STUDYING EXTREMAL DEPENDENCE IN CLIMATE USING COMPLEX NETWORKS

Imme Ebert-Uphoff[1], Whitney Huang[2], Adway Mitra[3], Dan Cooley[4], Singdhansu B. Chatterjee[5], Chen Chen[6], Zhonglei Wang[7]

***Abstract*—This manuscript investigates how the framework of complex networks can be applied to the analysis of extremal dependencies, in order to gain new insights into extreme weather and climate events. We identify the most suitable existing type of climate network (event synchronization network), introduce a new network type based on extreme value theory (χ-network), calculate a sample χ-network and suggest future research directions.**

I. INTRODUCTION

Weather and climate extremes play an increasingly important role for society [1] due to their impact on food and water security [2] and increased risks of flooding, heat waves, etc. Here we investigate how the framework of complex networks can be applied to gain a deeper understanding of such extremes. Section II provides a quick review of complex networks in climate science. Section III reviews which existing network type is most suitable to study extremes. Section IV proposes a new network type based on extreme value theory, the χ-network. Section V provides results for a χ-network. Section VI discusses similarities between the new χ-network and the existing network type. Section VII outlines limitations and future work.

II. COMPLEX NETWORKS IN CLIMATE SCIENCE

Tsonis and Roebber in 2004 introduced the idea of *climate networks* in their seminal paper [3], and brought the framework of complex network theory to climate science. Climate networks provide an important framework for the identification and visualization of connectivity between different geographic regions [4]. The basic idea is to derive a network that represents connectivity between different points on the globe, based on observations (gridded data product or station locations), namely on time series data at the individual locations. Each network consists of a set of nodes - each representing one location - and edges - each representing connections between a pair of nodes. The edges are typically based on pair-wise measures applied to the data at the two locations. For example, Tsonis and Roebber [3] defined their network to contain an edge between two grid points if and only if the correlation between the time series at the two points exceeds a chosen threshold. This type of network is called a *correlation network* and remains the most popular type of complex network used in climate science to date.

Other well established network types in climate science include *event synchronization networks* [5], [6], [7] - which define connections based on whether extreme events at one point are regularly followed by extreme events at another point; *phase synchronization networks* [8] - which view the signals at each point as oscillation and seek to measure the coupling between those oscillations; and *causal networks* [9], [10], [11] - which seek to identify potential cause-effect relationships. The resulting networks differ based not only on the network type, but also depending on the global field they represent (rainfall data yields a very different network structure than temperature data), the temporal scale (daily/monthly/annual), and many other factors. Once a climate network has been constructed its structure can be studied, including properties such as the degree of a node (number of edges at a node), clustering coefficients, etc.

III. MOST SUITABLE EXISTING NETWORK TYPE

It is well known that dependence measures such as correlation may not be suitable to analyze *extremal* behavior. The reason is that correlation measures dependence from the center (mean) of the distribution, and the dependence exhibited by the small number of observations occurring in the tail (the extreme values) have little influence. Thus correlation networks may

Corresponding author: Imme Ebert-Uphoff, iebert@colostate.edu. [1]Electrical and Computer Engineering, Colorado State University, USA. [2]Pacific Climate Impacts Consortium, University of Victoria, Canada. [3]School of Electrical Sciences, IIT Bhubaneswar, India. [4]Statistics, Colorado State University, USA. [5]Statistics, University of Minnesota, USA. [6]Statistics, University of Chicago, USA. [7]Wang Yanan Institute for Studies in Economics, Xiamen University, China.

misrepresent extremal dependencies. Careful review revealed that the event synchronization network [5], [6], [7] is the only type of existing network that specifically focuses on extremal dependencies.

Event synchronization (ES) measures were first developed by Quiroga et al. [12], who introduced two bi-variate measures, Q and q, to describe the amount of synchronization between two time series at different locations. Q is undirected and indicates whether an extreme event (that is, an observation exceeding a prespecified threshold) in one location tends to occur within a given time window (say within N days) of an extreme event at another location. q is a directed signal that indicates in which location the event usually occurs first. Malik et al. [5], [6] introduced event synchronization to climate networks, resulting in event synchronization networks, where points across the globe are connected if their time series are synchronized. The undirected measure, Q, leads to undirected networks, while the directed measure, q, leads to directed networks (edges have a direction).

Event synchronization measures are calculated using the following steps. (1) The times at which a location has events above a certain threshold (e.g. above $97th$ percentile) is extracted at each location. (2) A suitable synchronization time window, τ, is defined. (3) The count, $c(P_1|P_2)$, is defined to be the number of events that occur at P_1 shortly (within less than τ) after they appeared at P_2. (Note that events that occur at the same time are only counted as 1/2 occurrence, rather than 1.) (4) The ES measures, $Q(j,k)$ and $q(j,k)$, are calculated as the sum/difference of the terms $c(j|k)$ and $c(k|j)$, respectively, scaled by the number of events. For details, see [5], [6], [7].

IV. NEW NETWORK TYPE BASED ON EXTREME VALUE THEORY

Extreme value theory (EVT) [13], [14] provides a solid framework for the study of extremes, i.e. to focus on the behavior of samples in the tails of a distribution. However, to the best of our knowledge, no measure from EVT has ever been used in climate networks. We identified a commonly used measure from EVT, the *Upper Tail Dependence*, χ [15], that could be used as connectivity measure in climate networks. The χ measure is defined as follows:

$$\chi = \lim_{u \to 1^-} \mathbb{P}(F_{X_2}(X_2) > u | F_{X_1}(X_1) > u), \quad (1)$$

where X_1 and X_2 are random variables and F_{X_1}, F_{X_2} their corresponding cumulative distribution function (CDF). Thus the χ-measure is the limiting conditional probability of one random variable being extreme, given that the other random variable is extreme. Note that the χ measure is in fact symmetric in terms of which variable is chosen to be conditioned upon, due to the standardization of the marginal behavior by the CDFs.

A χ-network is defined by using the χ-measure as connectivity measure for any two points, thus indicating tail dependence. Namely, any two points in a χ-network are connected by an edge if and only if the χ-measure between the two points exceeds a chosen threshold, χ_{min}.

In order to construct a χ-network, one needs to 1) estimate χ for all pairs of locations, and 2) choose χ_{min}. For task 1) one must estimate the marginal distribution at each location, and to choose u sufficiently close to 1 to estimate χ, the limiting conditional probability, while at the same time retaining enough data to perform estimation. Choosing estimation thresholds such as u is a topic which continues to be of great interest in the extremes community. As simple empirical estimators of χ are known to be quite variable, in Section V we use an estimator for a different extremal dependence measure, the F-madogram [16], and convert this to its corresponding value for χ.

V. SAMPLE χ-NETWORK

Fig. 1 shows a χ-network for the study of extreme dependence of daily precipitation across different locations along the US Gulf Coast. This network is based on precipitation data from the *Global Historical Climatology Network (GHCN)* data set [17], which covers the US Gulf Coast (Texas, Louisiana, Mississippi, Florida and Georgia) from Jan. 1949 to Oct. 2017. We only use those stations of the GHCN that have less than 10% missing daily values. This results in 339 stations. To focus on the hurricane season only daily data from June to Oct are used.

Extremal dependence of daily data is likely to have very limited spatial extent. We were interested if extremes on longer time scales show long-range extremal dependence (this is discussed further in Section VI). Hence, we calculate the annual maximum for each location considering only the hurricane season (June-Oct.), and the χ-measure is calculated for all pairs of considered GHCN stations. As it has been suggested that there might be an increasing trend in precipitation extremes, we investigated fitting and removing trends, but found doing so did not appreciably affect the network.

8th International Workshop on Climate Informatics
September 19-21, 2018
Hosted by the National Center for Atmospheric Research in Boulder, CO

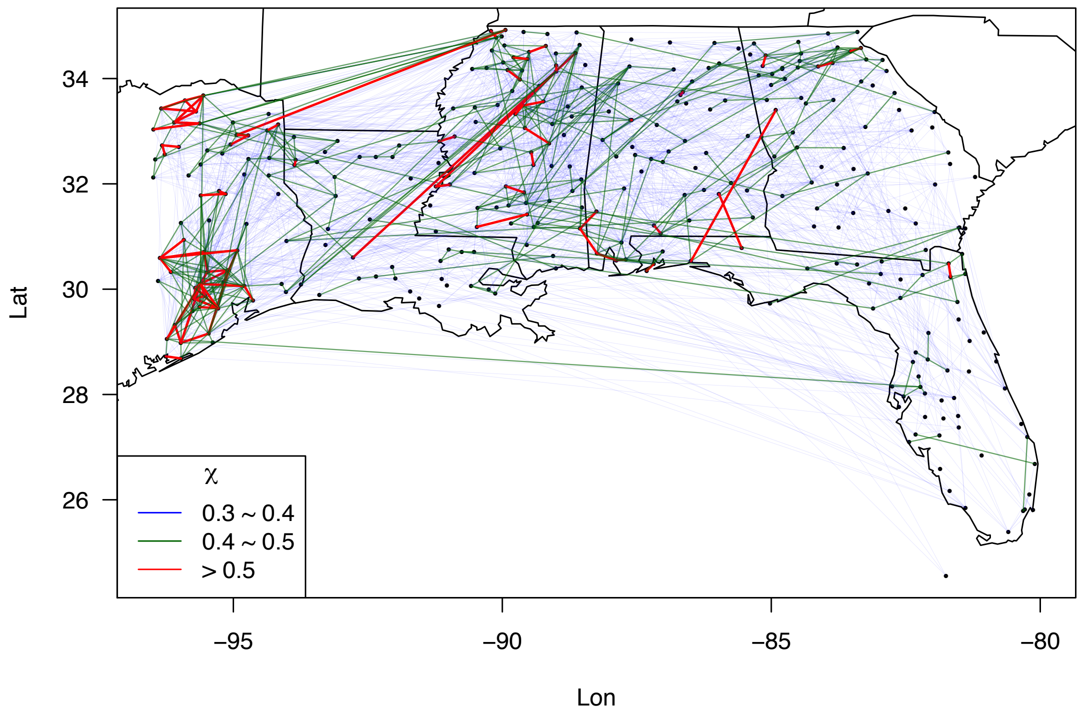

Fig. 1. χ network for analysis of extremal dependence of precipitation during hurricane seasons across US Golf Coast, showing connection with a chi-value of at least 0.3. Weakest connections are in blue, medium ones in green and the strongest in red. (Blue lines are best viewed in the digital (not printed) version, as many of them might not show up when printed.)

The resulting network in Fig. 1 shows lines connecting any two points with a χ-value of at least 0.3. The weakest such links are shown in blue, stronger ones in green and the very strongest in red. Recalling the definition of the χ-measure, each connection in Fig. 1 indicates that the probability of one point having an extreme value in a year is high, given that the other has an extreme value the same year. To be more precise, for any two points connected by a red line we know that if one location has an extreme (e.g. more extreme than 10-year return level) in a year, there is at least a 50% chance that the other location does, too (i.e., having an event that is more extreme than 10-year return level). The network provides a very powerful visualization of these dependencies.

It will also be interesting to study *properties* of this network in the future, such as the average distance of connections at each location, or the local degree. The irregular locations must be taken into account for these measures - average distance should be less affected, but the degree of a location with many points nearby would be biased toward having a high degree.

VI. χ-NETWORKS VERSUS ES NETWORKS

There are many similarities between event synchronization and the χ-measure. When applied to daily data – which is the predominant way in which ES measures are used to date –, both first extract extreme events from the time series, then seek to measure how often extreme events at one point are synchronized with events at another point. The key difference is that the ES measures use a synchronization window (τ), while χ does not. However, it turns out that preprocessing can fill this gap. Specifically, when using the simplifications described in Section IV to calculate the χ-measure, then careful comparison of χ to the undirected ES measure, Q, reveals that Q is actually extremely similar to combining χ with certain preprocessing steps. For example, for daily data, if one first takes the N-day mean (or maximum) of the data - as is commonly done in such applications - and then calculates the simplified χ-measure, the results are very similar to Q with a synchronization window of N days. This relationship provides a new interpretation, and thus some theoretical foundation, to the undirected ES measure, Q, which can now be understood as being very similar to the simplified version of the χ-measure with preprocessing.

In contrast, the directed ES measure, q, does not have a straight forward interpretation through the χ-measure. While it is possible to express q in terms of the χ-measure, that connection is not particularly helpful. (Namely, q can be expressed as the sum of two

χ-measure terms which result from applying χ twice to the same data but after different preprocessing methods have been applied,)

On the EVT side, on the other hand, we have only scratched the surface of possibilities, namely only considered the χ-measure and only with highly simplifying assumptions. Thus there is a large space of measures yet to be explored for network construction, most of which are not expected to be as similar to Q.

Note that both types of measures, based on EVT and ES, can be applied at different time scales. Using short time scales, e.g., daily data, as is commonly done in ES networks, allows us to explore *token-level* dependencies, addressing questions such as the following: *How do individual extreme events propagate? Is flooding in one region always followed by flooding in another region?* In contrast, using larger time scales, e.g., using seasonal or annual averages, as was done in Section V, instead focuses on *type-level* dependencies, addressing questions such as the following: *In years with severe flooding in one region, are there other regions (potentially far away in both time and space) that also exhibit strong flooding? Are there certain conditions that lead to certain groups of events?* The latter type of question is of interest for example to insurance companies.

VII. LIMITATIONS AND FUTURE WORK

Limitations of the current work include the fact that we have not yet demonstrated the robustness of the results in Fig. 1, i.e. we have not yet demonstrated that the results show physical, representative dependencies, rather than just representing a series of historical events. Furthermore, we have not yet provided an interpretation of the results obtained. We are currently working on both of these topics, and are also exploring the effect of high/low annual SST on network structure. We also plan to investigate other EVT measures and their corresponding networks. Nevertheless, we have provided new ideas and concepts that link extreme value theory and complex networks for the first time, and that provide a sound theoretical foundation for future research in this area.

ACKNOWLEDGMENTS

We thank Ken Kunkel for providing the GHCN data set. This material was based upon work partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute (SAMSI). Support was also provided by NSF grants AGS-3891445978 (Ebert-Uphoff), and DMS-1243102 (Cooley). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] J. Sillmann, T. Thorarinsdottir, N. Keenlyside, N. Schaller, L. V. Alexander, G. Hegerl, S. I. Seneviratne, R. Vautard, X. Zhang, and F. W. Zwiers, "Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities," *Weather and Climate Extremes*, vol. 18, pp. 65 – 74, 2017.

[2] R. R. Heim Jr, "An overview of weather and climate extremes–products and trends," *Weather and Climate Extremes*, vol. 10, pp. 1–9, 2015.

[3] A. A. Tsonis and P. J. Roebber, "The architecture of the climate network," *Physica A: Statistical Mechanics and its Applications*, vol. 333, pp. 497–504, 2004.

[4] K. Steinhaeuser, N. V. Chawla, and A. R. Ganguly, "Complex networks in climate science: Progress, opportunities and challenges.," in *CIDU*, pp. 16–26, 2010.

[5] N. Malik, N. Marwan, and J. Kurths, "Spatial structures and directionalities in monsoonal precipitation over south asia," *Nonlinear Processes in Geophysics*, vol. 17, no. 5, pp. 371–381, 2010.

[6] N. Malik, B. Bookhagen, N. Marwan, and J. Kurths, "Analysis of spatial and temporal extreme monsoonal rainfall over south asia using complex networks," *Climate dynamics*, vol. 39, no. 3-4, pp. 971–987, 2012.

[7] N. Boers, A. Rheinwalt, B. Bookhagen, H. M. Barbosa, N. Marwan, J. Marengo, and J. Kurths, "The south american rainfall dipole: a complex network analysis of extreme events," *Geophysical Research Letters*, vol. 41, no. 20, pp. 7397–7405, 2014.

[8] K. Yamasaki, A. Gozolchiani, and S. Havlin, "Climate networks based on phase synchronization analysis track el-niño," *Progress of Theoretical Physics Supplement*, vol. 179, pp. 178–188, 2009.

[9] I. Ebert-Uphoff and Y. Deng, "Causal discovery from spatio-temporal data with applications to climate science," in *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*, pp. 606–613, IEEE, 2014.

[10] T. Zerenner, P. Friederichs, K. Lehnertz, and A. Hense, "A gaussian graphical model approach to climate networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 24, no. 2, p. 023103, 2014.

[11] M. Kretschmer, D. Coumou, J. F. Donges, and J. Runge, "Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation," *Journal of Climate*, vol. 29, no. 11, pp. 4069–4081, 2016.

[12] R. Q. Quiroga, T. Kreuz, and P. Grassberger, "Event synchronization: a simple and fast method to measure synchronicity and time delay patterns," *Physical review E*, vol. 66, no. 4, p. 041904, 2002.

[13] S. Coles, *An introduction to statistical modeling of extreme values*. Springer series in statistics, London ; New York: Springer, 2001.

[14] J. Yan and D. K. Dey, *Extreme Value Modeling and Risk Analysis*. New York: Chapman and Hall/CRC, 1st ed., 2015.

[15] S. Coles, J. Heffernan, and J. Tawn, "Dependence measures for extreme value analyses," *Extremes*, vol. 2, no. 4, pp. 339–365, 1999.

[16] D. Cooley, P. Naveau, and P. Poncet, "Variograms for spatial max-stable random fields," in *Dependence in probability and statistics*, pp. 373–390, Springer, 2006.

[17] M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, "An overview of the global historical climatology network-daily database," *Journal of Atmospheric and Oceanic Technology*, vol. 29, no. 7, pp. 897–910, 2012.