Inference of directed climate networks: role of instability of causality estimation methods

J. Hlinka D. Hartman M. Vejmelka M.Paluš

Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic

This study was supported by the Czech Science Foundation project No. P103/11/J068.

THE ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

GAČR GRANTOVÁ AGENTURA ČESKÉ REPUBLIKY

Introduction

Climate data are increasingly analyzed by complex network analysis methods, including graph-theoretical approaches [1]. For such analysis, links between localized nodes of climate network are typically quantified by some statistical measures of dependence (connectivity) between measured variables of interest. To obtain information on the directionality of the interactions in the networks, a wide range of methods exists. These can be broadly divided into linear and nonlinear methods, with some of the latter having the theoretical advantage of being model-free, and principally a generalization of the former [2]. However, as a trade-off, this generality comes together with lower accuracy – in particular if the system was close to linear. In an overall stationary system, this may potentially lead to higher variability in the nonlinear network estimates. Therefore, with the same control of false alarms, this may lead to lower sensitivity for detection of real changes in the network structure. We aimed to assess and compare the reliability of common methods for directed (causal) graph inference.

Data

Data: NCEP/NCAR reanalysis dataset

- surface air temperatures
- daily data (years 1948 - 2007; 21900 timepoints)
- global grid 73 × 144 points (2.5 deg ×2.5 deg sampling)
- yearly cycle removed (anomalies)
- seasonal variability in variance removed
- dimensionality reduction:
  - VARIMAX-rotated PCA on monthly data (years 1948 - 2007; 720 timepoints) + projection into daily data OR
  - resampling to equidistant grid of 162 spatial locations

Methods

- Two methods compared:
  - linear Granger causality anlaysis/index (GCA, [5])
  - nonlinear conditional mutual information (aka Transfer Entropy, TE [6])
- Linear model order p=1 was selected to allow direct comparability of GCA and TE.
- For TE, two standard algorithms were applied: 1) based on discretization of variables into Q equiquantal bins (EQQ, [Paluš]) and 2) k-nearest neighbour search (kNN, [Vejmelka]). Both require setting an additional parameter, we used a range of parameters for both methods to track the dependence on parameter settings.

Discussion and conclusions

- Both linear and nonlinear causality methods provided reproducible graphs.
- Linear Granger Causality outperformed all the tested nonlinear algorithms in terms of stability of results.
- Results of linear and nonlinear methods relatively similar, in agreement with earlier results [3,4].
- For parameter settings providing stable results, the nonlinear methods converged ever closer to the linear Granger causality.
- Robustness was tested with respect to different spatial and temporal subsampling, generalizability of results to other climatic datasets is under investigation.

References

[1] Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M. & Hwang, D. U.: Complex networks: Structure and dynamics, Physics Reports, 2006, 424, 175-308

[2] Barnett, L.; Barrett, A. B. & Seth, A. K.: Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables, Physical Review Letters, 2009, 103, 238701

[3] Hlinka, J.; Hartman, D.; Vejmelka, M.; Novotná, D.; Paluš, M.: Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity, preprint available (http://arxiv.org/abs/1211.6688)

[4] Hlinka, J., Hartman, D., Vejmelka, M., Runge, J., Marwan, N., Kurths, J. and Paluš, M.: Reliability of inference of directed climate networks using conditional mutual information. Submitted

[5] Granger, C. W. Investigating causal relations by econometric model and cross-spectral methods, Econometrica, 1969, 37, 414–

[6] Schreiber, T. Measuring information transfer Physical Review Letters, 2000, 85, 461-464

Results

Both linear and nonlinear methods provided reproducible causality matrices, with the linear Granger causality outperforming the Transfer Entropy in reproducibility (both in a stationary model and real data).

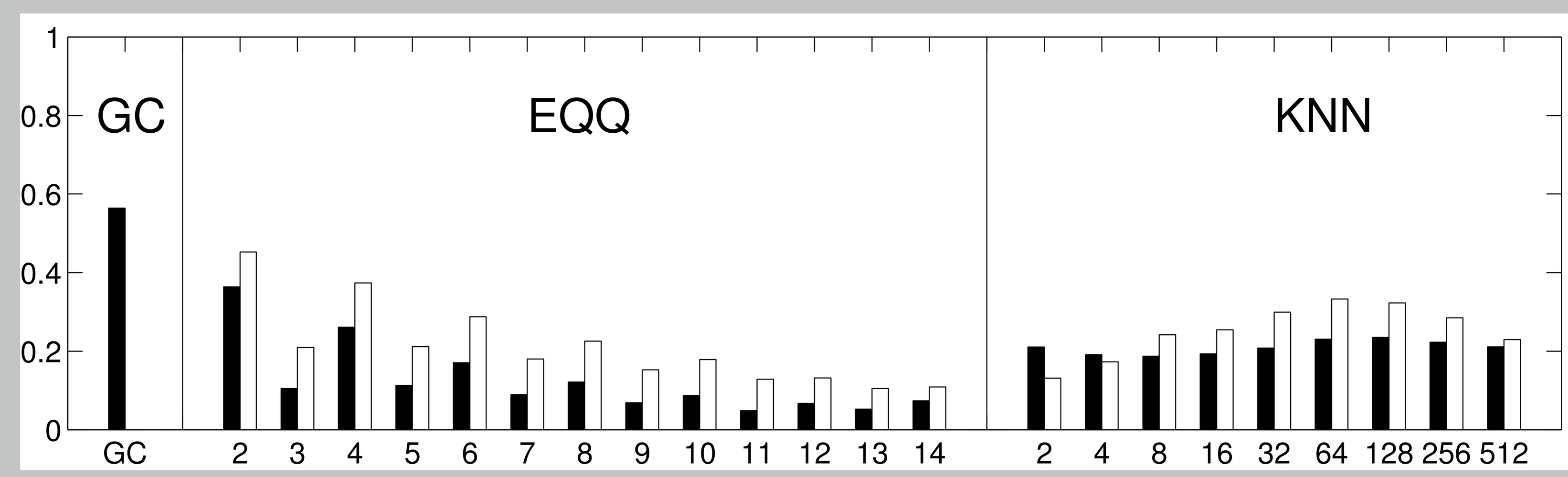
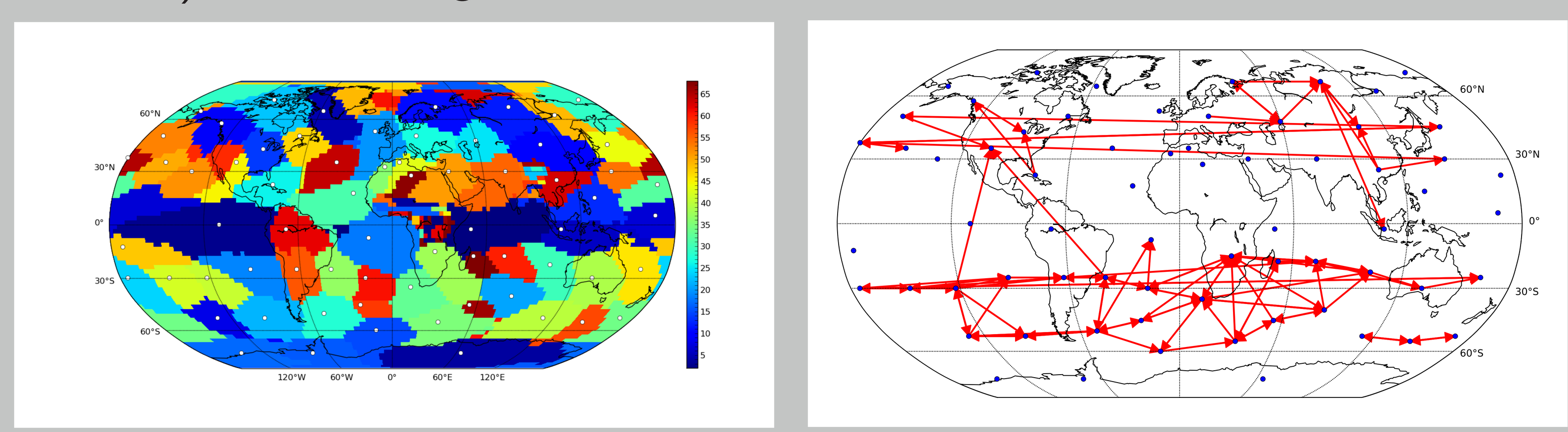
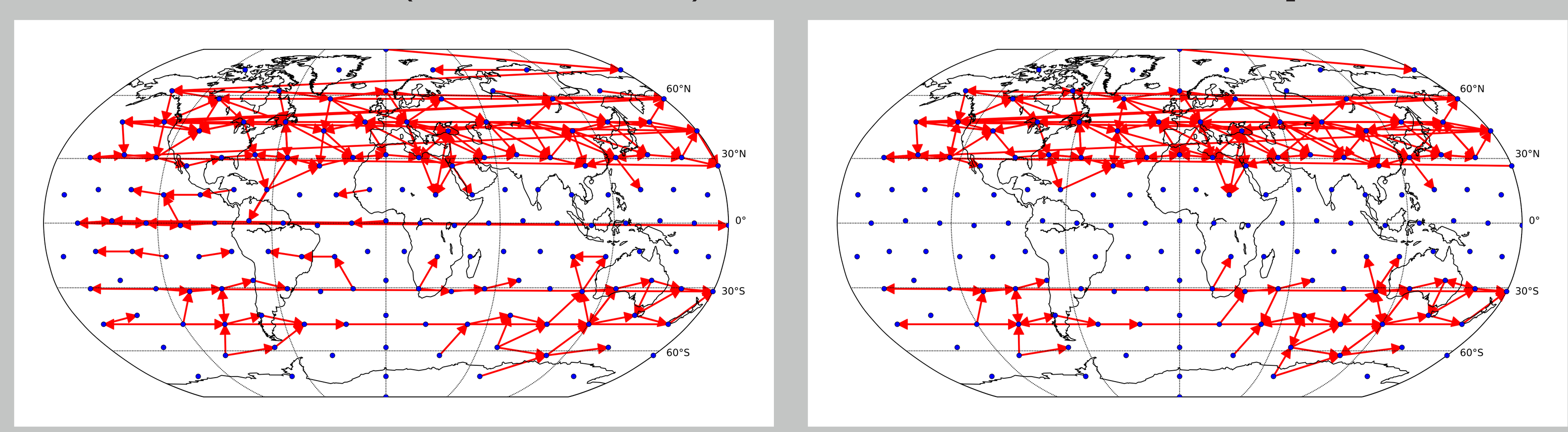


Figure: Reliability of causality network detection using different causality estimators, and the similarity to linear causality network estimates: Fourier surrogates model. For each estimator, six causality networks are estimated, one for each decade-long section of model stationary data (a Fourier surrogate realization of the original data). Black: the height of the bar corresponds to the average Spearman's correlation across all 15 pairs of decades. White: the height of the bar corresponds to the average Spearman's correlation of nonlinear causality network and linear causality network across 6 decades.

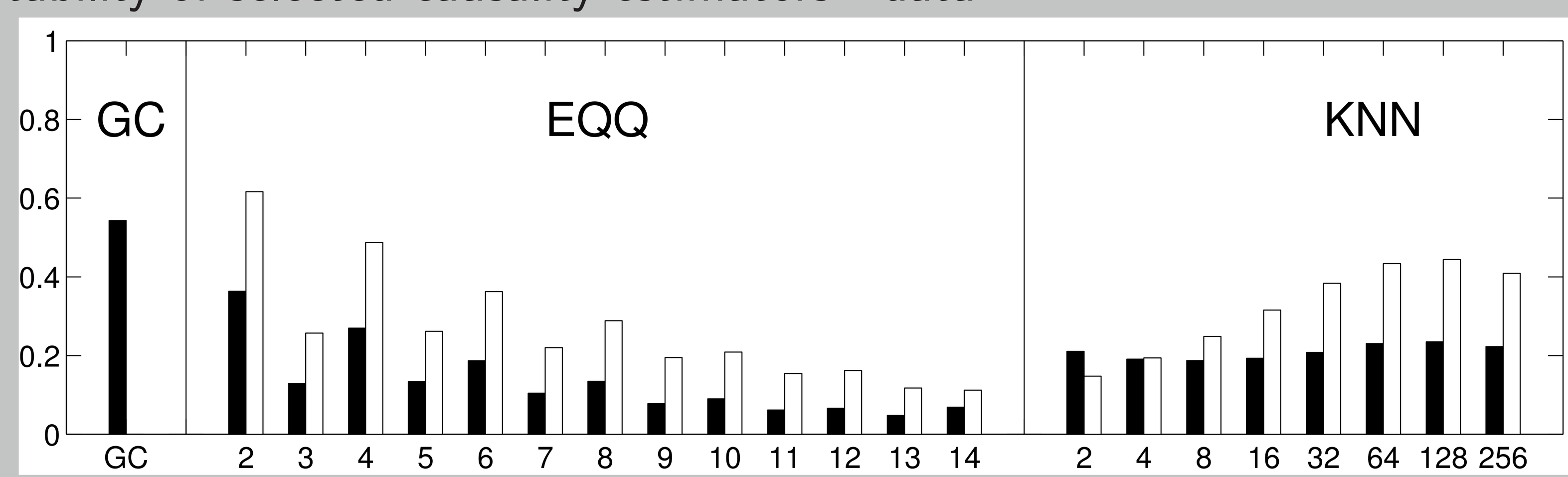
- Left: components - areas of dominance for each of 67 components; Right: Detected causality graph using linear Granger Causal Analysis (average over 6 decades) - 100 strongest links shown



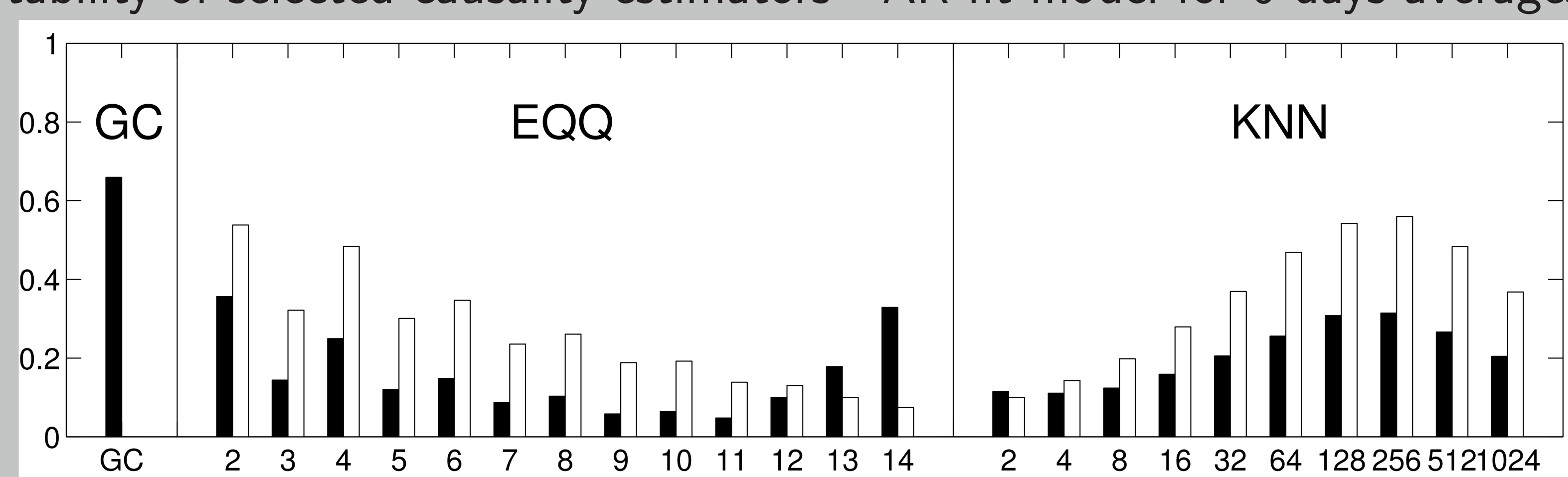
- Detected causality graph (average over 6 decades) - equidistant grid: [Left: using linear Granger Causal Analysis, Right: using nonlinear conditional mutual information (EQQ, $Q = 2$), 200 strongest links shown]



- Stability of selected causality estimators - data



- Stability of selected causality estimators - AR fit model for 6-days averages



- Reliability using Jaccard similarity instead of Spearman correlation