The ESA Earth System Lab: A light-weight data cube approach

Miguel Mahecha\textsuperscript{1}, Gunnar Brandt\textsuperscript{2}, Fabian Gans\textsuperscript{1}, Norman Fomferra\textsuperscript{2}, Carsten Brockmann\textsuperscript{2}, Markus Reichstein\textsuperscript{1} \textbf{et al.}

\textsuperscript{1}Max Planck Institute for Biogeochemistry
\textsuperscript{2}Brockmann Consult GmbH

November 15, 2018
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

→ 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 $L = \{\text{lat}, \text{lon}, \text{time}, \text{var}\}$
- $G$ are the corresponding grids
  
  $$G = \prod_{l \in L} \text{grid}(l)$$
  $$= \text{grid(lat)} \times \text{grid(lon)} \times \text{grid(time)} \times \text{grid(var)}$$

- $X$ is a collection of univariate data $\{(X_i)_{i \in G} \subseteq \mathbb{R}_{\text{NA}} : = \mathbb{R} \cup \text{NA}\}$
A pretty radical data cube idea . . .

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

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

$$G = \prod_{l \in L} \text{grid}(l)$$

$$= \text{grid(lat)} \times \text{grid(lon)} \times \text{grid(time)} \times \text{grid(var)}$$

- $X$ is a collection of univariate data $\{(X_i)_{i \in G} \subseteq \mathbb{R}_\text{NA} := \mathbb{R} \cup \text{NA}\}$
A pretty radical data cube idea . . .

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

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

$$G = \prod_{l \in L} \text{grid}(l)$$

$$= \text{grid(lat)} \times \text{grid(lon)} \times \text{grid(time)} \times \text{grid(var)}$$

- $X$ is a collection of univariate data $\{(X_i)\}_{i \in G} \subseteq \mathbb{R}_{\text{NA}} := \mathbb{R} \cup \text{NA}$
to apply user defined functions on a cube

- Example: calculating a scalar with $f_s$

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

- Example: spectral decomposition

  $$f_d : C(\{\text{time}\}) \rightarrow C(\{\text{time}, \text{freq}\})$$
... to apply user defined functions on a cube

- Example: calculating a scalar with $f_s$

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

- Example: spectral decomposition

  $$f_d : C(\{\text{time, lat, lon, time, var}\}) \to C(\{\text{time, freq, lat, lon, time, var}\})$$
Implementation

Earth System Data Cube (ESDC)

The Earth System Data Cube (ESDC) is a multi-scale data cube that models the Earth system on a common grid and shares a common data model. The ESDC also supports a Data Analytics Model for use. The ESDC provides a computing environment to access, analyze, and visualize the data set.

This project is to foster a cross-disciplinary understanding of the Earth system by providing an advanced and flexible framework at the planetary science data center accessed at the international community data set of spatial and temporal data.
The overall approach
The apparent approach
Data cube flavors

**Coverage**  Global \((\frac{1}{12} \degree, \frac{1}{4} \degree, 1 \degree)\); 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. Revisiting the 2010 Russian Heatwave

with workflows capable of dealing with multivariate correlated data,

- Kernel Density Estimation
- $k$-nearest neighbors
- Recurrences
- Mahalanobis Distance
- . . .

1. Revisiting the 2010 Russian Heatwave

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

1. Revisiting the 2010 Russian Heatwave

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

2. Model optimization

\[ R(i) = R_b(i)Q_{10}^{\left( T(i) - T_{ref} \right)/10K} \]

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

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
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

European Space Agency

for excellent support!