

ESD-review - Earth system data cubes unravel global multivariate dynamics

The authors present a new data cube approach, called Earth System Data Cube, where multiple spatiotemporal data streams are treated as one singular, very high-dimensional data stream.

This paper is of high quality and clearly outlines the authors' reasoning of implementing the ESDC in the way it is implemented and the advantages it brings to Earth System dynamics studies.

The ESDC datacube approach brings different data streams to a common grid, which is beneficial for specific scientific Earth System studies. Common operations, as e.g. resampling and bringing the data to the same spatial and temporal resolution, are just taken care of and users can focus more on science.

Yet, the unified grid from the beginning might be a limitation for other application areas.

Some comments and areas for minor revision:

What is not so clear for me so far is the overall implementation strategy of the ESDC. I understand that I can now use the ESDC with the datasets as outlined in the paper appendix. Is there the option that users can extend the ESDC data list? Or will it be similar to the OpenDataCube concept where multiple implementations of the ESDC concept can be set up with different data? Where is the current ESDC data hosted? And how long does the data preparation take? I suggest to bring in these additional aspects.

What are the bottlenecks in gathering the data? I understand that in order to implement the ESDC, data has to be moved from the respective data repositories. Is this correct? Would be good to elaborate on this aspect as well.

It is without doubt that Julia is a powerful and efficient programming language. However, the fact that the ESDC package has been developed in Julia could be a restrictive factor in the uptake, as I argue that most potential users of the ESDC use Python or R and might be less motivated to take up a new programming language. Are there plans to extend it to Python or R?

The authors argue that the ESDL is closest to the Climate Data Store. In my current understanding of both data systems, I would disagree. Do the authors talk about the Climate Data Store, which is primarily a data dissemination system, or the Climate Data Store toolbox, which is the processing editor on top of the CDS?

It would be necessary to elaborate more on what aspects both systems are close and why the authors come to this assessment. It is also important that the authors differentiate between the CDS and the CDS toolbox.

Some additional typing errors discovered:

page 2 line 21: remove 'the' many key processes

page 4 line 109: remove 'are' ... variable only can be considered

page 8 line 174: form instead of from

page 11 line 230: remove 'computing' ... scientific programming language

page 12 line 284 don't understand the last part of the sentence 'data speed up data access

page 12 line 287 add 'be' 'would be to treat every ...

page 13 line 303 behaviour is doubled 'long-term system behaviour'

page 17 line 370 'confirm' instead of confirms

page 19 line 428 'these findings' instead of finding

page 21 line 456 a unique data grid instead of an unique datagrid

page 21 line 475 rephrase sentence, hard to understand

page 22 line 510 'several' is double