Issues in Estimating the Probabilities of Extremes

Prashant Sardeshmukh, Gil Compo, and Cécile Penland

CIRES Climate Diagnostics Center, University of Colorado
and Physical Sciences Division/ERL/NOAA
Boulder, Colorado, U.S.A

Abstract

Reliable estimation of extreme anomaly statistics from observational or reanalysis datasets of limited length is a prerequisite for estimating any changes in those statistics under climate change. Extreme anomalies are by definition rare, which makes direct estimation of their statistics prone to sampling errors, and some form of statistical modeling a necessity. An important issue is whether changes of extremes can be more reliably derived indirectly from changes in the mean and standard deviation of a climate variable or through the direct statistical modeling of extremes using say, Generalized Extreme Value (GEV) distributions. The former approach is physically attractive, but assumes that climate PDFs are characterized completely by their first two moments. The chief virtue of the GEV approach is its quasi-universal applicability associated with limiting behavior. This strength is, however, offset by concerns about the extent to which the limiting behavior obtains in the problem at hand, and the practical weakness of making use of only the extreme and not all values in a data sample to estimate the parameters of the GEV distribution.

We are developing a new and complementary approach to this problem using the first four statistical moments (mean, standard deviation, skewness, and kurtosis) to estimate the entire PDF of a climate variable using the Stochastically Generated Skewed (SGS) probability distribution theory of Sardeshmukh and Sura (2009), which has been validated in both atmospheric and oceanic contexts. An important virtue of this approach is that it uses a physically based model to derive the entire PDF and not just its tails, and also makes use of all data values in a sample to estimate the parameters of that PDF. All of the model's parameters can be estimated from knowledge of the first four moments of the climate variable and its temporal correlation scale. The model can then be run forward in time to generate Monte Carlo statistics of extremes and their sampling error bars, and also to estimate more complicated statistics such as the durations of extreme events that are challenging to estimate using standard Extreme Value approaches. Our approach is particularly well suited for investigating the statistics of precipitation extremes.