Report for project 1083 "Climate Informatics: New machine learning methods for climate data and climate model evaluation"

Project:	**1083**
Project title:	**Climate Informatics: New machine learning methods for climate data and climate model evaluation**
Project leader:	**Dr Jakob Runge**
Allocation period:	**2018-07-01 to 2019-06-30**
Report period:	**2018-07-01 to 2019-04-30**

1 Overview of work during the reporting period

In the reporting period, we have started to set up our group on Climate Informatics and conducted first analyses. Due to delays in recruiting however, only a fraction of the requested resources were used to date and not all tasks have been started. We have just recently successfully hired new Master and PhD students as well as PostDocs who will work on this project at DKRZ. We therefore expect to consume additional resources until the end of the allocation period on 30 June 2019 and to significantly enhance calculations in the next allocation period for the coming year.

In the reporting period, therefore, the analysis mainly focused on one study to detect hurricane activity with the Maximum Divergent Interval (MDI) Algorithm (Barz et al., 2018) as well as some additional first test simulations on emergent constraints. The studies on hurricane with the MDI algorithm so far concentrated on exploring the method with meteorological reanalyses data. We therefore restricted it to rather coarse horizontal and temporal resolution. We have also not yet processed model output from the Coupled Model Intercomparison Project Phase 6 (Eyring et al., 2016) which only now starts to become available. The final analysis of this work will therefore require substantially more computing resources. In addition, other analyses are planned for the next year (see proposal for the allocation period 2019-07-01 to 2020-06-30).

The tasks on developing and applying causal discovery methodology for CMIP model evaluation *(**Task i to iii in 2018 proposal**)* have only now begun with a Master student and an additional PhD student who will start on 1 June 2019. These tasks are therefore not covered in this report.

2 Development and application of deep learning techniques for CMIP model output *(Task iv in 2018 proposal)*

Considering the large amount of geophysical data available from Earth system models and satellite data, it is necessary to find fast and effective ways to analyze them. One important task is the detection of extreme weather events, e.g., tropical cyclones (TC), since they have a high potential for damage. Existing methods are mostly based on threshold criteria of variables and indices (Walsh and Watterson, 1997) which are arbitrary and based on expert assessments (Tang and Monteleoni, 2015). In contrast, anomaly detection methods identify data that is divergent to the remainder, independent of arbitrary thresholds. Since geophysical extreme events occur over space and time, it is necessary to find not only abnormal points but abnormal intervals. The unsupervised MDI method of Barz et al. (2017) fulfills this requirement. The MDI is an unsupervised machine learning algorithm and therefore does not need any training data to detect anomalies. It identifies various patterns of extreme weather events like droughts and mid-latitude cyclones besides tropical storms.

During the reporting period, we applied the MDI on ERA-interim reanalysis data to start evaluating with which settings and variables the MDI performs best in detecting tropical storms. The results shown in this report are taken from preliminary results of a Master thesis by Simon Zitzmann from DLR.

So far, ERA-interim reanalysis data (Dee et al., 2011) spanning the North Atlantic and North Pacific region as shown in Figure 1 were analyzed with a spatial resolution of 1°. The data series starts on 1 January 2000 and ends on 31 December 2010 with a 24-hour time step resulting in 4018 temporal slices. As one important pre-processing step, we keep only the data of the hurricane season starting on 1 June and ending on 30 November of each year. We have so far analyzed the variables geopotential (z in m^2/s^2), vertical velocity (w in pa/s), specific humidity (q in kg/kg), relative humidity (r in %), temperature (t in K), wind speed (computed of u and v in m/s), divergence of wind (d in 1/s), relative vorticity (v_o in 1/s), and potential vorticity (pv in m^2K/kgs). The analyzed height levels are 200, 300, 400, 500, 600, 700, 800, 900 and 1000hPa.

To evaluate whether a detection of the MDI is a tropical storm or not, we use the International Best Track Archive for Climate Stewardship (IBTrACS, Knapp et al. (2010)) storm database version v03r10 that was released in September 2018 (see Figure 2). It provides the most complete set of six-hourly information like position, wind speed and classification of all known historical tropical cyclones and subtropical storms worldwide since 1851.

To evaluate the performance of every single combination, the number of correctly detected storms called 'hits (H)' is counted. Since we want to tune the algorithm with the goal that every spatio-temporal detection box of the MDI contains exactly one storm, one detection box can only refer to one storm. Therefore, the counting algorithm works as following:

- For every single detection box, all ground truths laying in it are listed. In Figure 1 this would be two storms with 5 ground truths in total for detection box 1, and one storm with one ground truth for detection box 2.
- If there is more than one storm listed for one detection box, the storm with the most ground truths is associated with the detection box. The storm name is marked as 'taken' and cannot be given to another detection box. In Figure 1 detection box 1 would be associated with Storm A, detection box 2 would be associated with Storm B.

Detection boxes that are not associated with any storm are termed *false alarm (F)*. If a storm in the IBTrACS dataset is not linked to any detection box, this is called a *miss (M)*. Furthermore, some categorical statistical measures (Ebert and McBride, 2000) are computed. The *probability of detection (POD)* is defined as *POD = H / (H + M)*. This measure is the total number of hits divided by the total number of observed TCs and gives the probability by which an observed TC is detected by the MDI. The *false alarm ratio (FAR)* measures how many of the detected TCs are falsely detected. It is defined as *FAR = F / (H + F)*.

Figure 3 provides a snapshot of PODs and FARs for the MDI analysis with single variables done so far. The detection rates are still below 0.5 and are thus on an unsatisfactory level, even though there are promising variables like the vertical velocity, specific humidity, geopotential, relative vorticity and temperature. First results of multivariate runs lower the false alarm ratio significantly, but did not improve the probability of detection. Further work is required to fully exploit the potential of the multivariate capabilities of the algorithm to outperform the univariate counterparts. Another potential lies in further filtering the data.
As a next step, multivariate runs will be performed in order to check if they reach higher detection rates. As soon as the detection rate reaches a satisfactory level, it will potentially be possible to apply the MDI on CMIP model output to study changes of TC frequency in the historical period and in the future.

Figures

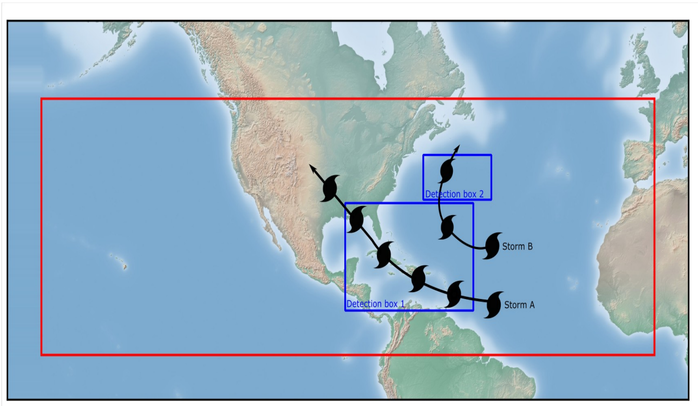

Figure 1: The red box shows the analyzed area of ERA-interim data spanning 180 °E to 360 °E and 0 °N to 40 °N. The black storm tracks and blue detection boxes schematically indicate how the detections of the MDI are matched to the IBTrACS storm positions.

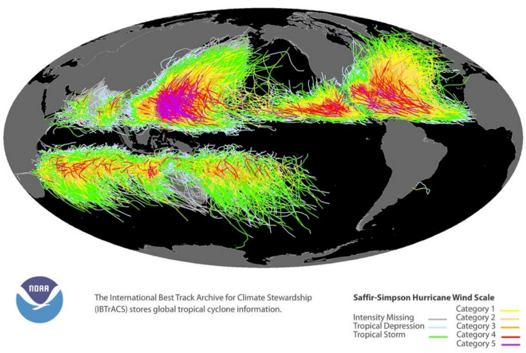

Figure 2: International Best Track Archive for Climate Stewardship (IBTrACS) storm positions.

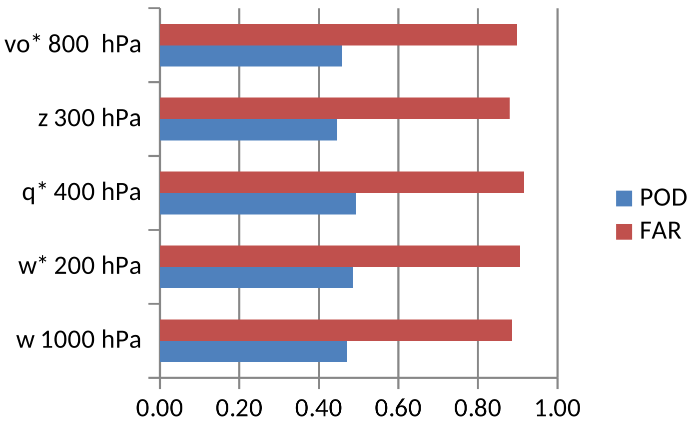

Figure 3: Current probability of detection (POD) and false alarm ratio (FAR) for selected test cases analyzed.

References:

Barz, B., Rodner, E., Guanche Garcia, Y. and Denzler, J., 2018. Detecting Regions of Maximal Divergence for Spatio-Temporal Anomaly Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.N. and Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656): 553-597.

Ebert, E.E. and McBride, J.L., 2000. Verification of precipitation in weather systems: determination of systematic errors. Journal of Hydrology, 239(1-4): 179-202.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev., 9(5): 1937-1958.

Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J. and Neumann, C.J., 2010. THE INTERNATIONAL BEST TRACK ARCHIVE FOR CLIMATE STEWARDSHIP (IBTrACS) Unifying Tropical Cyclone Data. Bulletin of the American Meteorological Society, 91(3): 363-+.

Tang, C. and Monteleoni, C., 2015. Can Topic Modeling Shed Light on Climate Extremes? Computing in Science & Engineering, 17(6): 43-52.

Walsh, K. and Watterson, I.G., 1997. Tropical cyclone-like vortices in a limited area model: Comparison with observed climatology. Journal of Climate, 10(9): 2240-2259.

Antrag auf Verlängerung des Projekts 1083 „Climate Informatics”

REVIEWER COMMENTS ON PREVIOUS PROPOSAL and RESPONSES:

Comment: This proposal is a very unusual one in several aspects. First, it seems to be based on an unreviewed project except that it is related to the recently established research group of the pi at the dlr institute for informatics in Jena, this should be clarified in any case.

Response: The project implements one of the core topics of the recently established Climate Informatics research group at the DLR institute for informatics in Jena. The group was established after strict internal peer review at DLR. Furthermore, the envisaged causal discovery methodology for CMIP model evaluation was proposed in a recently accepted Nature Communications paper (Runge et al. 2019). The ClimateIinformatics group focuses on the development of causal discovery and machine learning (ML) algorithms and works in close collaboration with the Earth System Model Evaluation and Analysis Department at DLR Institute of Atmospheric Physics (DLR-IPA) who focuses on applying these methods to Earth system data. This department led by the co-PI of this project, Prof. Veronika Eyring, has the application of ML techniques as main mandate to support the institute's strategy that was again internally reviewed and approved. The work is additionally anchored in a collaborative project with the co-PIs department on Climate Modelling at the University of Bremen (UBremen), funded by the Helmholtz society as part of the W2/W3 programme for female researchers.

Comment: Second, it is a proposal to analyse very large existing data sets from the cmip5 or upcoming cmip6 data without invoking explicit model simulations. The analysis method (causal network inference) is based on a recently and partly published paper by the pi with a very complicated structure apparently programmed in python, from the proposal it remains unclear if HLRE-3 Ressourcen are really required for this new project, Further the planned usage of CMIP5/CMIP6 simulations is too vague and needs to be specified in more detail.

Response: The causal discovery techniques have been published in several evolutions in a number of papers (Runge et al. 2012, Runge et al. 2014, Runge et al. 2015). Indeed, the last evolution is still under peer review and published only in the arXiv currently (Runge et al. 2018). Applying this method on the large-scale CMIP archive requires significant resources, as requested. A master thesis that has started early 2019 under the supervision of Prof. Eyring and Dr. Runge shows promising results and additionally verifies the requested resources. First experiments will show how many resources in detail are required to compare and evaluate climate models. For example, it might turn out that causal networks reconstructed from few atmospheric variables are sufficient to compare models and observations. The experiments are now specified in the new proposal. We focus on analysing pre-industrial control simulations to assess internal variability, historical simulations for model evaluation and intercomparison, and future projections to assess how dependency *structures for the major modes of variability are changing over time under different forcing scenarios.*

Comment: Thirdly, the pi plans to implement a deep learning neural network for further analyses if the network structure but implemented on the mistral gpu nodes, it is hard to believe that they really need to requested amount.

Response: We asked DKRZ beforehand about how much GPU resources we can realistically apply for. The answer was that all GPUs are mostly idle and we can request a high amount. Note that training neural networks always benefits from more computational power to fit larger networks or test more parameters.

Comment: The time plan for the proposed work needs to be specified in more detail, too, as many CMIP6 simulations have not been performed, yet.

Response: Several CMIP6 model simulations are already available and the ESMValTool is already used to analyse several CMIP6 models routinely at DKRZ, so this is not an issue for the current proposal. Also, as written in the last years proposal, the entire work could have also be done with CMIP5 simulations why we do not have any constraints regarding the availability of model simulations to perform the project.

Comment: The suggestion for the a grant is to give the pi very limited access to dkrz resources to allow implementation of the code and effeviency measures of it on standard mistral nodes and on gpu nodes, such that there are sustainable numbers available for the next grant application including the requested clarifications.

Response: Summarizing, we know that this is very much outside the standard computing requests, but we hope that DKRZ is open to this project since the field of Climate Informatics tries to enrich climate science with a plethora of methods from ML and related fields. We also note that due to large delays in recruiting suitable PhDs and Postdocs for the new group, this project is only starting now. However, already today at DLR-IPA/UBremen we have one Master and one PhD student working on applying causal discovery methods to CMIP data, one Master student working on hurricane detection with anomaly detection techniques, a

PostDoc working on droughts detection with anomaly detection techniques, a PhD student working on ML based parametrizations for climate models, and a PhD student who is using ML based techniques to explore the value of weighting multi-model projections based on model performance. At DLR-DW, a new Postdoc started only in March and will work on the causal discovery methods and a PhD student works on emergent constraints. If significant cuts are made to the requested resources, this would have large negative impacts on these important studies.

References:

Runge, Jakob, Sebastian Bathiany, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Marlene Kretschmer, Miguel Mahecha, et al. 2019. "Inferring Causation from Time Series in Earth System Sciences." Nature Communications Perspective article, in press.

Runge, Jakob, Jobst Heitzig, Vladimir Petoukhov, and Jürgen Kurths. 2012. "Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy." Physical Review Letters 108 (25): 258701. https://doi.org/10.1103/PhysRevLett.108.258701.

Runge, Jakob, Vladimir Petoukhov, and Jürgen Kurths. 2014. "Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models." Journal of Climate 27 (2): 720–39. https://doi.org/10.1175/JCLI-D-13-00159.1.

Runge, Jakob, Vladimir Petoukhov, Jonathan F. Donges, Jaroslav Hlinka, Nikola Jajcay, Martin Vejmelka, David Hartman, Norbert Marwan, Milan Paluš, and Jürgen Kurths. 2015. "Identifying Causal Gateways and Mediators in Complex Spatio-Temporal Systems." Nature Communications 6: 8502. https://doi.org/10.1038/ncomms9502.

Runge, Jakob, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. 2018. "Detecting Causal Associations in Large Nonlinear Time Series Datasets." ArXiv:1702.07007v2 [Stat.ME].