5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO



# USING CAUSAL DISCOVERY TO STUDY CONNECTIONS BETWEEN TOA RADIATIVE FLUX AND SURFACE TEMPERATURE

Christian Rodriguez[1], Imme Ebert-Uphoff[1], Yi Deng[2]

***Abstract*—The energy budget of the earth accounts for energy entering from the sun, energy lost to space, and energy stored in the atmosphere and the planet. The exchange of energy between space, the atmosphere and the planet is a very complex process affected by many factors, including surface and atmospheric temperature, surface albedo, the amounts of clouds, aerosols and various trace gases, such as water vapor and carbon dioxide in the atmosphere. A thorough understanding of the earth's energy budget is essential to predict how the climate responds to perturbations in external forcing. In this project we seek to develop a better understanding of the interactions between radiative flux measurements at the top of atmosphere and air/surface temperatures, using methods from causal discovery. This project is still in its initial stages, so this abstract focuses on the basic methodology and illustrates it with initial results from some first test runs.**

## I. BACKGROUND AND MOTIVATION

*Causal discovery* is a machine learning technique that seeks to identify potential cause-effect relationships from observational data. We use *probabilistic graphical models* for this purpose [1], [2]. (Related methods include *Gaussian graphical models* [3] and *Granger graphical models* [4], [5].) The output of the probabilistic graphical model approach is a graph structure that indicates the potential causal connections between the observed variables. While originally developed for applications in economics and the social sciences, causal discovery has yielded many important insights in the area of bioinformatics [6] and, more recently, in climate science, primarily to identify interactions between different compound indices [7], [8] and to track interaction pathways around the globe based on geopotential height observations and other data [9], [10], [3].

The key idea of this project is to gain new insights into the complex dynamics governing the interactions between the radiative flux at the top of the atmosphere (TOA) and air/surface temperatures, by applying causal discovery algorithms on observational and reanalysis data for these variables. The radiative flux and air/surface temperatures are physically related through numerous dynamical and thermodynamical processes that cannot be perfectly accounted for even by the most complicated climate models we have now. The ultimate goal of this project is to (hopefully) find key variables or variables at key temporal and spatial locations that establish feedback loops connecting TOA radiation and surface temperature, as it is fundamental for understanding climate feedback processes active for $CO_2$-induced warming of the Earth's climate.

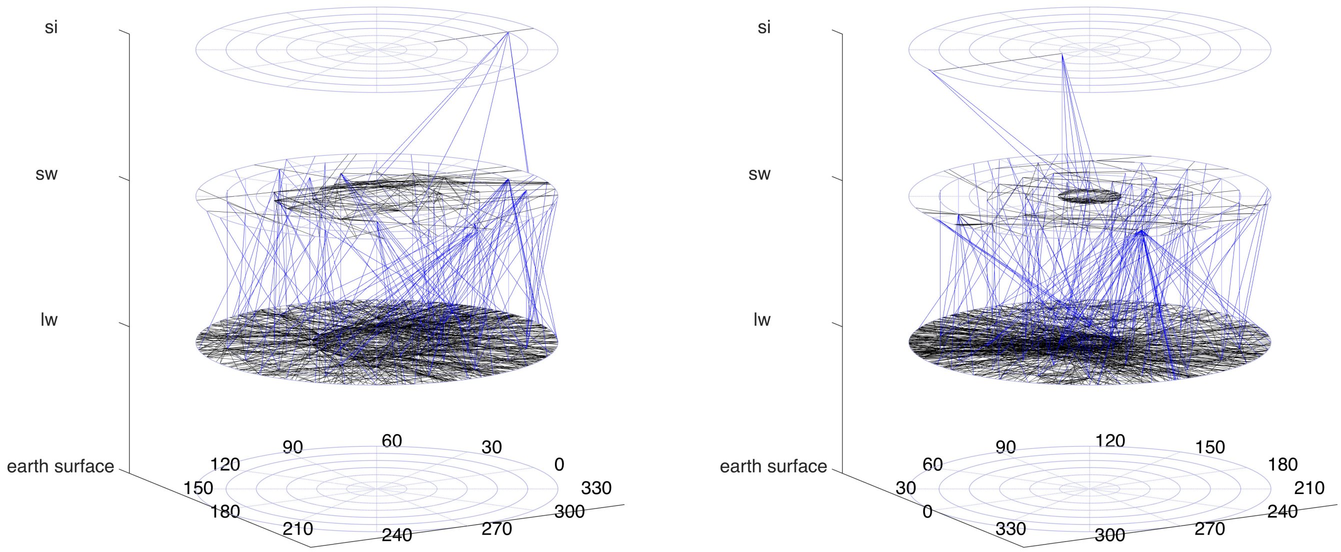
## II. DATA

We are using daily data from March 1, 2000 to December 31, 2013 (5054 days) from two different sources. Namely, we are using the NASA CERES [http://ceres.larc.nasa.gov/index.php] data for shortwave flux at TOA (sw), long wave flux at TOA (lw) and solar insolation at TOA (si), and the NASA MERRA [http://gmao.gsfc.nasa.gov/merra/] daily air temperature at 850, 500 and 50hPa and at the surface. For these first experiments we use a very low spatial resolution, namely 20 x 20 degrees, resulting in 19 longitude and 10 latitude values, i.e. 190 different locations around the globe.

## III. SPECIFIC METHOD USED

We use the well established framework of *structure learning* for *probabilistic graphical models* [1], [2], specifically constraint-based structure learning based on the well-known *PC algorithm*, which yields graph structures that indicate interactions between the observed variables. Details for applying this method to climate applications are given in [8]. In our first experiments we explore static models, and plan to explore temporal models later. Furthermore, our framework allows us to make use of expert knowledge as constraints to the algorithms, e.g. we can impose that solar insolation can only be a cause of the other variables, but not the effect.

Corresponding author: I. Ebert-Uphoff, iebert@engr.colostate.edu. [1] Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, USA. [2] School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA.




(a) Northern hemisphere

(b) Southern hemisphere

Fig. 1. Initial experiment tracking interactions between only three fields, si, sw and lw. Black lines indicate interactions between different locations for a single field, while blue lines indicate interactions between different fields.

## IV. First Runs and Future Work

Our first runs track interactions only between the TOA flux variables (sw, lw, si), and without using any expert knowledge, primarily to test our procedure. Figure 1 shows the results in form of layered stereographic plots. Each layer represents one of the three fields in stereographic projection and the lines show interactions identified by the algorithm. These test results appear to be consistent with what we know about the physical processes, namely (1) there are *direct* connections between si and sw, but not between si and lw; (2) there are tighter connections between si and sw in the polar regions, which is likely due to the relatively large ice cover in these regions, resulting in a stronger coupling between si and sw; (3) the weak sw-sw and si-sw connections over lower latitudes are likely due to large spatial/temporal variability in planetary albedo over tropical oceans.

What makes the interpretation of these initial results hard is the low resolution of the current model and the fact that the temperature fields are not yet included, although temperature is an important link between sw and lw. Thus the next steps are (1) including temperature fields; (2) performing runs with increasing resolution and varying levels of expert knowledge as constraints; (3) analyzing the results to identify any new and interesting interaction patterns; (4) working closely with the climate scientist on our team (last author of this abstract) to continuously revise the experiments, visualization tools, etc., and to further study any interesting effects detected in Step 3 and to answer specific scientific questions.

## Acknowledgments

This work was supported by two grants of the NSF Climate and Large-Scale Dynamics (CLD) program, namely Grant AGS-1147601 awarded to Deng, and a collaborative grant (AGS-1445956 and AGS-1445978) awarded to Deng and Ebert-Uphoff. Rodriguez is supported through an REU supplement associated with grant AGS-1445978.

## References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 2nd ed., Morgan Kaufman Publishers, 552 pp, 1988.

[2] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, Springer Lecture Notes in Statistics. 1st ed. Springer Verlag, 526 pp., 1993.

[3] T. Zerenner, P. Friedrichs, K. Lehnerts and A. Hense, *A Gaussian graphical model approach to climate networks*, Chaos, 23, 023103, 2014.

[4] A. Arnold, Y. Liu, and N. Abe, *Temporal causal modeling with graphical Granger methods*, Proc. 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (SIGKDD07), San Jose, CA, 10 pp, 2007.

[5] M. Gong, K. Zhang, B. Scholkopf, D. Tao, P. Geiger, *Discovering Temporal Causal Relations from Subsampled Data*, ICML 2015, pp. 1898-1906, 2015.

[6] N. Friedman, M. Linial, I. Nachman and D. Peer, *Using Bayesian networks to analyze expression data*, J. Comput. Biol., 7 (34), 601620, 2000.

[7] T. Chu, D. Danks, and C. Glymour, *Data Driven Methods for Nonlinear Granger Causality: Climate Teleconnection Mechanisms*, Tech. Rep. CMU-PHIL-171, Dep. of Philos., Carnegie Mellon Univ., Pittsburgh, Pa., 2005.

[8] I. Ebert-Uphoff, Y. Deng, *Causal Discovery for Climate Research Using Graphical Models*, Journal of Climate, Vol. 25, No. 17, Sept 2012, pp. 5648-5665.

[9] I. Ebert-Uphoff and Y. Deng, *A New Type of Climate Network based on Probabilistic Graphical Models: Results of Boreal Winter versus Summer*, Geophysical Research Letters, vol. 39, L19701, 7 pages, doi:10.1029/2012GL053269, 2012.

[10] Y. Deng and I. Ebert-Uphoff, *Weakening of Atmospheric Information Flow in a Warming Climate in the Community Climate System Model*, Geophysical Research Letters, 7 pages, doi: 10.1002/2013GL058646, Jan 2014.