Estimation and visualization of quantile surfaces in flexible space-time settings

Dana Sylvan[1] **and Constantin Tarcolea**[2]
[1] Hunter College of The City University of New York, USA
[2] University "Politehnica" Bucharest, Romania

The study of quantiles of probability distribution functions is fundamental in mathematical statistics, with widespread applicability. Many real-life processes have been observed to have complex structures generated by their temporal and spatial dependencies, often times showing significant departures from Gaussianity and/or and stationary behavior and thus yielding serious metholodgical challenges. The practical motivation of our study stems from environmentally-related health concerns. Generally, people are adversely affected only by very high values of temperature, precipitation, pollution, thus the need for modeling effects of high order quantiles rather than modeling mean effects. Maximum readings would give a more relevant statistic for monitoring than average values, however, high order quantiles are preferred to maximum values for increased statistical stability. Environmental standards in Europe and North America are set based on various distributional characteristics, therefore information on quantiles is valuable to policy makers. More examples of relevant applications may be found in biomedical research, climatology, ecology, education, finance, sports. Here we present statistical methodology for modeling and prediction of quantile fields in flexible classes of space-time processes.

Specifically, we compare and contrast two situations: (1) observations consist of long time series collected at smaller number of random spatial locations, and (2) data are observed on space-time lattices. The problem of interest is to predict the quantile field at any time and location in the domain where there are no observations. Denote by $X(t, s)$ a random field observed at n time points $\{t_1, \ldots, t_n\}$ and m spatial locations $\{s_1, \ldots, s_m\} \in D \subset \mathbf{R}^2$. For $x \in \mathbf{R}$, let $F_{t,s}(x) = P[X(t, s)) \leq x]$ be the probability distribution function of the process, yielding the space-time-varying conditional quantile function $q_\alpha(t, s) = \inf \{x : F_{t,s}(x) \geq \alpha\}$ for fixed $\alpha \in [0, 1]$.

The first situation is common for geostatistical applications, where typically $m \ll n$, so we may think of a collection of spatially correlated time series observed at random locations in a fixed domain. In this case, guided by the data collection mechanism, it is reasonable to assume space-time separability and model the space-time quantile surface sequentially in two steps. First, we fix the spatial location and estimate the time-varying quantile functions based on the observed time series. To accommodate complex temporal dependencies and allow flexibility across locations, we opt for a nonparametric, data-driven approach, for example by using the smoothed moving window quantile curve estimator introduced in Draghicescu et al. (2009). In the second step we fix t and employ spatial interpolation to predict the quantile field of interest at any location $s_0 \in D$ as $q^*_\alpha(t, s_0) = \sum_{i=1}^m \lambda_i \hat{q}_\alpha(t, s_i)$. Here $\hat{q}_\alpha(t, s)$ is an estimator of the time-varying quantile curve, and the interpolation weights λ_i, $1 \leq i \leq m$ are completely specified by the parameters describing the second order spatial structure. Assuming that the spatial quantile field is isotropic, we model its spatial covariance parametrically as $\mathrm{cov}\Big(\hat{q}_\alpha(t, s), \hat{q}_\alpha(t, s')\Big) = C_\alpha(\theta_t, h)$, where $h = ||s - s'||$ is the Euclidean distance between s, s' and the function C_α is chosen from a flexible class of spatial covariance models, such as the Matérn class. This procedure is known as universal kriging (Stein 1999), and, accordingly, $q^*_\alpha(t, s_0)$ is a best linear unbiased predictor. A similar sequential two-step approach was used in Cameletti et al. (2013) for modeling space-time threshold exceedance probabilities.

In case (2) we assume that the data are collected on space-time lattices and thus the observations are distributed in equally-spaced blocks. The sequential method described before can be used in this scenario as well. However, we may now take advantage of the large number of spatial points where the process is observed and relax the spatial assumptions. Thus, in the second step, instead of kriging, we smooth over space through a bivariate kernel, for instance by adapting methodology in Biau (2003). As an alternative, we introduce a direct (one-step) procedure based on moving space-time blocks of sample quantiles.

For the cases presented above we give theoretical considerations, discuss implementation procedures, and illustrate the findings on applications to environmental space-time data and on Monte Carlo simulations. We conclude by indicating some open problems that will be addressed in future work.

Keywords: space-time quantile maps, smoothing, kriging, random fields

References:

Biau, G. (2003). Spatial kernel density estimation. *Mathematical Methods of Statistics, 12*, 371–390.

Cameletti, M., Ignaccolo, R., Sylvan, D. (2013). Assessment and visualization of threshold exceedance probabilities in complex space-time settings: A case study of air quality in Northern Italy. *Spatial Statistics, 5*, 57–68.

Draghicescu, D., Guillas, S., Wu, W. B. (2009). Quantile curve estimation and visualization for non-stationary time series. *Journal of Computational and Graphical Statistics*, *18 (1)*, 1–20.

Stein, M. L. (1999). *Interpolation of Spatial Data. Some Theory of Kriging.* Springer.