Modelling the entire range of daily precipitation using mixture distributions

Hyunwoo Rho

1. Introduction

To measure and predict the precipitation are the most oldest concerns of mankind. It is an essential factor in agriculture, forest management, hydrology, and preparing for disasters like flood and drought. For generating stochastic precipitaion model, there are some feasible time intervals as a base period of modelling, but taking a daily basis is considerably natural **?**. Constructing daily stochastic precipitaion model can be extended to precipitation analytics like monthly or yearly basis approach and spatial analysis.

Many of previous studies tried to model daily rainfall data through parametric approaches e.g. **????** and also non-parameteric approaches e.g. **??**Along some studies in hydrology community, the evidence of heavy-tailed phenomenon on the distribution of high precipitation amount **?**. In this perspective, various methods were introduced more focused on tail behavior of daily precipitation **????**.

In this paper, we compare various type of parametric models and introdcue a particular type of mixture distribution called phase-type distribution. Through this model we tried to fit the whole range of continuous daily precipitation.

2. Data Sets

The main data set used to analyse is United States Historical Climatology Network(USHCN) Daily data set. Raw data set contains daily record of precipitation, snowfall, snow depth, maximum temperature, minumum temperature, and information about flag. Among 48 files of each sontiguous states, According as **?**, Texas was selected. For following their work, the same data selection criteria were adopted which are all nonzero precipitation of 1940 to 2009 without taking care of missings.

Additionally, Daily Global Historical Climatology Network(GHCN-DAILY) was used to do extra analyses **??** which has identical data format with USHCN. This set contains about 100 thousand stations and also encompasses USHCN as its subset. By filtering stations along **?**, (a) record length of over 50 years, (b) percentage of missing values less than 20%, data assigned with suspicious "quality flags" less than 0.1%. The screen values of quality flags are two, one with "G"(failed gap check), and another with "X"(failed bound check). For more information about data set, see **?**.

Finally we handled 49 stations in USHCN-Texas data, and 19328 stations in GHCN data after filtering.

3. Existing models

3.1. One component models

Fundamental characteristrics of daily precipitation data are described as follows; non-negative, continuous except at the spike on zero, right-skewed. Assuming we only focus on the continuous part of the data, such distributions include exponential **?**, gamma **???**, and kappa **?**, and so on.

$$f(x;\theta) = \theta e^{-\theta x}, \quad x \geq 0, \theta > 0, \tag{1}$$

3.2. Mixture models

3.3. Hybrid models

4. Phase-type distribution class

5. USHCN Texas daily precipitation data set

6. World data

7. Conclusion

References

Furrer, E. M. and Katz, R. W. (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, **44**(12).

Harrold, T. I., Sharma, A., and Sheather, S. J. (2003). A nonparametric model for stochastic generation of daily rainfall amounts. *Water Resources Research*, **39**(12).

Ison, N., Feyerherm, A., and Bark, L. D. (1971). Wet period precipitation and the gamma distribution. *Journal of Applied Meteorology*, **10**(4), 658–665.

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: I. theoretical investigation/statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. recherche théorique. *Hydrological sciences journal*, **49**(4).

Li, C., Singh, V. P., and Mishra, A. K. (2012). Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water resources research*, **48**(3).

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, **29**(7), 897–910.

Mielke Jr, P. W. and Johnson, E. S. (1973). Three-parameter kappa distribution maximum likelihood estimates and likelihood ratio tests. *Monthly Weather Review*, **101**(9), 701–707.

Papalexiou, S., Koutsoyiannis, D., and Makropoulos, C. (2013). How extreme is extreme? an assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, **17**(2), 851–862.

Papalexiou, S. M. and Koutsoyiannis, D. (2012). Entropy based derivation of probability distributions: A case study to daily rainfall. *Advances in Water Resources*, **45**, 51–57.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, **17**(1), 182–190.

Schoof, J., Pryor, S., and Surprenant, J. (2010). Development of daily precipitation projections for the united states based on probabilistic downscaling. *Journal of Geophysical Research: Atmospheres*, **115**(D13).

Sharma, A. and Lall, U. (1999). A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation*, **48**(4), 361–371.

Stern, R. and Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General)*, pages 1–34.

Todorovic, P. and Woolhiser, D. A. (1975). A stochastic model of n-day precipitation. *Journal of Applied Meteorology*, **14**(1), 17–24.

Wilks, D. S. (1999). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, **93**(3), 153–169.