Project: **1083**

Project title: **Climate Informatics: New machine learning methods for climate data and climate model evaluation**

Principal investigator: **Jakob Runge**

Allocation period: **2018-07-01 to 2019-06-30**

Abstract

The German Aerospace Center (DLR) and Friedrich-Schiller-University (FSU) Jena in collaboration with the Max-Planck Institute for Biogeochemistry established a research group on Climate Informatics at the new DLR Institute of Data Science in Jena, Germany, led by Dr. Jakob Runge (http://www.climateinformaticslab.com/). The group works in close collaboration with the Earth System Model Evaluation (ESMVal) group of the DLR Institute of Atmospheric Physics in Oberpfaffenhofen and with the Climate Modelling Department of the University of Bremen, both led by Prof. Veronika Eyring. Climate Informatics is a challenging and promising research field where a concentrated effort will have a high impact both to advance science and to address climate change topics of critical importance for the society (Monteleoni et al., 2013).

The Earth system is one of the best-observed complex dynamical systems with satellite observations and weather stations providing almost global coverage for the past decades. These datasets are complemented by the output of global climate models that simulate the basic physical processes underlying the atmosphere, oceans, and terrestrial domain, and provide projections of how future climate looks like under different plausible scenarios of anthropogenic influence. Yet, our tools to analyze and understand Earth system data, observations as well as climate model output, are still mainly based on simple descriptive statistics, correlations, and regression analysis.

At the same time, rapid methodological progress has been made in computer science, machine learning, and statistics allowing for much more informative analyses of large-scale datasets. Climate data poses several challenges to such new methods: (1) Climate data is big data on the Petabyte scale, e.g., the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al. (2012)) and upcoming Phase 6 (CMIP6, Eyring et al. 2016, Balaji et al. (2018)) model output, (2) processes interact on vastly different spatio-temporal scales, (3) highly interconnected subprocesses with (4) partially nonlinear interactions.

The goal and mission of the Climate Informatics group is to develop tools that address these challenges with a core methodological expertise on causal discovery (Spirtes et al., 2000, Runge et al., 2015, Runge et al., 2017, nonlinear time series analysis, and deep learning (Goodfellow et al., 2016).

For this project the focus will be on developing, evaluating, and applying methods related to causal discovery utilizing DKRZ's CPUs as well as deep learning requiring DKRZ's GPU nodes. In particular, we plan four subprojects: (i) Developing causal discovery methodology for CMIP models, (ii) application to CMIP model evaluation and intercomparison, (iii) application to systematically finding emergent constraints, and (iv) development and application of deep learning techniques to CMIP model output.

References

Monteleoni, C., Schmidt, G.A., Alexander, F., Niculescu-Mizil, A., Steinhaeuser, K., Tippett, M., Banerjee, A., Blumenthal, M.B., Ganguly, A.R., Smerdon, J.E. and Tedesco, M., 2013. Climate Informatics. In: N.C.

T. Yu, S. Simoff (Editor), Com-putational Intelligent Data Analysis for Sustainable Development; Data Mining and Knowledge Discovery Series. CRC Press, Taylor & Francis Group.

Taylor, K. E., et al. (2012). "An Overview of CMIP5 and the Experiment Design." Bulletin of the American Meteorological Society **93**(4): 485-498.

Eyring, V., et al. (2016). "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization." Geosci. Model Dev. 9(5): 1937-1958.

Balaji, V., et al. (2018). "Requirements for a global data infrastructure in support of CMIP6." Geosci. Model Dev. Discuss. 2018: 1-28.

Spirtes, P., Glymour, C. & Scheines, R. Causation, Prediction, and Search, vol. 81 (The MIT Press, Boston, 2000).

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., Kurths, J., 2015. Identifying causal gateways and mediators in complex spatio-temporal systems. Nature Communications, 6, 8502.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D., 2017. Detecting causal associations in large nonlinear time series datasets. Http://Arxiv.Org/Abs/1702.07007.

Goodfellow, I., Bengio, Y. & Courville, A. Deep learning . (MIT press, 2016).