Extreme events in climate and weather–an interdisciplinary workshop

Peter Guttorp, University of Washington/Norwegian Computing Center
Montserrat Fuentes, North Carolina State University

August 22- 27, 2010

August 22-27 an interdisciplinary workshop was held at the Banff International Research Station on Extreme events in climate and weather. The workshop had 32 participants from 7 countries. The goal of the workshop was to set a research agenda for statistical analysis of extreme climate events. The format was two lectures in the mornings, and group discussions in the afternoons. The lectures and posters, as well as a list of the participants, are available at the meeting website http://temple.birs.ca/ 10w5016

This document contains an overview of the lectures (section 1) and a summary of the discussions of research directions held at the workshop (sections 2-6). Each research direction has some concrete needs in boldface.

1 Lectures

1.1 Introduction

In the opening lecture Peter Guttorp described how general circulation models work, going from a simple energy balance model to the modern gridded solutions to a system of partial differential equations. He discussed the various sources of uncertainty in climate analysis, including uncertainties in forcings and data, and how well the models cope with features such as El Niño. The downscaling of global models to regional ones was demonstrated, and a comparison of data to a regional model was illustrated.

The second lecture of the first day was given by Eric GIlleland, who demonstrated the different scales of extreme weather events. He described the various definitions of droughts and heat waves, and illustrated some analyses based on extreme value theory. Analyses based on global models benefit from developing large-scale indicators of extreme weather, such as the product of maximum wind speed and wind shear, illustrated with various approaches to analysis and forecasting/projection.

1.2 Time series extremes

On the second day, Georg Lindgren posed the question whether a fixed seasonal model is adequate for analysis of weather related extremes in the presence of strong seasonal effects, and demonstrated in a simulation study that high quantiles may be underestimated by a factor of 2-3. He also discussed peaks over threshold analysis, and the effect of a temporal trend in the seasonal amplitude, suggesting use of nonparametric quantile regression in the latter case.

In the second lecture RIck Katz described how to use extreme value theory to model nonstationary weather phenomena. The effect of scaling and aggregation was illustrated. For clustered events, such as heat spells, he suggested modeling the clustering, rather than the common declustering approach. Models of damage from extreme weather events were presented and related to insurance issues. He closed with a discussion of risk communication under climate change, suggesting that we stop using the return level terminology under a changing climate, and rather use a probabilistic language.

1.3 Spatial extremes

Zhenyung Zhan opened the third day by presenting an analysis of US precipitation, with the emphasis on tail dependency between spatial pairs of stations. The main tool for testing for tail dependency was the tail quotient correlation coefficient with random thresholds, for which asymptotic theory was developed and checked using simulation studies. Strong tail dependency was found in many cases between stations quite far apart,

Dan Cooley gave the second lecture, making a distinction between weather and climate in the context of spatial dependence, where climate effects are changes of marginal distribution by location, while weather effects is the joint behavior of multiple locations. He suggested to replace he classical regional frequency analysis from hydrology with a Bayesian hierarchical model. The relatively standard conditional independence assumption made in such models was shown not to hold because of weather effects. The statistical difficulties with max-stable processes casts some doubt over whether this is the right approach. He showed an attempt to make approximate Bayesian inference using composite likelihood replacing the (uncomputable) real likelihood, showing improvement over the conditional independence model.

1.4 Forests and observing networks

Charmaine Dean described a mixture-modeling approach to Canadian forest fires, where the probability of a fire in a given location is a mixture of normal, extreme, and zero-heavy components. The question of interest is trends in this type of model, where the trend is both in the parameters of the different components and in the mixing proportions. The results show a movement from zero-heavy to normal risk, while the probability of the extreme component is relatively stable. She discussed data issues, such as changing detection efficiency and fire management strategies.

In the final lecture, Paul Whitfield illustrated the usefulness of networks of stations to study precipitation patterns as well as extreme precipitation (defined as at least one station in the network having extreme precipitation). The Pineapple Express, delivering moisture from the Pacific to the northwestern Americas, was used as a particularly interesting illustration. The effect of climate change on jet stream paths is still somewhat uncertain.

2 Climate models and extremes

Climate models and regional models produce outputs for several meteorological variables at predetermined spatial scale in terms of averages over a grid cell or areas over a grid cell. An interesting area of research is to develop models to downscale the outputs of such models to point level, or even to develop models to downscale predictions for extremes obtained from climate or regional models (Mannshardt-Shamseldin et al., 2010). Adapting to climate models methods already adopted to downscale outputs from air quality models, a possible approach would use historical station data and regress it on the regional model output for the grid cell where the meteorological station lies using coefficients that vary in space and time (Berrocal et al., 2011). The coefficients would then be in turn modeled using appropriate statistical models. Possibilities include Gaussian processes, appropriate transformation of Gaussian processes, Gaussian copulas, or Dirichlet processes, all with an autoregressive structure in time.

Output from climate models at different scales can be used to generate testable hypotheses about changes in weather.

There is a need to develop models to downscale the outputs of such models to point level, or even to develop models to downscale predictions for extremes obtained from climate or regional models.

2.1 Comparison of regional models to data

The interpretation of a grid square value depends on several factors, such as the numerical solution scheme, the boundary values and forcings chosen, etc. In fact, a regional model value may be a better descriptor of the area around the grid square than the precise grid square itself. To accommodate the fact that the output of a regional and a climate model really refers to a neighborhood of a grid cell, it might be more appropriate to actually regress the historical station data on the regional/climate model output at neighboring grid cells to the one where the meteorological station lies.

The distribution of observed (or reanalyzed) weather needs to be compared to distributions from the climate model. One can borrow spatial strength from nearby stations to predict grid square observation, or use spatial regression tools to predict station observations from model output. Statistical issues here include developing tools for multivariate two-sample comparisons,

Seasonal-to-decadal predictions (10-30yr) form an active field of research in the course of CMIP5 (Coupled Model Intercomparison Project 5). Decision makers that need to account for climate change adaptation and mitigation measures are particularly interested in these predictions, especially in terms of extreme climate events. The potential for skillful decadal predictions depends largely on the initialization of the GCM1 (Global Climate Model) runs and whether the GCMs simulate sufficient decadal climate variability, both in magnitude and structure (Meehl et al. 2009). It is, in this context, very important to investigate GCM ensembles since multiple initial conditions with contrasting parameter values and model structure are needed in order to capture extreme events in transient systems.

There is a need to develop more appropriate methods for validating the representation of extreme events in models.

2.2 Skill scores for climate models

With the new CMIP5 database becoming available in the next years, we will have multiple decadal model predictions available, for which we have to find appropriate statistical methods to analyze various aspects regarding extreme values. These include (1) skill assessment of the GCM predictions, (2) comparison of multi-model ensemble distributions to observations, and (3) determination of uncertainties in the predictions.

Concerning the skill assessment of GCMs, it will be important to first identify mechanisms (climate and/or weather patterns such as El Nino and atmospheric blocking, topography, etc.) that contribute to the occurrence of extreme events but can also be well captured with the available model resolution. Based on that knowledge we need to extend the current practice (i.e. described in Tebaldi and Knutti. 2007; Ferr0, 2007) and develop objective skill measures, that could be region specific, to rank GCMs within the multi-model ensemble. Numerous methods have been proposed in the forecasting literature to spatially verify weather forecasts on small scales, see Gilleland et al. (2009) and Gilleland et al. (2010). To be used in the current context, these methods need to be adapted both to account for the large scale of the GCMs and to focus on the specific mechanisms of interest.

Once a ranking of the multi-model GCM ensemble has been established, it can be used to obtain a probabilistic distribution from the appropriately weighted ensemble members (Friedrichs and Hense, 2007). Further statistical tools are then needed to evaluate the skill of the probabilistic distribution obtained from the weighted ensemble as compared to observations/reality. Such tools should also involve the estimation of the uncertainties (e.g. model based, scenario based) in the predictions. The usefulness of commonly applied skill measures including correlation coefficients or MSE (Kharin et al. 2009) should be analyzed in this context. To ensure propriety in the evaluation, the procedure should be based on proper scoring rules for probability distributions, such as the logarithmic score or the continuous rank probability score (Gneiting and Raftery, 2007; Stephenson et al., 2008).

There is a need for appropriate statistical methods to analyze various aspects regarding extreme values. These include (1) skill assessment of the GCM predictions, (2) comparison of multi-model ensemble distributions to observations, and (3) determination of uncertainties in the predictions

3 Multivariate extreme value tools

It would be desirable to have a peaks-over-threshold approach for multivariate extreme values. One possible approach would be to use the point process representation of max-stable processes to develop more tractable multivariate extremes models. We also need tools for dealing with extremes in vector block extremes that occur at different times during the block.

There is a need for appropriate statistical methods to analyze time series of multivariate extreme values.

4 Spatial and space-time models for extreme values

4.1 Tractable spatial models

The max-stable processes, while mathematically seemingly well suited to the analysis of spatial extremes, are statistically not very tractable, and only ad hoc approaches to their statistical inference are currently available.

Another approach to temporal extremes of space-time processes is to assign a spatial prior for the GEV-parameters for annual or seasonal extremes over a network of station while treating the stations as conditionally independent. However, considering, for example, temperature data, it is quite common in upper latitudes for extremely cold weather to arise from Arctic air masses in a high pressure situation. Hence there is a tendency for annual minima to appear simultaneously at several stations. The appropriate likelihood (as long as separated minima can be considered independent) would be the product of conditional densities of the nonextreme sites, given the values at the extreme sites. Calculating these densities can of course be a daunting task in itself, but the approximation due to Heffernan and Tawn (2004), appropriately extended to the situation at hand, would be a possibility.

There is a need for methods that are well suited to the analysis of spatial extremes and that are statistically tractable.

4.2 Temporal nonstationarity

It seems that the most usual approach to deal with 'non-stationarities' in extremes is to allow for parametric changes in time for a GEV distribution. Linear trends on GEV parameters are a first step and had shown practical use to assess long term changes in climate/weather extremes. On the other hand, there has been some work in using Generalized Additive Models and state-space models to accommodate smooth/non-linear parameter change (Davison and Ramesh, 2000; Yee and Stephenson, 2007). An interesting idea is to apply Hidden Markov Models to represent change points in time and cluster structure. A state-space model can account for seasonalities with time-varying amplitudes. In general, it is unclear that in analyzing time series of extremes that arise in weather and climate, these perhaps more flexible models are more useful than simple linear trends. What model comparison tools are available to learn about this?

There is a need for models with distributional changes and in particular with different shape parameters, but one has to be careful about estimating this shape parameter. It was recommended to first model changes in location, then location/scale, and finally consider location/scale/shape. Beyond parameter changes, we may also need to consider temporal dependence in extremes.

One main problem with nonstationarity of extreme data is our poor understanding of natural low frequency climate oscillations. For instance, the Atlantic Multi-Decadal Oscillation (AMO), arising from the slow oscillation in the strength of the North Atlantic thermohaline circulation, has a period of about 70 years, with a profound effect on the number and strength of Atlantic hurricanes. Our short climate records make it difficult to detect/understand very low frequency climate oscillations, and their contribution to nonstationarity in our relatively short records of extreme data.

There is a need for methods that are suitable for series that are non-stationary, whether that non-

stationarity is in location, distribution, etc., and such methods need to be able to address the connection to climate time series.

5 Climate and weather extremes

5.1 Heat waves

Heat waves are a complex form of extreme climate event with substantial health impacts. Yet extreme value theory has rarely been applied. Challenges include how to model the temporal clustering of temperatures at high levels and whether multivariate extreme value theory can be used to model climate variables that can contribute to heat waves (e.g., maximum and minimum temperature, dew point or humidity, wind speed, cloud cover) (Coles et al.,1994; Smith et al., 1997; Meehl and Tibaldi, 2004; Furrer et al., 2010). Such a research effort is needed to compare the statistics of observed heat waves (frequency, duration, severity) with those simulated by climate models, as well as to detect trends in heat wave statistics.

There is a need to improve the statistical modelling used for heat waves and other meteorological extremes.

5.2 Forest fires

One specific application area considered at the workshop was the analysis of fire events with a view to detecting trends in extremes. Spatio-temporal methods for this important application area have not utilized methodology from extreme value theory. In the forest fire context, increasing temperatures could lead to an increase in the number of ignitions, an increase in the length of the fire season, and an increase in the amount of severe fire weather (Schoenberg et al., 2003). Some additional challenges with quantifying extremes in this context is the need for homogenization of data from long records, incorporating information about changes in suppression activities, and fire management strategies. Given the challenges with climate predictions, it is also unclear what is the best way to accommodate weather variables to evaluate impacts under future climate scenarios; a sensible approach may be to focus on accessing how large a change in weather would lead to specific forestry vulnerabilities.

There is a need for methods that allow reconstruction, restoration, infilling of incomplete records.

There is a need for more robust methods to detect changes in environmental time series that are rich in zeros such as forest fires and ephemeral streamflows.

5.3 Extreme events that are not modeled by extreme value theory

Not all extreme climate events are extreme in the statistical sense. For example, Heavy rain on frozen ground can lead to severe flooding, or high winds following heavy snow and temperatures just around freezing can lead to severe forest destruction. One may be able to use climate model output to get an idea of future frequencies of particularly dangerous combination of factors, not all of which need to be extreme. From a modeling point of view it would be important to estimate conditional joint distributions of variables, given that one is extreme (Heffernan and Tawn, 2004), is one approach to this. Quantities such as trends in the onset of frost appear not directly amenable to extreme value theory, but would rather need nonstationary time series tools for directional data.

There is a need for robust methods that allow the separation of extreme events in relation to the generating processes. Floods being one example of an event with multiple generating mechanisms.

6 Index numbers

Climate and environmental indices need to be [1] robust, [2] specific, [3] relevant, and [4] comparable. There are many indices that may be useful in reducing the dimensions of climate and ecological studies, but many of them are problematic; some such as those that attempt to define the end of drought or the start and end of floods are particularly difficult in practice. Others such as FRICH for comparing global models may be more useful in the model comparison perspective than as test of reality. Indices may be more valuable when considering a change in the index as opposed to its absolute value.

Technically one can view indices as exceedances outside convex manifolds, and perhaps develop functionals that assess the degree of severity of the exceedance.

Water supply forecast models share many of the characteristics of indices as they are generally linear functions of several environmental variables. However, their output is either an estimate of future flow volumes or of their distribution function. Water supply forecasts are only of great importance for low values. EVT may be able to contribute to more rigorous forecasts of low flows which may occur due to the interaction of non-minimal variables. EVT may also be able to quantify the additional uncertainty of low flow volumes due to non-stationarity.

There is a need for methodology that can be used to assess properties of environmental indices, and determine if the are robust, specific, and comparable.

Further, the potential for indices that are not linear combinations, but of non-linear combinations would be useful in filed such as hydrology and climatology where non-linear processes are common, should be explored.

References

[1] V. J. Berrocal,A. E. Gelfand and D. M. Holland, A bivariate space-time downscaler under space and time misalignment. *Annals of Applied Statistics*, in press, (2011).

[2] S. G. Coles, J. A. Tawn and R .L. Smith, A seasonal Markov model for extremely low temperatures. *Environmetrics*, **5** (1994), 221-239.

[3] A. C. Davison and N. I. Ramesh, Local likelihood smoothing of sample extremes. *J. Royal Statistical Society, Series B*, **62** (2000) 191-208.

[4] C. A. T. Ferro, A probability model for verifying deterministic forecasts of extreme events. *Weather and Forecasting*, **22** (2007) 1089-1100.

[5] P. Friederichs and A. Hense, Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly Weather Review*, **135** (2007) 2365-2378.

[6] E. M. Furrer, R. W. Katz, M. D. Walter and R. Furrer, Statistical modeling of hot spells and heat waves. *Climate Research*, **43** (2010) 191-205.

[7] E. Gilleland, D. Ahijevych, B. Casati and B. Ebert, Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, **24pt** (2009) 1416-1430.

[8] E. Gilleland, J. Lindström and F. Lindgren, Analyzing the image warp forecast verification method on precipitation fields from the ICP. *Weather and Forecasting*, **25** (2010) 124pt9-1262.

[9] T. Gneiting and A. E. Raftery, Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102** (2007) 359-378.

[10] J. E. Heffernan and J. A. Tawn, A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society, Series B*, **66** (2004) 497546.

[11] V. V. Kharin, Q. Teng, F. W. Zwiers, G. J. Boer, J. Derome and J. S. Fontecilla, Skill assessment of seasonal hindcasts from the Canadian Historical Forcast Project. *Journal of Geophysical Research Atmosphere-Ocean*, **47** (2009) 204-223.

[12] E. C. Mannshardt-Shamseldin, R. L. Smith, S. R. Sain, L .O. Mearns, and D. Cooley, Downscaling extremes: a comparison of extreme value distributions in point-source and gridded precipitation data. *Annals of Applied Statistics*, **4** (2010) 484-502.

[13] G. A. Meehl, L. Goddard, J. Murphy, . J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M. A. Giorgetta, A. M. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer and T. Stockdale. Decadal Prediction: Can It Be Skillful? *Bulletin of the American Meteorological Society*, **90** (2009) 14671485.

[14] G. A. Meehl and C. Tebaldi, More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **305** (2004) 994-997.

[15] F. P. Schoenberg, R. Peng and J. Woods, On the distribution of wildfire sizes. *Environmetrics*, **14** (2003) 583-592.

[16] R. L. Smith, J. A. Tawn and S. G. Coles, Markov chain models for threshold exceedances. *Biometrika*, **84** (1997) 24pt9-268.

[17] D. B. Stephenson, B. Casati, C. A. T. Ferro and C. A. Wilson, The extreme dependency score: A non-vanishing measure for forecasts of rare events. *Meteorological Applications*, **15** (2008) 41-50.

[18] C. Tebaldi and R. Knutti, The use of the multimodel ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A*, **365** (2007) 2053-2075.

[19] T. W. Yee and A. G. Stephenson, Vector generalized linear and additive extreme value models. *Extremes*, **10** (2007) 1-19.