Geophysical Research Abstracts
Vol. 20, EGU2018-8609, 2018
EGU General Assembly 2018

Towards an optimised environmental data compression method for structured model output

Ugur Cayoglu (1), Peter Braesicke (2), Tobias Kerzenmacher (2), Jörg Meyer (3), and Achim Streit (3)

(1) Steinbuch Centre for Computing (SCC) & Institute of Meteorology and Climate Research - Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe Institute of Technology (KIT), Eggenstein-Leopoldshafen, Germany (ugur.cayoglu@kit.edu), (2) Institute of Meteorology and Climate Research - Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe Institute of Technology (KIT), Eggenstein-Leopoldshafen, Germany, (3) Steinbuch Centre for Computing (SCC), Karlsruhe Institute of Technology (KIT), Eggenstein-Leopoldshafen, Germany

New climate models and more powerful high-performance computers make it possible to generate ever higher resolution output in the environmental sciences. A dedicated compression procedure for climate data can help to reduce the amount of space required for data storage, transfer and processing. In order to develop such a procedure, it is necessary to analyse which temporal and spatial information can help to determine redundant data. For this, we looked at methods from different (related) fields of science and applied them to climate data: Recurrence analysis used in neurosciences, Shannon and sample entropy from an information-theoretical point of view, trend analysis commonly used in environmental science, and classical variance analysis methods such as Empirical Orthogonal Functions or Principal Component Analysis.

Applied to climate data, we can use recurrence, trend and variance analysis to identify the temporal dependency in the data. In addition to temporal dependency, the variance analysis also provides us with information about spatial dependency in the data. The Shannon and sample entropy provides us knowledge about the information content in the data and sets a limit for possible lossless compression.

Our results show that the optimal information space for the identification of redundant data depends on the type of the analyzed variable, the physical location on the earth (latitude, longitude and altitude) for which the information space is to be identified and the current season with the respective temporal resolution of the data set. The analyses carried out promise to provide a good basis for the development of a dedicated compression procedure for climate data.