Published in final edited form as:

J Am Stat Assoc. 2018 ; 113(523): 1327–1340. doi:10.1080/01621459.2017.1341838.

Confidence regions for spatial excursion sets from repeated random field observations, with an application to climate

Max Sommerfeld*,

FBMS, Universität Göttingen

Stephan Sain,

The Climate Corporation

Armin Schwartzman

Division of Biostatistics, University of California, San Diego

Abstract

The goal of this paper is to give confidence regions for the excursion set of a spatial function above a given threshold from repeated noisy observations on a fine grid of fixed locations. Given an asymptotically Gaussian estimator of the target function, a pair of data-dependent nested excursion sets are constructed that are sub- and super-sets of the true excursion set, respectively, with a desired confidence. Asymptotic coverage probabilities are determined via a multiplier bootstrap method, not requiring Gaussianity of the original data nor stationarity or smoothness of the limiting Gaussian field. The method is used to determine regions in North America where the mean summer and winter temperatures are expected to increase by mid 21st century by more than 2 degrees Celsius.

1 Introduction

Our motivation comes from the following problem. Faced with a global change in temperature within the next century, it is important to assess which geographical regions are particularly at risk of extreme temperature change. The data used here, obtained from the North American Regional Climate Change Assessment Program (NARCCAP) project (Mearns et al., 2009, 2012, 2013), consists of two sets of 29 spatially registered arrays of mean seasonal temperatures for summer (June–August) and winter (December–February) evaluated at a fine grid of fixed locations 0.5 degrees apart in geographic longitude and latitude over North America over two time periods: late 20th century (1971–1999) and mid

M.S. acknowledges support by the “Studienstiftung des Deutschen Volkes” and the SAMSI 2013–2014 program on Low-dimensional Structure in High-dimensional Systems. A.S. and S.S. were partially supported by NIH grant R01 CA157528. S.S. began working on this research while he was a Scientist with the Institute for Mathematics Applied to the Geosciences, National Center for Atmospheric Research, Boulder, CO. All authors wish to thank the North American Regional Climate Change Assessment Program (NARCCAP) for providing the data used in this paper. NARCCAP is funded by the National Science Foundation (NSF), the U.S. Department of Energy (DoE), the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Environmental Protection Agency Office of Research and Development (EPA).

SUPPLEMENTARY MATERIAL

Supplementary material Additional simulations and data analysis.

R-package narccapdata: R-package containing the climate data analyzed in the article.

21st century (2041–2069). Specifically, the data was produced by the WRFG climate model (Michalakes et al., 2004) using boundary conditions from the CGCM3 global model (Flato, 2005). We would like to determine the regions whose difference in mean summer or winter temperature between the two periods is greater than the 2°C benchmark (Rogelj et al., 2009, Anderson and Bows, 2011). However, the observed differences may be confounded by the natural year-to-year temperature variability. Can we set confidence bounds on such regions that reflect the year-to-year variability in the data?

In mathematical terms, suppose that for some unknown target function $\mu: S \rightarrow \mathbb{R}$ on a spatial domain $S \subset \mathbb{R}^d$ (in our case the difference in mean temperature between the two time periods), we are interested in the excursion set of $\mu(\mathbf{s})$ above a fixed threshold c , defined as $A_c := A_c(\mu) := \{\mathbf{s} \in S: \mu(\mathbf{s}) > c\}$. More specifically, based on an estimate $\hat{\mu}_n(\mathbf{s})$ of $\mu(\mathbf{s})$ we wish to obtain confidence regions $\hat{A}_c^{\pm}$ for which the probability that

$$\hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \quad (1)$$

holds is asymptotically above a desired level, say 90%.

We obtain such sets $\hat{A}_c^{\pm}$ as excursion sets of the estimate $\hat{\mu}_n(\mathbf{s})$ and we call them Coverage Probability Excursion (CoPE) sets. Assuming that $\mu(\mathbf{s})$ is continuous in the proximity of the level set $A_c = \{\mathbf{s} \in S: \mu(\mathbf{s}) = c\}$ and that the estimator $\hat{\mu}_n(\mathbf{s})$ satisfies a functional central limit theorem (CLT), we show that the probability that the inclusions (1) hold is given asymptotically by the distribution of the supremum of the limiting Gaussian random field on the level set A_c .

To apply this general framework to our data, we model the yearly average temperature fields $Y_i(\mathbf{s})$, $i = 1, \dots, n$, as realizations of a general linear model indexed by $\mathbf{s} \in S$

$$\mathbf{Y}(\mathbf{s}) = \mathbf{X}\mathbf{b}(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (2)$$

with $\mathbf{X}$ indicating the two time periods. Setting a simple contrast vector $\mathbf{w}$, the target function $\mu(\mathbf{s}) = \mathbf{w}^T \mathbf{b}(\mathbf{s})$ represents the difference of means to be estimated. For the general model (2) with arbitrary $\mathbf{X}$ and $\mathbf{w}$, we show that under proper conditions on design and error fields $\epsilon(\mathbf{s})$ a least-squares fit at each location $\mathbf{s}$ will produce a consistent and asymptotically Gaussian estimator $\hat{\mu}_n(\mathbf{s}) = \mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s})$ thereby allowing the construction of CoPE sets, as described above.

For illustration, Figure 1 shows CoPE sets for the temperature data. We target the region where the mean temperature in the future period (2041–2069) exceeds the mean temperature in the past (1971–1999) by 2°C or more. With simultaneous approximate probability of 90%

this region is no smaller than the region given by the red boundary ($\hat{A}_c^+$) and no larger than indicated by the green boundary ($\hat{A}_c^-$).

Climate model output is a new high-dimensional type of geostatistical data. With the increasing interest in climate modeling, there is a large amount of data of this type, e.g. from NARCCAP and its European counterpart ENSEMBLES (van der Linden and Mitchell, 2009). Unlike the traditional data setup of spatial statistics, climate model output data is produced at all locations (in practice, on a fine grid) without measurement error (Bukovsky, 2012). In this sense, our setup is more similar to that of population studies in brain imaging, where a difference map between two conditions is estimated from repeated co-located image observations at a fine spatial grid under those conditions (see e.g. Worsley et al. (1996), Genovese et al. (2002), Taylor and Worsley (2007), Schwartzman et al. (2010)). In the discussion we suggest an approach to apply our ideas in the setting of dense discrete observed locations and in the presence of measurement error, inspired by functional data analysis.

The problem of finding the threshold for our CoPE sets involves computation of the quantiles of the supremum of a limiting Gaussian random field. While the general theory is valid whenever a functional CLT can be established for the estimator $\hat{\mu}_n(s)$, we offer a practical procedure in the case of a linear model. In French (2014) a similar task was performed by Monte Carlo simulation assuming a parametric class for the covariance structure of the field. More generally for unknown and typically non-stationary covariance function, as we attempt here, we propose a simple and efficient multiplier bootstrap procedure (Wu, 1986, Hardle and Mammen, 1993, Mammen, 1992, 1993), that does not require estimating the unknown correlation function. The validity of this procedure for very high-dimensional data has been shown by Chernozhukov et al. (2013).

This problem has also been addressed and elegantly solved by Taylor and Worsley (2007) using the Gaussian kinematic formula. However, this method requires that the observations themselves be Gaussian and requires the field to be differentiable. The multiplier bootstrap allows us to avoid both these assumptions while being extremely fast to compute. We compare the finite sample performance of the Gaussian kinematic formula method and the multiplier bootstrap in a simulation.

Related Work

The problem of finding confidence regions for spatial excursion sets (exceedance regions) or level sets (contours), has been studied in other contexts. In the geostatistics literature, the target function is itself a Gaussian field, partially observed at relatively few spatial locations. The limited amount of information available is typically compensated by assuming that the field is stationary, which we do not. The uncertainty on the estimation of the random level contours and excursion sets has been addressed from a frequentist perspective for the former (Lindgren and Rychlik, 1995, Wameling, 2003, French and Sain, 2013) and the latter (French and Sain, 2013), as well as from a Bayesian perspective for both (Bolin and Lindgren, 2014). Incidentally, our techniques share some similarities with French (2014),

although we will show that distinguishing between level contours and excursion sets is important.

In non-parametric density estimation and regression the target function is a probability density or regression function, estimated from realizations random vector or observations at a discrete set of locations, respectively. While the estimation of both level sets and contours have been well studied (Tsybakov, 1997, Cavalier, 1997, Cuevas et al., 2006, Willett and Nowak, 2007, Singh et al., 2009, Rigollet and Vert, 2009), there is less literature on confidence statements. Mason and Polonik (2009) showed asymptotic normality of plug-in level set estimates with respect to the measure of symmetric set difference. Mammen and Polonik (2013) proposed a bootstrapping scheme to obtain confidence regions analogous to our CoPE sets from vector-valued samples.

Other settings involving excursions over thresholds (albeit not necessarily excursion sets in our sense) include the following. Extreme value theory studies the occurrence and size of exceedances over high thresholds (Davison and Smith, 1990). The spatial cumulative distribution function (SCDF) of a random field measures the proportion of the area where the field takes values below a certain threshold (Majure et al., 1995, Lahiri et al., 1999, Zhu et al., 2002). In a Bayesian framework, special loss functions have been proposed to obtain estimates of maxima or near-maxima or exceedance regions from the posterior (Wright et al., 2003, Zhang et al., 2008). A related approach is Bayesian hotspot detection (De Oliveira and Ecker, 2002). Scan statistics are widely used to detect hotspots or clusters (Patil and Taillie, 2004, Tango and Takahashi, 2005, Duczmal et al., 2006).

Software

All required functions for computation and visualization of CoPE sets developed in this paper are available in the R-package cope (Sommerfeld, 2015) in R (R Core Team, 2015).

2 Error control for excursion sets - CoPE sets

2.1 Setup

The domain $S \subset \mathbb{R}^N$ on which all our functions and processes are defined, is assumed to be a compact (but not necessarily connected) subset of Euclidean space.

Assumption 1. *We assume that*

- (a) *every open ball around a point in the contour or level set $A_c = \{\mathbf{s} \in S : \mu(\mathbf{s}) = c\}$ has non-empty intersection with $\{\mathbf{s} : \mu(\mathbf{s}) > c\}$ and with $\{\mathbf{s} : \mu(\mathbf{s}) < c\}$. Further, there exists $\eta_0 > 0$ such that the target function $\mu(\mathbf{s})$ is continuous on the set $U = \{\mathbf{s} \in S : |\mu(\mathbf{s}) - c| \leq \eta_0\}$.*
- (b) *the estimators $\hat{\mu}_n(\mathbf{s})$ are continuous on U for all $n \in \mathbb{N}$ large enough and there exist a sequence of numbers $\tau_n \rightarrow 0$ and a continuous function $\sigma : U \rightarrow \mathbb{R}^+$ bounded from above and below such that*

$$\left\{ \frac{\hat{\mu}_n(\mathbf{s}) - \mu(\mathbf{s})}{\tau_n \sigma(\mathbf{s})} \right\}_{\mathbf{s} \in U} \rightarrow \{G(\mathbf{s})\}_{\mathbf{s} \in U}, \quad (3)$$

weakly in $C(U)$ as $n \rightarrow \infty$. Here, G is a Gaussian field on U with mean zero, unit variance and continuous sample paths with probability one.

- (c) there is a constant $C > 0$ such that for every $\mathbf{s} \notin U$ we have

$$(\hat{\mu}_n(\mathbf{s}) - c) \cdot \text{sgn}(\mu(\mathbf{s}) - c) \geq C + \sigma(\mathbf{s})Z_n(\mathbf{s}), \text{ a.s.}, \quad (4)$$

with $Z_n(\mathbf{s})$ a sequence of stochastic processes on S such that $\sup_{\mathbf{s} \in S} |Z_n(\mathbf{s})| = \mathcal{O}_P(\tau_n)$.

Remark 1. The above assumptions need some explanation.

- (a) The first part of Assumptions 1 (a) ensures that $\mu(\mathbf{s})$ is not flat at the level c such that A_c is in fact of lower dimension than N and excludes pathological cases where c is an isolated point in the image of $\mu(\mathbf{s})$ such that $U = A_c$ for positive η_0 . The second part guarantees that $\mu(\mathbf{s})$ is continuous whenever it has values close to c .
- (b) Part (b) says that $\hat{\mu}_n(\mathbf{s})$ needs to be a good estimator of $\mu(\mathbf{s})$ around the level c in the sense that it satisfies a functional central limit theorem.
- (c) The last assumption concerns the quality of $\hat{\mu}_n(\mathbf{s})$ away from the level c . It need not be a good estimator there, not even consistent. We merely require that whenever $\mu(\mathbf{s})$ is smaller/larger than c that also $\hat{\mu}_n(\mathbf{s})$ is (eventually with high probability) smaller/larger than c . This flexibility can be useful when the target function $\mu(\mathbf{s})$ is not continuous outside of U since in this case it might not be possible to find an estimator that is consistent everywhere.

Remark 2.

- (a) The Assumptions 1 reflect the properties of the temperature data motivating the theory. Notably, we have disconnectedness of the domain in question (cf. Figure 1) due to islands and discontinuities can appear on the boundaries of large lakes.
- (b) While we require continuity of the target function around the level of interest c for our theory, simulations suggest (see the supplement) that this assumption is not essential to attain the desired coverage.
- (c) In general, Assumptions 1 may include estimators based on an increasing number of repeated observations (like in our data example) or an increasing number of sampling spatial points (like in the spatial statistics and nonparametric regression problems).

Finite Locations—In practice, we might check inclusion only on a subset of locations D_n , e.g. a grid of points, which may change with n . That is, we may want to control the probability that $(\hat{A}_c^+ \cap D_n) \subset (A_c \cap D_n) \subset (\hat{A}_c^- \cap D_n)$. The following quantities of the set D_n will be relevant. Define for every $\mathbf{s} \in S$ the projections

$$\mathbf{p}_n^\pm(\mathbf{s}) = \operatorname{argmin}_{\mathbf{s}' \in D_n \cap \{\mathbf{s}' : \pm(\mu(\mathbf{s}') - c) > 0\}} |\mathbf{s} - \mathbf{s}'|.$$

Note that $\mathbf{p}_n^\pm(\mathbf{s})$ are functions of the point $\mathbf{s}$. Further, let

$$\Delta_n^\pm = \sup_{\mathbf{s} \in \partial A_c} |\mathbf{p}_n^\pm(\mathbf{s}) - \mathbf{s}|, \quad \nabla_n^\pm = \sup_{\mathbf{s} \in \partial A_c} \left| \frac{\mu(\mathbf{p}_n^\pm(\mathbf{s})) - c}{\sigma(\mathbf{p}_n^\pm(\mathbf{s}))} \right|.$$

Assumption 2. We assume that the sequence of sets $D_n \subset S$ is such that $\Delta_n^\pm \rightarrow 0$ and

$$\nabla_n^\pm / \tau_n \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Remark 3.

- (a) The quantity $\Delta_n^\pm$ controls how dense D_n is around the contour A_c . For example, if D_n is a regular grid with spacing δ_n , then $\Delta_n^\pm \leq \delta_n$.
- (b) On the other hand, $\nabla_n^\pm$ describes the regularity of the function μ around A_c . If $\mu(\mathbf{s})$ is Lipschitz continuous with Lipschitz constant L in a neighborhood of A_c then $\nabla_n^\pm \leq L \sup_{\mathbf{s} \in \partial A_c} |\sigma^{-1}(\mathbf{s})| \Delta_n^\pm$. Hence, in this case, Assumption 2 is satisfied if $\Delta_n^\pm = o(\tau_n^{-1})$.

Combining both observations, Assumption 2 holds, for example, if $\mu(\mathbf{s})$ is continuously differentiable in a neighborhood of the contour and D_n is a regular grid with spacing $o(\tau_n^{-1})$.

2.2 CoPE Sets

We will obtain nested estimates by thresholding the surface $\hat{\mu}_n(\mathbf{s})$ as follows:

$$\hat{A}_c^+ := \hat{A}_c(+a) := \left\{ \mathbf{s} : \frac{\hat{\mu}_n(\mathbf{s}) - c}{\tau_n \sigma(\mathbf{s})} \geq +a \right\}, \quad \hat{A}_c^- := \hat{A}_c(-a) := \left\{ \mathbf{s} : \frac{\hat{\mu}_n(\mathbf{s}) - c}{\tau_n \sigma(\mathbf{s})} \geq -a \right\}, \quad (5)$$

where a is an appropriate non-negative constant to be determined. Note that $\hat{A}_c = \hat{A}_c(0)$ and $\hat{A}_c^\pm$ are themselves excursion sets. Moreover, for any choice of $a \geq 0$ we have the inclusions $\hat{A}_c^+ \subset \hat{A}_c \subset \hat{A}_c^-$, and hence the estimates obtained via (5) are nested. We refer to $\hat{A}_c^+$ and $\hat{A}_c^-$ as

inner and outer confidence region, respectively. The function used to define the excursion sets is similar to the test statistic used in French and Sain (2013).

The following main result shows how the constant a in (5) can be chosen such that the desired inclusions hold with a predefined probability.

Theorem 1. *If the Assumptions 1 hold, then*

(a)

$$\lim_{n \rightarrow \infty} P \left[\hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \right] = P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \leq a \right].$$

(b) for any subset $D \subset S$,

$$\liminf_{n \rightarrow \infty} P \left[(D \cap \hat{A}_c^+) \subset (D \cap A_c) \subset (D \cap \hat{A}_c^-) \right] \geq P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \leq a \right].$$

(c) for a sequence of sets D_n that additionally satisfy Assumption 2,

$$\lim_{n \rightarrow \infty} P \left[(D_n \cap \hat{A}_c^+) \subset (D_n \cap A_c) \subset (D_n \cap \hat{A}_c^-) \right] = P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \leq a \right].$$

A direct consequence of Theorem 1 is that $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$ with asymptotic probability at least $1 - \alpha$ if we choose a such that $P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \geq a \right] \leq \alpha$. The determination of a poses a computational challenge since the distribution of the supremum of $|G(\mathbf{s})|$ and the set A_c are unknown. In Section 3.2 we propose an easy and fast way to approximate this distribution by a multiplier bootstrap.

The second and third part of Theorem 1 address situations in which the fields are only observed on a, usually finite, set of locations D , e.g. a finite grid. In this case, the theorem asserts that, independent of the set D , the desired inclusions hold at all observed locations D asymptotically with probability *at least* $1 - \alpha$. This is in agreement with the intuition that the desired inclusion is easier to obtain if one only considers a small set of locations. The third part of the Theorem asserts that for a sequence of sets D_n that become dense around the contour sufficiently fast, the inclusion even holds with probability *equal to* $1 - \alpha$. We address this issue further in the simulations section and in the discussion.

Confidence regions for the excursion set A_c yield confidence sets for the contour A_c .

Corollary 1. *Under the assumptions of Theorem 1, we have*

$$\lim_{n \rightarrow \infty} P \left[\partial A_c \subset \text{cl} \left(\hat{A}_c^- \setminus \hat{A}_c^+ \right) \right] = P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \leq a \right],$$

where cl denotes the topological closure.

Note that, conversely, confidence sets for the contour do not automatically give confidence regions for the excursion set. In Figure 2 we show a simple schematic example of a pair of nested sets $\hat{A}_c^\pm$ for which $\partial A_c \subset \hat{A}_c^- \setminus \hat{A}_c^+$ holds but $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$ does not. In fact, excluding these cases is the more laborious part of the proof of Theorem 1. The key is to divide the region S into a close-range zone where $\mu(s)$ is close to c , given by the inflated boundary $A_c^\eta = \{s \in S : c - \eta\sigma(s) \leq \mu(s) \leq c + \eta\sigma(s)\}$ and a long-range zone. Then, the strategy of the proof is to let the parameter η go to zero at an appropriate rate as $n \rightarrow \infty$ such that, eventually, the probability of a part of $\hat{A}_c^+$ falsely appearing in the long-range zone $S \setminus A_c^\eta$ (as shown in Figure 2) vanishes. The probability of making an error remains in the close-range zone, and is asymptotically given by $P \left[\sup_{\partial A_c} |G(s)| > a \right]$.

3 CoPE sets for general linear models

For concreteness and later application to the climate data, we here present how CoPE sets are obtained, in theory and in practice, when the target function is a parameter function in a general linear model and n is the number of repeated observations.

3.1 Asymptotic coverage probabilities

We consider the model

$$\mathbf{Y}(s) = \mathbf{X}\mathbf{b}(s) + \epsilon(s), \quad s \in S \subset \mathbb{R}^N, \quad (6)$$

where $\mathbf{Y}(s)$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is a known $n \times p$ design matrix, $\mathbf{b}(s) = (b_1(s), \dots, b_p(s))$ is an unknown $p \times 1$ vector of parameters and $\epsilon(s) = (\epsilon_1(s), \dots, \epsilon_n(s))$ an unknown stochastic process. CLTs of the kind presented here are known in the case of independent errors (see e.g. Eicker (1963)). We show and prove versions tailored to our specific purpose for coherence and convenience. Specifically, for the noise ϵ , we assume that there exists a positive definite function $\mathbf{c} : S \times S \rightarrow (-1, 1)$ with $\mathbf{c}(s, s) = 1$ for all $s \in S$, and a spatially varying positive definite matrix $\mathbf{\Omega}(s) \in \mathbb{R}^{n \times n}$ such that

$$\text{cov} [\epsilon(s_k), \epsilon(s_l)] = \mathbf{c}(s_k, s_l) \mathbf{\Omega}^{1/2}(s_k) \mathbf{\Omega}^{1/2}(s_l).$$

The correlation function $\mathbf{c}(\cdot, \cdot)$ describes the (non-stationary) spatial correlation structure of the error field, while $\mathbf{\Omega}(s)$ describes the dependence structure between errors in different observations at each location. Note that it will *not* be necessary to estimate this function.

To estimate $\mathbf{b}(s)$ we use the generalized least squares estimator (GLS) (Rao and Toutenburg, 1995)

$$\hat{\mathbf{b}}(\mathbf{s}) = (\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{Y}(\mathbf{s}).$$

The matrix $\boldsymbol{\Omega}(\mathbf{s})$ is unknown in general but can be replaced by a consistent estimate when a sufficiently simple model is assumed. We will use an autoregressive model for our data application in Section 5.

In the notation of Section (2), the target function $\mu(\mathbf{s})$ is now a linear combination $\mathbf{w}^T \mathbf{b}(\mathbf{s})$ for some vector of coefficients $\mathbf{w} \in \mathbb{R}^p$. Naturally, $\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s})$ now plays the role of the estimator $\hat{\mu}_n(\mathbf{s})$.

Next, we establish a CLT for the parameter estimators in our setting, generalizing well-known CLTs in the case of independent errors (see e.g. Eicker (1963)).

Additional notation—Recall that for $1 \leq p < \infty$, the p -norm $\|A\|_p$ of a matrix A is defined to be $\|A\|_p = \sup_{\|x\|_p=1} \|Ax\|_p$. Hence, by definition $\|Ax\|_p \leq \|A\|_p \|x\|_p$ for all x . When $p = \infty$ the matrix norm $\|A\|_\infty$ is the maximum absolute row sum of A , i.e.

$$\|A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|.$$

Definition 1. We denote the Lebesgue measure of a set $A \subset S$ by $|A|$. For vectors $\mathbf{s}, \mathbf{t} \in \mathbb{R}^N$ define the *block* $(\mathbf{s}, \mathbf{t}) = (s_1, t_1] \times \cdots \times (s_N, t_N] \subset \mathbb{R}^N$ and for a stochastic process $(\mathbf{s})$ with index set containing $(\mathbf{s}, \mathbf{t})$ define the *increment* of $(\mathbf{s})$ around $(\mathbf{s}, \mathbf{t})$ (cf. Bickel and Wichura (1971)) as

$$\epsilon((\mathbf{s}, \mathbf{t})) = \sum_{\kappa_1=0,1} \cdots \sum_{\kappa_N=0,1} (-1)^{N - \sum_{j=1}^N \kappa_j} \epsilon(s_1 + \kappa_1(t_1 - s_1), \dots, s_N + \kappa_N(t_N - s_N)).$$

In dimension $N=1$ the definition of an increment yields $\epsilon((s_1, t_1]) = \epsilon(t_1) - \epsilon(s_1)$, while in dimension $N=2$, $\epsilon((\mathbf{s}, \mathbf{t})) = \epsilon(t_1, t_2) - \epsilon(s_1, t_2) - \epsilon(t_1, s_2) + \epsilon(s_1, s_2)$.

Assumption 3. Write $\mathbf{Z}(\mathbf{s}) = \boldsymbol{\Omega}^{-1/2}(\mathbf{s}) \mathbf{X}(\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X})^{-1/2}$ for brevity. Assume that

- (a) the noise field $(\mathbf{s})$ has continuous sample paths with probability one. Moreover, a centered unit variance Gaussian field with the same correlation function $c(\mathbf{s}_1, \mathbf{s}_2)$ as the noise field $(\mathbf{s})$ also has continuous sample paths with probability one.
- (b) the maps $\mathcal{T} \rightarrow \boldsymbol{\Omega}(\mathbf{s})$ is continuous as a function $S \rightarrow \mathbb{R}^{n \times n}$.
- (c) the de-correlated error field $\tilde{\epsilon}(\mathbf{s}) = \boldsymbol{\Omega}^{-1/2}(\mathbf{s}) \epsilon(\mathbf{s})$ has independent components.
- (d) there exists $\delta > 0$ and a constant $K_\delta > 0$ such that $\sup_{\mathbf{s} \in S} E |\tilde{\epsilon}_j(\mathbf{s})|^{2+\delta} \leq K_\delta$, independent of j and n and such that $\sup_{\mathbf{s} \in S} n \|\mathbf{Z}(\mathbf{s})\|_\infty^{2+\delta} \rightarrow 0$ as $n \rightarrow \infty$.

- (e) *there exist $\gamma \geq 0$ and $\beta > 0$ such that for a constant $K_{(\beta, \gamma)} > 0$ independent of n and $\mathbf{s}$ we have $E\|[\mathbf{Z}(\cdot)\epsilon(\cdot)](B)\|_1^{2+\gamma} \leq K_{(\beta, \gamma)}|B|^{1+\beta}$.*
- (f) *the function $\mathbf{v}_n(\mathbf{s}) = \mathbf{w}^T(X^T\boldsymbol{\Omega}^{-1}(\mathbf{s})X)^{-1/2}$ satisfies $\mathbf{v}_n(\mathbf{s})/\|\mathbf{v}_n(\mathbf{s})\|_2 \rightarrow \mathbf{v}_0(\mathbf{s})$ uniformly in $\mathbf{s}$ as $n \rightarrow \infty$ for some continuous function $\mathbf{v}_0 : S \rightarrow \mathbb{R}$.*

Remark 4. These assumptions will ensure that $\mathbf{w}^T \mathbf{b}(\mathbf{s})$ enjoys a functional CLT as required in Assumptions 1. Let us give some explanation.

- (a) Conditions (a) and (b) ensure the continuity of the relevant fields, such that convergence can be considered in the space of continuous functions on S .
- (b) Condition (c) is, for example, satisfied for autoregressive moving average (ARMA) models, which we will employ in our data application. However, we note that condition (c) could be replaced by an appropriate mixing condition.
- (c) Conditions (d) and (e) are assumptions on the moments and increments of the error field, coupled with conditions on the design matrix $\mathbf{X}$ and the correlation structure $\boldsymbol{\Omega}(\mathbf{s})$. In the simple case of independent noise and a first order linear model with equally spaced design points these conditions decouple and reduce to standard conditions on moments and increments of the noise field (see Remark 6).

Theorem 2 (CLT for general linear model). *Under Assumption 3 (a)-(d) the weak convergence*

$$\sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X}} (\hat{\mathbf{b}}(\mathbf{s}) - \mathbf{b}(\mathbf{s})) \rightarrow G^{\otimes p}(\mathbf{s}),$$

holds as $n \rightarrow \infty$, where $G^{\otimes p}$ is an $\mathbb{R}^p$ -valued mean zero, unit variance Gaussian random field with correlation function $[G^{\otimes p}(\mathbf{s}_1), G^{\otimes p}(\mathbf{s}_2)] = \boldsymbol{\epsilon}(\mathbf{s}_1, \mathbf{s}_2) \mathbf{I}_p$.

If additionally part (e) of Assumptions 3 holds, then we have

$$\|\mathbf{v}_n(\mathbf{s})\|_2^{-1} (\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) - \mathbf{w}^T \mathbf{b}(\mathbf{s})) \rightarrow G(\mathbf{s}),$$

weakly as $n \rightarrow \infty$, where G is a mean zero, unit variance Gaussian field on S with correlation function $\text{cov}[G(\mathbf{s}_1), G(\mathbf{s}_2)] = \boldsymbol{\epsilon}(\mathbf{s}_1, \mathbf{s}_2)$.

The statement of Theorem 2 guarantees that parts (b) and (c) of Assumptions 1 hold. Therefore, we can obtain CoPE sets for $\mathbf{w}^T \mathbf{b}(\mathbf{s})$ as follows.

Corollary 2 (CoPE sets for general linear model). *If the function $\mu(\mathbf{s}) = \mathbf{w}^T \mathbf{b}(\mathbf{s})$ satisfies part (a) of Assumptions 1 and if Assumptions 2 and 3 hold, then*

$$\hat{A}_c^\pm = \left\{ \mathbf{s} : \frac{\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) - c}{\|\mathbf{v}_n(\mathbf{s})\|_2} \geq \pm a \right\} \quad (7)$$

are CoPE sets for $A_c = \{\mathbf{s} \in S : \mathbf{w}^T \mathbf{b}(\mathbf{s}) = c\}$, that is,

$$\lim_{n \rightarrow \infty} P \left[(D_n \cap \hat{A}_c^+) \subset (D_n \cap A_c) \subset (D_n \cap \hat{A}_c^-) \right] = P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \leq a \right].$$

Remark 5. Note that Theorem 2 and hence also Corollary 2 will (by Slutsky's Theorem) continue to hold if $\mathbf{\Omega}(\mathbf{s})$ is replaced by a consistent estimate $\hat{\mathbf{\Omega}}(\mathbf{s})$. This is what one will typically obtain from a GLS estimate with a fixed correlation structure, e.g. an ARMA process with fixed orders but unknown coefficients.

3.2 Approximating the tail probabilities of G

3.2.1 Multiplier bootstrap—In practice, the distribution of G in Corollary 2 is unknown, because it depends on the unknown (non-stationary) covariance function. In our motivating application the only information we have about G is contained in the de-correlated residuals

$$(R_1(\mathbf{s}), \dots, R_n(\mathbf{s})) = \mathbf{R}(\mathbf{s}) = \mathbf{\Omega}^{-1/2}(\mathbf{s}) (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\mathbf{b}(\mathbf{s}))$$

of the linear regression.

A way of approximating the distribution of the limiting Gaussian field G is given by the multiplier or wild bootstrap first introduced by Wu (1986) and later studied by Mammen (1992, 1993), Hurdle and Mammen (1993). Let $g_1, \dots, g_n$ be i.i.d standard Gaussian random variables independent of the data. Then, conditional on the residuals $\{R_j(\mathbf{s})\}_{j=1}^n$, the field

$\tilde{G}(\mathbf{s}) = n^{-\frac{1}{2}} \sum_{j=1}^n g_j R_j(\mathbf{s})$ is Gaussian and has covariance

$$\text{cov} [\tilde{G}(\mathbf{s}_1), \tilde{G}(\mathbf{s}_2) | \mathbf{R}(\mathbf{s})] = \frac{1}{n} \sum_{i,j=1}^n R_i(\mathbf{s}_1) R_j(\mathbf{s}_2) \text{cov} [g_i, g_j] = \frac{1}{n} \sum_{j=1}^n R_j(\mathbf{s}_1) R_j(\mathbf{s}_2), \quad (8)$$

the sample covariance. The idea is to approximate $P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \leq a \right]$, needed in Theorem

1 by $P \left[\sup_{\mathbf{s} \in \partial A_c} |\tilde{G}(\mathbf{s})| \leq a \mid \{R_j(\mathbf{s})\}_{j=1}^n \right]$. The latter can be efficiently computed by

generating a large number M of i.i.d. copies $\tilde{G}_1(\mathbf{s}), \dots, \tilde{G}_M(\mathbf{s})$ of $\tilde{G}(\mathbf{s})$, conditional on

$\{R_j(\mathbf{s})\}_{j=1}^n$ and evaluating $M^{-1} \sum_{j=1}^M \mathbf{1} \left[\sup_{\mathbf{s} \in \partial A_c} |\tilde{G}_j(\mathbf{s})| \leq a \right]$.

The sample covariance (8) itself is not a good estimator of the true covariance function in our high dimensional setting, where the number of locations is much higher than the sample size n (about ten thousand grid points vs. 58 field realizations in the climate data). However, the claim of the approximation is about the distribution of the supremum of the process instead. For a discrete set of locations, the supremum $\sup_{s \in A_c} G(s)$ has the distribution of the maximum of a high-dimensional Gaussian random vector. Extending the results of Mammen (1993) in the high dimensional setting, Chernozhukov et al. (2013) show that the distribution of the maximum can be well approximated by the Gaussian multiplier bootstrap using realizations of a not necessarily Gaussian random vector with the same covariance matrix. In this sense, the multiplier bootstrap is valid in our setting. This is confirmed by simulations in Section 4 below.

3.2.2 An alternative method for smooth Gaussian fields—If the limiting field is twice differentiable and the error field $\epsilon(s)$ in (6) is Gaussian itself then the Gaussian Kinematic Formula (GKF) (e.g. Taylor (2006), Taylor et al. (2005), Taylor and Worsley (2007)) offers an elegant and accurate way of computing tail probabilities of Gaussian fields that does not require simulations.

The GKF is based on two properties of smooth Gaussian fields. First, for such fields the exceedance probability for high thresholds can be approximated by the expected Euler characteristic: for any set $B \subset S$ with smooth boundary,

$$P \left[\sup_{s \in B} G(s) \geq a \right] = E[\chi(A_a(G) \cap B)] + \mathcal{O}(\exp(-a^2/2)), \quad (9)$$

where χ is the Euler characteristic (see e.g. Adler and Taylor (2007)). Second, the expected Euler characteristic in (9) has the closed formula

$$E[\chi(A_a(G) \cap B)] = \sum_{d=0}^N \mathcal{L}_d(B, \Lambda) \rho_d(a), \quad (10)$$

with $\mathcal{L}_d$ the d -th order *Lipschitz-Killing curvature (LKC)* (Taylor, 2006) dependent on the covariance $\Lambda(s)$ of the partial derivatives of G and known functions $\rho_d(a)$. The GKF shows that, at least for smooth fields, the tail probability of the supremum is intrinsically lowdimensional, as it depends only (up to an exponential error term) on $N+1$ parameters. This gives an additional justification for the ability of the multiplier bootstrap method to estimate the tail probability despite the high dimensionality of the field.

We can also use the GKF to obtain CoPE sets: If the Assumptions (1) are satisfied then Theorem (1) implies in conjunction with (9) and (10) that

$$\lim_{n \rightarrow \infty} P \left[\hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \right] = 1 - \sum_{d=0}^{N-1} \mathcal{L}_d(\partial A_c, \Lambda) \rho_d(a) + \mathcal{O}(\exp(-a^2/2)). \quad (11)$$

To use (11), the problem amounts to estimating the LKCs of the boundary A_c . Taylor and Worsley (2007) propose a method to estimate LKCs based on a finite number of realizations of the Gaussian field G and a triangulation of a plug-in estimate $\hat{A}_c$ of the boundary A_c . In our application, however, we only have realizations of a generally non-Gaussian field (the residuals in of the linear model, cf. Section 5) with asymptotically the same covariance as G . For completeness, we compare this method to the multiplier bootstrap method in the simulations section below.

3.3 Algorithm

Combining the results of the previous sections we can give the exact procedure for obtaining CoPE sets for the parameters of the linear model $\mathbf{Y}(\mathbf{s}) = \mathbf{X}\mathbf{b}(\mathbf{s}) + \epsilon(\mathbf{s})$.

Algorithm 1. *Given a design matrix $\mathbf{X}$ and observations $\mathbf{Y}(\mathbf{s})$ following the linear model (6). If the assumptions of Corollary 2 hold, the following yields CoPE sets for $\mathbf{w}^T \mathbf{b}(\mathbf{s})$.*

- (a) *Compute the estimate $\hat{\mathbf{b}}(\mathbf{s}) = (\mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1}(\mathbf{s}) \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1}(\mathbf{s}) \mathbf{Y}(\mathbf{s})$ and the corresponding de-correlated residuals $\mathbf{R}(\mathbf{s}) = \hat{\boldsymbol{\Omega}}^{-1/2}(\mathbf{s}) (\mathbf{Y}(\mathbf{s}) - \mathbf{X} \hat{\mathbf{b}}(\mathbf{s}))$, with the empirical covariance matrix $\hat{\boldsymbol{\Omega}}(\mathbf{s})$ determined according to the model used for $\boldsymbol{\Omega}(\mathbf{s})$.*
- (b) *Determine a such that approximately $P \left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \geq a \right] \leq \alpha$. For example, use the multiplier bootstrap procedure with the residuals $\{R_j(\mathbf{s})\}_{j=1}^n$ to generate i.i.d. copies of a Gaussian $\tilde{G}$ field with covariance structure given by the sample covariance. Then, use these and the plug-in estimate $\partial \hat{A}_c = \{\mathbf{s} : \mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) = c\}$ for the boundary A_c to determine a such that $P \left[\sup_{\mathbf{s} \in \partial \hat{A}_c} |\tilde{G}(\mathbf{s})| \geq a \right] \leq \alpha$.*
- (c) *Obtain the nested CoPE sets defined by*

$$\hat{A}_c^\pm = \left\{ \mathbf{s} : \frac{\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) - c}{\|\hat{\mathbf{v}}_n(\mathbf{s})\|_2} \geq \pm a \right\},$$

$$\text{with } \|\hat{\mathbf{v}}_n(\mathbf{s})\|_2 = \sqrt{\mathbf{w}^T \mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1}(\mathbf{s}) \mathbf{X} \mathbf{w}}.$$

3.4 Model for the climate data

In our application we have a total of $n = n^{(a)} + n^{(b)}$ observations, the first $n^{(a)}$ observations are the 'past', the last $n^{(b)}$ are the 'future'. Within each period we model the change in mean temperature linearly in time. Model (6) takes the form

$$\begin{aligned} Y_j(\mathbf{s}) &= T^{(a)}(\mathbf{s}) + m^{(a)}(\mathbf{s})t_j^{(a)} + \epsilon_j(\mathbf{s}), & j = 1, \dots, n^{(a)} \\ Y_j(\mathbf{s}) &= T^{(b)}(\mathbf{s}) + m^{(b)}(\mathbf{s})t_j^{(b)} + \epsilon_j(\mathbf{s}), & j = n^{(a)} + 1, \dots, n^{(a)} + n^{(b)}. \end{aligned} \quad (12)$$

Without loss of generality, we may assume that $\sum_{j=1}^{n^{(a)}} t_j^{(a)} = 0$ and $\sum_{j=n^{(a)}+1}^{n^{(a)}+n^{(b)}} t_j^{(b)} = 0$. Our goal is to give CoPE sets for the excursion sets of the difference $T^{(b)}(\mathbf{s}) - T^{(a)}(\mathbf{s})$. Therefore, we define $\mathbf{w} = (-1, 1, 0, 0)^T$. The next statement is a direct application of Corollary 2.

Proposition 1. *If the design and the noise field in model (12) satisfy Assumptions 3 then Algorithm 1 yields asymptotically valid CoPE sets for $T^{(b)}(\mathbf{s}) - T^{(a)}(\mathbf{s})$.*

Remark 6. In the case of independent noise fields $\epsilon_1(\mathbf{s}), \dots, \epsilon_n(\mathbf{s}) \sim \epsilon(\mathbf{s})$ we can replace parts (d)-(f) of Assumptions 3 by the following simpler conditions. A proof is given in the appendix.

- There exist numbers $\delta_s > 0$ and $\gamma \geq 0$ and constants $K_\delta > 0$ and $K_{(\beta, \gamma)} > 0$ such that $\sup_{\mathbf{s} \in S} E|\epsilon(\mathbf{s})|^{1+\delta} \leq K_\delta$ and $E\|\epsilon(B)\|_1^{2+\gamma} \leq K_{(\beta, \gamma)}|B|^{1+\beta}$ for every block $B \subset \mathbb{R}^D$.
- $n^{(a)} = n^{(b)} = n/2$ and both sets of design points $t_j^{(a)}$ and $t_j^{(b)}$ are equally spaced (possibly with different spacing for the periods (a) and (b)).

4 Simulations

The simulations in this section show that the proposed methods provide approximately the right coverage in practical non-asymptotic situations with non-smooth, non-stationary and non-Gaussian noise. Our objective in the design of the error fields with these properties is not to imitate the data but to introduce non-stationarity and non-Gaussianity in a transparent and reproducible way, showing the full potential of the method. In fact, the error field that we encounter in the data (cf. Section (5)) is better behaved as far as smoothness and stationarity are concerned than the artificial fields we investigate here.

4.1 Setup

We consider a one-group version of model (12) ($n = n^{(a)}$, $n^{(b)} = 0$) on $t_j = 0, 1/(n-1), 2/(n-1), \dots, 1$, with non-stationary and non-Gaussian noise over a square region $S = [0, 1]^2$ sampled at a uniform 64×64 grid. The signal $\mu = T^{(a)}$ is a linear combination of three Gaussians and the slope field $m(\mathbf{s})$ is linear in the y-coordinate and constant in the x-coordinate (Figure 3a, 3b). We consider IID and AR(1) processes $(\epsilon_1(\mathbf{s}), \dots, \epsilon_n(\mathbf{s}))$ (according to the spatial pattern in Figure 3c) for three different base fields:

Noise 1—Independent standard normal random variables are assigned to each pixel in the left half of S and to blocks of 4-by-4 pixels in the right half (Figure 3d). The entire picture is convolved with a Gaussian kernel with bandwidth 0.1 and all values are multiplied by a scaling factor of 50 (Figure 3e).

Noise 2—Identical to Noise 1 except the image is smoothed by a Laplace kernel with bandwidth 0.1 instead of a Gaussian and the scaling factor is 100 (Figure 3f).

Noise 3—Each pixel in the upper half is assigned a Laplace random variable with mean zero and variance two. In the lower half, pixels are assigned independent Student t -distributed random variables with 10 degrees of freedom. The entire picture is convolved with a Gaussian kernel of bandwidth 0.1 and scaled with a factor of 25 (Figure 3g).

The noise fields Noise 1–3 are intentionally designed to have non-homogeneous variance and scaling factors are chosen ad-hoc such that all three fields can be conveniently displayed on a common scale (Figure 3 d-f).

Our target is the excursion set of the mean function $\mu(\mathbf{s})$ at level $c = 4/3$. Note that the spatially varying AR parameter was not assumed to be known but was rather estimated in each trial from the data. We controlled the probability of coverage at the level $1 - \alpha = 0.9$ using Corollary 2. The threshold a was computed using the multiplier bootstrap procedure proposed in Section 3.2.1 using either the true boundary A_c or the plug-in estimate $\hat{A}_c$. These boundaries were determined as piece-wise linear splines with the R-function `contourLines` where values of fields on non-grid locations are determined via linear interpolation. The results using $\hat{A}_c$ for each one run with the three noise fields and sample sizes $n = 40, 60, 80$ are shown in Figure 4.

4.2 Performance of CoPE sets

We analyzed the performance of our method on 5000 runs of the toy examples in Section 4.1. Figure 5 shows the percentage of trials in which coverage $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$ was achieved, if either the true boundary A_c or the plug-in estimator $\hat{A}_c$ was used to determine the threshold. The empirical coverage is seen to approach the nominal level of 90% as the sample size increases, suggesting asymptotic unbiasedness. Comparing the coverage for the true and estimated boundaries, we see that the non-asymptotic bias is not caused by the lack of knowledge of the true boundary. It may be a consequence of the bootstrap procedure instead.

We conducted several additional experiments with different choices of parameters for the noise fields and obtained results very similar to those reported here.

Computational performance—As already noted in Section 3.2.1, the multiplier bootstrap allows for a very fast computation of CoPE sets. In the simulations, the CoPE sets for a sample of size $n = 120$, each on a grid of $64 \times 64 = 4096$ locations could be computed in less than two minutes on a standard laptop for all noise models. We found that only a negligible fraction of the total runtime of the method is spent on the multiplier bootstrap - with the pointwise generalized least squares estimation taking far longer.

4.3 Estimation of exceedance probabilities

In this Section we compare the multiplier bootstrap with the method proposed by Taylor and Worsley (2007), as described in Section 3.2.2, where $(\mathbf{s})$ is distributed according to Noise 1

(see Section 4.1 above), $\sigma^2(\mathbf{s}) = \text{var}[(\mathbf{s})]$ and A_c is the contour $A_c(\mu)$ of the function $\mu(\mathbf{s})$ shown in Figure 3a at level $c = 4/3$. The true cumulative density function for $\sup_{\mathbf{s} \in \partial A_c} |\epsilon(\mathbf{s})/\sigma(\mathbf{s})|$ and its empirical approximations based on both methods are shown in Figure 6. The empirical cdfs are each based on a single i.i.d. sample $\epsilon_1(\mathbf{s}), \dots, \epsilon_n(\mathbf{s})$ for $n = 10, 30$ and 60 . For the multiplier bootstrap we generated 5,000 bootstrap realizations. The true cdf was calculated empirically using 10,000 i.i.d. samples of $\epsilon(\mathbf{s})$.

Both methods give a remarkably good approximation of the true distribution of the supremum, particularly for sample sizes of $n = 30$ and higher. However, while Taylor's method only gives a valid approximation in the tail of the distribution, the multiplier bootstrap approximates all parts of the cdf.

4.4 Effect of Grid Coarseness

In order to assess the influence of the coarseness of the grid we repeated the experiment of Section 4.1 on grids of sizes 16×16 , 32×32 and 64×64 , in the same setting as in the top left panel of Figure 5, with 500 trials.

The results are shown in Figure 7 (top row). For the lower resolutions 16×16 and 32×32 , coverage is above the nominal level of 90%, but it approaches the nominal level as the resolution increases. This is consistent with parts b) and c) of Theorem 1 stating that the procedure may produce over-coverage if D_n is small but coverage will approach the nominal level for increasing grid size.

In order to check whether over-coverage for lower resolutions is due to the CoPE sets being overly conservative, Figure 7 (bottom row) shows the ratio $\# \hat{A}_c^+ / \# \hat{A}_c^-$ of number of grid points contained in the inner CoPE set over the number of grid points contained in the outer CoPE set. We see that there is hardly any difference between the resolutions. We conclude that CoPE sets do not become more conservative as the resolutions decreases.

5 Application to the climate data

Data processing

Since the climate model used to generate the data does not give meaningful values over water we applied a mask to only include temperature over land. This also has the benefit that the data shows no apparent discontinuities over land, in agreement with Assumptions 1.

For step (a) of Algorithm 1 we used the R-package `nlme` to obtain estimates $\hat{\beta}(\mathbf{s})$ and $\hat{\Omega}(\mathbf{s})$ assuming an AR(1) correlation within past and future time periods and no correlation between the two periods. For step (b) we used the multiplier bootstrap to obtain the threshold a . The entire analysis for either the summer or winter data including the pointwise GLS and the multiplier bootstrap to obtain the CoPE sets was performed in under seven minutes on a regular laptop.

Results

The results for the climate data, shown in Figure 1, correspond to CoPE sets for $T^{(b)}(s) - T^{(a)}(s)$. The target level is $c = 2^\circ\text{C}$ and nominal coverage probability is fixed at $1 - \alpha = 0.9$.

For the mean summer temperature, it may be stated with 90% confidence that the Rocky Mountains and the Sierra Madre Occidental mountains of Mexico are at risk of exhibiting a warming of 2°C or more in the given time period, while the Florida Peninsula, parts of the Mexican Gulf, large parts of the Canadian Northwest and the northern part of the Labrador Peninsula are not at risk.

For the mean winter temperature, some regions around the Hudson Bay and in the Canadian Shield are identified to be at a high risk while a comparatively small region north of the Mexican Gulf is considered not at risk for extreme warming.

In the supplement to this paper we show CoPE sets when the domain is restricted to the contiguous USA. These sets are only marginally different from the ones shown here.

6 Discussion

Discrete Locations and Measurement Errors

Consistent with our data, we have assumed that we have repeated observations of the entire fields contaminated with a continuous noise field and without nugget effect. Our theory for CoPE sets may be extended following Degras (2011) who considered a model of the form

$$Y_{ij} = \mu(x_j) + \epsilon_i(x_j) + \eta_{ij}, \quad 1 \leq i \leq n; \quad 1 \leq j \leq p \quad (13)$$

with locations x_j in the domain of interest, the ϵ_i independent copies of an unknown random process and η_{ij} i.i.d. Gaussian noise (nugget effect). They show that (in dimension 1 or 2) under suitable assumptions on the target function and the model, a local linear estimate $\hat{\mu}$ of μ , depending on a smoothing parameter, satisfies a functional CLT, that is $\sqrt{n}(\hat{\mu} - \mu)$ converges weakly to a Gaussian random field with the same covariance as ϵ . The functional CLT gives the major part of our assumptions 1 needed to obtain CoPE sets. However, it still depends on the design locations x_j becoming dense quickly. It is an interesting question for further research how the threshold a may be obtained in this setting and how it may be affected by the choice of the smoothing parameter.

More Efficient Estimators

In view of techniques from geostatistics, it is natural to ask whether more efficient estimators than pixel-wise regression can be used to estimate the parameter functions in the model (6). There is a trade-off in the feasibility of such estimators. While local pooling may increase the accuracy, optimal bandwidth will vary spatially and depend on the unknown covariance structure of the field. Obtaining an estimator in this fashion is feasible if one makes strong parametric assumptions on the covariance structure, but not so if one only assumes non-stationarity and general moment and increment conditions (as we do here).

A Proofs

Proof of Theorem 1. In part I of the proof, we show that

$$\liminf_{n \rightarrow \infty} P \left[\hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \right] \geq P \left[\sup_{\partial A_c} |G(s)| \leq a \right]. \quad (14)$$

Note that (14) immediately gives the second part of the theorem since

$$P \left[D \cap \hat{A}_c^+ \subset D \cap A_c \subset D \cap \hat{A}_c^- \right] \geq P \left[\hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \right].$$

In part II we show that for a sequence of sets D_n that satisfy Assumption 2 we have

$$\limsup_{n \rightarrow \infty} P \left[D_n \cap \hat{A}_c^+ \subset D_n \cap A_c \subset D_n \cap \hat{A}_c^- \right] \leq P \left[\sup_{s \in \partial A_c} |G(s)| \leq a \right]. \quad (15)$$

This and part b) of the Theorem immediately give part c). Part a) also follows from (14) and (15) since $D_n \equiv S$ trivially satisfies Assumption 2.

Part I

For $\eta > 0$ define the inflated boundary $A_c^\eta = \{s \in S : c - \eta\sigma(s) \leq \mu(s) \leq c + \eta\sigma(s)\}$. Points outside of A_c^η become irrelevant in the limit $n \rightarrow \infty$ since their values are far from c and, if η goes to zero at an appropriate rate, we finally end up with the boundary A_c .

Assume $\eta < \eta_0$ and let $C' = \inf_{s \in U} \sigma(s)$. We first show that

$$\sup_{s \in A_c^\eta} \left| \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \right| \leq a, \quad \sup_{s \in U} \left| \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \right| \leq a + C' \eta \tau_n^{-1} \quad \text{and} \quad \sup_{s \in S} \left| \tau_n^{-1} Z_n(s) \right| \leq a + C \tau_n^{-1}$$

implies that $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$. To see this, assume the above conditions hold and consider $s \in A_c$. If $s \in A_c \cap A_c^\eta$ we have that

$$\frac{\hat{\mu}_n(s) - c}{\tau_n \sigma(s)} \geq \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \geq -a$$

and hence $s \in \hat{A}_c^+$. If $s \in A_c \cap (U \setminus A_c^\eta)$ then

$$\frac{\hat{\mu}_n(s) - c}{\sigma(s) \tau_n} \geq \frac{\hat{\mu}_n(s) - \mu(s)}{\sigma(s) \tau_n} + C' \eta \tau_n^{-1} \geq -a - C' \eta \tau_n^{-1} + C' \eta \tau_n^{-1} = -a$$

and therefore again $\mathbf{s} \in \hat{A}_c^-$. If finally $\mathbf{s} \in A_c \setminus U$ we have

$$\frac{\hat{\mu}_n(\mathbf{s}) - c}{\tau_n \sigma(\mathbf{s})} \geq C\tau_n^{-1} + \tau_n^{-1} Z_n(\mathbf{s}) \geq C\tau_n^{-1} - a - C\tau_n^{-1} = -a$$

and hence $\mathbf{s} \in \hat{A}_c^-$ as well. Therefore we have shown that the above properties imply $A_c \subset \hat{A}_c^-$.

An analogous argument works for the inclusion $\hat{A}_c^+ \subset A_c$.

Now, let $\{\eta_n\}_{n \in \mathbb{N}}$ be a sequence of positive numbers such that $\eta_n \rightarrow 0$ and $\eta_n \tau_n^{-1} \rightarrow \infty$.

With what we showed above, we can write

$$\begin{aligned} P \left[\hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \right] &\geq P \left[\sup_{\substack{\mathbf{s} \in A_c \\ \eta_n}} \left| \frac{\hat{\mu}_n(\mathbf{s}) - \mu(\mathbf{s})}{\tau_n \sigma(\mathbf{s})} \right| < a \right] \\ &+ P \left[\sup_{\mathbf{s} \in U} \left| \frac{\hat{\mu}_n(\mathbf{s}) - \mu(\mathbf{s})}{\tau_n \sigma(\mathbf{s})} \right| \leq a + C' \eta_n \tau_n^{-1} \right] - 1 + P \left[\sup_{\mathbf{s} \in S} \left| \tau_n^{-1} Z_n(\mathbf{s}) \right| < a + C \tau_n^{-1} \right] - 1. \end{aligned} \quad (16)$$

First, note that the last two probabilities in this expression go to one since $\sup_{\mathbf{s} \in U} (\hat{\mu}_n(\mathbf{s}) - \mu(\mathbf{s})) / (\tau_n \sigma(\mathbf{s}))$ and $\sup_{\mathbf{s} \in S} |\tau_n^{-1} Z_n(\mathbf{s})|$ are tight and $\eta_n \tau_n^{-1} \rightarrow \infty$.

To prove convergence of the first term we need the following

Lemma 1. *Under Assumptions 1 part (a) if $\eta_n \rightarrow 0$ then the Hausdorff distance $\delta_n := d_H(A_c^{\eta_n}, \partial A_c) \rightarrow 0$.*

Proof. Define the set $(A_c)_\varepsilon := \{\mathbf{s} \in S : d(\mathbf{s}, A_c) < \varepsilon\}$. We prove the assertion by showing that for any $\varepsilon > 0$ there exists an $\eta > 0$ such that $A_c^\eta \subset (A_c)_\varepsilon$. To this end, assume the contrary. Then, there exists $\varepsilon > 0$ such that for any $\eta > 0$ we find $\mathbf{s}_\eta \in A_c^\eta$ with $d(\mathbf{s}_\eta, A_c) \geq \varepsilon$. The sequence $(\mathbf{s}_\eta)_{\eta > 0}$ is contained in the compact set S and hence has a convergent subsequence with limit $\mathbf{s}^*$, say. By construction and continuity of μ on A_c^η for η small enough, we have $\mathbf{s}^* \in \cap_{\eta > 0} A_c^\eta = \partial A_c$. On the other hand, $0 = d(\mathbf{s}^*, A_c) = \lim_{\eta \rightarrow 0} d(\mathbf{s}_\eta, A_c) \geq \varepsilon$, a contradiction.

Recall that for a function $f: S \rightarrow \mathbb{R}$ and some number $\delta > 0$ the modulus of continuity is defined as

$$w_U(f, \delta) = \sup_{s_1, s_2 \in U, |s_1 - s_2| \leq \delta} |f(s_1) - f(s_2)|. \quad (17)$$

Since $\{\tau_n^{-1}\sigma(s)^{-1}(\hat{\mu}_n(s) - \mu(s))\}_{s \in U, n \in \mathbb{N}}$ is weakly convergent, we have

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P \left[w_U \left(\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}, \delta \right) \geq \zeta \right] = 0 \quad (18)$$

for all positive ζ (Khoshnevisan, 2002, Prop. 2.4.1 and Exc. 3.3.1). Together with Lemma 1 this implies that for n large enough

$$\left| \sup_{s \in A_c} \frac{|\hat{\mu}_n(s) - \mu(s)|}{\tau_n \sigma(s)} - \sup_{s \in \partial A_c} \frac{|\hat{\mu}_n(s) - \mu(s)|}{\tau_n \sigma(s)} \right| \leq w_U \left(\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}, \delta_n \right) \rightarrow 0,$$

in probability. Since $\sup_{s \in \partial A_c} |(\hat{\mu}_n(s) - \mu(s))/(\tau_n \sigma(s))|$ converges in distribution to $\sup_{s \in \partial A_c} |G(s)|$ this yields

$$\sup_{s \in A_c} \frac{|\hat{\mu}_n(s) - \mu(s)|}{\tau_n \sigma(s)} \rightarrow \sup_{s \in \partial A_c} |G(s)|$$

in distribution. In view of (16) this completes the proof of (14).

Part II

To show (15) we first prove that the inclusion $(\hat{A}_c^+ \cap D_n) \subset (A_c \cap D_n)$ does *not* hold if for some $\delta > 0$ we have $\sup_{s \in \partial A_c} (\hat{\mu}_n(s) - c)/(\tau_n \sigma(s)) \geq a + \delta$ and $\nabla_n^-/\tau_n + w_n^- \leq \delta$.

To this end, assume the latter holds and let $s_0 \in A_c$ such that $(\hat{\mu}_n(s_0) - c)/(\tau_n \sigma(s_0)) \geq a + \delta$. Then, with (17) and the abbreviation

$$w_n^\pm := w_U \left(\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}, \Delta_n^\pm \right),$$

we have for large enough n for the projection $\mathbf{p}_n^-(s_0)$ of s_0 that

$$\begin{aligned}
\frac{\hat{\mu}_n(\mathbf{p}_n^-(s_0)) - c}{\tau_n \sigma(\mathbf{p}_n^-(s_0))} &= \frac{\mu(\mathbf{p}_n^-(s_0)) - c}{\tau_n \sigma(\mathbf{p}_n^-(s_0))} + \frac{\hat{\mu}_n(\mathbf{p}_n^-(s_0)) - \mu(\mathbf{p}_n^-(s_0))}{\tau_n \sigma(\mathbf{p}_n^-(s_0))} \\
&\geq \frac{\mu(\mathbf{p}_n^-(s_0)) - c}{\tau_n \sigma(\mathbf{p}_n^-(s_0))} + \frac{\hat{\mu}_n(s_0) - \mu(s_0)}{\tau_n \sigma(s_0)} - w_n^- \\
&\geq a + \delta - \nabla_n^- / \tau_n - w_n^- \geq a.
\end{aligned}$$

This shows $\mathbf{p}_n^-(s_0) \in \hat{A}_c^+$. But $\mathbf{p}_n^-(s_0) \in (S \setminus A_c) \cap D_n$ by definition and hence $(\hat{A}_c^+ \cap D_n) \subset (A_c \cap D_n)$ does not hold.

Since an analogous calculation works for the inclusion $(A_c \cap D_n) \subset (\hat{A}_c^- \cap D_n)$ we obtain the bound

$$P[D_n \cap \hat{A}_c^+ \subset D_n \cap A_c \subset D_n \cap \hat{A}_c^-] \leq P\left[\sup_{\mathbf{s} \in \partial A_c} \left| \frac{\hat{\mu}_n(\mathbf{s}) - \mu(\mathbf{s})}{\tau_n \sigma(\mathbf{s})} \right| < a + \delta\right] + P[\nabla_n^\pm / \tau_n + w_n^\pm > \delta].$$

By assumptions 1 and because $\nabla_n^\pm \rightarrow 0$ and $w_n^\pm \rightarrow 0$ in probability by (18) the right hand side of this bound converges to $P\left[\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})| \leq a + \delta\right]$. Since $\delta > 0$ was arbitrary and $\sup_{\mathbf{s} \in \partial A_c} |G(\mathbf{s})|$ has a continuous distribution the claim follows.

Proof of Corollary 1. For any pair of nested sets $\hat{A}_c^\pm$, we have that $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$ implies $\partial A_c \subset \text{cl}(\hat{A}_c^- \setminus \hat{A}_c^+)$. On the other hand, the latter will certainly fail to hold if $\sup_{\mathbf{s} \in \partial A_c} |\hat{\mu}_n(\mathbf{s}) - c| / |\tau_n \sigma(\mathbf{s})| > a$. Combining these two observations yields

$$P[\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-] \leq P[\partial A_c \subset \text{cl}(\hat{A}_c^- \setminus \hat{A}_c^+)] \leq P\left[\sup_{\mathbf{s} \in \partial A_c} \left| \frac{\hat{\mu}_n(\mathbf{s}) - c}{\tau_n \sigma(\mathbf{s})} \right| \leq a\right].$$

Taking the limit $n \rightarrow \infty$ of this inequality and using Theorem 1 gives the assertion.

Proof of Theorem 2. Define $\mathbf{A} = \mathbf{X}^T \mathbf{\Omega}^{-1}(\mathbf{s}) \mathbf{X}^{-1} \mathbf{X}^T \mathbf{\Omega}^{-1/2}(\mathbf{s})$, giving $\hat{\mathbf{b}}(\mathbf{s}) - \mathbf{b}(\mathbf{s}) = \mathbf{A} \tilde{\epsilon}(\mathbf{s})$. To prove weak convergence of the process we first show convergence of the finite dimensional distributions and then tightness of the sequence (Khoshnevisan, 2002, Prop. 3.3.1).

For the former, let $\mathbf{s}_1, \dots, \mathbf{s}_K \in S$ be arbitrary. We need to show that with

$$\begin{aligned}
U &= \left(\sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X}} (\hat{\mathbf{b}}(\mathbf{s}_1) - \mathbf{b}(\mathbf{s}_1)), \dots, \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X}} (\hat{\mathbf{b}}(\mathbf{s}_K) - \mathbf{b}(\mathbf{s}_K)) \right) \\
&= \text{diag} \left[\sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_1) \mathbf{X} \mathbf{A}}, \dots, \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_K) \mathbf{X} \mathbf{A}} \right] (\tilde{\mathbf{e}}(\mathbf{s}_1), \dots, \tilde{\mathbf{e}}(\mathbf{s}_K)),
\end{aligned}$$

(here, ‘ $\otimes$ ’ denotes the Kronecker product of two matrices) we have convergence $U \rightarrow \mathcal{N}(0, [\mathbf{c}(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1}^K \otimes \mathbf{I}_p)$ in distribution. We readily see that $E[U] = 0$ and for the covariance we compute

$$\begin{aligned}
\text{cov}[U] &= \text{diag} \left[\sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_1) \mathbf{X} \mathbf{A}}, \dots, \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_K) \mathbf{X} \mathbf{A}} \right] E \left\{ (\tilde{\mathbf{e}}(\mathbf{s}_1), \dots, \tilde{\mathbf{e}}(\mathbf{s}_K)) (\tilde{\mathbf{e}}(\mathbf{s}_1), \dots, \tilde{\mathbf{e}}(\mathbf{s}_K))^T \right\} \\
&\times \text{diag} \left[\mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_1) \mathbf{X}}, \dots, \mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_K) \mathbf{X}} \right] \\
&= \left\{ \text{diag} \left[\sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_1) \mathbf{X} \mathbf{A}}, \dots, \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_K) \mathbf{X} \mathbf{A}} \right] \right\} \left\{ [\mathbf{c}(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1}^K \otimes \mathbf{I}_n \right\} \\
&\times \left\{ \text{diag} \left[\mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_1) \mathbf{X}}, \dots, \mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_K) \mathbf{X}} \right] \right\} = [\mathbf{c}(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1}^K \otimes \mathbf{I}_p.
\end{aligned}$$

We employ the Cramér-Wold device to show convergence of U . Indeed, let and $(a_1, \dots, a_K) \in \mathbb{R}^{K \times p}$ be some fixed arbitrary vector and compute

$$\begin{aligned}
\langle U, \alpha \rangle &= \left\langle (\tilde{\mathbf{e}}(\mathbf{s}_1), \dots, \tilde{\mathbf{e}}(\mathbf{s}_K)), \text{diag} \left[\mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_1) \mathbf{X}}, \dots, \mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_K) \mathbf{X}} \right] \alpha \right\rangle \\
&= \sum_{i=1}^K \langle \tilde{\mathbf{e}}(\mathbf{s}_i), \mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_i) \mathbf{X}} \alpha_i \rangle = \sum_{i=1}^K \sum_{j=1}^n \tilde{e}_j(\mathbf{s}_i) \left(\mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_i) \mathbf{X}} \alpha_i \right)_j.
\end{aligned}$$

By interchanging the sums and defining $W_j = \sum_{i=1}^K \tilde{e}_j(\mathbf{s}_i) \left(\mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_i) \mathbf{X}} \alpha_i \right)_j$, we have by Assumption 3 part (c) managed to write $\langle U, \alpha \rangle$ as a sum of independent random variables $\langle U, \alpha \rangle = \sum_j W_j$. The goal is now to use the CLT in the form of Lyapunov for the random variables W_j . To this end compute $\text{var} \left[\sum_{j=1}^n W_j \right] = \text{var}[\langle U, \alpha \rangle] = \alpha^T \text{cov}[U] \alpha$ and note that since we have already showed U to have the right covariance the claimed convergence will follow once we establish the Lyapunov condition. For this purpose let δ be as in Assumption 3 to give

$$\begin{aligned}
\sum_{j=1}^n E|W_j|^2 + \delta &\leq K^2 \sum_{j=1}^n \sum_{i=1}^K E|\tilde{e}_j(\mathbf{s}_i)|^2 + \delta \left\| \mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_i) \mathbf{X}} \alpha_i \right\|^2 + \delta \\
&\leq K^2 K_\delta \sum_{j=1}^n \sum_{i=1}^K \left\| \mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_i) \mathbf{X}} \alpha_i \right\|_\infty^2 + \delta \\
&\leq K^3 K_\delta \|\alpha\|_\infty^n \left\| \mathbf{A}^T \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}_i) \mathbf{X}} \right\|_\infty^2 + \delta \rightarrow 0,
\end{aligned}$$

as $n \rightarrow \infty$. This concludes the proof of convergence for the finite dimensional distributions.

Tightness of the sequence $\{\sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X}} \mathbf{A} \tilde{\epsilon}(\mathbf{s})\}$ follows from Assumption 3 (e) and (Bickel and Wichura, 1971, Thm. 3).

The second statement follows from the fact that

$$\begin{aligned} \|\mathbf{v}_n(\mathbf{s})\|_2^{-1} (\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) - \mathbf{w}^T \mathbf{b}(\mathbf{s})) &= \|\mathbf{v}_n(\mathbf{s})\|_2^{-1} \mathbf{w}^T (\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X})^{-1/2} \sqrt{\mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{s}) \mathbf{X}} (\hat{\mathbf{b}}(\mathbf{s}) - \mathbf{b}(\mathbf{s})) \\ &\rightarrow \mathbf{v}_0(\mathbf{s})^T G \otimes P(\mathbf{s}) \stackrel{\mathcal{D}}{=} G(\mathbf{s}), \end{aligned}$$

where the weak convergence from the previous part is used.

Proof of Corollary 2. This is an application of Theorem 1. Part (a) of Assumption 3 is tantamount to part (a) of Assumption 1. Theorem 2 gives part (b) of Assumption 1. Part (c) of the Assumptions follows from

$$\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) - c = (\mathbf{w}^T \mathbf{b}(\mathbf{s}) - c) + (\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) - \mathbf{w}^T \mathbf{b}(\mathbf{s}))$$

where the supremum of the absolute value of the second summand is $\mathcal{O}_P(\|\mathbf{v}_0(\cdot)\|_\infty)$ by part (b) of the Theorem. To conclude the proof, note that $\mathbf{w}^T \mathbf{b}(\mathbf{s}) - c \geq \eta \operatorname{sgn}(\mathbf{w}^T \mathbf{b}(\mathbf{s}) - c)$ for $\mathbf{s} \notin A_c^\eta(\mathbf{w}^T \mathbf{b})$.

Proof of Remark 6. We compute

$$\mathbf{X}^T \mathbf{X} = \operatorname{diag}\left(n/2, n/2, \omega^{(a)}, \omega^{(b)}\right), \quad \omega^{(a)} = \sum_{j=1}^{n^{(a)}} \left(t_j^{(a)}\right)^2, \quad \omega^{(b)} = \sum_{j=n^{(a)}+1}^{n^{(a)}+n^{(b)}} \left(t_j^{(b)}\right)^2.$$

It follows that

$$(\mathbf{X}^T \mathbf{X})^{-1/2} = \operatorname{diag}\left(\sqrt{2/n}, \sqrt{2/n}, 1/\sqrt{\omega^{(a)}}, 1/\sqrt{\omega^{(b)}}\right).$$

With this we obtain

$$\|\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1/2}\|_\infty \leq \frac{2}{\sqrt{n}} + \frac{\max_{1 \leq j \leq n/2} |t_j^{(a)}|}{\sqrt{\omega^{(a)}}} + \frac{\max_{n/2+1 \leq j \leq n} |t_j^{(b)}|}{\sqrt{\omega^{(b)}}}.$$

Now we note that since the design points $t_j^{(a)}$ and $t_j^{(b)}$ are equally spaced by assumption we have $\max_{1 \leq j \leq n} |t_j^{(a)}| = \mathcal{O}(n)$ and $\omega^{(a)} = \mathcal{O}(n^3)$, and the same is true for the (b) -counterparts. This shows that $\|\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1/2}\|_\infty = \mathcal{O}(n^{-1/2})$ and therefore Assumptions 3 d) -f) are valid.

Moreover, in the notation of Theorem 2, we have $\mathbf{v}_n = \sqrt{\frac{2}{n}}(1, -1, 0, 0)$ so that $\mathbf{v}_n / \|\mathbf{v}_n\|_2^{-1} = (-1, 1, 0, 0)^T$. This finally gives $\frac{\sqrt{n}}{2}\sigma(\mathbf{s})^{-1}(\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{s}) - \mathbf{w}^T \mathbf{b}(\mathbf{s})) \rightarrow G(\mathbf{s})$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Adler RJ and Taylor JE (2007). Random fields and geometry. Springer, New York.
- Anderson K and Bows A (2011). Beyond ‘dangerous’ climate change: emission scenarios for a new world. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1934):20–44.
- Bickel PJ and Wichura MJ (1971). Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *The Annals of Mathematical Statistics*, 42(5):1656–1670.
- Bolin D and Lindgren F (2014). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages n/a–n/a.
- Bukovsky MS (2012). Temperature Trends in the NARCCAP Regional Climate Models. *Journal of Climate*, 25(11):3985–3991.
- Cavalier L (1997). Nonparametric Estimation of Regression Level Sets. *Statistics*, 29(2):131–160.
- Chernozhukov V, Chetverikov D, and Kato K (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6): 2786–2819.
- Cuevas A, González-Manteiga W, and Rodríguez-Casal A (2006). Plug-in Estimation of General Level Sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19.
- Davison AC and Smith RL (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442.
- De Oliveira V and Ecker MD (2002). Bayesian hot spot detection in the presence of a spatial trend: application to total nitrogen concentration in Chesapeake Bay. *Environmetrics*, 13(1):85–101.
- Degras DA (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4).
- Duczmal L, Kulldorff M, and Huang L (2006). Evaluation of Spatial Scan Statistics for Irregularly Shaped Clusters. *Journal of Computational and Graphical Statistics*, 15(2):428–442.
- Eicker F (1963). Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions. *The Annals of Mathematical Statistics*, 34(2):447–456.
- Flato GM (2005). The third generation coupled global climate model (CGCM3). Available on line at <http://www.cccma.bc.ec.gc.ca/models/cgcm3.shtml>.
- French JP (2014). Confidence regions for the level curves of spatial data. *Environmetrics*, 25(7):498–512.
- French JP and Sain SR (2013). Spatio-temporal exceedance locations and confidence regions. *The Annals of Applied Statistics*, 7(3):1421–1449.
- Genovese CR, Lazar NA, and Nichols T (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, 15(4):870–878. [PubMed: 11906227]
- Hardle W and Mammen E (1993). Comparing Nonparametric Versus Parametric Re-gression Fits. *The Annals of Statistics*, 21(4):1926–1947.
- Khoshnevisan D (2002). Multiparameter Processes: an introduction to random fields. Springer.
- Lahiri SN, Kaiser MS, Cressie N, and Hsu N-J (1999). Prediction of Spatial Cumulative Distribution Functions Using Subsampling. *Journal of the American Statistical Association*, 94(445):86–97.
- Lindgren G and Rychlik I (1995). How reliable are contour curves? Confidence sets for level contours. *Bernoulli*, 1(4):301–319.

- Majure JJ, Cook D, Cressie N, Kaiser MS, Lahiri SN, and Symanzik J (1995). Spatial CDF Estimation and Visualization with Applications to Forest Healthmonitoring. Iowa State University. Department of Statistics. Statistical Laboratory.
- Mammen E (1992). Bootstrap, wild bootstrap, and asymptotic normality. *Probability Theory and Related Fields*, 93(4):439–455.
- Mammen E (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *The Annals of Statistics*, 21(1):255–285.
- Mammen E and Polonik W (2013). Confidence regions for level sets. *Journal of Multi-variate Analysis*, 122:202–214.
- Mason DM and Polonik W (2009). Asymptotic normality of plug-in level set estimates. *The Annals of Applied Probability*, 19(3):1108–1142.
- Mearns LO, Arritt R, Biner S, Bukovsky MS, McGinnis S, Sain S, Caya D, Correia J, Flory D, Gutowski W, Takle ES, Jones R, Leung R, Moufouma-Okia W, McDaniel L, Nunes AMB, Qian Y, Roads J, Sloan L, and Snyder M (2012). The North American Regional Climate Change Assessment Program: Overview of Phase I Results. *Bulletin of the American Meteorological Society*, 93(9):1337–1362.
- Mearns LO, Gutowski W, Jones R, Leung R, McGinnis S, Nunes A, and Qian Y (2009). A Regional Climate Change Assessment Program for North America. *Eos, Transactions American Geophysical Union*, 90(36):311–311.
- Mearns LO, Sain S, Leung LR, Bukovsky MS, McGinnis S, Biner S, Caya D, Arritt RW, Gutowski W, Takle E, Snyder M, Jones RG, Nunes AMB, Tucker S, Herzmann D, McDaniel L, and Sloan L (2013). Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Climatic Change*, 120(4):965–975.
- Michalakes J, Dudhia J, Gill D, Henderson T, Klemp J, Skamarock W, and Wang W (2004). The weather research and forecast model: software architecture and performance. In *Proceedings of the 11th ECMWF Workshop on the Use of High Performance Computing In Meteorology*, volume 25, page 29 World Scientific.
- Patil GP and Taillie C (2004). Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics*, 11(2):183–197.
- R Core Team (2015). R: A language and environment for statistical computing.
- Rao CR and Toutenburg H (1995). *Linear models*. Springer.
- Rigollet P and Vert R (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15(4):1154–1178.
- Rogelj J, Hare B, Nabel J, Macey K, Schaeffer M, Markmann K, and Meinshausen M (2009). Halfway to Copenhagen, no way to 2 °C. *Nature Reports Climate Change*, (0907):81–83.
- Schwartzman A, Dougherty RF, and Taylor JE (2010). Group Comparison of Eigenvalues and Eigenvectors of Diffusion Tensors. *Journal of the American Statistical Association*, 105(490):588–599.
- Singh A, Scott C, and Nowak R (2009). Adaptive Hausdorff estimation of density level sets. *The Annals of Statistics*, 37(5B):2760–2782.
- Sommerfeld M (2015). R-Package cope: Coverage Probability Excursion (CoPE) Sets.
- Tango T and Takahashi K (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, 4:11. [PubMed: 15904524]
- Taylor J, Takemura A, and Adler RJ (2005). Validity of the expected Euler characteristic heuristic. *The Annals of Probability*, 33(4):1362–1396.
- Taylor JE (2006). A Gaussian kinematic formula. *The Annals of Probability*, 34(1):122–158.
- Taylor JE and Worsley KJ (2007). Detecting Sparse Signals in Random Fields, with an Application to Brain Mapping. *Journal of the American Statistical Association*, 102(479):913–928.
- Tsybakov AB (1997). On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969.
- van der Linden P and Mitchell JF, editors (2009). *ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project*. Met Office Hadley Centre, Exeter, UK.

- Wameling A (2003). Accuracy of geostatistical prediction of yearly precipitation in Lower Saxony. *Environmetrics*, 14(7):699–709.
- Willett R and Nowak R (2007). Minimax Optimal Level-Set Estimation. *IEEE Transactions on Image Processing*, 16(12):2965–2979. [PubMed: 18092596]
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC, and others (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73. [PubMed: 20408186]
- Wright DL, Stern HS, and Cressie N (2003). Loss functions for estimation of extrema with an application to disease mapping. *Canadian Journal of Statistics*, 31(3):251–266.
- Wu CFJ (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*, 14(4):1261–1295.
- Zhang J, Craigmile PF, and Cressie N (2008). Loss Function Approaches to Predict a Spatial Quantile and Its Exceedance Region. *Technometrics*, 50(2):216–227.
- Zhu J, Lahiri SN, and Cressie N (2002). Asymptotic inference for spatial CDFs over time. *Statistica Sinica*, 12(3):843–861.

Figure 1:

Output of our method for the increase of the mean summer (June-August) and winter (December-February) temperatures. Shown are heat maps of the estimate of the difference. The uncertainty in the excursion set estimate $\hat{A}_c$ (purple boundary) for $c = 2^\circ\text{C}$ is captured by the CoPE sets $\hat{A}_c^+$ (red boundary) and $\hat{A}_c^-$ (green boundary). The threshold was obtained according to Theorem 1 to guarantee inclusion $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$ with confidence $1 - \alpha = 0.9$. The horizontal and vertical axes are indexed in degrees longitude and latitude, respectively.

Figure 2:

A simple example of nested sets $\hat{A}_c^\pm$ that bound the contour A_c but do not satisfy the inclusion $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$

Figure 3:

The spatial signal, temporal slope, temporal AR coefficient, and one realization of each of the noise fields in the toy example.

Figure 4:

Single realizations of the output of our method at confidence level 90% for the three noise fields (rows) and for sample sizes $n = 40, 60, 80$ (columns) with the target function $\mu(\mathbf{s})$ from Figure 3a. In all pictures we show a heat map of the estimator $\hat{\mu}_n(\mathbf{s})$, the boundary of $A_c(\mu)$ (solid purple) and $A_c(\hat{\mu})$ (dashed purple) and the boundaries of $\hat{A}_c^+$ (red) and $\hat{A}_c^-$ (green). The threshold a was obtained using the multiplier bootstrap and the plug-in estimate of the boundary A_c .

Figure 5:

Percentage of trials in which $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$. The nominal coverage probability is 90%.

Figure 6:

The probability $P \left[\sup_{\mathbf{s} \in \partial A_c} |\epsilon(\mathbf{s})/\sigma(\mathbf{s})| \leq a \right]$ (the horizontal axis shows a) and approximations of it via the multiplier bootstrap or Taylor's method based on a sample of size n . Here, the error field $\epsilon(\mathbf{s})$ has the distribution described in Noise 1 and $A_c = A_c(\mu)$ for the function μ shown in Figure 3a.

Figure 7:
Empirical coverage for CoPE sets (top row) and relative size of inner and outer CoPE sets for different grid sizes (bottom row).