

PERSPECTIVE OPEN

Advancing climate science with knowledge-discovery through data mining

Annalisa Bracco^{1,2}, Fabrizio Falasca¹, Athanasios Nenes^{1,3,4,5}, Ilias Fountalis⁶ and Constantine Dovrolis⁶

Global climate change represents one of the greatest challenges facing society and ecosystems today. It impacts key aspects of everyday life and disrupts ecosystem integrity and function. The exponential growth of climate data combined with Knowledge-Discovery through Data-mining (KDD) promises an unparalleled level of understanding of how the climate system responds to anthropogenic forcing. To date, however, this potential has not been fully realized, in stark contrast to the seminal impacts of KDD in other fields such as health informatics, marketing, business intelligence, and smart city, where big data science contributed to several of the most recent breakthroughs. This disparity stems from the complexity and variety of climate data, as well as the scientific questions climate science brings forth. This perspective introduces the audience to benefits and challenges in mining large climate datasets, with an emphasis on the opportunity of using a KDD process to identify patterns of climatic relevance. The focus is on a particular method, δ -MAPS, stemming from complex network analysis. δ -MAPS is especially suited for investigating local and non-local statistical interrelationships in climate data and here is used to elucidate both the techniques, as well as the results-interpretation process that allows extracting new insight. This is achieved through an investigation of similarities and differences in the representation of known teleconnections between climate reanalyses and climate model outputs.

npj Climate and Atmospheric Science (2018)1:20174; doi:10.1038/s41612-017-0006-4

INTRODUCTION

Many of the greatest scientific challenges involve problems of vast complexity and interconnectedness that transcend traditional disciplinary boundaries. Climate science is one such example, requiring an interdisciplinary approach to advance scientific knowledge. The fast growing availability of observations from remote sensing platforms (space-borne, aircraft-based and ground-based), and detailed outputs from global-scale earth system models provide an overwhelming flow of spatio-temporal data that far exceeds data analysis capacity. While the development of statistical tools applied to climate fields is mature, the big data-induced revolution seen in health-care, financial banking, advertising or biology have yet to be duplicated in climate science. In the last decade, however, several groups attempted to apply the so-called Knowledge-Discovery through Data mining (KDD)^{1,2} to climate. KDD refers to the overall process of using data mining algorithms that autonomously identify patterns from various data sources to find, extract and identify what is qualified as knowledge, and interpret the outcomes. It begins with choosing the tools for the data mining steps, as well as the preprocessing steps, and concludes with the evaluation and interpretation of the patterns resulting from the chosen algorithms. The KDD process, therefore, while encompassing data mining, adds several important steps.

Here we discuss key big-data challenges facing climate science, with an overview of recent efforts to apply KDD to this field, and we provide concrete examples from ongoing research. We focus on knowledge discovery using complex network analysis³ coupled

to dimensionality reduction techniques with the objective of extracting and analyzing statistical interrelationships in fields of climatic interest.

COMPLEX NETWORK ANALYSIS AND CLIMATE SCIENCE

It is widely recognized that anthropogenic emissions contribute to the observed rates of temperature increase, as ratified at the 2015 Paris Agreement.⁴ The fundamental scientific mechanism behind greenhouse gas-induced climate warming is straightforward and indisputable but many uncertainties remain on the extent, patterns, and implications of changes in climate fields over space and time. We have reasonably constrained global mean trends and rates of changes of heat and carbon dioxide reservoirs in the ocean, atmosphere and land over the past forty years, but we struggle to provide robust regional assessments, diagnose how modes of natural climate variability and global warming are interlinked, deduce ecosystem responses, or infer how climate change may affect weather events.^{5–11} The spatial and temporal scales involved in climate-relevant interactions are daunting. For example, greenhouse gases and aerosol modify the millimeter-scale size of cloud droplets and ice crystals, which in turn modulates the ability of clouds reflective sunlight and planetary heat, with global feedbacks on temperature and precipitation,^{12–14} while decadal or longer-scale climate changes are felt by society through changes in the character of weather-like extreme events.^{15–17} A second challenge is associated with the inadequacy of the available observing system to sample thoroughly the

¹School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA; ²Institute of Geosciences and Earth Resources, National Research Council of Italy, Pisa, PI 56124, Italy; ³Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA; ⁴Institute of Chemical Engineering Sciences, Foundation for Research and Technology Hellas, Patras GR-26504, Greece; ⁵Institute for Environmental Research and Sustainable Development, National Observatory of Athens, P. Penteli 1GR-5236, Greece and ⁶College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA
Correspondence: Annalisa Bracco (abracco@gatech.edu)

Received: 12 April 2017 Revised: 26 July 2017 Accepted: 3 August 2017

Published online: 09 January 2018

spatio-temporal scales on which climate varies. Remote sensing platforms have revolutionized climate science, but satellite records effectively start in the late 1970s, while technological challenges hamper sensing key areas of climatic interest such as high latitudes or the deep portions of the oceans.^{18–20}

As in other areas of science and engineering, numerical models have become indispensable for understanding climate science. In the past thirty years, they have evolved to account for an increasing number of physical, chemical, and biological processes. The end result is better numerical simulations with codes that use finer grids and include more interacting processes. Climate modelers, however, are faced by challenges that include the multiplicity and nonlinearity of the processes contributing to the climate system, the high-dimensionality of the problem, and the computational requirements.^{13,21} Despite substantial improvements in the representation of large-scale averages, climate models remain difficult to constrain at regional scales. The uncertainties about linkages between subgrid processes, regional scale changes and large scale dynamics in both observations and model outputs hamper the confidence in regional-scale attribution of on-going changes and future projections.^{13,22–24}

Evaluating climate datasets and model outputs in an efficient and robust way, while gathering information about linkages between fields, geographical regions, or time intervals is therefore a priority. This can be achieved through complex network analysis which premise is that the underlying topology or network structure of a system has a strong impact on its dynamics and evolution. Applications to climate science have received growing attention since 2004,²⁵ when graph theory was applied to the investigation of global geopotential height. Network analysis has been since applied to studies of numerous climate modes,^{26–31} of atmospheric and oceanic circulation drivers,^{32–35} of precipitation in different time periods,^{36–38} and of Rossby wave dynamics.³⁹

Generally networks are constructed as undirected, binary graphs. A graph is a set of vertices or nodes that, in the case of climate variables, represent geographical locations and, for gridded data-sets, grid points. The *edges* or links between the nodes are bidirectional (undirected), commonly do not carry information about the weight of the links (binary), and are inferred using simultaneous linear or non-linear similarity measures such as Pearson correlation, mutual information, or phase synchronization.^{27,29,39,40} Often, two nodes that are not linked according to the chosen criterion have their correlations deleted or *pruned*. In the case of climate fields, however, cell-level pruning can cause loss of robustness in the network inference, and methods that adopt pruning should not be used for intercomparison studies.^{41,42} Community detection (clustering) algorithms are commonly used to reduce the dimensionality of graphs.⁴³

Recent developments in network analysis applications to climate have focused on three issues. First, it has been noted that detecting communities in climate variables requires separating between dynamical links and autocorrelations^{44,45} because of teleconnections between non-adjacent regions and autocorrelations over different spatio-temporal scales. Second, multivariate networks⁴⁶ and networks with links that account for lagged interactions³⁸ have been developed to explore interactions between different variables and characterize time-lagged relationships. Finally, new methodologies that uncover directed or even causal relationships have been proposed.^{47,48}

δ-MAPS

Here we focus on a network-based methodology, δ-MAPS, that we developed to robustly compare spatial consistent gridded fields. Our goal is to exemplify how data mining methods can assist with discovering important linkages, or their absence, in climate data.

δ-MAPS identifies the spatially contiguous components of a system, or *domains*, that contribute in a homogenous way to the system' dynamics, and then infers their connections accounting for autocorrelations. It refines a previously proposed methodology^{41,42} and allows for overlapping domains and weighted links at a temporal lag, both relevant to climate fields. After the domains are identified, δ-MAPS infers a functional network between them by examining the statistical significance of each lagged cross-correlation between any two domains, calculating a range of potential lag values for each edge, and assigning a weight that is based on the covariance of the signal of the corresponding two domains. While a temporally ordered correlation does not imply causation, it provides information on the plausible directionality of interactions. Finally each domain has a '*strength*' calculated as the sum of the absolute weights of all links ignoring their directionality. The greater the strength, the larger is the domain influence on the system at the temporal scales considered.

Details about the methodology are provided as a Supplementary file (Supplementary Methods) and illustrations of advantages of δ-MAPS compared to standard techniques such as principal component analysis, clustering and community detection are presented in Fountalis et al.⁴⁹

We present a sample of networks from two global monthly sea surface temperature (SST) reanalysis datasets, the HadISST⁵⁰ and COBE-SST2,⁵¹ from the fractional ice content within clouds from the MERRA-2 project⁵² available from 1980 onward and corresponding variables from a representative member of the Community Earth System Model (CESM) large ensemble.⁵³ The resolution is 1.25°x1° and the focus on the latitudinal range [60°S–60°N] for SST and [55°S–55°N] for clouds to avoid regions where the correlation across reanalyses is widely low⁵¹ or data are not continuously available. All networks are built using detrended monthly anomalies.

Figure 1 presents strength maps over the period 1971–2015. Domains are similar in the reanalyses, but generally weaker in COBE. The strongest domain covers the El Niño Southern Oscillation (ENSO) region extending to 60°N with a pattern reminiscent of the Pacific Decadal Oscillation (PDO) footprint. Strong domains include the horseshoe areas north and south of the equator, the eastern portion of the South Pacific, the tropical Indian Ocean, the north Tropical Atlantic, and in the reanalyses the south Tropical Atlantic. A domain occupies the Warm Pool only in HadISST. We verified that also the ERSSTv4⁵⁴ reanalysis network and the MERRA-2 cloud fields presented later do not include it. In the randomly chosen CESM member no domain occupies the Warm Pool region and the south Tropical Atlantic area is extremely weak. Both features are common to all other CESM runs analyzed.

The connections between the strongest domains including the Warm Pool for HadISST, and their lags are shown in Fig. 2. In the reanalyses the ENSO/PDO area is linked to all others at zero or positive lags except for the south Tropical Atlantic, which is anticorrelated and leads by 8 to 10 months. Positive (negative) spring SST anomalies in the Equatorial Tropical Atlantic and in the Gulf of Guinea indeed strengthen (weaken) the Walker circulation, modifying the equatorial winds and the eastern Pacific upwelling and favoring La Niña (El Niño) conditions the following winter^{55,56} through a Gill-Matsuno-type response.⁵⁷ Such connection is only partially counteracted by the thermodynamic link from the ENSO area into the Tropical Atlantic through the warming of the entire tropical troposphere following El Niños^{58,59} and by the dynamical response of the tropical Atlantic trades to the Pacific warming.^{59–61} In CESM links from the Pacific to the Indian Ocean and north Tropical Atlantic are stronger than observed, while the connection from the south Tropical Atlantic is missing. The relation between ENSO and south Atlantic domains is indeed weak and opposite in sign.

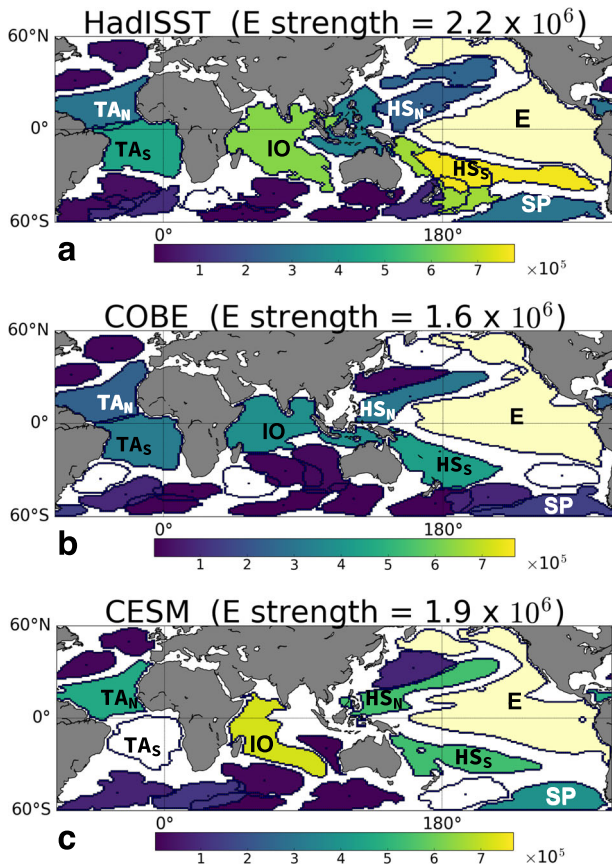


Fig. 1 SST domains identified by δ -MAPS and their strength in **a** HadISST, **b** COBE, and **c** one member of the CEM ensemble over the 1971–2015 period. The strength of the domain occupying the ENSO region (E) is off-scale and indicated atop of each panel

The network analysis of cloud fields can contribute to diagnose this common model bias.⁶² Despite the higher level of noise and intermittency of cloud fields compared to SST, the δ -MAPS outcome is insightful. Figure 3 presents maps of strength for all domains and links from the ENSO area for the ice cloud fraction. Focusing on the Equatorial and south Tropical Atlantic, two domains are identified in MERRA-2, with the first negatively connected to the Equatorial Pacific, and the southern one positively correlated as expected in the thermodynamic response to ENSO; in SST these domains are merged due to the oceanic circulation. In the CEM ice cloud fraction network there is only one domain, positively, but statistically insignificantly, linked to ENSO; a weakly anticorrelated one is found entirely shifted into the northern hemisphere. The domains in MERRA-2 are used to define boxes to evaluate correlograms of SST anomalies with respect to those from the E domain (Fig. 3e–f). In HadISST (or COBE) both the thermodynamic feedback, lead by ENSO and mostly effective into the southern box, and the dynamical Gill-Matsuno teleconnection, lead by the Equatorial Atlantic, are identified. The second dominates the total domain signal. In CEM the dynamical connection is mostly absent, the Equatorial Atlantic evolves independently of ENSO and the thermodynamic link is stronger than observed⁶³ but not sufficient to achieve statistical significance. All other 29 members of the large-ensemble confirm that CEM overestimates the thermodynamic feedback and underestimates the dynamic teleconnection, which prevails only in one run. In several integrations the thermodynamic feedback is

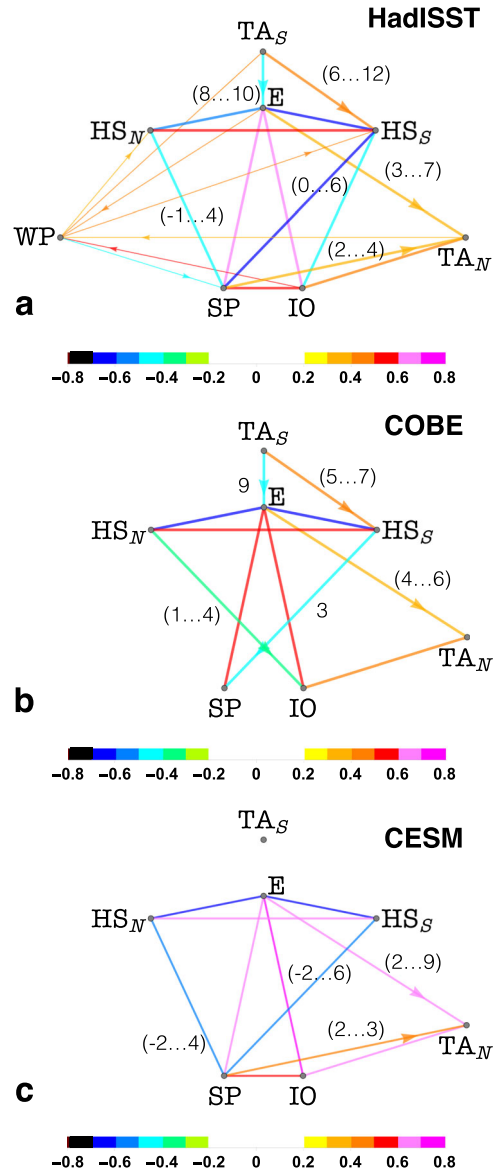


Fig. 2 SST network across the **a** seven of the strongest domains in HadISST (the Warm Pool domain is excluded), **b** the seven strongest domains in COBE, and **c** six strongest domains in CEM, where TA_S has no links. The color of each link represents the corresponding cross-correlation. Arrows indicate signed definitive (positive or negative) lags. The absence of arrow indicates that connections are significant also at zero lags. Some (not all for clarity) lags are indicated

so strong that a significant link from ENSO to the south Tropical Atlantic domain characterizes the SST network.

DISCUSSION: A WAY FORWARD

In seeking to understand past, present and future changes in our climate is mandatory to leverage advances in KDD research while accounting for the characteristics of climate data. KDD methods that account for the characteristics of climate data can effectively aid scientific theory and should be integral to any interdisciplinary framework to quantify uncertainties in climate projections or to unveil linkages between perturbations to the climate system and its response. δ -MAPS, for example, infers the high-level abstract linkages across components of the climate system,⁴⁹ highlights

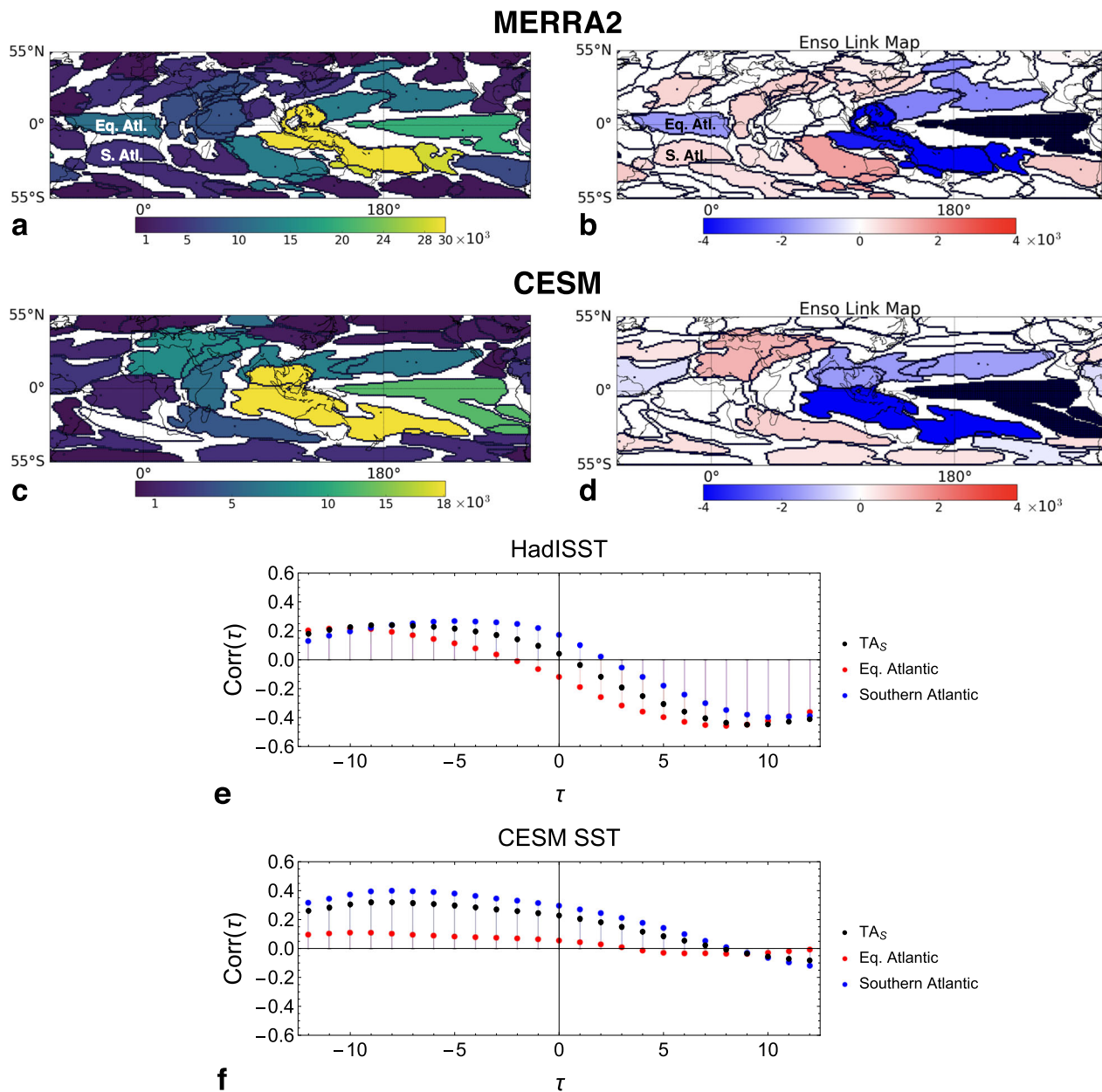


Fig. 3 Cloud ice fraction domains identified by δ -MAPS over the period 1980–2015 with their strength (left) and link maps (right) from the Equatorial Pacific (ENSO-related) domain in **a–b** MERRA-2, **c–d** CESM. **e–f**: correlograms between the SST domain signal of E and TA_S and E and the signal calculated over the Eq. Atl. and S. Atl. boxes identified in the MERRA-2 network

quantifiable differences across datasets, and provides a reduced form model that can be continuously informed from data updates. It is therefore uniquely suited to assess impacts, evaluate model performances and biases, and characterize pathway scenarios, climate trajectories, and the propagation of perturbations from local forcing agents (e.g., aerosols) across climate fields.

Immediate applications range from diagnosing representation and changes in teleconnections—or connectivity in the case of ecosystems—over space and time, to aiding adjoint models in a general framework for regional or global attribution studies.

Data availability

All data sets used are publicly available. The software for δ -MAPS is available at <https://github.com/FabriFalasca/delta-MAPS>

ACKNOWLEDGEMENTS

We thank two anonymous reviewers for their insightful feedbacks. δ -MAPS development was supported by the Department of Energy through grant DE-SC0007143, and by the National Science Foundation (grant DMS1049095). A.N. and F.F. received support from the Department of Energy (grant DE-SC0007145) and from the NASA MAP program (grant NNX13AP63G). A.B. was supported by a Faculty Development Grant from the Georgia Institute of Technology during her stay at the CNR Institute of Geosciences and Earth Resources where this work was completed.

AUTHOR CONTRIBUTIONS

A.B. wrote the manuscript. A.B., F.F., I.F. conceived the study. F.F. performed the analyses. C.D. and I.F. developed the δ -MAPS methodology. A.N. contributed to writing and interpretation.

ADDITIONAL INFORMATION

Supplementary Information accompanies the paper on the *npj Climate and Atmospheric Science* website (<https://doi.org/10.1038/s41612-017-0006-4>).

Competing interests: The authors declare no competing financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Change history: The original version of this Article had an incorrect Article number of 4 and an incorrect Publication year of 2017. These errors have now been corrected in the PDF and HTML versions of the Article.

REFERENCES

- Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. From data mining to knowledge discovery: an overview. In *Advances in Knowledge Discovery and Data Mining* (eds Fayyad, U. M., Piatetsky-Shapiro G., Smyth P. & Uthurusamy R.) 1–34 (MIT Press, Cambridge, MA, 1996).
- Chakrabarti, D. & Faloutsos, C. Graph mining: Laws, generators and algorithms. *ACM Comput. Surv.* **38**, Art. 2, <https://doi.org/10.1145/1132952.1132954> (2006).
- Newman, M., Barabasi, A. L. & Watts, D. J. *The structure and dynamics of networks*. (Princeton University Press, 2006).
- Adoption of the Paris Agreement FCCC/CP/2015/L.9/Rev.1* (UNFCCC, 2015). http://unfccc.int/paris_agreement/items/9485.php
- Knutti, R. & Sedláček, J. Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Chang.* **3**, 369–373 (2013).
- Kharin, V. V., Zwiers, F. W., Zhang, X. & Wehner, M. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Chang.* **119**, 345–357 (2013).
- Shepherd, T. G. Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* **7**, 703–708 (2014).
- Wang, C., Zhang, L., Lee, S.-K., Wu, L. & Mechoso, C. R. A global perspective on CMIP5 climate model biases. *Nat. Clim. Chang.* **4**, 201–205 (2014).
- Anderson, B. T. et al. Sensitivity of terrestrial precipitation trends to the structural evolution of sea surface temperatures. *Geophys. Res. Lett.* **42**, 1190–1196 (2015).
- Bopp, L. et al. Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. *Biogeosciences* **10**, 6225–6245 (2013).
- Shaw, T. A. et al. Storm track processes and the opposing influences of climate change. *Nat. Geosci.* **9**, 656–664 (2016).
- Ramanathan, V. et al. Aerosols, climate, and the hydrological cycle. *Science* **294**, 2119–2124 (2001).
- Seinfeld, J. H. et al. Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system. *Proc. Natl Acad. Sci. USA* **113**, 5781–5790 (2016).
- Schneider, T. et al. Climate goals and computing the future of clouds. *Nat. Clim. Chang.* **7**, 3–5 (2017).
- Kunkel, K. E. et al. Monitoring and understanding trends in extreme storms. *Bull. Am. Meteor. Soc.* **94**, 499–514 (2013).
- Orlowsky, B. & Seneviratne, S. I. Global changes in extreme events: regional and seasonal dimension. *Clim. Chang.* **110**, 669–696 (2012).
- Diffenbaugh, N. S. & Ashfaq, M. Intensification of hot extremes in the United States. *Geophys. Res. Lett.* **37**, L15701 (2010).
- The Global Observing System for Climate: Implementation needs (GCOS-200, GOOS-214, August 2010) (2010).
- Rhein, M. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 3, 255–315 (IPCC, Cambridge Univ. Press, Cambridge, UK and New York, NY, USA, 2013).
- Durack, P. J., Gleckler, P. J., Landerer, F. W. & Taylor, K. E. Quantifying underestimates of long-term upper-ocean warming. *Nat. Clim. Chang.* **4**, 999–1005 (2014).
- Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C. & Meyerson, J. E. Considerations for parameter optimization and sensitivity in climate models. *Proc. Natl Acad. Sci. USA* **107**, <https://doi.org/10.1073/pnas.1015473107> (2010).
- Sarojini, B. B., Stott, P. A. & Black, E. Detection and attribution of human influence on regional precipitation. *Nat. Clim. Chang.* **6**, 669–675 (2016).
- Stevens, B. & Feingold, G. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature* **461**, 607–613 (2009).
- Cohen et al. Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.* **7**, 627–637 (2014).
- Tsonis, A. A. & Roebber, P. J. The architecture of the climate network. *Phys. A* **333**, 497–504 (2004).
- Yamasaki, K., Gozolchiani, A. & Havlin, S. Climate networks around the globe are significantly affected by El Niño. *Phys. Rev. Lett.* **100**, 228501 (2008).

- Donges, J. F., Zou, Y., Marwan, N. & Kurths, J. The backbone of the climate network. *EPL* **87**, 48007 (2009).
- Van Der Mheen, M. et al. Interaction network based early warning indicators for the Atlantic MOC collapse. *Geophys. Res. Lett.* **40**, 2714–2719 (2013).
- Tsonis, A. A., Swanson, K. & Kravtsov, S. A new dynamical mechanism for major climate shifts. *Geophys. Res. Lett.* **34**, L13705 (2007).
- Guez, O., Gozolchiani, A., Berezin, Y., Brenner, S. & Havlin, S. Climate network structure evolves with North Atlantic Oscillation phases. *EPL* **98**, 38006 (2012).
- Tantet, A. & Dijkstra, H. A. An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature. *Earth Syst. Dynam.* **5**, 1–14 (2014).
- Berezin, Y., Gozolchiani, A., Guez, O. & Havlin, S. Stability of climate networks with time. *Sci. Rep.* **2**, 666 (2012).
- Deza, I., Barreiro, M. & Masoller, C. Assessing the direction of climate interactions by means of complex networks and information theoretic tools. *Chaos* **25**, 033105 (2015).
- Tirabassi, G. & Masoller, C. Unravelling the community structure of the climate system by using lags and symbolic time-series analysis. *Sci. Rep.* **6**, 29804 (2016).
- Hlinka, J., Jajcay, N., Hartman, D. & Paluš, M. Smooth information flow in temperature climate network reflects mass transport. *Chaos* **27**, 035811 (2017).
- Malik, N., Bookhagen, B., Marwan, N. & Kurths, J. Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Clim. Dyn.* **39**, 971–987 (2011).
- Rehfeld, K., Marwan, N., Breitenbach, S. F. M. & Kurths, J. Late Holocene Asian summer monsoon dynamics from small but complex networks of paleoclimate data. *Clim. Dyn.* **41**, 3–19 (2012).
- Wang, Y. et al. Dominant imprint of Rossby waves in the climate network. *Phys. Rev. Lett.* **111**, 138501 (2013).
- Wiedermann, M., Donges, J. F., Handorf, D., Kurths, J. & Donner, R. V. Hierarchical structures in Northern Hemispheric extratropical winter ocean-atmosphere interactions. *Int. J. Climatol.* <https://doi.org/10.1002/joc.4956> (2016).
- Scarsoglio, S., Laio, F. & Ridolfi, L. Climate dynamics: a network-based approach for the analysis of global precipitation. *PLoS One* **8**, e71129 (2013).
- Fountalis, I., Bracco, A. & Dovrolis, C. Spatio-temporal network analysis for studying climate patterns. *Clim. Dyn.* **42**, 879–899 (2014).
- Fountalis, I., Bracco, A. & Dovrolis, C. ENSO in CMIP5 simulations: network connectivity from the recent past to the twenty-third century. *Clim. Dyn.* **45**, 511–538 (2015).
- Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
- Kramer, M. A., Eden, U. T., Cash, S. S. & Kolaczyk, E. D. Network inference with confidence from multivariate time series. *Phys. Rev. E* **79**, 061916 (2009).
- Guez, O. C., Gozolchiani, A. & Havlin, S. Influence of autocorrelation on the topology of the climate network. *Phys. Rev. E* **90**, 062814 (2014).
- Steinhäuser, K., Ganguly, A. R. & Chawla, N. V. Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Clim. Dyn.* **39**, 889–895 (2012).
- Runge, J. et al. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nat. Comm.* **6**, 8502 (2015).
- Ebert-Uphoff, I. & Deng, Y. Causal discovery in the geosciences—Using synthetic data to learn how to interpret results. *Comput. Geosci.* **99**, 50–60 (2017).
- Fountalis, I., Bracco, A., Dilkina, B. & Dovrolis, C. δ -MAPS: From Spatio-temporal Data to a Weighted and Lagged Network Between Functional Domains, In *Proceedings of the Workshop on Mining Big Data in Climate and Environment (MBDCE 2017)* 17th SIAM International Conference on Data Mining (SDM 2017), 27–29 April 2017, Houston, Texas, USA (in the press).
- Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407 (2003).
- Hirahara, S., Ishii, M. & Fukuda, Y. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **27**, 57–75 (2014).
- Molod, A., Takacs, L., Suarez, M. & Bacmeister, J. Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geosci. Model. Dev.* **8**, 1339–1356 (2015).
- Kay, J. E. et al. The community earth system model (CESM) large ensemble project. A community resource for studying climate change in the presence of internal climate variability. *Bull. Am. Meteor. Soc.* **96**, 1333–1349 (2015).
- Huang, B. et al. Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part I. Upgrades and intercomparisons. *J. Climate* **28**, 911–930 (2015).
- Rodriguez-Fonseca, B. et al. Are Atlantic Niños enhancing Pacific ENSO events in recent decades? *Geophys. Res. Lett.* **36**, L20705 (2009).
- Kucharski, F., Kang, I.-S., Farneti, R. & Feudale, L. Tropical Pacific response to 20th century Atlantic warming. *Geophys. Res. Lett.* **38**, L03702 (2011).

57. Wang, C., Kucharski, F., Barimalala, R. & Bracco, A. Teleconnections of the tropical Atlantic to the tropical Indian and Pacific Oceans: a review of recent findings. *Meteorol. Z.* **18**, 445–454 (2009).
58. Nnamchi et al. Thermodynamic controls of the Atlantic Niño. *Nat. Comm.* **6**, 8895 (2015).
59. Chang, P., Fang, Y., Saravanan, R., Ji, L. & Seidel, H. The cause of the fragile relationship between the Pacific El Niño and the Atlantic Niño. *Nature* **443**, 324–328 (2006).
60. Latif, M. & Barnett, T. P. Interactions of the tropical oceans. *J. Clim.* **8**, 952–964 (1995).
61. Lübbecke, J. F. & McPhaden, M. J. On the inconsistent relationship between Pacific and Atlantic Niños. *J. Clim.* **25**, 4294–4303 (2012).
62. Kucharski, F., Syed, F. S., Burhan, A., Farah, I. & Gohar, A. Tropical Atlantic influence on Pacific variability and mean state in the twentieth century in observations and CMIP5. *Clim. Dyn.* **44**, 881–896 (2015).
63. He, J., Deser, C. & Soden, B. J. Atmospheric and oceanic origins of tropical precipitation variability. *J. Clim.* **30**, 3197–3217 (2017).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018