Multivariate dependence among extremes, abrupt change and anomalies in space and time for climate applications

Auroop R. Ganguly
Oak Ridge National Laboratory

Tailen Hsing
Ohio State University

Rick Katz
Nat'l Ctr. for Atmospheric Research

David J. Erickson III
Oak Ridge National Laboratory

George Ostrouchov
Oak Ridge National University

Thomas J. Wilbanks
Oak Ridge National Laboratory

Noel Cressie
Ohio State University

ABSTRACT
This paper discusses multivariate spatio-temporal dependence between extremes or abrupt change and unusual values or anomalies in the context of climate dynamics and climate change. In climate, as in many other applications, anomalies (or extremes) in one variable like sea surface temperature may be a precursor for extremes (or abrupt change) in another variable like regional precipitation. In addition, this multivariate dependence may be spatially or temporally lagged, owing to climate "teleconnections". However, the anomalies may not be easily detectable and their dependence with extremes and rapid change may be difficult to quantify. This paper provides a brief review of the literature, which is followed by a description of critical gaps, both in the data or computational sciences as well as in the climate sciences. The quantification and visualization of multivariate dependence among extreme values and anomalies in highly nonlinear or stochastic systems is an emerging research area in theoretical statistics, with limited development in application areas and/or for massive or disparate space-time data. Further development is needed in these areas for multiple domains ranging from climate sciences and geography to sensor networks and national security.

Categories and Subject Descriptors
I.5.4 [**Pattern Detection**]: Applications; G.3 [**Probability & Statistics**]: Multivariate statistics; Statistical computing; J.2 [**Physical Sciences & Engineering**]: Earth/atmospheric sciences.

General Terms
Algorithms.

Keywords
Extremes, Anomalies, Dependence, Spatio-temporal, Climate.

Workshop on "Data Mining for Anomaly Detection" n conjunction with Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '05, August 21, 2005, Chicago, IL, USA. Copyright 2005 ACM 1-58113-000-0/00/0004…$5.00.

1. INTRODUCTION
Multivariate dependence among extremes and anomalies in climate variables is important for a number of reasons. Climate anomalies like unusual sea surface temperatures in one part of the globe may cause imbalances in the atmospheric system leading to phenomena like the El Nino Southern Oscillation, which in turn may cause extremes in precipitation at regional scales. One consequence of global warming may be a preponderance of regional climate extremes like heat waves or intense storms. The inherently nonlinear climate system may be triggered to a different behavioral mode or equilibrium state due to sudden or gradual disruptions induced by human activities. There are several aspects to this problem. Extremes may be defined in terms of static or dynamic thresholds (e.g., exceedence over fixed threshold or thresholds that depend on location and time) and/or may be context-specific (e.g., temperature for human comfort levels, precipitation for optimal agricultural productivity). Large anomalies like the El Nino phenomena may be relatively easily detectable from data, while other anomalies may be more elusive. In certain situations, the anomalies may not be directly detectable from data and may have to be defined ex-post or from historical data analysis, for example, as a set of conditions among multiple variables that lead to extremes in another variable of interest. Thus, the problem of multivariate dependence among extremes, abrupt change, anomalies and unusual values is arguably more important in climate than the detection of anomalies from data.

2. PROBLEM STATEMENT
The ability to understand and eventually predict climate extremes and abrupt change is critical for science [1-2, 12-13, 16-17] and policy [17-20], and can help answer questions like the following: (a) Are heat waves or precipitation extremes likely to grow more intense in the next century? (b) Is an increase in Atlantic hurricane activity caused by warming of the earth at global or regional scales? (c) Is Sahel Africa likely to experience more extensive droughts in the next few decades? (d) Does an observed strong anomaly in sea surface temperatures in the Eastern Pacific imply extreme rainfall in South America within a few months? (e) Can abrupt change in historical (paleo-) climates be related to leading indicator variables, and can these be used to assign likelihoods of future change? (f) What are the likelihoods, risks and impacts of global or regional scale abrupt change or extremes in climate, and how can the adverse impacts be mitigated?

3. STATE OF THE ART

The current generation of general circulation models (GCM) yields precise, but not necessarily accurate, simulations of future climate scenarios. The Oak Ridge National Laboratory (ORNL) has vast quantities of climate simulations based on "IPCC runs" (*IPCC: Inter-governmental Panel on Climate Change*): the latest runs are global 3-hourly outputs for the atmosphere available from 2000 to 2100 at roughly 100 km spatial resolutions. Government agencies like NASA and NOAA have significant amount of observed data based on remote or in-situ sensors of climate variables. Our understanding of climate extremes and abrupt change can dramatically improve if the vast quantities of observed and simulated data can be mined using focused methodologies. The climate system involves potentially nonlinear dependence among multiple variables dimensioned by space and time, and exhibits strong teleconnections, or geographically dispersed dependence. Thus, methodologies for multivariate and potentially nonlinear dependence in space and time, geared towards extreme values, abrupt change or anomalous behavior, are key requirements. However, the literature in climate extremes [16; 1-2, 17] rely on simplistic statistics. Time series anomaly detection methods have been developed in statistics [5] and nonlinear dynamics [8], and applied to environmental problems [11-14]. However, methods for multivariate extremal and nonlinear dependence, especially in space and time [4, 9], are not well developed. Extreme value theory in statistics [3], and its environmental applications [6, 12-13], are well-established. However, the literature on multivariate dependence among extreme values and their visualization [15, 7, 10] is beginning to emerge. Novel methods are needed for massive, space-time data.

4. NEW DIRECTIONS

The focus of our ongoing research at ORNL, with collaborators at NCAR and OSU, is to develop new approaches for multivariate dependence among extremes, abrupt change and anomalies for potentially large data sets that are dimensioned by space and time, and then implementing these new approaches in the context of regional and global climate change and climate teleconnections. We are further developing and implementing recent advances in statistical theory of extreme values, including specialized probabilistic models for temporal and spatial extremal patterns and lagged dependence, as well as approaches to relate extremes in the dependent variable with temporally or spatially lagged anomalies or extremes in the independent variables. In addition, we are developing new theories and measures for multivariate dependence among extreme values and anomalies. The theories developed for extremes are applicable to abrupt change, following the differencing operation. The extensions and new formulations are being designed for easy visualization and quantification of the multivariate dependence among extremes, abrupt changes and anomalies, as well as for applications to massive data. The approaches are expected to yield the uncertainty associated with the anticipated extremes or unusual events at multiple scales, and relate these uncertainties to risks and economic/societal impacts.

ACKNOWLEDGMENTS

Research sponsored by the Laboratory Directed Research and Development Program (SEED funds) of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U. S. DOE under Contract No. DE-AC05-00OR22725. Workshop participation sponsored by ORNL's SensorNet® R&D program and Knowledge Discovery initiative.

REFERENCES

[1] American Institute of Physics, *Rapid climate change*, http://www.aip.org/history/climate/rapid.htm, 2004.

[2] Alley, R.B. et al, Abrupt climate change, *Science*, 299, 2005-10, 2003.

[3] Coles S., *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, 2001.

[4] Cressie, N., *Statistics for Spatial Data*, Rev. Edition, Wiley, 1993.

[5] Downing, D.J., V.V. Fedorov, W.F. Lawkins, M. D. Morris, and G. Ostrouchov, Large data series: Modeling the usual to identify the unusual, *Computational Statistics & Data Analysis*, 32, 245, 2000.

[6] Erickson, D.J. III, and J.A. Taylor, Non-Weibull behavior observed in a model-generated global surface wind speed frequency distribution, *Journal of Geophysical Research*, 94, 12, 693-8, 1989.

[7] Gilleland, E., R. Katz, and G. Young, *Extremes toolkit (extRemes): weather and climate applications of extreme value statistics*, National Center for Atmospheric Research, 2004.

[8] Heffernan, J.E., and J.A. Tawn, A conditional approach for multivariate extreme values, *Journal of the Royal Statistical Society*, Serial B, 66(3): 497-546, 2004.

[9] Hively, L.M., P.C. Gailey, and V. Protopopescu, Detecting Dynamical Change in Nonlinear Time Series, *Physics Letters A*, 258, 103, 1999.

[10] Hoffman, F.M., W.W. Hargrove, D.J. Erickson, III, and R. Oglesby, Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models, to appear in *Earth Interactions*, 2005.

[11] Hsing, T., Kluppelberg, C., and Kuhn, G., Dependence estimation and visualization in multivariate extremes with applications to financial data, to appear in *Extremes*, 7, 99-121, 2005.

[12] Jin, Y.-H., A. Kawamura, K. Jinno, and R. Berndtsson, Nonlinear multivariable analysis of SOI and local precipitation and temperature, *Nonlinear Processes in Geophysics*, 12: 67-74, 2005.

[13] Katz, R.W., G.S. Brush, and M.B. Parlange, Statistics of extremes: Modeling ecological disturbances, *Ecology*, 86: 1124-1134, 2005.

[14] Katz, R.W., M.B. Parlange, and P. Naveau, Statistics of extremes in hydrology, *Advances in Water Resources*, 25: 1287-1304, 2002.

[15] Khan, S., A.R. Ganguly, S. Saigal, Detection and predictive modeling of chaos in finite hydrological time series, *Nonlinear Processes in Geophysics*, 12(1): 41-53, 2005.

[16] Ledford, A.W., and J.A. Tawn, Diagnostics for dependence within time series extremes, *Journal of the Royal Statistical Society*, Serial B, 65(2): 521-543, 2003.

[17] Meehl, G.A., and C. Tebaldi, More intense, more frequent, & longer lasting heat waves in the 21st century, *Science*, 305: 994-7, 2004.

[18] NRC, *Abrupt Climate Change: Inevitable Surprises*, Committee on Abrupt Climate Change, National Research Council, Washington, DC, USA, 230 pp., 2002.

[19] Perrings, C., The economics of abrupt climate change, *Philosophical Transactions of the Royal Society*, London, A 361, 2043-59, 2003.

[20] Schneider, S. H., B. L. Turner, and H. Morehouse Gariga, Imaginable Surprise in Global Change Science, *Journal of Risk Research*, 1 (2): 165-185, 1998.

[21] Schneider, S. H., K. Kuntz-Duriset, and C. Azar, Costing Nonlinearities, Surprises, and Irreversible Events, *Pacific and Asian Journal of Energy*, 10 (1): 81-106, 2000.