A Nested Kernel Density Estimator for Improved Characterization of Precipitation Extremes

Chao Li (lichao@stanford.edu) and Anna M. Michalak (michalak@stanford.edu)
Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, 94305

Research Motivation
The number and intensity of short-term precipitation extremes has recently been a topic of much interest, with record-setting events occurring in the United States, Europe, Asia, and Australia. These events show the importance of characterizing the behavior of short-term (daily and sub-daily) precipitation intensity so as to properly understand and predict the occurrence and magnitude of extreme precipitation events.

One such characterization method is the use of kernel density estimators (KDE), which avoid parametric assumptions, and can therefore uncover complex properties such as multimodality. State-of-the-art kernel density estimators have two major recognized drawbacks, however. The first is that KDEs that use unbounded kernels cannot enforce the fact that precipitation is strictly non-negative, because they are subject to “probability leakage” at the boundary. The second is that they tend to produce spurious fluctuations in the tails of the distribution.

In view of these two issues, the current research aims at presenting a method which represents an improvement over existing ones in more accurately characterizing the behavior of precipitation extremes without strict assumptions, while also being computationally tractable for large datasets.

Model Overview
We present here a nested transformation kernel density estimator (NTKDE), consisting of one or two transformation steps, as described below.

1. Transformation Parameter Identification
To identify the optimal transformation parameters, we adopt the following objective function to direct the optimization search [Wand and Jhon, 1995; Yang and Marron, 1999]:

\[ D_{\text{M}}(p_X) = c \sum \left( f_Y(x) - f_X(x) \right)^2 \]

where \( f_Y(x) \) is the density of the transformed precipitation.

2. Objective Function Minimization
To minimize the objective function \( D_{\text{M}} \), the MATLAB built-in function \( fminbnd \) and a grid search procedure are used to determine the power parameter \( a \) and the Johnson parameter \( \lambda \), respectively.

3. Computational Efficiency
Direct evaluation of the objective function \( D_{\text{M}} \) is computationally very expensive, prohibiting its efficient application to large dataset. To address this issue, we employ the fast \( \alpha \)-exact algorithm [Raykar and Duraiswami, 2006] to approximate the objective function \( D_{\text{M}} \). We have adapted the original algorithm to accommodate the heavy-tailed distributional property of precipitation.

Model Evaluation
We applied NTKDE to 1217 precipitation records from 1961-2008 at daily and monthly scale. The dataset was downloaded from http://ccl.dacc.noaa.gov. Figure 3 shows the locations of these stations with the color code representing the sample size of each daily record.

4. Data Transformation
In the first step, the power transformation is used to transform precipitation observations \( p : x(p) \rightarrow p^a \) where \( p \) is the power parameter. In the second step, if it is needed, NTKDE implements the Johnson transformation \( x(\cdot) \rightarrow \phi^{-1}(\cdot) \), where \( x \sim \phi^{-1}(\cdot) \) being the Johnson parameter to be identified.

5. Investigation to Point 1:
To see the first point, we designed a joint measure of accuracy and smoothness of a tail estimate (MSD). Figure 4 shows MSD evaluated in within-sample (MSD1) and out-of-sample (MSD2) contexts. It is evident that NTKDE performed significantly better than KDE in providing a smooth tail estimate, yet slightly worse than KDE with logarithmic transformation (LOGKDE).

6. Conclusions
Major conclusions from this research can be summarized as follows:

- Compared with KDE, NTKDE can suppress spurious bumps in the upper tail domain.
- Compared with LOGKDE, NTKDE can be more flexibly applied to precipitation of different time scales.
- In practice, NTKDE can be used as a reliable tool to reconstruct the probability density of precipitation across different time scales.

Acknowledgements & Selected References