NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

02-20

Great Expectations and even Greater Exceedances from Spatially Referenced Data

Noel Cressie and Thomas Suesse

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone +61 2 4221 5076, Fax +61 2 4221 4998.
Email: karink@uow.edu.au

Great expectations and even greater exceedances from spatially referenced data

Noel Cressie[a,*], Thomas Suesse[a,**]

[a]*National Institute for Applied Statistics Research Australia, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522 AUSTRALIA*

Abstract

We clear land for agricultural purposes, we draw water from streams and aquifers, and we build houses in coastal regions for their ready access to the sea. Our need for food, water, and shelter is basic, but so is variability in our natural environment. Understanding this is key to the long-term sustainability of the anthropocene. At any location in the environment, long-term temporal averages may indicate regions that exceed sustainability thresholds. For example, environmental thresholds might be used by insurance companies to set the prices of insurance premiums. In this article, we present the problem of spatial-statistical inference on exceedance regions, defined as a set of locations whose long-term environmental condition is above a given threshold. An example is given of rainfall in Paraná, Brazil, averaged over a period of three decades.

Keywords: Paraná-rainfall data; kriging; loss function; predicted exceedance sets; spatial hypothesis tests

*Corresponding author: Noel Cressie Email: ncressie@uow.edu.au

**Email: tsuesse@uow.edu.au

1. Spatial data and latent exceedance regions: How high, how low, how extensive, and how often?

In the environmental sciences, these questions about exceedances are fundamental. In this article, we concentrate on answering the first three questions for spatial processes that result from long-term temporal averaging. We model a generic spatial process as the stochastic process

$$\{Y(\mathbf{s}) : \mathbf{s} \in D\}, \tag{1}$$

where $D \subset \mathbb{R}^d$ is a subset of the d-dimensional Euclidean space with Lebesgue measure $|D| > 0$, $\mathbf{s}$ is a d-dimensional vector in $\mathbb{R}^d$, and $Y(\mathbf{s})$ is a random variable. Assume that $\sup\{\text{var}(Y(\mathbf{s})) : \mathbf{s} \in D\} < \infty$; then for $\mathbf{s} \in D$, the mean of $Y(\mathbf{s})$ is well defined and denoted by $\mu_Y(\mathbf{s}) \equiv E(Y(\mathbf{s}))$; and for $\mathbf{s}, \mathbf{v} \in D$, the covariance of $Y(\mathbf{s})$ and $Y(\mathbf{v})$ is well defined and denoted by $C_Y(\mathbf{s}, \mathbf{v}) \equiv \text{cov}(Y(\mathbf{s}), Y(\mathbf{v}))$.

Note that (1), sometimes written as $Y(\cdot)$, is determined by all its finite-dimensional distributions, namely the joint distributions of $\{Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_a) : \mathbf{s}_1, \ldots, \mathbf{s}_a \in \mathbb{R}^d \text{ and } a = 1, 2, \ldots\}$. The ergodic property (e.g., Cressie, 1993, pp. 53-58) equates empirical probabilities over space or time with theoretical probabilities of a random variable or a random vector. That is, knowing the set of finite-dimensional distributions of (1) addresses the question, "How extensive?" Spatial statistics has overwhelmingly concentrated on prediction

of $Y(\cdot)$ from data,

$$\mathbf{Z}(\mathcal{s}^o) \equiv (Z(\mathbf{s}_1^o), \ldots, Z(\mathbf{s}_n^o))^\top,$$

at known locations $\mathcal{s}^o \equiv \{\mathbf{s}_1^o, \ldots, \mathbf{s}_n^o\}$ in $\mathbb{R}^d$, where $Z(\mathbf{s}_i^o)$ is a "noisy version" of $Y(\mathbf{s}_i^o)$. Specifically, conditional on $Y(\cdot)$, $Z(\mathbf{s}_i^o)$ is assumed to depend on $Y(\mathbf{s}_i^o)$ but not on the other $\{Y(\mathbf{s}_j^o) : j \neq i\}$. Jointly, we assume that

$$[\mathbf{Z}(\mathcal{s}^o)|Y(\cdot)] = \prod_{i=1}^{n} [Z(\mathbf{s}_i^o)|Y(\mathbf{s}_i^o)], \tag{2}$$

which says that the measurements are conditionally independent. In this article, $[A|B]$ denotes the conditional distribution of A given B (and $[A, B]$ and $[A]$ denote the joint and marginal distributions, respectively). Since $[A, B] = [A|B][B]$, the combination of (1) and (2) defines a joint probability measure over $Y(\cdot)$ and $\mathbf{Z}(\mathcal{s}^o)$, which henceforth we shall write simply as $Pr(\cdot)$.

Note that it is typically the case that the conditional mean of the measurement is the value of the process that it is measuring. That is,

$$Z(\mathbf{s}_i^o)|Y(\mathbf{s}_i^o) \sim Dist(Y(\mathbf{s}_i^o)), \tag{3}$$

where "$Dist(\mu)$" denotes a generic distribution, $Dist$, with mean μ, and $Dist$ may depend on other parameters such as a scale parameter, a skewness parameter, and so forth.

Let $-\infty < u(\cdot) < \infty$ denote a threshold that is a known function of spatial location in D and that can be varied by the resource manager or

decision-maker. The question, "How high and to what extent?" is answered by inferring the "greater than" random set defined for the latent process $Y(\cdot)$:

$$G_Y \equiv \{\mathbf{s} \in D : Y(\mathbf{s}) > u(\mathbf{s})\} \tag{4}$$

and, for any locations $\mathbf{s}_1, \ldots, \mathbf{s}_a$ in D and $a \in \{1, 2, \ldots\}$, by estimating

$$Pr(Y(\mathbf{s}_1) > u(\mathbf{s}_1), \ldots, Y(\mathbf{s}_a) > u(\mathbf{s}_a)). \tag{5}$$

Note that in (4), we could have emphasized dependence of G_Y on the given $u(\cdot)$, but for ease of notation, we have chosen not to. A special case of $u(\cdot)$ is a constant threshold that does not depend on spatial location. Also note that in (5), attention is often given to the cases $a = 1$ and $a = 2$. An area of research known as multi-point geostatistics considers values of $a > 2$ (e.g., Daly and Caers, 2010).

The question, "How low and to what extent?" is correspondingly answered by inferring the "less than" random set,

$$L_Y \equiv \{\mathbf{s} \in D : Y(\mathbf{s}) < u(\mathbf{s})\}, \tag{6}$$

and by estimating

$$Pr(Y(\mathbf{s}_1) < u(\mathbf{s}_1), \ldots, Y(\mathbf{s}_a) < u(\mathbf{s}_a)). \tag{7}$$

Note that in (4) and (6), using $u(\cdot)$ as a threshold for both G_Y and L_Y is just a notational convenience; when the situation demands it, a distinction should be made between a "greater than" threshold and a "less than" threshold.

In what follows, we shall assume that the finite-dimensional distributions of $Y(\cdot)$ are absolutely continuous with respect to Lebesgue measure, and hence the strict inequalities in (4) and (6) could be any combination of strict and non-strict inequalities. This will hold if $Y(\cdot)$ has continuous sample paths (e.g., $Y(\cdot)$ in (1) is a Gaussian process on $D \subset \mathbb{R}^d$).

In the rest of this article (unless stated otherwise), we make the assumption that $Y(\cdot)$ is a Gaussian process on $D \subset \mathbb{R}^d$ with mean function $\mu_Y(\cdot) \equiv \{\mu_Y(\mathbf{s}) : \mathbf{s} \in D\}$ and covariance function $C_Y(\cdot,\cdot) \equiv \{C_Y(\mathbf{s},\mathbf{v}) : \mathbf{s},\mathbf{v} \in \mathbb{R}^d\}$. The long-term temporal averaging that addresses the sustainability question means that this assumption is reasonable. It remains so if shorter-term averaging is enough to remove high-frequency temporal variation. There are also instances when $Y(\cdot)$ is a latent Gaussian process after using a link function in a generalised linear model of the measurement process (Diggle et al., 1998).

A Gaussian model would not be appropriate for determining, for example, the potential for flash-flooding after a high-rainfall event. In that case, a max-stable spatial process (Reich and Shaby, 2012; Bacro et al., 2016; Stephenson et al., 2016), or the spatial analogue to the generalized-extreme-value distributions (Huser and Wadsworth, 2019), or other non-Gaussian distributions (e.g., Naveau et al., 2009; Cooley et al., 2010, 2012) would be better models. There is a very comprehensive review of software for the statistical modelling

of extreme events in Gilleland et al. (2013) over a wide range of problems, although inference on extreme events is different to the problem of inference on exceedance regions.

Measurement of the underlying process $Y(\cdot)$ is modelled here through additive Gaussian errors. The measurement is:

$$Z(\mathbf{s}_i^o) = Y(\mathbf{s}_i^o) + \varepsilon(\mathbf{s}_i^o);\ i = 1, \ldots, n, \tag{8}$$

where $\varepsilon(\cdot)$ is a Gaussian white-noise process with mean zero and variance σ_ε^2 (i.e., the finite-dimensional distributions of $\varepsilon(\cdot)$ are defined by independent and identically distributed random variables distributed as $Gau(0, \sigma_\varepsilon^2)$). Note that (8) can be equivalently expressed as follows: $[Z(\mathbf{s}_i^o)|Y(\mathbf{s}_i^o)]$ is the density of a $Gau(Y(\mathbf{s}_i^o), \sigma_\varepsilon^2)$ random variable, independently for $i = 1, \ldots, n$; see (3).

Consider a data set of average rainfall in the May-June period in the state of Paraná, Brazil, where the average is taken over approximately three decades (Diggle and Ribeiro Jr, 2002). After a spatial statistical model is fitted to the data, the kriging predictor $Y_{kr}(\mathbf{s})$, for $\mathbf{s}$ in $D \equiv$ Paraná, can be obtained; see (10) below. One quick and easy way to infer G_Y given by (4) is the "plug-in" predictor

$$G_{Y,pl} \equiv \{\mathbf{s} : Y_{kr}(\mathbf{s}) > u\} \equiv G_{Y,kr}, \tag{9}$$

where it is convenient on occasions to use both $G_{Y,pl}$ and $G_{Y,kr}$ to denote the same exceedance-set predictor. For the Paraná-rainfall data set, the

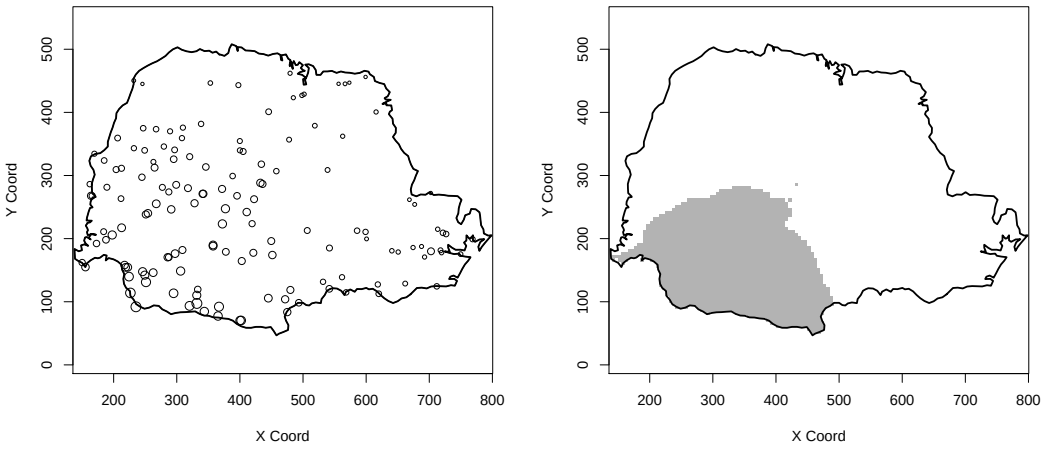
predicted exceedance set (9) with $u = 300$mm is shown in Figure 1 along with the data. In this article, we show how inference on G_Y needs special attention beyond the predictor given by (9).

The organization of the paper is as follows. Section 2 reviews in some detail the approaches that have been taken to infer exceedance regions, as befitting this special issue. In Section 3, new results emphasize spatial predictors that are not the posterior mean. French and Sain (2013) define outer and inner exceedance sets that we adapt in Section 3 to these new exceedance-set predictors. Section 4 compares the usual approach that "plugs in" the posterior mean ("great expectations") to others that recognize the special nature of inference on G_Y ("even greater exceedances"), for the Paraná-rainfall data. Section 5 gives a discussion and conclusions, and an Appendix contains a small simulation study that compares the validity and efficiency of the approaches presented in Section 3.

2. Inference for exceedance sets

In this section, we review research related to the problem of predicting the greater exceedance set G_Y that depends on a given threshold u (which may or not be a function of spatial location) from spatial data $\mathbf{Z}(\mathcal{s}^o)$, where recall that the process $Y(\cdot)$ and the data $\mathbf{Z}(\mathcal{s}^o)$ are assumed jointly Gaussian, and the spatial mean function μ_Y and the spatial covariance function C_Y are the "parameters" that summarize the spatial dependence in the joint probability $Pr(\cdot)$. Clearly, spatial dependence in the data set and the underlying process

Figure 1: Average rainfall in the state of Paraná, Brazil. Left panel: State of Paraná with weather stations shown by circles, where the average amount of rainfall in mm is indicated by the size of the circles. Right panel: Predicted greater exceedance set $G_{Y,pl} = G_{Y,kr}$ given by (9).



should be incorporated into a solution to the problem,

In what follows, we assume that μ_Y and C_Y are known; they usually need to be estimated (and they will be in the example given in Section 4). The (simple) kriging predictor, $Y_{kr}(\cdot)$, presented in Section 1 is given by (e.g., Cressie, 1993, Ch. 3),

$$Y_{kr}(\mathbf{s}) \equiv E(Y(\mathbf{s})|\mathbf{Z}(\mathcal{s}^o)) = \mu_Y(\mathbf{s}) + \mathbf{c}_Y(\mathbf{s})^\top \boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z}(\mathcal{s}^o) - \boldsymbol{\mu}_Z). \qquad (10)$$

In (10), $\mathbf{c}_Y(\mathbf{s}) \equiv (C_Y(\mathbf{s}, \mathbf{s}_1^o), \ldots, C_Y(\mathbf{s}, \mathbf{s}_n^o))^\top$ and $\boldsymbol{\mu}_Z \equiv E(\mathbf{Z}(\mathcal{s}^o)) = (\mu_Y(\mathbf{s}_1^o), \ldots, \mu_Y(\mathbf{s}_n^o))^\top$, and note that (10) is the best linear predictor of $Y(\mathbf{s})$ in that it minimizes the mean squared prediction error (MSPE). That minimized MSPE is the

(simple) kriging variance and is given by (e.g., Cressie, 1993, Ch. 3),

$$\sigma_{kr}(\mathbf{s})^2 = C_Y(\mathbf{s},\mathbf{s}) - \mathbf{c}_Y(\mathbf{s})^\top \boldsymbol{\Sigma}_Z^{-1} \mathbf{c}_Y(\mathbf{s}). \tag{11}$$

An intuitive but ad hoc predictor of the exceedance set G_Y is given by (9) in Section 1. That is, anywhere $Y(\mathbf{s})$ appears in (4), $Y_{kr}(\mathbf{s})$ is "plugged in." Once the spatial summaries μ_Y and C_Y are known, and provided there are no computational issues in obtaining the inverse of $\boldsymbol{\Sigma}_Z$ in (10), computation of (9) and (10) is straightforward. However, kriging gives a surface $Y_{kr}(\cdot)$ that is statistically "smoother than" the original process $Y(\cdot)$, which one can see by observing that

$$\text{var}(Y(\mathbf{s})) = \text{var}(E[Y(\mathbf{s})|\mathbf{Z}(\mathcal{A}^o)]) + E(\text{var}[Y(\mathbf{s})|\mathbf{Z}(\mathcal{A}^o)]) \geq \text{var}(Y_{kr}(\mathbf{s})).$$

Further discussion of the smoothness properties of kriging can be found in Zammit-Mangion et al. (2018).

Consequently, for thresholds u that are greater than the mean function, we expect that the predictor, $G_{Y,pl}$, will be "smaller than" the predictand, G_Y. That is, while $Y(\mathbf{s})$ and $Y_{kr}(\mathbf{s})$ have the same mean $\mu_Y(\mathbf{s})$, the smaller variance of $Y_{kr}(\mathbf{s})$ will lead to having fewer locations for which $Y_{kr}(\cdot)$ is above a large threshold than the number of locations for which $Y(\cdot)$ is above a large threshold (Zhang et al., 2008).

Kriging has a number of variants, many of them devoted to fast computation of the matrix-inverse component $\boldsymbol{\Sigma}_Z^{-1}$, for example reduced-rank

methods (e.g., Cressie and Johannesson, 2006, 2008; Eidsvik et al., 2012; Banerjee et al., 2008), spatial precision-matrix methods (e.g., Lopés et al., 2011; Lindgren et al., 2011; Nychka et al., 2015), and nearest "neighbour" approximations (e.g., Tadić et al., 2015; Datta et al., 2016). Some of these are compared using a common notation in Bradley et al. (2016), where their statistical and computational efficiencies are evaluated on data sets of different sizes. There is one version of kriging that attempts to "roughen up" the pointwise minimum MSPE predictor, $Y_{kr}(\cdot)$, in order to make statistical properties of the predicted surface more like those of the target surface $Y(\cdot)$. This approach, called covariance matching constrained kriging (CMCK), minimizes the MSPE but puts constraints on the spatial predictor so that its variances and some of its covariances match those of $Y(\cdot)$ (Aldworth and Cressie, 2003). The resulting predictor, $Y_{cm}(\cdot)$, was applied to the problem of inference on $g(Y(\cdot))$, where $g(y)$ is a real, non-linear function (e.g., $g(y) = \exp(y)$). While retaining desirable MSPE properties, the predictor $g(Y_{cm}(\mathbf{s}))$ is approximately unbiased for $g(Y(\mathbf{s}))$, which Aldworth and Cressie (2003) established using the delta method. They also applied it to our case of $g(y) = I(y > u)$, where u is known; however, since CMCK was developed for situations where $g(y)$ is a smooth function of y, it did not perform very well when g was an indicator function.

Rather than looking for a spatial predictor that does universally well for any $g(Y(\cdot))$, a contrasting strategy is to recognise that each functional $g(\cdot)$ needs to be considered on a case-by-case basis. Suppose we want to predict

$g(Y(\mathbf{s}))$ with a generic predictor $\delta(\mathbf{Z}(\mathscr{s}^o))$, and we use a loss function,

$$L(g(Y(\mathbf{s})), \delta(\mathbf{Z}(\mathscr{s}^o))), \tag{12}$$

to guide the prediction. Then the optimal predictor, δ^*, minimizes the posterior expected loss (Cressie, 1993, p.108); that is,

$$\delta^*(\mathbf{Z}(\mathscr{s}^o); \mathbf{s}) = \underset{\delta(\cdot)}{\arg\inf}\, E\left[L(g(Y(\mathbf{s})), \delta(\cdot)|\mathbf{Z}(\mathscr{s}^o)\right], \tag{13}$$

where $\delta(\cdot)$ is a function of the data $\mathbf{Z}(\mathscr{s}^o)$. In the case of squared-error loss, namely $(g(Y(\mathbf{s})) - \delta(\mathbf{Z}(\mathscr{s}^o)))^2$, it is straightforward to see that

$$\delta^*(\mathbf{Z}(\mathscr{s}^o); \mathbf{s}) = E(g(Y(\mathbf{s}))|\mathbf{Z}(\mathscr{s}^o)), \tag{14}$$

which is the conditional (on the data) mean of $g(Y(\mathbf{s}))$. For example, putting $g(y) = I(y > u)$, one obtains

$$\delta^*(\mathbf{Z}(\mathscr{s}^o); \mathbf{s}) = Pr(Y(\mathbf{s}) > u|\mathbf{Z}(\mathscr{s}^o)).$$

Note that the optimal spatial predictor (14) adapts to the functional $g(\cdot)$. However, squared-error loss may not be appropriate for the problem under investigation, particularly where exceedances are involved. For example, if a temperature threshold $u(\cdot)$ indicates when crops are stressed resulting in a lower yield, then a far greater loss is typically felt if the prediction were

wrong when $g(Y(\mathbf{s})) \equiv I(Y(\mathbf{s}) > u(\mathbf{s})) = 1$ than if the prediction were wrong when $g(Y(\mathbf{s})) \equiv I(Y(\mathbf{s}) > u(\mathbf{s})) = 0$.

From the general result (13), we see that δ^* is adaptable to both $g(\cdot)$ and the loss function L given in (12). Motivated by (13) and the opportunity to choose a loss function tailored for the specific purpose of predicting $g(Y(\mathbf{s}))$, a number of authors have treated the problem of predicting exceedances as one of defining an appropriate (usually asymmetric) loss function; see Stern and Cressie (1999), Wright et al. (2003), Craigmile et al. (2006), and Zhang et al. (2008). Then the problem becomes one of (analytic or numerical) optimization as specified in (13).

In one of these articles (Wright et al., 2003), a weighted ranks squared error loss function for spatial lattice data was proposed. When adapted to spatial prediction for geostatistical data, it is possible to bypass the generalized ranks that were defined by Craigmile et al. (2006) and define weights directly on the predictor. Define the weighted squared error loss as:

$$L_{wt}(Y, \delta) = w(Y)(Y - \delta)^2, \tag{15}$$

where $w(\cdot) \geq 0$, and Y is a generic random variable such that $Pr(w(Y) > 0) > 0$. Then from (13) and (15), the optimal spatial predictor of $Y(\mathbf{s})$ is easily seen to be

$$\delta^*_{wt}(\mathbf{Z}(\mathcal{s}^o); \mathbf{s}) = \frac{E(w(Y(\mathbf{s}))Y(\mathbf{s})|\mathbf{Z}(\mathcal{s}^o))}{E(w(Y(\mathbf{s}))|\mathbf{Z}(\mathcal{s}^o))}. \tag{16}$$

Another way to look at this result is that the posterior distribution, $[Y(\mathbf{s})|\mathbf{Z}(\mathcal{I}^o)]$, has been modified to:

$$f_{wt}(Y(\mathbf{s})|\mathbf{Z}(\mathcal{I}^o)) \equiv \frac{w(Y(\mathbf{s}))[Y(\mathbf{s})|\mathbf{Z}(\mathcal{I}^o)]}{\int w(y)[y|\mathbf{Z}(\mathcal{I}^o)]dy}. \tag{17}$$

Clearly, for every unweighted loss function involving $Y(\mathbf{s})$ and weight function $w(\cdot)$, there is a weighted version whose optimal spatial predictor depends on the weighted posterior distribution, $f_{wt}(\cdot|\mathbf{Z}(\mathcal{I}^o))$.

When there is no measurement error, $Z(\mathbf{s}_i^o) = Y(\mathbf{s}_i^o)$, and hence the spatial process $Y(\cdot)$ is observed directly through the data, $Y(\mathbf{s}_1^o), \ldots, Y(\mathbf{s}_n^o)$. Then Marchini and Presanis (2004) give approximations for predicting the exceedance set G_Y by applying random-field theory (e.g., Adler, 1981). However, an assumption of no measurement error is one of mathematical convenience and not something that can be defended in environmental-science applications. In a substantial advance, Bolin and Lindgren (2015) include and then account for measurement error when inferring excursions and contours.

Because the predictand G_Y can be equivalently written as $G_Y = \{\mathbf{s} : I(Y(\mathbf{s}) > u) = 1\}$, indicator kriging (IK) (e.g., Journel and Posa, 1990) could be used to infer the components of G_Y. Note that the IK methodology assumes no measurement error. It also assumes that the means and covariances of the indicator process, $\{I(Y(\mathbf{s}) > u) : \mathbf{s} \in D\}$, are stationary. These assumptions are used to derive a method-of-moments predictor, the IK predictor of $I(Y(\mathbf{s}) > u)$, for a given $\mathbf{s} \in D$. The form of the predictor

is a linear combination of $\{I(Z(\mathbf{s}_i^o) > u) : i = 1, \ldots, n\}$; however, Papritz (2009) showed that IK is probabilistically incoherent and should be avoided. We concur with this conclusion.

When computing the predictor of G_Y, the Lebesgue-measurable domain D typically has to be discretized within the computational algorithm. Replacing D with a regular grid of pixels is a common strategy, and hence a hypothesis test could be applied to test whether or not $Y(\mathbf{s})$ exceeds u at each grid node or pixel location $\mathbf{s}$. The well known multiple-hypothesis-testing problem asks what significance level should be used at each pixel location to ensure a family-wise error rate (FWER) less than or equal to α, for α small.

When the multiple hypothesis tests involve test statistics that are independent, using a different criterion, namely the false discovery rate (FDR) of Benjamini and Hochberg (1995), results in powerful signal detection. Note that here, an exceedance at $\mathbf{s} \in D$ is considered to be a "discovery" or (part of) a "signal". Now, in the spatial setting, the pixel-by-pixel hypothesis tests based on (perhaps nearby subsets of) the spatial data $\mathbf{Z}(\mathcal{J}^o)$ are statistically dependent. Shen et al. (2002) and later Huang et al. (2019) decorrelated these tests by transforming from geographical space to wavelet space, where the problem is then one of multiple hypothesis testing on wavelet coefficients. They then applied FDR methods to detect the signal in wavelet space and transformed back to geographical space to obtain the spatial signal; in doing so, careful calibration is needed to obtain a pre-specified FWER (Pavlicova

et al., 2008).

Working directly with hypothesis tests on a gridded D, French and Sain (2013), French (2014), and French and Hoeting (2016) used spatial statistical methodology to show that contours of $Y(\cdot)$ could be bounded above and bounded below, and probabilities could be specified that a predicted set contains (or is contained in) G_Y with those probabilities. It is this latter approach that we shall review and then generalize in Section 3.

3. Exceedance-set prediction using spatial hypothesis testing

In this article, inference on G_Y (and analogously on L_Y) focuses on statistics that can be used both for predicting G_Y and for providing inner and outer sets that, respectively, are contained in and contain G_Y with a pre-specified probability (e.g., 90%, 95%). This is analogous to the one-dimensional problem of finding one-sided prediction intervals for a random variable.

At location $\mathbf{s} \in D$, we formulate a hypothesis test in terms of the hidden process $Y(\cdot)$; the null hypothesis is $H_0(\mathbf{s})$, and the alternative hypothesis is $H_1(\mathbf{s})$. That is, we test

$$H_0(\mathbf{s}) : Y(\mathbf{s}) > u \;\text{ versus }\; H_1(\mathbf{s}) : Y(\mathbf{s}) \leq u, \tag{18}$$

where recall that more generally u could be a function of $\mathbf{s} \in D$ (French and Sain, 2013). A couple of slightly unusual things to note are that the null $H_0(\mathbf{s})$ (and not the alternative $H_1(\mathbf{s})$) is "the presence of signal," and

the hypotheses are formed in terms of a latent random variable and not a deterministic unknown parameter. As will be seen below, these non-standard features result in a different prediction methodology than those we reviewed in Section 2 for inferring the random set G_Y. Another feature to note is that, due to the spatial dependence, the hypotheses are statistically dependent so that in general, $Pr(Y(\mathbf{s}) > u, Y(\mathbf{v}) > u) \neq Pr(Y(\mathbf{s}) > u) \cdot Pr(Y(\mathbf{v}) > u)$.

3.1. Optimal prediction

When testing generic hypotheses, H_0 versus H_1, the choice of test statistic determines how powerful the test is. We have already noted that (18) is non-standard, and hence other considerations may be used to choose a test statistic. In our spatial setting where we are testing $H_0(\mathbf{s})$ versus $H_1(\mathbf{s})$, we could simply replace $Y(\mathbf{s})$ in (18) with its optimal kriging predictor $Y_{kr}(\mathbf{s})$, given by (10), and use it as the test statistic: Then we prefer $H_0(\mathbf{s})$ to $H_1(\mathbf{s})$ if $Y_{kr}(\mathbf{s})$ is "large" (French and Sain, 2013).

As a justification for this intuition, consider the following prediction problem at $\mathbf{s} \in D$: Predict the random variable,

$$I(Y(\mathbf{s}) > u) = \begin{cases} 1, & \text{if } Y(\mathbf{s}) > u \\ 0, & \text{if } Y(\mathbf{s}) \leq u. \end{cases}$$

Assuming a squared-error loss function and using (14), we showed in Section

2 that this binary random variable is optimally predicted by

$$
\begin{aligned}
E\left(I(Y(\mathbf{s}) > u)|\mathbf{Z}(\mathcal{s}^o)\right) &= Pr\left(Y(\mathbf{s}) > u|\mathbf{Z}(\mathcal{s}^o)\right) \\
&= Pr\left(\frac{Y(\mathbf{s}) - Y_{kr}(\mathbf{s})}{\sigma_{kr}(\mathbf{s})} > \frac{u - Y_{kr}(\mathbf{s})}{\sigma_{kr}(\mathbf{s})}\middle|\mathbf{Z}(\mathcal{s}^o)\right) \\
&= \Phi\left(\frac{Y_{kr}(\mathbf{s}) - u}{\sigma_{kr}(\mathbf{s})}\right), \qquad (19)
\end{aligned}
$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard $Gau(0,1)$ random variable. The Gaussian assumptions made earlier are important here because then, $\sigma_{kr}(\mathbf{s})^2 = \text{var}\left(Y(\mathbf{s})|\mathbf{Z}(\mathcal{s}^o)\right)$; more generally for non-Gaussian data, $\sigma_{kr}(\mathbf{s})^2 = E(\text{var}(Y(\mathbf{s})|\mathbf{Z}(\mathcal{s}^o))$ (Cressie and Wikle, 2011, p. 141). Clearly, the result (19) can be generalized to joint probabilities given by (5). For example, for $a = 2$ and the random variable $I(Y(\mathbf{s}_1) > u$ and $Y(\mathbf{s}_2) > u)$, its optimal predictor is the (conditional) bivariate probability, $\delta^*(\mathbf{Z}(\mathcal{s}^o)) \equiv Pr(Y(\mathbf{s}_1) > u, Y(\mathbf{s}_2) > u|\mathbf{Z}(\mathcal{s}^o))$, a result upon which we shall build in Section 3.3.

Since $I(Y(\mathbf{s}) > u) \in \{0, 1\}$ and its optimal predictor (19) is a probability $\in (0, 1)$, the final predictor of $I(Y(\mathbf{s}) > u)$ could be set equal 1 if (19) is greater than or equal to a cut-off such as 0.5 (and otherwise set equal to 0). Thus, the statistic (19) used for prediction of an exceedance of the hidden process $Y(\cdot)$ at $\mathbf{s}$ is a monotonic function of

$$
T_{kr}(\mathbf{s}) \equiv \frac{Y_{kr}(\mathbf{s}) - u}{\sigma_{kr}(\mathbf{s})}, \qquad (20)
$$

for all $\mathbf{s} \in D$; it is (20) that French and Sain (2013) used as a test statistic for testing $H_0(\mathbf{s})$ versus $H_1(\mathbf{s})$. That is, they chose $H_0(\mathbf{s})$ if $T_{kr}(\mathbf{s}) > K(\mathbf{s})$, where $K(\mathbf{s})$ is a threshold that depends on $\mathbf{s} \in D$; otherwise $H_1(\mathbf{s})$ was chosen. We now discuss how $K(\mathbf{s})$ might be chosen.

We noted above that this hypothesis-testing formulation is somewhat non-standard, and we see below how it goes in a different direction to the usual approach that controls $Pr(\text{type I error})$ and optimises the power $= 1 - Pr(\text{type II error})$. But first, observe that spatial dependence has been explicitly incorporated into the test statistic $T_{kr}(\mathbf{s})$ given by (20). From (10), it is straightforward to see that

$$\text{cov}(T_{kr}(\mathbf{s}), T_{kr}(\mathbf{v})) = \begin{cases} 1; & \mathbf{s} = \mathbf{v} \\ c_Y(\mathbf{s})^\top \boldsymbol{\Sigma}_Z^{-1} c_Y(\mathbf{v}) / \sigma_{kr}(\mathbf{s}) \sigma_{kr}(\mathbf{v}); & \mathbf{s} \neq \mathbf{v}, \end{cases}$$

which in general is not a function of $\mathbf{v} - \mathbf{s}$ (i.e., is not stationary). When D is replaced with a (regular or irregular) grid, the full correlation matrix for all kriged values has row and column dimension given by the number of grid nodes in D and, for spatially dependent data, the matrix is not diagonal.

Since (4) can be equivalently written as $G_Y \equiv \{\mathbf{s} : I(Y(\mathbf{s}) > u) = 1\}$, one could consider predicting the random set by performing the multiple hypothesis tests, $\{H_0(\mathbf{s}) \text{ versus } H_1(\mathbf{s}) : \mathbf{s} \in D\}$. As discussed in Section 2, multiple hypothesis testing is notoriously problematic, even more so in the presence

of spatial dependence. However, French and Sain (2013) exploited this non-standard formulation to turn one-at-a-time inference at each location $\mathbf{s} \in D$, into inference for the whole exceedance set. That inference comes in the form of outer and inner predicted exceedance sets that contain or are contained in G_Y with a given probability (e.g., 90%), as shown below.

Define a $(1-\alpha)100\%$ outer predicted exceedance set G_O^α as a set in $\mathbb{R}^d$ that satisfies

$$Pr\left(G_Y \subset G_O^\alpha | \mathbf{Z}(\mathcal{s}^o)\right) = 1 - \alpha; \tag{21}$$

and define a $(1-\alpha')100\%$ inner predicted exceedance set $G_I^{\alpha'}$ as a set in $\mathbb{R}^d$ that satisfies

$$Pr\left(G_I^{\alpha'} \subset G_Y | \mathbf{Z}(\mathcal{s}^o)\right) = 1 - \alpha'. \tag{22}$$

Note that the sets G_O^α and $G_I^{\alpha'}$ are "statistics" in the sense that they are functions of $\mathbf{Z}(\mathcal{s}^o)$, and the probabilities in (21) and (22) are obtained from the conditional distribution of $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ given the data $\mathbf{Z}(\mathcal{s}^o)$.

We now describe the methodology developed by French and Sain (2013). They based it on the test statistics $\{T_{kr}(\mathbf{s}) : \mathbf{s} \in D\}$, but it is easy to see that the approach is quite general and can be carried out in terms of generic test statistics $\{T(\mathbf{s}) : \mathbf{s} \in D\}$, which we now proceed to do.

We wish to find critical values C_O^α and $C_I^{\alpha'}$ (both functions of $\mathbf{Z}(\mathcal{s}^o)$) such

that

$$Pr\left(\inf_{\mathbf{s}\in G_Y}\{T(\mathbf{s})\} \leq C_O^{\alpha}|\mathbf{Z}(\mathcal{J}^o)\right) = \alpha \tag{23}$$

$$Pr\left(\sup_{\mathbf{s}\notin G_Y}\{T(\mathbf{s})\} \geq C_I^{\alpha'}|\mathbf{Z}(\mathcal{J}^o)\right) = \alpha'. \tag{24}$$

Note that the test statistics and the critical values in (23) and (24) depend on $\mathbf{Z}(\mathcal{J}^o)$ but not on $Y(\cdot)$; hence, the conditional (on $\mathbf{Z}(\mathcal{J}^o)$) probabilities in (23) and (24) derive from the random exceedance set G_Y. The special cases where $G_Y = \emptyset$ or $G_Y = D$ occur with probability zero under the Gaussian assumptions, although in the computational algorithm they can occur. From the large-deviation principle (e.g., Dembo and Zeitouni, 1995), we set

$$\inf_{s\in\emptyset}\{T(\mathbf{s})\} = \infty; \text{ and } \sup_{s\in\emptyset}\{T(\mathbf{s})\} = -\infty.$$

In the computations, ∞ and $-\infty$ are replaced with respectively, large positive and negative numbers.

Then

$$G_O^{\alpha} \equiv \{\mathbf{s} \in D : T(\mathbf{s}) > C_O^{\alpha}\} \tag{25}$$

$$G_I^{\alpha'} \equiv \{\mathbf{s} \in D : T(\mathbf{s}) \geq C_I^{\alpha'}\}, \tag{26}$$

are outer and inner predicted exceedance sets, respectively, and they are both functions only of the data. (Any unknown parameters of the Gaussian

distributions are estimated separately and substituted into the inference procedure as if they were known, as was done in the Paraná-rainfall application in Section 4.)

Intuitively, if $\alpha' = \alpha$, then we should see $G_I^\alpha \subset G_O^\alpha$, but this is not guaranteed. For the events in (23) and (24), there is in general a positive probability that $\inf\{T(\mathbf{s}) : \mathbf{s} \in G_Y\} > \sup\{T(\mathbf{s}) : \mathbf{s} \notin G_Y\}$. However, it is straightforward to show that as α and α' increase, the outer sets nest and get smaller, while the inner sets nest and get bigger.

3.2. Computational aspects of the outer and inner predicted exceedance sets

Henceforth in this article, suppose that the Lebesgue-measurable domain D has been discretized into a fine-resolution grid of pixels, finite in number. With a slight abuse of notation, we use the same notation D for this finite-grid representation of the original continuous index set; now D consists of the nodes of the grid. Furthermore, the grid is assumed to be regular and square, although the methodology does not require that. This discretization of D has an ignorable effect provided the grid is fine enough (French and Hoeting, 2016).

The critical values C_O^α and $C_I^{\alpha'}$ in (25) and (26), respectively, can be obtained via a type of Monte Carlo testing (Hope, 1968). Write the nodes of the grid as $D \equiv \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ and, conditional on $\mathbf{Z}(\mathcal{I}^o)$, simulate B realizations of $\{Y(\mathbf{s}_i) : i = 1, \ldots, N\}$, which is called conditional simulation in the geostatistics literature (e.g., Cressie, 1993, Ch. 3). This is straightforward to

do, because of the Gaussian assumptions made in Section 1, and we denote the realizations as:

$$\{Y^{(b)}(\mathbf{s}_i) : i = 1, \ldots, N; b = 1, \ldots, B\}.$$

To obtain C_O^α and hence the outer predicted exceedance set G_O^α, for each $b = 1, \ldots, B$ we define the set of locations $\mathbf{s}$ for which $H_0(\mathbf{s})$ is chosen: $S_O^{(b)} \equiv \{\mathbf{s}_i : Y^{(b)}(\mathbf{s}_i) \geq u; i = 1 \ldots, N\}$. Now we find a location $\mathbf{s}_O^{(b)}$ in $S_O^{(b)}$ at which $T(\cdot)$ is a minimum (and if there are several we choose one). That is, $\mathbf{s}_O^{(b)}$ satisfies: $T(\mathbf{s}_O^{(b)}) = \min\{T(\mathbf{s}) : \mathbf{s} \in S_O^{(b)}\}$. Then C_O^α is the α-th quantile of $\{T(\mathbf{s}_O^{(1)}), \ldots, T(\mathbf{s}_O^{(B)})\}$, and

$$G_O^\alpha = \{\mathbf{s}_i : T(\mathbf{s}_i) > C_O^\alpha; i = 1, \ldots, N\}. \tag{27}$$

For the inner predicted exceedance set $G_I^{\alpha'}$, we proceed in a similar manner, and for each $b = 1, \ldots, B$ we define $S_I^{(b)} \equiv \{\mathbf{s}_i : Y^{(b)}(\mathbf{s}_i) < u; i = 1 \ldots, N\}$. Now we find a location $\mathbf{s}_I^{(b)}$ at which $T(\cdot)$ is a maximum (and if there are several we choose one). That is, $\mathbf{s}_I^{(b)}$ satisfies: $T(\mathbf{s}_I^{(b)}) = \max\{T(\mathbf{s}) : \mathbf{s} \in S_I^{(b)}\}$. Then $C_I^{\alpha'}$ is the $(1-\alpha')$-th quantile of $\{T(\mathbf{s}_I^{(1)}), \ldots, T(\mathbf{s}_I^{(B)})\}$, and

$$G_I^{\alpha'} = \{\mathbf{s}_i : T(\mathbf{s}_i) \geq C_I^{\alpha'}; i = 1, \ldots, N\}. \tag{28}$$

3.3. Test statistics for inference on exceedance sets

The methodology of French and Sain (2013) for inference on exceedance sets, is based on the spatial hypothesis test (18) and the test statistic $T_{kr}(\cdot)$ given by (20). In this subsection, we choose other test statistics for which the computations of Section 3.2 can be applied. Our emphasis here is to develop test statistics $T(\cdot)$ that are better tuned than $T_{Y,kr}$ to inferences on extremes. (We have mentioned already that $\{Y_{kr}(\mathbf{s}) : \mathbf{s} \in D\}$ is more appropriate for predicting the overall surface of $Y(\cdot)$ than for predicting its extremes.)

We saw in Section 2 that the squared-error loss function leads to the predictor $Y_{kr}(\cdot)$, but we also argued that this loss function is not realistic for predicting exceedances. We also saw in (15) how to modify the squared-error loss function with weights, resulting in the optimal predictor (16). In line with more weight going to instances where $Y(\mathbf{s}) > u$, define

$$w(y) \equiv \frac{1}{1 + \exp(-k(y-u))}; k \geq 0. \tag{29}$$

Note that when $k = 0$, $w(y) = 1/2$ (equal weights) and, as $k \to \infty$, $w(y) \to I(y > u)$, which says that all the losses occur when there is an exceedance. The curve given by (29) is sigmoid in shape and values of k in $(0, \infty)$ give a smooth transition from a low-loss to a high-loss situation as y moves from being less than u to being greater than u. On the unitless probability scale, Craigmile et al. (2006) chose a value of $k = 9.15$ around a probability of 0.9. Since the threshold u will generally be larger than the mean function $\mu_Y(\mathbf{s})$,

we would expect $\Phi((u - \mu_Y(\mathbf{s}))/C_Y(\mathbf{s},\mathbf{s})^{1/2})$ to be in the upper tail of the $Gau(0,1)$ distribution. Therefore, we suggest a choice of

$$k = c \cdot \phi\left((u - \mu_Y(\mathbf{s}))/C_Y(\mathbf{s},\mathbf{s})^{1/2}\right)/C_Y(\mathbf{s},\mathbf{s})^{1/2}; \mathbf{s} \in D, \tag{30}$$

where $\phi(\cdot)$ is the density function of a standard $Gau(0,1)$ random variable, and c is a constant that specifies how the loss is "spread out" in (15) around $Y = u$. Note that in (30), k depends on $\mathbf{s}$ through the mean and variance functions; if the process $Y(\cdot)$ is modeled as second-order stationary, k would be constant over D. This illustrates a strength of using a weighted squared-error loss function, that by using weights like (29), (30), the methodology is adaptive to the location and scale of $Y(\mathbf{s})$ at each $\mathbf{s} \in D$.

From (16), the optimal predictor is the weighted posterior mean,

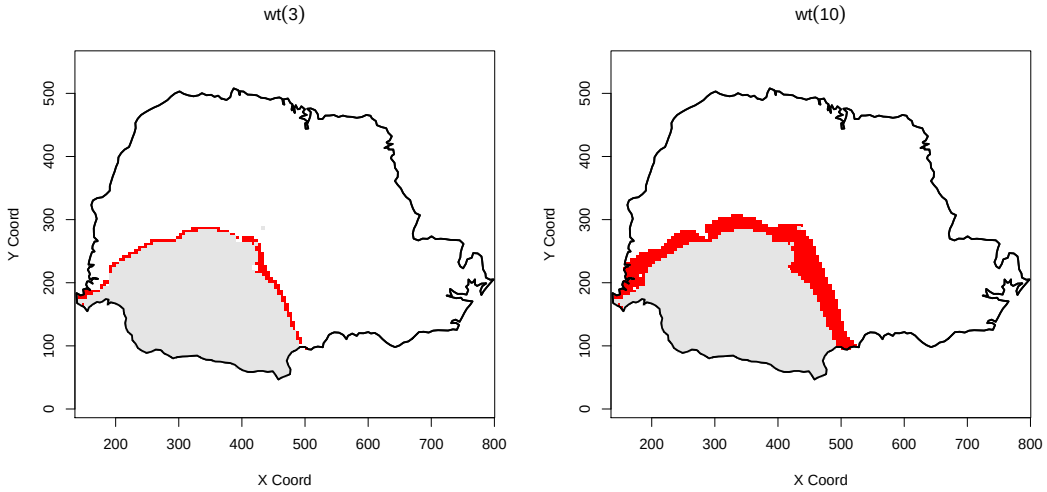
$$Y_{wt(c)}(\mathbf{s}) \equiv \frac{E(w(Y(\mathbf{s}))Y(\mathbf{s})|\mathbf{Z}(\mathcal{J}^o))}{E(w(Y(\mathbf{s}))|\mathbf{Z}(\mathcal{J}^o))} \equiv E_{wt(c)}(Y(\mathbf{s})|\mathbf{Z}(\mathcal{J}^o)), \tag{31}$$

where $w(Y(\mathbf{s}))$ is given by (29), (30). In Figure 2, for $u = 300$mm, we give two predictions of G_Y defined by

$$G_{Y,wt(c)} \equiv \{\mathbf{s} : Y_{wt(c)}(\mathbf{s}) > u\}, \text{for } c = 3 \text{ and } c = 10, \tag{32}$$

from the Paraná-rainfall data set introduced in Section 1. The choice of $c = 3$ and 10 was guided by the simulation study in the Appendix. Hence,

Figure 2: Predicted greater exceedance sets for the Paraná-rainfall data set. Shown in light gray is $G_{Y,kr}$ given by (9). The light gray and red pixels jointly show $G_{Y,wt(c)}$ defined by (32). Left panel: c=3. Right panel: $c = 10$.



we propose the test statistic,

$$T_{wt(c)}(\mathbf{s}) \equiv \frac{Y_{wt(c)}(\mathbf{s}) - u}{C_Y(\mathbf{s},\mathbf{s})^{1/2}}; \mathbf{s} \in D, \tag{33}$$

where recall that the constant c in (30) determines the spread of the loss around the threshold u.

The next test statistic is motivated by the observation that the methodology proposed in Section 3.1 is the same whether inference is on $Y(\mathbf{s})$ using squared-error loss or is on $I(Y(\mathbf{s}) > u)$ using squared-error loss. However, now we wish to incorporate more spatial dependence in the problem, which results in going from $a = 1$ to $a = 2$ in (5). Spatial covariance functions do this, but we would like to adapt this idea of joint dependence to the problem

of inference on exceedance sets.

Now, for $\mathbf{s}, \mathbf{v} \in D$,

$$I(Y(\mathbf{s}) > u \text{ and } Y(\mathbf{v}) > u) = I(Y(\mathbf{s}) > u) \cdot I(Y(\mathbf{v}) > u). \tag{34}$$

For squared-error loss, the optimal predictor is the conditional expectation of the left-hand side of (34), namely $Pr(Y(\mathbf{s}) > u, Y(\mathbf{v}) > u | \mathbf{Z}(\mathcal{S}^o))$, something we observed in Section 3.1. Define the test statistic based on these bivariate probabilities for vectors $\mathbf{v} = \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, and $\mathbf{v}_4$ displaced from $\mathbf{s}$ in the four principal directions on the square grid D:

$$T_{jt(h)} \equiv \underset{j}{\text{ave}}\{Pr(Y(\mathbf{s}) > u, Y(\mathbf{v}_j) > u | \mathbf{Z}(\mathcal{S}^o))\}, \text{ for } h = 1, 2, \ldots, \tag{35}$$

where $h = ||\mathbf{v}_j - \mathbf{s}||/(\text{length of grid cell})$ and $j = 1, 2, 3, 4$. The choice $h = 1$ results in the four nearest neighbours being chosen, and $h = 2, 5$, and 10 were also chosen in the simulation study in the Appendix, to gauge what "nearby" means in this context of predicting exceedance regions. Note that (35) could be generalized to include all nearby grid cells rather than just those in a "ring" around $\mathbf{s}$.

The probabilities in (35) are based on bivariate orthant probabilities of a standard bivariate Gaussian random vector (W_1, W_2) with joint probability

P, with means 0, variances 1, and correlation ρ. Then,

$$\begin{aligned} Pr(Y(\mathbf{s}) > u, Y(\mathbf{v}) > u | \mathbf{Z}(\mathcal{J}^o)) &= P\left(W_1 < \frac{u - Y_{kr}(\mathbf{s})}{\sigma_{kr}(\mathbf{s})}, W_2 < \frac{u - Y_{kr}(\mathbf{v})}{\sigma_{kr}(\mathbf{s})}\right) \\ &= P(W_1 < T_{kr}(\mathbf{s}), W_2 < T_{kr}(\mathbf{v})) \\ &\equiv \Phi_\rho\left(T_{kr}(\mathbf{s}), T_{kr}(\mathbf{v})\right), \end{aligned} \tag{36}$$

where $\Phi_\rho(\cdot, \cdot)$ is the bivariate cumulative distribution function of standard Gaussian random variables with correlation ρ. In this setting,

$$\rho = \frac{\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{v}) | \mathbf{Z}(\mathcal{J}^o))}{\mathrm{var}(Y(\mathbf{s}) | \mathbf{Z}(\mathcal{J}^o))^{1/2} \mathrm{var}(Y(\mathbf{v}) | \mathbf{Z}(\mathcal{J}^o))^{1/2}}. \tag{37}$$

To evaluate (37), we need the conditional covariance function of $Y(\cdot)$ given $\mathbf{Z}(\mathcal{J}^o)$. Now, $C(\mathbf{u}, \mathbf{v}) = E(\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{v}) | \mathbf{Z}(\mathcal{J}^o))) + \mathrm{cov}(Y_{kr}(\mathbf{s}), Y_{kr}(\mathbf{v}))$ and hence, because of Gaussianity,

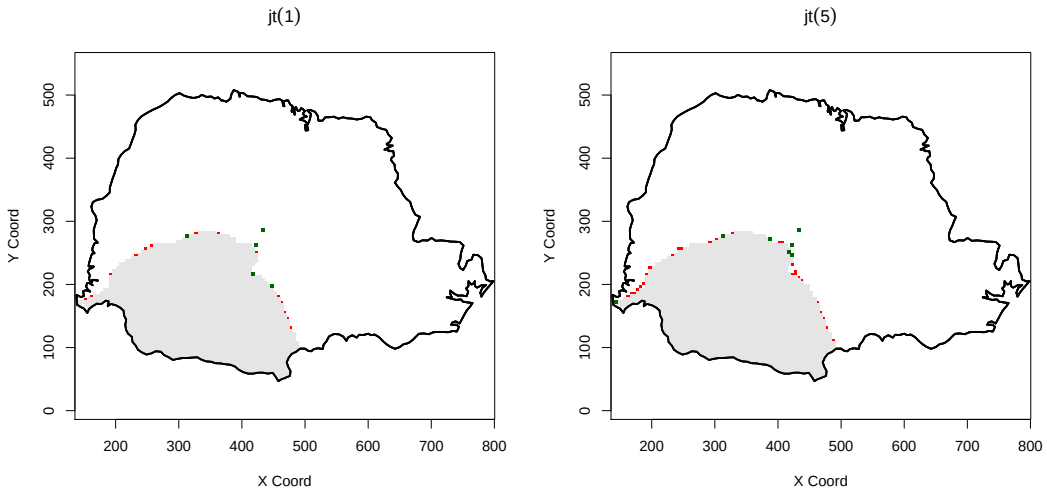
$$\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{v}) | \mathbf{Z}(\mathcal{J}^o)) = C_Y(\mathbf{s}, \mathbf{v}) - \mathbf{c}_Y(\mathbf{s})^\top \mathbf{\Sigma}_Z^{-1} \mathbf{c}_Y(\mathbf{v}). \tag{38}$$

Consequently, $T_{jt(h)}(\mathbf{s})$ given by (35) is:

$$T_{jt(h)}(\mathbf{s}) = \underset{j}{\mathrm{ave}} \left\{ \Phi_\rho(T_{kr}(\mathbf{s}), T_{kr}(\mathbf{v}_j) \right\}, \tag{39}$$

where ρ is given by (37), (38).

Figure 3: Predicted greater exceedance sets for the Paraná-rainfall data set. Shown in light gray is $G_{Y,kr}$ given by (9). The red pixels show the subset of $G_{Y,jt(h)}$, defined by (40), that is not contained in $G_{Y,kr}$, and the dark-green pixels show the subset of $G_{Y,kr}$ that is not contained in $G_{Y,jt(h)}$. Left panel: $p(1) = 0.30$. Right panel: $p(5) = 0.25$.



In Figure 3, we show two predictions of G_Y given by

$$G_{Y,jt(h)} \equiv \{\mathbf{s} : T_{jt(h)} > p(h)\}, \text{ for } h = 1 \text{ and } 5, \tag{40}$$

for the Paraná-rainfall data set introduced in Section 1, where $p(1) = 0.3$ and $p(5) = 0.25$, reflecting the weaker dependence that will occur at a larger lag. These joint probability thresholds were obtained by writing a bivariate joint probability as a product of a marginal and a conditional probability. The marginal probability threshold is 0.5, and the fitted correlation function implied a conditional-probability threshold of 0.6 when $h = 1$ and 0.5 when $h = 5$. Hence the joint-probability thresholds, $p(h)$, are adaptive to the spatial-dependence present in the data.

In this subsection, we have generalized the test statistic $T_{kr}(\cdot)$ of French and Sain (2013) to a weighted version, $T_{wt(c)}(\cdot)$. We have also generalized the test statistic, $\Phi(T_{kr}(\cdot))$ (which is equivalent to $T_{kr}(\cdot)$ due to the monotonic relationship between them), to a test statistic $T_{jt(h)}(\cdot)$ obtained by averaging the bivariate probabilities, $\Phi_\rho(T_{kr}(\cdot), T_{kr}(\cdot))$. The generic methodology for any test statistic $T(\cdot)$ presented in Section 3.1, shows that these will result in, respectively, outer predicted exceedance sets: $G^{\alpha}_{O,kr}$, $G^{\alpha}_{O,wt(c)}$, $G^{\alpha}_{O,jt(h)}$; and inner predicted exceedance sets: $G^{\alpha'}_{I,kr}$, $G^{\alpha'}_{I,wt(c)}$, $G^{\alpha'}_{I,jt(h)}$. The plug-in predictor (9) of G_Y, denoted as $G_{Y,pl}$ in this context, can be adapted to yield outer and inner predicted exceedance sets as follows:

$$G^{\alpha}_{O,pl} = \{\mathbf{s} : Y_{kr}(\mathbf{s}) + \Phi^{-1}(1-\alpha)\sigma_{kr}(\mathbf{s}) > u\} \tag{41}$$
$$G^{\alpha'}_{I,pl} = \{\mathbf{s} : Y_{kr}(\mathbf{s}) - \Phi^{-1}(1-\alpha')\sigma_{kr}(\mathbf{s}) > u\}.$$

Comparisons of these four (i.e., kr, $wt(c)$, $jt(h)$, and pl) outer and inner sets are given in a simulation study in the Appendix and, we have already seen earlier in this section, applications of the different test statistics to spatial data of average May-June rainfall over three decades in the Brazilian state of Paraná. A full spatial analysis of this data set will now be given.

4. Rainfall data from the state of Paraná in Brazil

To illustrate the methodology presented in the previous section, we focus on the spatial data set shown in Figure 1 with distances in kilometers. It

is freely available through the `R` package `geoR` (Ribeiro Jr and Diggle, 2018) and refers to 143 stations throughout the state of Paraná, Brazil, whose rainfall during the May-June (dry season) period was averaged over three decades (Diggle and Ribeiro Jr, 2002). The vector $\mathbf{x}(\mathbf{s}_i^o) \equiv (1, x_1(\mathbf{s}_i^o), x_2(\mathbf{s}_i^o))^\top$ represents the covariates, the last two of which are the spatial coordinates (in km) from an arbitrary origin shown in Figure 1. Here $Z(\mathbf{s}_i^o)$ is the averaged recorded May-June rainfall, and $\{Y(\mathbf{s}) : \mathbf{s} \in \text{Paraná}\}$ is the true rainfall field to be inferred from data $\mathbf{Z}(\mathcal{s}^o) = (Z(\mathbf{s}_1^o), \ldots, Z(\mathbf{s}_{143}^o))^\top$. As in Diggle and Ribeiro Jr (2002), we use a rainfall threshold of $u = 300$mm, and we wish to predict the exceedance set $G_Y \equiv \{\mathbf{s} \in D : Y(s) > 300\text{mm}\}$. Note that their inferences on exceedances were based on posterior probability contours at levels of 0.10, 0.50, and 0.90 for the random set G_Y.

We follow the `R` code provided by `geoR` package author Paulo Ribeiro Jr, which is freely available at `http://www.leg.ufpr.br/geoR/PASI/`, to process the data and fit a spatial model. The exponential covariance function and the spherical covariance function are two commonly used models; for lags greater than 10km, we chose the latter for its finite range of spatial dependence.

A spherical semivariogram was fitted using weighted least squares (Cressie, 1985) to the empirical semivariogram of Cressie and Hawkins (1980) at those lags: $\gamma_{sp}(h) = \tau_Z^2 + \sigma_Z^2(1 - \rho(h;\phi))$ where $\rho(h;\phi) = \{1 - 1.5(h/\phi) + 0.5(h/\phi)^3\}I(h \leq \phi)$; τ_z^2 is the intercept at $h = 0$; and $\tau_Z^2 + \sigma_Z^2$ is the "sill" (i.e., the total variance), $\text{var}(Z(\mathbf{s})) = \tau_Z^2 + \sigma_Z^2$. A separate analysis of semi-

variogram lags less than 2km, allowed us to estimate the measurement-error variance σ_ε^2 of $\varepsilon(\cdot)$ in (8).

Since an average over three decades of rainfall will likely produce quite a smooth spatial surface, it is of interest to model the smaller-scale spatial dependence at lags smaller than the first lag h_1, used to fit the spherical model. To do this, we fitted a cubic function $\{\gamma_{cu}(h) : 0 \leq h \leq h_1\}$, such that $\gamma_{cu}(0) = \sigma_\varepsilon^2$; and at $h = h_1$, we matched the values $\gamma_{cu}(h_1) = \gamma_{sp}(h_1)$ as well as the first and second derivatives. Thus, the fitted variogram is:

$$\hat{\gamma}_Z(h) = \begin{cases} 0; & h = 0 \\ \gamma_{cu}(h); & 0 < h \leq h_1 \\ \gamma_{sp}(h); & h > h_1, \end{cases}$$

where the parameters were estimated as $\hat{\sigma}_\varepsilon^2 = 10.07(\text{mm})^2$, $\hat{\sigma}_Z^2 = 1384.19(\text{mm})^2$, $\hat{\tau}_Z^2 = 378.34(\text{mm})^2$, $\hat{\phi} = 513.72$km.

An estimate of the covariance function of $Z(\cdot)$ is

$$\hat{C}_Z(h) = \hat{\tau}_Z^2 + \hat{\sigma}_Z^2 - \hat{\gamma}_Z(h), \tag{42}$$

and recall that C_Y satisfies:

$$C_Z(h) = C_Y(h) + \sigma_\varepsilon^2 I(h = 0).$$

Hence,

$$\hat{C}_Y(h) = \hat{C}_Z(h) - \hat{\sigma}_\varepsilon^2 I(h = 0). \tag{43}$$

A little algebra reveals that the fitted covariance function (43) is continuous at $h = 0$, namely $\lim_{h\to 0} \hat{C}_Y(h) = \hat{C}_Y(0)$; this absence of a nugget effect for $Y(\cdot)$ is expected, since our model $\gamma_{cu}(h)$ for small h, has replaced it.

Then, using (42), the following regression coefficients were estimated via generalized least squares: $\hat{\beta}_0 = 399.61$, $\hat{\beta}_1 = -0.088$ (latitude), and $\beta_2 = -0.426$ (longitude). Based on these estimates, the methodology outlined in Section 3.2 was applied using $B = 10,000$ conditional simulations.

Importantly, the square grid superimposed on the state of Paraná was calibrated to be of a resolution that matched the resolution of the grid on $[0, 1] \times [0, 1]$ in the simulation study in the Appendix. Figure 4 shows the outer and inner predicted exceedance sets based on four of the statistics compared in the Appendix; $T_{kr}(\cdot)$ given by (20), $T_{wt(3)}$ given by (33), $T_{jt(5)}$ given by (35), and the plug-in exceedance sets (41) were also included to give a baseline.

From the Appendix, the outer and inner predicted exceedance sets derived from $T_{jt(5)}$ have the required coverage and, under perturbations of the process mean $\mu_Y(\cdot)$, have the least deterioration of coverage; see Tables 1 and 2. The performance of $T_{wt(3)}$ in Figure 4 appears similar, yet $T_{jt(5)}$ does better in the simulation when taking both tables into consideration. A close inspection of Figure 3 shows that $G_{Y,jt(5)}$ contains $G_{Y,kr}$ which is a desirable feature.

Figure 4: Inner (purple) and outer (blue) predicted exceedance sets with $\alpha = \alpha' = 0.10$ for the rainfall data from the state of Paraná in Brazil. Left top panel: Inference based on 'pl' where the transparent-gray overlay shows $G_{Y,pl}$. Right top panel: Inference based on 'kr' where the transparent-gray overlay shows $G_{Y,kr}$. Left bottom panel: Inference based on 'wt(3)' where the transparent-gray overlay shows $G_{Y,wt(3)}$. Right bottom panel: Inference based on 'jt(5)', where the transparent-gray overlay shows $G_{Y,jt(5)}$.

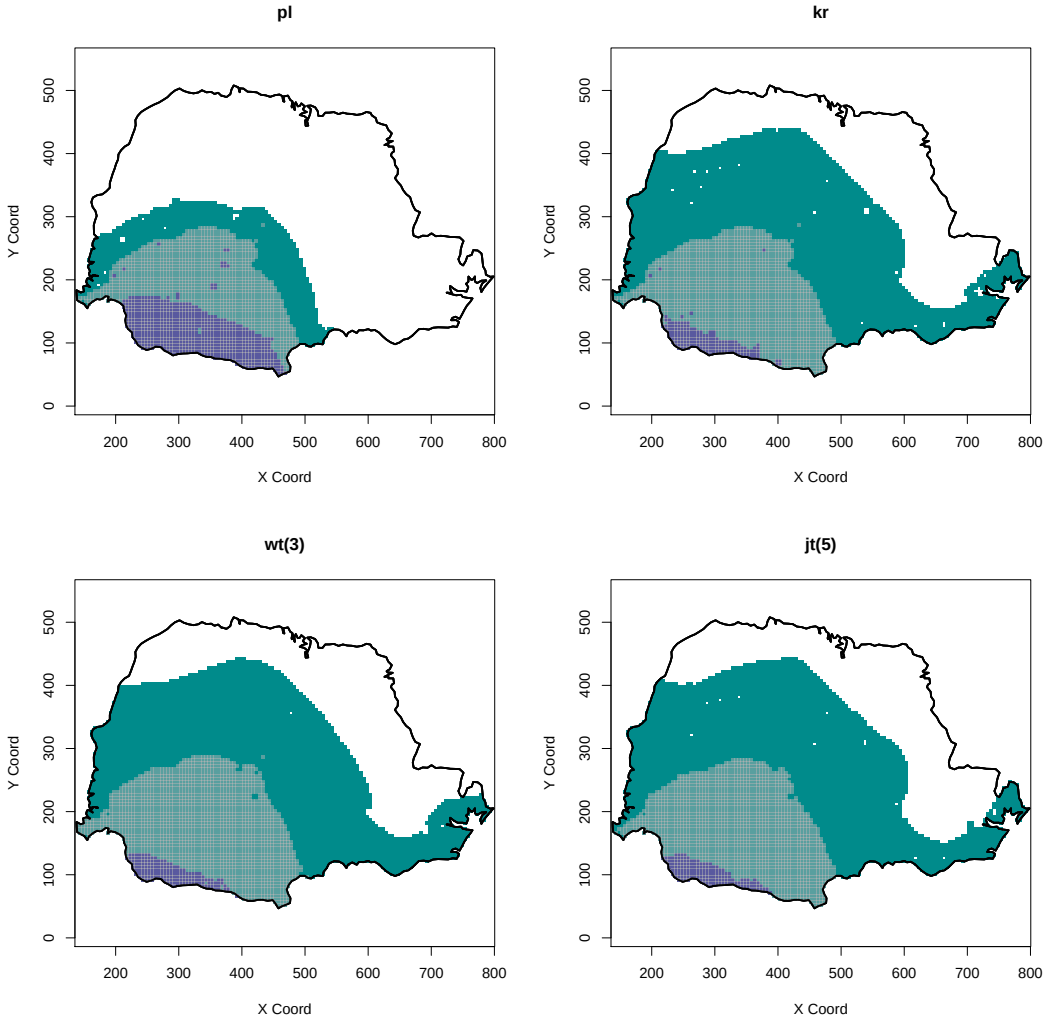
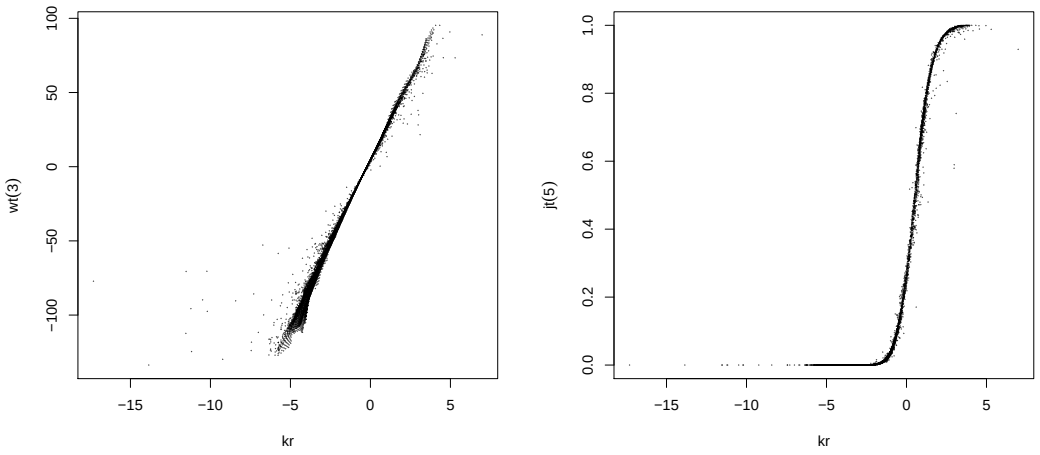


Figure 5: Relationships between test statistics for the Paraná-rainfall data set and $u =$ 300mm. Left panel: $T_{wt(3)}(\cdot)$ versus $T_{kr}(\cdot)$. Right panel: $T_{jt(5)}(\cdot)$ versus $T_{kr}(\cdot)$.



The price paid for the perturbation advantage shown in the Appendix is an outer set $G^{0.1}_{O,jt(5)}$, where typically $|G^{0.1}_{O,jt(5)}| > |G^{0.1}_{O,kr}|$, and an inner set $G^{0.1}_{I,jt(5)}$ where typically $|G^{0.1}_{I,jt(5)}| < |G^{0.1}_{I,kr}|$; see Tables 3 and 4 in the Appendix. Similar conclusions apply to $T_{wt(3)}$. However, for the Paraná-rainfall data set, the price is very small, as the sets only differ by a relatively small number of pixels: $|G^{0.1}_{I,kr}| = 174$, $|G^{0.1}_{I,wt(3)}| = 167$, and $|G^{0.1}_{I,jt(5)}| = 161$; $|G^{0.1}_{O,kr}| = 5,537$, $|G^{0.1}_{O,wt(3)}| = 5,575$, and $|G^{0.1}_{O,jt(5)}| = 5,584$.

5. Discussion and conclusions

The kriging predictor is a conditional expectation, and it is very effective at filtering out noise in the data and filling in gaps between observations. However, it is not explicitly designed to predict when the underlying spatial

process exceeds a threshold; we have called such regions "greater than" (G_Y) and "less than" (L_Y) exceedance sets and, without loss of generality, in this article we have concentrated on G_Y.

The kriging predictor is optimal from a decision-theoretic viewpoint, when the loss function is squared-error loss, which penalizes under- and over-prediction equally. However, when predicting exceedances, pressing the "alarm button" when the true process is below the threshold, typically incurs a much smaller loss than not pressing this button when the true process is above the threshold.

In this article, we developed a new weighted kriging predictor, where the weights come directly from a weighted-squared-error loss function. We also considered a spatial predictor of G_Y based on the joint event that the process at a given location and at its neighbour(s) both exceed the threshold. These three exceedance-set predictors, $G_{Y,kr}$, $G_{Y,wt}$, and $G_{Y,jt}$ are compared in a number of ways.

To obtain uncertainty quantification for the predictors, we adopted the methodology of French and Sain (2013), which defines outer (G_O) and inner (G_I) exceedance sets in an analogous way that upper and lower prediction intervals are defined for a random variable. French and Sain (2013) used a test statistic that was itself a spatial process and depended directly on the kriging predictor, which is derived from a least-squares loss function. We have generalized their approach to test statistics that are derived from a weighted-least-squares approach and from a joint-probability approach. This has led

to outer and inner pairs $(G_{O,kr}, G_{I,kr})$, $(G_{O,wt}, G_{I,wt})$, and $(G_{O,jt}, G_{I,jt})$, and to further comparison of the three inferences in a simulation study in the Appendix: We concluded that the joint-probability approach, 'jt', was preferred. Finally, we applied our methodology to a data set of average (over three decades) winter rainfall in the state of Paraná, Brazil.

The simulation study indicated that the uncertainty quantifications were quite similar (apart from the plug-in approach, which was clearly worse). When the different test statistics are monotonically related, this is to be expected; see (19). Figure 5 shows this relationship for the Paraná-rainfall data set, where the points $(T_{jt(5)}(\mathbf{s}), T_{kr}(\mathbf{s}))$ (left panel) and $(T_{wt(3)}(\mathbf{s}), T_{kr}(\mathbf{s}))$ (right panel) are plotted.

The methodology of French and Sain (2013) controls the FWER of the multiple hypothesis tests $\{H_0(\mathbf{s})$ versus $H_1(\mathbf{s})\}$. It results in very conservative "bounds" for the set G_Y, in that the symmetric set difference between G_I^α and G_O^α is always quite big; see Figure 4 for $\alpha = 0.1$. This is because there is no tolerance in the methodology for G_Y to be "almost" contained in G_O^α.

There are several avenues to explore that will bring the inner and outer sets closer and result in more precise inferences. The 'inf' and 'sup' in (23) and (24) might be replaced with a small quantile and a large quantile, respectively; if so, what would be the corresponding probability statement about the relationship of the inner and outer sets to G_Y? The FWER criterion might be changed to the criterion k-FWER (Korn et al., 2004; Lehmann and

Romano, 2005), where the probability of falsely rejecting at least k hypotheses is controlled, or to controlling directly the false discovery proportion and not only its expectation (i.e., the FDR); if so, what would the effect of spatial dependence be on a statistical procedure that controls k-FWER? The choice of α, α' in (23) and (24) could be increased from the conservatively small values (e.g., $\alpha = \alpha' = 0.1$) traditionally chosen; if so, what would be a good choice of α, α'? In response to the last question, we have investigated a criterion call Higher Criticism (HC) proposed by Donoho et al. (2004). While the presence of spatial dependence invalidates their theoretical results, the intuitive idea behind HC remains: One wishes to test whether there is a small fraction of grid points whose mean is perturbed. To do so, individual p-values associated with each hypothesis test are standardized and an optimal a cut-off defined that is adaptive to the data.

6. Acknowledgements

Noel Cressie's research was supported by ARC Discovery Project DP190100180.

References

Adler, R. J., 1981. The Geometry of Random Fields. Wiley, Chichester, UK.

Aldworth, J., Cressie, N., MAR 1 2003. Prediction of nonlinear spatial functionals. Journal of Statistical Planning and Inference 112 (1-2), 3–41.

Bacro, J.-N., Gaetan, C., Toulemonde, G., 2016. A flexible dependence model for spatial extremes. Journal of Statistical Planning and Inference 172, 36–52.

Banerjee, S., Gelfand, A. E., Finley, A. O., Sang, H., 2008. Stationary process approximation for the analysis of large spatial datasets. Journal of the Royal Statistical Society, Series B 70 (4), 825–848.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57 (1), 289–300.

Bolin, D., Lindgren, F., JAN 2015. Excursion and contour uncertainty regions for latent Gaussian models. Journal of the Royal Statistical Society, Series B 77 (1), 85–106.

Bradley, J., Cressie, N., Shi, T., 2016. A comparison of spatial predictors when datasets could be very large. Statistics Surveys 10, 100–131.

Cooley, D., Davis, R. A., Naveau, P., 2010. The pairwise beta distribution:

A flexible parametric multivariate model for extremes. Journal of Multivariate Analysis 101 (9), 2103–2117.

Cooley, D., Davis, R. A., Naveau, P., et al., 2012. Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data. The Annals of Applied Statistics 6 (4), 1406–1429.

Craigmile, P. F., Cressie, N., Santner, T. J., Rao, Y., 2006. A loss function approach to identifying environmental exceedances. Extremes 8 (3), 143–159.

Cressie, N., 1985. Fitting variogram models by weighted least squares. Journal of the international Association for mathematical Geology 17 (5), 563–586.

Cressie, N., 1993. Statistics for Spatial Data, rev. edn. John Wiley and Sons, New York, NY.

Cressie, N., Hawkins, D. M., 1980. Robust estimation of the variogram: I. Journal of the International Association for Mathematical Geology 12 (2), 115–125.

Cressie, N., Johannesson, G., 2006. Spatial prediction for massive data sets. In: Mastering the Data Explosion in the Earth and Environmental Sciences: Proceedings of the Australian Academy of Science Elizabeth and Frederick White Conference. pp. 1–11.

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. Journal of the Royal Statistical Society, Series B 70 (1), 209–226.

Cressie, N., Wikle, C. K., 2011. Statistics for Spatio-Temporal Data. John Wiley and Sons, Hoboken, NJ.

Daly, C., Caers, J., 2010. Multi-point geostatistics – an introductory overview. First Break 28 (9), 39–47.

Datta, A., Banerjee, S., Finley, A. O., Gelfand, A. E., 2016. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. Journal of the American Statistical Association 111 (514), 800–812.

Dembo, A., Zeitouni, O., 1995. Large deviations via parameter dependent change of measure, and an application to the lower tail of gaussian processes. In: Seminar on Stochastic Analysis, Random Fields and Applications. Springer, pp. 111–121.

Diggle, P. J., Ribeiro Jr, P. J., 2002. Bayesian inference in Gaussian model-based geostatistics. Geographical and Environmental Modelling 6 (2), 129–146.

Diggle, P. J., Tawn, J. A., Moyeed, R., 1998. Model-based geostatistics. Journal of the Royal Statistical Society, Series C 47 (3), 299–350.

Donoho, D., Jin, J., et al., 2004. Higher criticism for detecting sparse heterogeneous mixtures. The Annals of Statistics 32 (3), 962–994.

Eidsvik, J., Finley, A. O., Banerjee, S., Rue, H., JUN 2012. Approximate Bayesian inference for large spatial datasets using predictive process models. Computational Statistics and Data Analysis 56 (6), 1362–1380.

French, J. P., NOV 2014. Confidence regions for the level curves of spatial data. Environmetrics 25 (7), 498–512.

French, J. P., Hoeting, J. A., 2016. Credible regions for exceedance sets of geostatistical data. Environmetrics 27 (1), 4–14.

French, J. P., Sain, S. R., SEP 2013. Spatio-temporal exceedance locations and confidence regions. Annals of Applied Statistics 7 (3), 1421–1449.

Gilleland, E., Ribatet, M., Stephenson, A. G., 2013. A software review for extreme value analysis. Extremes 16 (1), 103–119.

Hope, A., 1968. A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society, Series B 30 (3), 582–598.

Huang, H.-C., Cressie, N., Zammit-Mangion, A., Huang, G., 2019. False discovery rates to detect signals from incomplete spatially aggregated data. arXiv:1905.06268, 1–40, .

Huser, R., Wadsworth, J. L., 2019. Modeling spatial processes with unknown extremal dependence class. Journal of the American Statistical Association 114 (525), 434–444.

Journel, A. G., Posa, D., 1990. Characteristic behavior and order relations for indicator variograms. Mathematical Geology 22 (8), 1011–1015.

Korn, E. L., Troendle, J. F., McShane, L. M., Simon, R., 2004. Controlling the number of false discoveries: Application to high-dimensional genomic data. Journal of Statistical Planning and Inference 124 (2), 379–398.

Lehmann, E. L., Romano, J. P., 06 2005. Generalizations of the familywise error rate 33 (3), 1138–1154.

Lindgren, F., Martins, T., Rue, H., Simpson, D., MAR 2011. Discussion of "Spatial prediction in the presence of positional error". Environmetrics 22 (2), 127.

Lopés, H. F., Gamerman, D., Salazar, E., MAR 1 2011. Generalized spatial dynamic factor models. Computational Statistics and Data Analysis 55 (3), 1319–1330.

Marchini, J., Presanis, A., JUL 2004. Comparing methods of analyzing fMRI statistical parametric maps. Neuroimage 22 (3), 1203–1213.

Naveau, P., Guillou, A., Cooley, D., Diebolt, J., 2009. Modelling pairwise dependence of maxima in space. Biometrika 96 (1), 1–17.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., Sain, S., APR 3 2015. A multiresolution Gaussian process model for the analysis of large spatial datasets. Journal of Computational and Graphical Statistics 24 (2), 579–599.

Papritz, A. J., 2009. Why indicator kriging should be abandoned. Pedometron March 2009 (26), 4–7.

Pavlicova, M., Santner, T., Cressie, N., 2008. Detecting signals in FMRI data using powerful FDR procedures. Statistics and its Interface 1, 23–32.

Reich, B. J., Shaby, B. A., 2012. A hierarchical max-stable spatial model for extreme precipitation. Annals of Applied Statistics 6, 1430–1451.

Ribeiro Jr, P. J., Diggle, P. J., 2018. geoR: Analysis of Geostatistical Data. R package version 1.7-5.2.1.
URL `https://CRAN.R-project.org/package=geoR`

Shen, X., Huang, H.-C., Cressie, N., DEC 2002. Nonparametric hypothesis testing for a spatial signal. Journal of the American Statistical Association 97 (460), 1122–1140.

Stephenson, A. G., Lehmann, E. A., Phatak, A., 2016. A max-stable process model for rainfall extremes at different accumulation durations. Weather and Climate Extremes 13, 44–53.

Stern, H., Cressie, N., 1999. Disease Mapping and Risk Assessment for Public Health. Wiley, Chichester, UK, Ch. Inference for extremes in disease mapping, pp. 63–84.

Tadić, J. M., Qiu, X., Yadav, V., Michalak, A. M., 2015. Mapping of satellite Earth observations using moving window block kriging. Geoscientific Model Development 8 (10), 3311–3319.

Wright, D. L., Stern, H. S., Cressie, N., 2003. Loss functions for estimation of extrema with an application to disease mapping. Canadian Journal of Statistics 31 (3), 251–266.

Zammit-Mangion, A., Cressie, N., Shumack, C., 2018. On statistical approaches to generate Level 3 products from remote sensing retrievals. Remote Sensing 10, 155.

Zhang, J., Craigmile, P. F., Cressie, N., 2008. Loss function approaches to predict a spatial quantile and its exceedance region. Technometrics 50, 216–227.

Appendix: Simulation study

In this Appendix, we investigate the performance of outer and inner predicted exceedance sets for a number of test statistics used to test $\{H_0(\mathbf{s}) : Y(\mathbf{s}) > u$ versus $H_1(\mathbf{s}) \leq u : \mathbf{s} \in D\}$. The original idea, due to French and Sain (2013), was to make inference on G_Y in an analogous way that one-sided prediction intervals are used to make inference on a random variable. Their test statistic was based on the kriging predictor, and in this study we compare it to several new ones that were developed in the main body of the paper.

Consider the unit-square $\{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\} \subset \mathbb{R}^2$, and discretize it into 31×31 grid points as follows: $D \equiv \{(x, y) : x = \frac{0}{M}, \frac{1}{M} \ldots, \frac{M}{M}; y = \frac{0}{M}, \frac{1}{M} \ldots, \frac{M}{M}\}$ with $M = 30$, which yields 961 grid points. We assume $\mu_Y(\mathbf{s}) = 0$ and an isotropic, mean-zero Gaussian process with an exponential covariance function $C_Y(\mathbf{s} - \mathbf{v}) \equiv \sigma^2 \exp(-||\mathbf{s} - \mathbf{v}||/\phi)$, where $||\cdot||$ is the Euclidean distance. Without loss of generality, we set the variance $\sigma^2 = 1$. In our study, $\phi = 0.5$, which yields an equivalent range of approximately 1.5 and represents moderate spatial dependence; the Paraná-rainfall data set analyzed in Section 4 has slightly weaker spatial dependence. We set the measurement-error variance $\sigma^2_\varepsilon = 0.5$ corresponding to a signal-to-noise ratio $\sigma^2/\sigma^2_\varepsilon = 2$; the Paraná-rainfall dataset has a much higher signal-to-noise ratio. Then we produced $10,000$ Monte Carlo (MC) simulated data sets $\mathbf{Y} = (Y(\mathbf{s}) : \mathbf{s} \in D)^\top$ for the hidden process and likewise for the measurements $\mathbf{Z} = (Z(\mathbf{s}) : \mathbf{s} \in D)^\top$. Then we randomly selected $n = 106$

(approximately 10%) locations $\{\mathbf{s}_i^o : i = 1, \ldots, 106\}$, which were fixed for all simulated data sets (i.e., $\mathcal{J}^o$ was fixed). Then, for each observed data set $\mathbf{Z}(\mathcal{J}^o)$, we produced $B = 10,000$ conditional simulations from $[Y(\mathbf{s})|\mathbf{Z}(\mathcal{J}^o)]$ for all $\mathbf{s} \in D$, by sampling from a Gaussian distribution with mean given by the kriging predictor (10) and variance-covariances given by the kriging variance (11) and kriging covariance, (38). Then we computed the statistics, $T_{kr}(\cdot)$ given by (20), $T_{wt(c)}(\cdot)$ given by (33), and $T_{jt(h)}(\cdot)$ given by (39), where the choice of threshold $u \in \{0.67, 1.5\}$ allowed us to see the effect of smaller-versus-larger thresholds. Finally, the exceedance-set methodology presented in Section 3 was applied to each of the test statistics.

We considered the following "responses": For G_O and G_I, we computed the coverage (Tables 1 and 2) and the average area (Tables 3 and 4) obtained for each test statistic. We also computed these responses under deviations from the working model $\mu_Y(\mathbf{s}) = 0$, by adding an extra signal to $\mu_Y(\cdot)$, although in the data analysis mean of zero was assumed. Thus, we generated data $\mathbf{Z}(\mathcal{J}^o)$ with mean,

$$\mu_Z(\mathbf{s}) = \mu_Y(\mathbf{s}) = H \cdot I(\mathbf{s} \in \Delta); \mathbf{s} \in D,$$

where Δ is a square region of area $10/M \times 10/M$ in the middle of the $(M+1) \times (M+1)$ grid ($M = 30$ here) and $H \in \{0, 1, 2\}$ is a perturbation parameter. However, our spatial analyses assumed that $\mu_Z(\cdot) = \mu_Y(\cdot) \equiv 0$.

For $H = 0$ (where no perturbation was added to the mean function μ_Y),

Tables 1 and 2 show that the coverage is excellent for all methods for G_O, except 'jt(10)', and conservative for all methods for G_I. Closer inspection reveals that the discreteness of the grid is causing the over-coverage for G_I, with the remedy being a finer-resolution grid (e.g., $M = 100$) and a considerable lengthening of computation time. Coverage worsens slightly for the larger u, which again could be restored with a finer-resolution grid (and more conditional-simulation samples).

The parameter H adds a sensitivity-to-mean-assumption factor to our simulation study; we did this since it is rare in practice that the fitted model captures all the trend in the data. It is true that a spatial model has inherent stability since what is missed in the mean can be captured (perhaps imperfectly) in the covariance function. Nevertheless, a comparison of the coverages and average areas of G_O and G_I for the different methods indicate which statistic is stable when the model mean is perturbed.

For $H > 0$ and increasing, the coverage deteriorates for all test statistics but some more than others. In particular, 'kr' is sensitive and, of the others, 'jt(5)' appears to be preferred and enjoys better coverage than its brethren, 'jt(1)' and 'jt(2)'. For the weighted statistics, both 'wt(1)' and 'wt(3)' appear to be prefered over 'kr' and also over 'wt(10)'. From Tables 3 and 4, the average area of the set G_O is smallest mostly for 'kr', and the average area of the set G_I is largest mostly for 'kr'. Considering this response by itself, 'kr' is hard to beat, since the most desirable result would be to have an outer predicted exceedance set that is "small" and an inner predicted exceedance

set that is "large." However, average area is a relevant response when the coverage probability remains high. Taking the responses of the study into account for all H, the methodology associated with 'jt(5)' and 'wt(3)' appear to be an improvement over that for 'kr'.

We also investigated the performance of the pointwise-plug-in outer and inner predicted exceedance sets, $G^{\alpha}_{O,pl}$ and $G^{\alpha'}_{I,pl}$, defined by (41). It is apparent from the results in Tables 1 and 2 that they have considerable under-coverage, even when $H = 0$, and hence they are not suitable as outer and inner predicted exceedance sets.

Diggle and Ribeiro Jr (2002) considered the probability contours $C^p \equiv \{\mathbf{s} \in D : Pr(Y(\mathbf{s}) > u|\mathbf{Z}(\jmath^o)) \geq p\}$, for $p \in (0,1)$ and chose $p = 0.1, 0.5, 0.9$ for illustration. Now, their equation that defines the probability contours C^p, namely $p \leq Pr(Y(\mathbf{s}) > u|\mathbf{Z}(\jmath^o)) = \Phi\left(\frac{Y_{kr}(\mathbf{s})-u}{\sigma_{kr}(\mathbf{s})}\right)$, can be re-written as $Y_{kr}(\mathbf{s}) - \Phi^{-1}(p)\sigma_{kr}(\mathbf{s}) > u$. When comparing this latter expression with (41), it is apparent that C^p is equivalent to $G^{\alpha'}_{I,pl}$ with $\alpha' = 1 - p$ and also to $G^{\alpha}_{O,pl}$ with $\alpha = p$, upon using the relation, $\Phi^{-1}(p) = -\Phi^{-1}(1-p)$. For example, suppose $\alpha = \alpha' = 0.1$, then $G^{0.1}_{I,pl}$ is equivalent to $C^{0.9}$ and $G^{0.1}_{O,pl}$ is equivalent to $C^{0.1}$. Based on these equivalences and our simulation results, we conclude that the probability contours used by Diggle and Ribeiro Jr (2002) should be avoided.

Table 1: Coverage, $Pr(G_Y \subset G_O)$, of the outer predicted exceedance set G_O with $\alpha = 0.1$ for the statistics 'pl', 'kr', 'wt(c)' and 'jt(h)' for threshold values of $u = 0.67, 1.28$ and perturbation parameter values of $H = 0, 1, 2$, based on 10,000 simulated data sets and exponential covariance function ($\phi = 0.5$, $\sigma^2 = 1$, $\sigma_\varepsilon^2 = 0.5$).

	$u = 0.67$			$u = 1.28$		
Method	$H = 0$	$H = 1$	$H = 2$	$H = 0$	$H = 1$	$H = 2$
pl	0.069	0.049	0.041	0.172	0.086	0.016
kr	0.901	0.821	0.670	0.901	0.771	0.441
jt(1)	0.900	0.824	0.689	0.901	0.777	0.474
jt(2)	0.901	0.833	0.713	0.902	0.790	0.515
jt(5)	0.895	0.843	0.764	0.893	0.817	0.621
jt(10)	0.858	0.819	0.751	0.856	0.788	0.620
wt(1)	0.900	0.845	0.717	0.896	0.816	0.538
wt(3)	0.899	0.839	0.706	0.897	0.812	0.525
wt(10)	0.901	0.827	0.675	0.900	0.796	0.485

Table 2: Coverage, $Pr(G_I \subset G_Y)$, of the inner predicted exceedance set G_I with $\alpha' = 0.1$ for the statistics 'pl', 'kr', 'wt(c)' and 'jt(h)' for threshold values of $u = 0.67, 1.28$ and perturbation parameter values of $H = 0, 1, 2$, based on 10,000 simulated data sets and exponential covariance function ($\phi = 0.5$, $\sigma^2 = 1$, $\sigma_\varepsilon^2 = 0.5$).

	$u = 0.67$			$u = 1.28$		
Method	$H = 0$	$H = 1$	$H = 2$	$H = 0$	$H = 1$	$H = 2$
pl	0.462	0.339	0.166	0.754	0.630	0.394
kr	0.926	0.850	0.456	0.956	0.917	0.661
jt(1)	0.933	0.870	0.601	0.961	0.925	0.758
jt(2)	0.936	0.892	0.703	0.968	0.940	0.815
jt(5)	0.948	0.919	0.785	0.976	0.961	0.882
jt(10)	0.964	0.939	0.827	0.984	0.970	0.883
wt(1)	0.935	0.917	0.777	0.976	0.968	0.902
wt(3)	0.938	0.920	0.787	0.977	0.968	0.903
wt(10)	0.940	0.923	0.790	0.980	0.970	0.907

Table 3: Average area, $|G_O|$, of the outer predicted exceedance set G_O with $\alpha = 0.1$ for the statistics 'pl', 'kr', 'wt(c)' and 'jt(h)' for threshold values of $u = 0.67, 1.28$ and perturbation parameter values of $H = 0, 1, 2$, based on 10,000 simulated data sets and exponential covariance function ($\phi = 0.5$, $\sigma^2 = 1$, $\sigma_\varepsilon^2 = 0.5$).

	$u = 0.67$			$u = 1.28$		
Method	$H = 0$	$H = 1$	$H = 2$	$H = 0$	$H = 1$	$H = 2$
pl	405.9	452.7	483.3	187.4	230.1	272.7
kr	634.8	668.2	683.1	396.7	441.6	470.4
jt(1)	637.2	670.4	685.1	399.4	444.2	472.7
jt(2)	639.9	673.1	687.9	402.5	447.3	476.1
jt(5)	662.5	695.1	711.0	431.7	478.0	509.6
jt(10)	723.8	752.9	770.2	528.6	579.1	628.8
wt(1)	669.1	700.1	712.8	437.9	481.8	507.7
wt(3)	659.3	690.8	704.1	431.7	475.7	502.0
wt(10)	637.9	671.0	685.6	411.6	456.2	483.7

Table 4: Average area, $|G_I|$, of the inner predicted exceedance set G_I with $\alpha' = 0.1$ for the statistics 'pl', 'kr', 'wt(c)' and 'jt(h)' for threshold values of $u = 0.67, 1.28$ and perturbation parameter values of $H = 0, 1, 2$, based on 10,000 simulated data sets and exponential covariance function ($\phi = 0.5$, $\sigma^2 = 1$, $\sigma_\varepsilon^2 = 0.5$).

	$u = 0.67$			$u = 1.28$		
Method	$H = 0$	$H = 1$	$H = 2$	$H = 0$	$H = 1$	$H = 2$
pl	101.0	135.7	179.7	27.2	44.4	80.9
kr	40.8	61.8	102.6	10.6	19.3	47.3
jt(1)	37.2	57.2	95.8	9.46	17.6	43.6
jt(2)	32.3	50.3	83.7	8.03	15.1	37.6
jt(5)	18.3	28.5	46.7	4.20	7.93	20.1
jt(10)	7.06	9.28	11.7	1.39	2.10	3.35
wt(1)	17.0	28.3	58.5	1.69	3.53	12.6
wt(3)	15.9	26.5	55.8	1.56	3.29	11.8
wt(10)	15.5	26.1	55.2	1.33	2.80	10.3