Space-time quantile surfaces of non-stationary random fields: a comparison study

D. Sylvan, C. Tarcolea, A. S. Paris

Abstract. We present statistical methodology for estimation and prediction of quantile fields in flexible settings. Specifically, we compare and contrast quantiles of space-time separable and non-separable processes, and illustrate the findings through Monte Carlo simulations. These quantile estimates make a versatile exploratory tool that may be used to describe various distributional characteristics for data with complex spatial and temporal dependencies and have countless applications in many disciplines.

M.S.C. 2010: 62G99, 62P99.
Key words: Space-time quantile maps; smoothing; kriging; random fields; Monte Carlo simulations.

1 Motivation and background

The study of quantiles of probability distribution functions is fundamental in mathematical statistics and has widespread applicability. Many real-life processes have been observed to have complex structures generated by their variations with time and space, often times showing significant departures from Gaussianity and stationary behavior and thus yielding serious methodological challenges. An example of practical motivation stems from environmentally-related health concerns. Generally, people are adversely affected by very high or very low values of temperature and precipitation, large amounts of pollutants, and so on, triggering the need for modeling effects of high or low order quantiles rather than focusing on mean effects alone. Readings of extreme values would also provide more informative statistics for monitoring than average values, however, higher or lower quantiles are preferred for increased statistical stability. Moreover, environmental standards in Europe and North America are set based on various distributional characteristics, and thus information on quantiles is valuable to policy makers. Alternative approaches involve extreme value theory and functional data analysis, see for example [8] for an El Niño application. For a finance

BSG Proceedings 23. The International Conference "Differential Geometry - Dynamical Systems" DGDS-2015, October 8-11, 2015, Bucharest-Romania, pp. 68-74.

example we refer to [7]. Other relevant applications may be found in biomedical research, climatology, ecology, economics, education, social sciences, sports.

Characterization of the dependence structure plays a crucial role for mathematical modeling and statistical inference of spatio-temporal data, and not much is understood about quantiles of stochastic processes in complex space-time settings. In addition, mathematically, it is very challenging to accurately and realistically describe space-time structure and dynamics. For instance, in the one-dimensional case (time series), one possibility is given by linear processes, discussed in detail in [15] and the references therein. This and subsequent articles analyze estimators of mean functions, however, the methodology can be extended for the estimation of probability distribution functions and consequently of quantiles. Another way to generate one-dimensional dependent processes is via a time-varying transformation $G(t, Z_t)$ of a stationary process Z_t. By allowing the unknown transformation G to vary with time, the probability distribution function of the resulting process may also change, and therefore the process may not be stationary. Moreover, these processes need not be Gaussian. This is a typical working assumption, that many times does not hold for real-life processes. In [9] the authors studied the asymptotic properties of a nonparametric conditional quantile estimator (obtained by inverting a kernel estimator of the probability distribution function) in this setting, where the underlying process Z_t was assumed to be long-range-dependent and Gaussian. A similar estimator was analyzed in [4] for the case when the underlying Gaussian process has short memory (under the general assumption that the correlations are summable). A data-driven procedure for optimal bandwidth selection for these kernel quantile estimators was proposed in [10]. In [6] the authors give asymptotic results for several smoothed quantile curve estimates for nearly stationary causal processes, and propose a data-driven scheme for the selection of optimal smoothing parameter. Regarding the multidimensional case, [1] discusses theoretical and practical aspects related to kernel estimation of the density of spatial fields observed on lattices. A spatial sampling scheme is also introduced, and showed to preserve the rate of the mean squared errors of the proposed estimators. Detailed considerations on nonparametric regression for spatial random fields are made in [2].

In this paper we investigate statistical methodology for modeling, prediction and visualization of quantile fields in flexible classes of space-time processes. Specifically, we compare and contrast novel quantile estimates for space-time separable and non-separable processes, quantify and assess their uncertainty, and illustrate the findings through extensive Monte Carlo simulations. The remainder of the paper is organized as follows. Section 2 presents the theoretical framework, simulations results are presented in Section 3, with concluding remarks given in Section 4.

2 Theoretical framework

We present a mathematical framework for modeling and prediction of quantile fields under two scenarios two: (1) observations consist of long time series collected at smaller number of random spatial locations in a fixed domain, and (2) data are observed on space-time lattices. The first situation is common for geostatistical applications, where typically m is much smaller than n and 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. Detailed analyses of smoothed quantile estimates for time series are given in [5] and [6]. In the second step we fix the time and predict the quantile of interest at any location in the the domain under consideration. We give below two options for this second step, namely by using smoothing in space yielding $\hat{q}_{1,\alpha}$ or spatial interpolation (kriging) yielding $\hat{q}_{2,\alpha}$, respectively.
Denote by $X(t,s)$ a space-time random field observed at n time points $\{t_1,\ldots,t_n\}$ and m spatial locations $\{s_1,\ldots,s_m\} \in D \subset \mathbf{R}^d$, where d is the spatial dimension. 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 $q_\alpha(t,s) = \inf\{x : F_{t,s}(x) \geq \alpha\}$ defined for fixed $\alpha \in [0,1]$. [5] proposed a nonparametric, completely data-driven approach to identify space-time patterns in extreme Swiss precipitation. Assuming space-time separability, an estimator of $F_{s,t}(x)$ is obtained by sequentially smoothing in time and space as

$$\hat{F}_{s,t}(x) = \frac{1}{nb}\sum_{i=1}^{n}\sum_{j=1}^{m} K_t\Big(\frac{t_i - t}{b}\Big) K_{s,B}(s_j - s) 1_{\{X(t_i,s_j)\leq x\}}.$$

Here time points and spatial locations are rescaled $t_i = i/n, s_j = j/m$, respectively, K_s is a d-variate kernel with $\int_D K_s(u)du = 1$, and K_t is a one-dimensional density kernel. Furthermore, $K_{s,B} = |B|^{-\frac{1}{2}} K_s(B^{-\frac{1}{2}}s)$, where B is a symmetric positive-definite $d \times d$ matrix (the bandwidth matrix). The space-time quantile estimator is then determined as

$$\hat{q}_{1,\alpha}(t,s) = \inf\{x : \hat{F}(t,s)(x) \geq \alpha. \tag{2.1}$$

Proof of the asymptotic behavior of these estimators, as well as small sample considerations are given in the aforementioned paper. The main motivation behind this research project was a conjecture that by switching to a parametric approach in the second step we may obtain better quantile estimates. This will be illustrated in the simulation study in next section.
As an alternative, we conjecture that when the temporal quantile curves may be assumed as realizations of a Gaussian process, the second step may be replaced by kriging (or spatial interpolation) yielding the second quintile estimator that we are going to use in our comparison.

$$\hat{q}_{2,\alpha}(t,s) = \frac{1}{ml^d}\sum_{j=1}^{m} \lambda_j \tilde{q}_\alpha(t,s_j), \tag{2.2}$$

where $\tilde{q}_\alpha(t,s_j)$ is a smooth estimator of the time-varying quantile curve, described below, and $\lambda_j, 1 \leq j \leq m$ are interpolation weights. We further assume here thatthe initial process is a time-varying transformation of a Gaussian filed $X(t_i,s) =: X_{i,n;s} = G_s(i/n; Z_i)$. Thus $Z_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots)$, $(\varepsilon_i)_{i\in\mathbf{Z}}$ are independent, identically distributed random variables, and G_s is a measurable function (unknown). Then

$F_{t,s}(x) = P[G_s(t; Z_i) \leq x],\ x \in \mathbf{R},\ 0 \leq t \leq 1$ is as before the probability distribution function of the process $(X_{i,n;s})_{i=1}^n$, yielding the quantile function $q_\alpha(t,s) = \inf\{x : F_s(x;t) \geq \alpha\}$. If there exists a constant $L < \infty$ such that for all $0 \leq t, t' \leq 1$, $\sup_{x \in \mathbf{R}} |F_s(x;t) - F_s(x;t')| \leq L|t - t'|$, the process is called *nearly stationary* ([6]). For rescaled time t_0, let $n_1 = \lfloor n(t_0 - b_n) \rfloor$, $n_2 = \lfloor n(t_0 + b_n) \rfloor$ and $p = n_2 - n_1 + 1$, where b_n is a sequence of bandwidths such that $b_n \to 0, nb_n \to \infty$ as $n \to \infty$. We can then estimate the space-time quantile $q_\alpha(t_0, s)$ by the sample quantile in this window, $\hat{q}_\alpha(t,s) = \inf\left\{x : \frac{1}{p}\sum_{i=n_1}^{n_2} \mathbf{1}_{\{X_{i,n;s} \leq x\}} \geq \alpha\right\}$. As t changes from 0 to 1, $\hat{q}_\alpha(t,s)$ may not be a continuous function of t, and thus an extra smoothing step is necessary. Consider the Nadaraya-Watson ([11], [14]) kernel estimator

$$\tilde{q}_\alpha(t,s) = \frac{\sum_{i=1}^n K\Big((t - i/n)/h_n\Big)\hat{q}_\alpha(i/n, s)}{\sum_{i=1}^n K\Big((t - i/n)/h_n\Big)},$$

where $h_n \to 0$ is another sequence of bandwidths such that $nh_n \to \infty$, and the kernel K is a nonnegative probability density function. Examples of nearly stationary processes, asymptotic properties of $\tilde{q}$, and a data-driven scheme for the selection of smoothing parameters are given in [6]. The smooth physical evolution of many real life processes makes nearly stationarity a natural working assumption. For details on kernel smoothing we refer to [12]. The interpolation weights λ_j in equation (2), $1 \leq j \leq m$ 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 $\mathrm{cov}\big((\tilde{q}_\alpha(t,s), \tilde{q}_\alpha(t,s)\big) = C_\alpha(\theta_t, h)$, where $h = ||s - s'||$ is the Euclidean distance between s, s' and the parametric 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 ([13]), and $\hat{q}_{2,\alpha}(t,s)$ is then a best linear unbiased predictor. A similar sequential two-step approach was used in [3] for modeling space-time threshold exceedance probabilities.

For **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. We introduce a direct (one-step) procedure based on moving space-time blocks of sample quantiles. The previous two approaches did not assume space-time interaction. By reindexing the observations as $u_l = (s_l; t_k), l = 1, \ldots, N = nm$, letting B be a $(d+1) \times (d+1)$ bandwidth matrix, K a $(d+1)$ - spherically symmetric kernel, the space-time-varying probability distribution function can be estimated by $\check{F}_{s,t}(x) = \sum_{l=1}^N |B|^{-1/2} K\left(B^{-1/2}(u_l - u)\right) 1_{\{X(u_l) \leq x\}}$ yielding, by inversion

$$\hat{q}_{3,\alpha}(t,s) = \inf\{x : \check{F}(t,s)(x) \geq \alpha\}. \tag{2.3}$$

Next section presents results of an extensive comparison study where we use Monte Carlo simulations to analyze the behavior of the three space-time-varying quantile surfaces given by equations (2.1), (2.2) and (2.3), respectively.

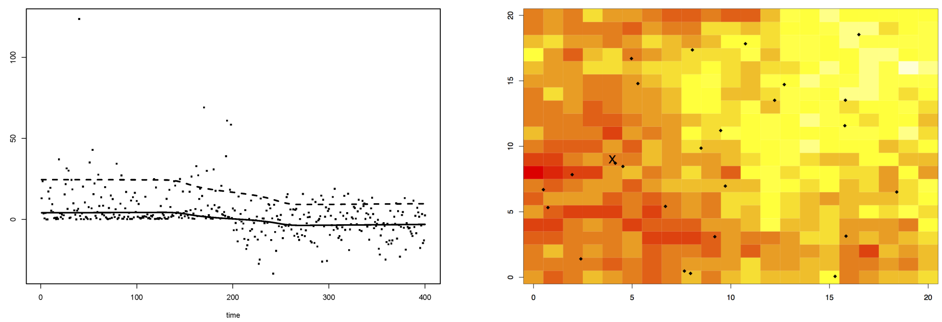

Figure 1: One realization of a non-stationary non-Gaussian temporal process with non-differentiable quantile curves of length $n = 400$, median: solid line, 0.9 quantile: dashed line (left); one realization of an isotropic Gaussian spatial process on a 20×20 spatial grid with 25 superimposed random locations (right).

	$\alpha = 0.5$			$\alpha = 0.9$		
n	100	400	1000	100	400	1000
$\hat{q}_{1,\alpha}$ (m=25)	1723	1215	1102	1912	2031	1869
$\hat{q}_{1,\alpha}$ (m=400)	1503	1428	1221	1822	1679	1555
$\hat{q}_{2,\alpha}$ (m=25)	1759	1671	1607	1827	1853	1715
$\hat{q}_{2,\alpha}$ (m=400)	1496	1388	1212	1661	1597	1483
$\hat{q}_{3,\alpha}$ (m=400)	887	765	511	981	763	677

Table 1: Integrated (over time) mean squared errors (multiplied by 10^3 and rounded) of the quantile estimates at location marked "X" in Figure 1 based on the three methods described in Section 2.

3 Simulations results

In order to compare and contrast the three space-time-varying quantile surfaces described above we simulated spatially correlated time series of length $n = 100, 400, 1000$, respectively assumed to be realizations of a non-stationary, non-Gaussian process with non-differentiable quantile functions. For an illustration see left panel in Figure 1. We assumed that the quantile curves associated with these processes are spatially correlated through an isotropic spatial covariance. The right panel in Figure 2 shows a realization of a Gaussian spatial random field with isotropic exponential covariance. All computations were done in R using the package `RandomFields`.

Table 1 presents mean squared errors $E(\hat{q}_{i,\alpha}(t,s) - q_{i,\alpha}(t,s))^2$, $i = 1,2,3$ associated with the three quantile estimates described in the previous section integrated over time, multiplied by 1000 and rounded. The spatial location s is fixed, it is the location with no data marked "X" in the right panel in Figure 1. It can be seen that the sequential methods yield similar results ($\hat{q}_1, \hat{q}_2$, respectively). It is known that kriging spatial predictors may overestimate uncertainty, however in our study these differences don't seem too large and the computational time is comparable. Remark that higher order quantiles ($\alpha = 0.9$) yield higher errors. This may be attributed to the fact that less data is used to compute these estimates. Note also that, as expected, increase in spatial sample size decreases the errors. The third, one-step method yields lower errors with a significant decrease in computational time. This method may be preferred as an initial exploratory tool without having to make too strong assumptions on the underlying process.

4 Conclusions

The increasing availability of high speed, inexpensive computing capabilities leads naturally to a growing demand for developing new flexible and informative statistical tools (such as summaries and graphs) for the exploration of large data sets with complex structures. The statistical methodology described in this paper is aimed to provide flexible, fast, accurate, and informative exploratory tools to be used for describing various distributional characteristics, such as the center (medians), extremes (high or low quantiles), and spread (interquartile ranges) for space-time data with complicated dependencies. Moreover, this methodology has a wide area of applicability in many fields, such as atmospheric sciences, demography, ecology, engineering, finance, medicine, psychology, public health.

References

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

[2] G. Biau and B. Cadre, *Nonparametric spatial prediction*, Statistical Inference for Stochastic Processes **7** (2004), 327–349.

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

[4] D. Draghicescu and S. Ghosh, *Smooth nonparametric quantiles*, Proceedings, MENP-2, Geometry Balkan Press 2003, 45–52.

[5] D. Draghicescu, *Smooth space-time quantile surfaces*, International Journal of Ecological Economics Statistics **8** (2007), 44–53.

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

[7] H. Drees, *Extreme quantile estimation for dependent data, with applications to finance,* Bernoulli, **9**(1), (2003), 617–657.

[8] F. Ferraty, A. Rabhi, P. Vieu, *Conditional quantiles for functional dependent data with application to the climatic El Niño phenomenon,* Sankhya, **67**, (2005), 378–398.

[9] S. Ghosh, J. Beran and J. Innes, *Nonparametric conditional quantile estimation in the presence of long memory*, Student **2** (1997), 109–117.

[10] S. Ghosh and D. Draghicescu, *An algorithm for optimal bandwidth selection for smooth nonparametric quantile estimation*, Statistical Data Analysis Based on the L_1-Norm and related methods (2002), 161–168.

[11] E. Nadaraya, *On estimating regression*, Theory of Probability and Its Applications, **9** (1964), 141–142.

[12] J. S. Simonoff, *Smoothing Methods in Statistics*, ger. (1998) ??? (traducere in engleza, cu adaugirea "(in German)" dupa titlu).

[13] M. L. Stein, *Interpolation of Spatial Data*, Theory of Kriging, Springer 1999.

[14] G. Watson, *Smooth regression analysis*, Sankhya, Series A, **26** (1964), 123–132.

[15] W. B. Wu and J. Mielniczuk, *Kernel density estimation for linear processes*, The Annals of Statistics **30** (2002), 1441–1459.

Authors' addresses:

Dana Sylvan
Department of Mathematics,
California State University Dominguez Hills,
1000 E Victoria St, Carson CA 90747, USA.
E-mail: dsylvan@csudh.edu

Constantin Tarcolea and Adrian Stere Paris
Faculty of Applied Sciences,
University Politehnica of Bucharest,
Splaiul Independentei 313, 060042 Bucharest, Romania.
E-mail: constantin_tarcolea@yahoo.com, sparis@rdslink.com