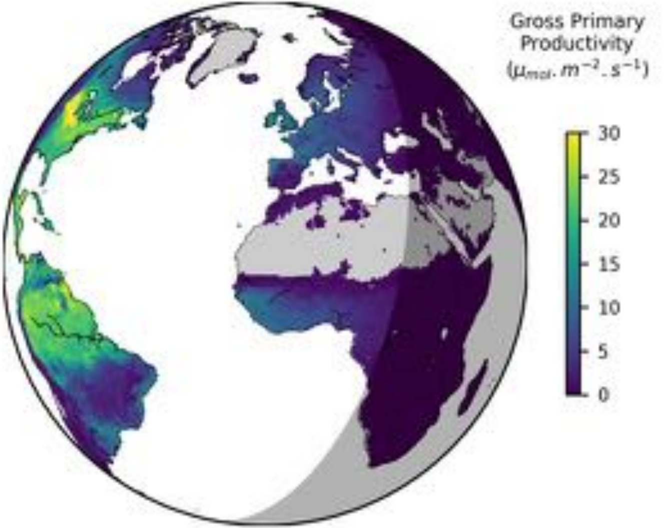
- *How does the metabolism of the Earth system work?*
- *What do we need to know about the couplings across all subsystems?*
- *What trajectory do we follow these days?*

Figure: Composite by the MPI-BGC

Multiple new relevant downstream data products



Multiple new relevant downstream data products



Bodesheim et al. (2018) *Earth System Science Data*, 10, 1327–1365

What we need for our research

Requirements emerged from"Integrated Land-Ecosystem Atmosphere Process Study (iLEAPS)" user workshops.

Required

- Wide range of analysis-ready downstream data
- Virtual laboratory for Earth system scientists

Not supported now

- Low-level Sentinel data analytics
- Sensor data processing
- Data generation

→ *A light weight data cube approach*

What we need for our research

Requirements emerged from“Integrated Land-Ecosystem Atmosphere Process Study (iLEAPS)” user workshops.

Required

- Wide range of analysis-ready downstream data
- Virtual laboratory for Earth system scientists

Not supported now

- Low-level Sentinel data analytics
- Sensor data processing
- Data generation

→ *A light weight data cube approach*

What we need for our research

Requirements emerged from"Integrated Land-Ecosystem Atmosphere Process Study (iLEAPS)" user workshops.

Required

- Wide range of analysis-ready downstream data
- Virtual laboratory for Earth system scientists

Not supported now

- Low-level Sentinel data analytics
- Sensor data processing
- Data generation

$\rightarrow$ *A light weight data cube approach*

A pretty radical data cube idea . . .

The ESDL cube C is a triplet $C = (L, G, X)$

- L is the set of axes labels $L = \{lat, lon, time, var\}$
- G are the corresponding grids

$$\begin{aligned} G &= \bigtimes_{l \in L} grid(l) \\ &= grid(lat) \times grid(lon) \times grid(time) \times grid(var) \end{aligned}$$

- X is a collection of univariate data $\{(X_i)\}_{i \in G} \subseteq \mathbb{R}_{NA} := \mathbb{R} \cup NA$

A pretty radical data cube idea ...

The ESDL cube C is a triplet $C = (L, G, X)$

- L is the set of axes labels $L = \{lat, lon, time, var\}$
- G are the corresponding grids

$$\begin{aligned} G &= \bigtimes_{l \in L} grid(l) \\ &= grid(lat) \times grid(lon) \times grid(time) \times grid(var) \end{aligned}$$

- X is a collection of univariate data $\{(X_i)\}_{i \in G} \subseteq \mathbb{R}_{NA} := \mathbb{R} \cup NA$

A pretty radical data cube idea . . .

The ESDL cube C is a triplet $C = (L, G, X)$

- L is the set of axes labels $L = \{lat, lon, time, var\}$
- G are the corresponding grids

$$\begin{aligned} G &= \bigtimes_{l \in L} grid(l) \\ &= grid(lat) \times grid(lon) \times grid(time) \times grid(var) \end{aligned}$$

- X is a collection of univariate data $\{(X_i)\}_{i \in G} \subseteq \mathbb{R}_{NA} := \mathbb{R} \cup NA$

. . . to apply user defined functions on a cube

- Example: calculating a scalar with f_s

$$f_s : C(\{lat, lon, time\}) \rightarrow C(\emptyset)$$

- Example: spectral decomposition

$$f_d : C(\{time\}) \rightarrow C(\{time, freq\})$$

. . . to apply user defined functions on a cube

- Example: calculating a scalar with f_s

$$f_s : C(\{lat, lon, time, var, model\}) \rightarrow C(\{var, model\})$$

- Example: spectral decomposition

$$f_d : C(\{time, lat, lon, time, var\}) \rightarrow C(\{time, freq, lat, lon, time, var\})$$

Idea and Concept

Implementation

Scientific Applications

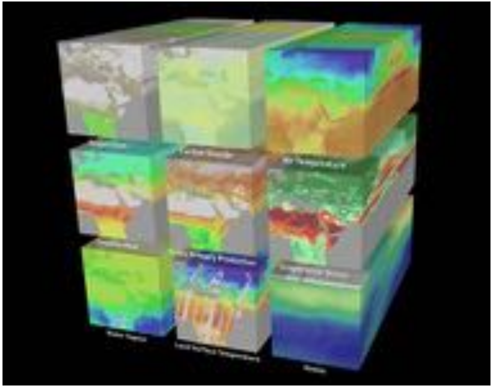
Participation

Outlook

Implementation

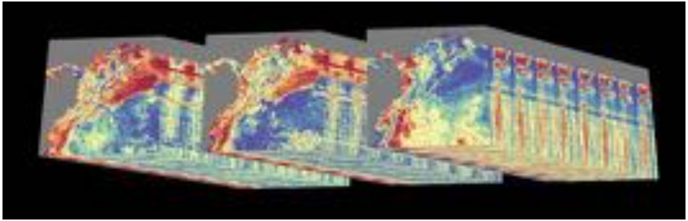
The overall approach

The apparent approach



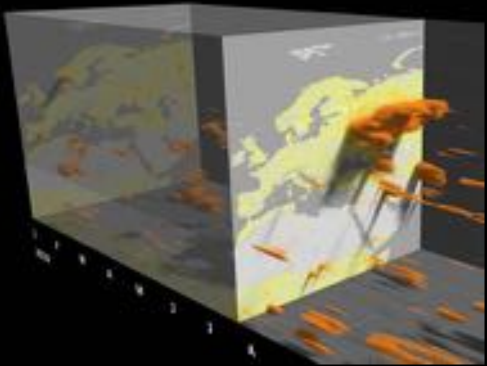
Data cube flavors

Coverage	Global ($\frac{1}{12}^{\circ}$, $\frac{1}{4}^{\circ}$, 1°); Regional (1km); 8d or monthly
Contents	$>$80 variables, focus on ESA CCI data
Optimization	For different use cases, for the cloud
Principle	Open source project

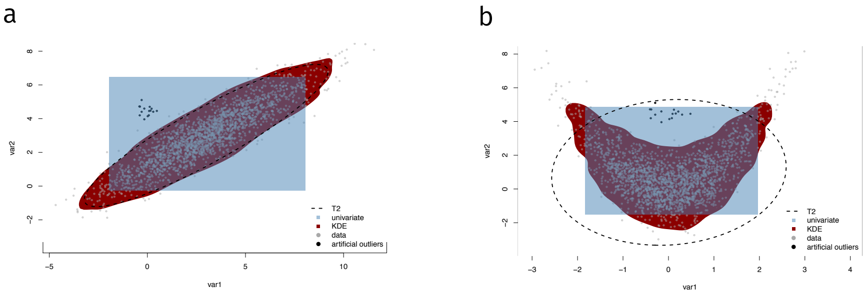


With the Alexander von Humboldt Institute, Bogota; GEO-BON Colombia, and Temple University, Philadelphia.

Research background



1. Revisting the 2010 Russian Heatwave



with workflows capable of dealing with multivariate correlated data,

- Kernel Densitiy Estimation
- *k*-nearest neighbors
- Recurrences
- Mahalanobis Distance
- ...

PhD thesis: Milan Flach; Flach et al. (2017) Earth System Dynamics, 8, 677-696.

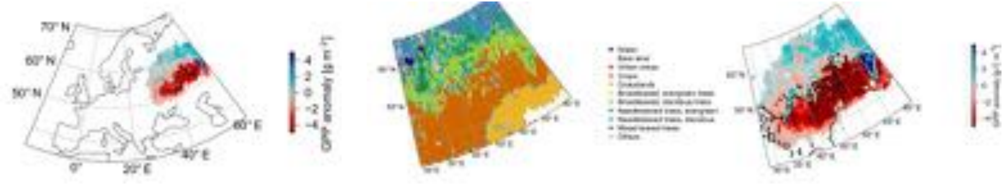
1. Revisting the 2010 Russian Heatwave

- Explaining the discrepancy between “atmospheric” vs. “biospheric” perspective on the same hydrometeorological extreme.

Flach et al. (2018) Biogeosciences, 15, 6067-608.

1. Revisting the 2010 Russian Heatwave

- Explaining the discrepancy between “atmospheric” vs. “biospheric” perspective on the same hydrometeorological extreme.



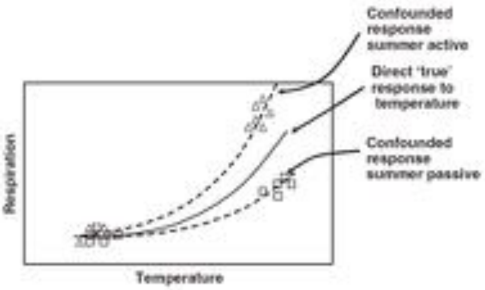
Flach et al. (2018) Biogeosciences, 15, 6067-608.

2. Model optimization

$$R(i) = R_{\mathrm{b}}(i) Q_{10}^{\frac{T(i) - T_{\mathrm{ref}}}{10\mathrm{K}}}$$

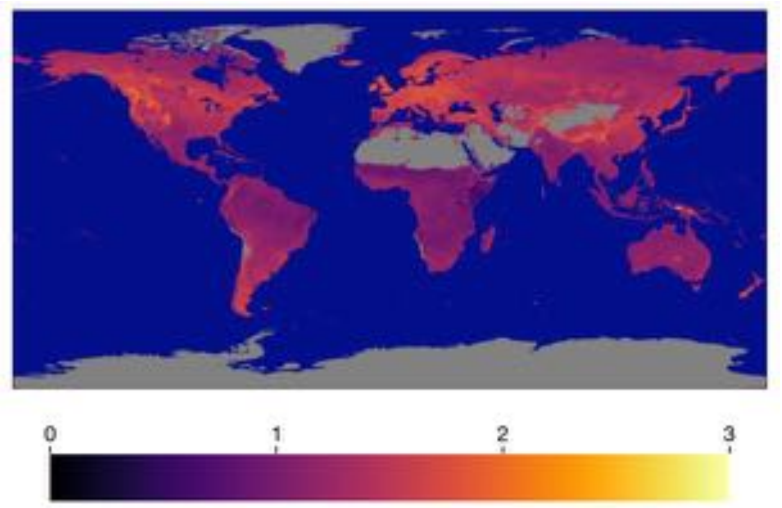
where

- R Respiration
- R_b Base respiration
- Q_{10} Sensitivity
- T Ambient temperature
- T_{ref} Reference temperature
- i Time index



R_{b} is covarying with T and therefore confounding the estimation of Q_{10} (Reichstein & Beer (2008), J. Plant Nutr. Soil Sci., 171, 344–354.).

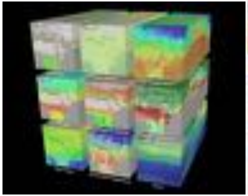
2. Model optimization



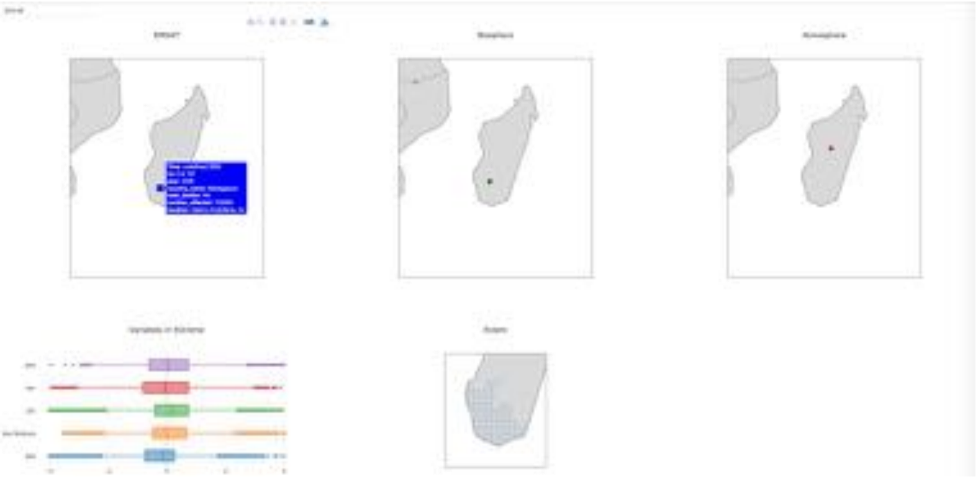
Nora Linscheid; method Mahecha et al. (2010) Science, 329, 838–840.

3. Human-environment nexus during climate extremes

Tap into the full potential of the existing data sources e.g.



3. Human-environment nexus during climate extremes



- Idea and Concept
- Implementation
- Scientific Applications
- Participation
- Outlook

3. Human-environment nexus during climate extremes

Participation

Three mechanisms

Register and try

- Registered users via GitHub account

Open early adopter call

- Supporting 30 early career scientists to realize own ideas/proposals

Champion users

- Center for Research on the Epidemiology of Disasters (CRED)
- Plymouth Marine Lab
- Alexander v. Humboldt Institute, Bogota

Three mechanisms

Register and try

- Registered users via GitHub account

Open early adopter call

- Supporting 30 early career scientists to realize own ideas/proposals

Champion users

- Center for Research on the Epidemiology of Disasters (CRED)
- Plymouth Marine Lab
- Alexander v. Humboldt Institute, Bogota

Three mechanisms

Register and try

- Registered users via GitHub account

Open early adopter call

- Supporting 30 early career scientists to realize own ideas/proposals

Champion users

- Center for Research on the Epidemiology of Disasters (CRED)
- Plymouth Marine Lab
- Alexander v. Humboldt Institute, Bogota

The Earth System Data Lab

Current project

- a virtual lab to explore global Earth system patterns
- fully user driven
- open source

Future ideas

- coupling to visual analytics
- expanding to model-archives
- . . . your input

We thank the

esa
European Space Agency

for excellent support!