Confidence regions for excursion sets in asymptotically Gaussian random fields, with an application to climate

Max Sommerfeld\textsuperscript{1}, Stephen Sain\textsuperscript{2}, Armin Schwartzman\textsuperscript{3}

\textsuperscript{1}FBMS, Universität Göttingen
\textsuperscript{2}The Climate Corporation
\textsuperscript{3}Department of Statistics, North Carolina State University

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.

Keywords: coverage probability, exceedance regions, general linear model, level sets

1 Introduction

Our motivation comes from the following problem. Faced with a global change in temperature over the globe 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 in geographic longitude and latitude apart over North America over two time periods: late 20th century (1971-1999) and mid 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?

Unlike the usual data setup of spatial statistics, the above data 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)). The methods in this paper are inspired by that kind of analysis.

In general, suppose that we observe $n$ random fields $Y_i(s), i = 1, \ldots, n$, over a spatial domain $S$, modeled as realizations of a general linear model indexed by $s \in S$. The target function $\mu : S \to \mathbb{R}$ could be one of the parameters in the model indexed by $s$, in our case the mean difference temperature field. With a proper design, fitting the linear model at each location $s$ will produce a consistent and asymptotically Gaussian estimator $\hat{\mu}_n : S \to \mathbb{R}$ as $n$ increases. Asymptotically Gaussian estimators indexed by $s$ also appear in nonparametric density estimation and regression. In those settings $n$ would be the number of sample points.
Let \( A_c \) be the excursion set of \( \mu \) above a fixed threshold \( c \), defined as \( A_c := A_c(\mu) := \{ s \in S : \mu(s) \geq c \} \), and denote the analog for \( \hat{\mu} \) by \( \hat{A}_c := A_c(\hat{\mu}_n) \). We wish to obtain confidence regions \( \hat{A}_c^\pm \) that are nested in the sense that \( \hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \) and for which the probability that

\[
\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-
\]

holds is asymptotically above a desired level, say 90\%. The sets \( \hat{A}_c^\pm \) here are obtained as excursion sets of the standardized observed field \( \hat{\mu} \) and we call them Coverage Probability Excursion (CoPE) sets. Assuming that the estimated field \( \hat{\mu} \) satisfies a central limit theorem (CLT), we show that the probability that \( \hat{A}_c^\pm \) holds is given asymptotically by the distribution of the supremum of the limiting Gaussian random field on the boundary \( \partial A_c \) of the true excursion set. Using a plug-in estimate for the unknown boundary, 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 (not necessarily stationary) correlation function of the limiting field. The validity of this procedure for very high-dimensional data has recently been shown by Chernozhukov et al. (2013).

For illustration, Figure 1 shows CoPE sets for the temperature data. The regions within the red boundary \( (\hat{A}_c^+)^\star \) have the highest confidence of being at risk, while the regions outside the green boundary \( (\hat{A}_c^-)^\star \) have the highest confidence of not being at risk. Over repeated sampling, there is a probability of about 90\% that the regions at risk include those within the red boundary and exclude those outside the green boundary.

The problem of finding confidence sets for spatial excursion sets, sometimes also called exceedence regions or level sets, has been studied in the past in two major contexts that substantially differ from the problem under consideration here. In the geostatistics literature, the target function is itself a Gaussian field. In consequence, the excursion and the contour sets are random themselves. The data in this setting is a partial realization of the field, that is, the values of a realization of the field at relatively few spatial locations. This severe limitation of available information is compensated by assuming that the covariance structure of the field is known. This problem has been addressed from a frequentist perspective in terms of confidence regions for level contours (Lindgren and Rychlik 1995, Wameling 2003, French 2014) and for excursion sets (French and Sain 2013). Incidentally, our techniques share some similarities with French (2014), although we will show that distinguishing between level contours and excursion sets is important. In a Bayesian setting for latent Gaussian models, Bolin and Lindgren address uncertainty in both, contours and excursion sets.

The second setting in which the problem has received attention is non-parametric density estimation and regression. Here, the target function is a probability density or regression function, estimated from realizations of a random variable with values in \( \mathbb{R}^q \) for some \( q \). 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 sets analogous to our CoPE sets from vector-valued samples.

The problem of finding the threshold for our CoPE sets involves computation of the tail probability of the supremum of a limiting Gaussian random field. In French (2014) this computation was done by Monte Carlo simulation assuming that the covariance structure of the field is known. More generally for unknown covariance function, as we attempt here, this problem was solved elegantly 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.

All computations in this paper were performed using R (R Core Team 2014). All required functions for computation and visualization of CoPE sets and in particular an implementation of the Algorithm 1 are available in the R-package cope (Sommerfeld 2015).

**Outline of the paper**

In Section 2, we propose a thresholding scheme to obtain CoPE sets \( \hat{A}_c^\pm \) as in (1) from an estimator \( \hat{\mu} \) of \( \mu \), only requiring continuity of \( \mu \) and, most importantly, that \( \hat{\mu} \) is asymptotically Gaussian. We show that the
asymptotic coverage probability is equal to the tail probability of the limiting Gaussian field on the boundary \( \partial A_c \) of the excursion set \( A_c \).

Section 3 is devoted to presenting results and algorithms for the construction of CoPE sets when the target function is the parameter function in a general linear model. First, in Section 3.1 we derive central limit theorems for these quantities. Then, in Section 3.2 we show how to obtain the threshold for the construction of CoPE sets by an efficient multiplier bootstrap. We compare it with a method for Gaussian smooth noise based on Taylor and Worsley (2007). Section 3.3 combines the previous results in a concise algorithm for the construction of CoPE sets.

Section 4 shows a toy example to investigate the non-asymptotic performance of CoPE sets and finally, in Section 5, we apply our methods for a general linear model to the climate data. All proofs are in the appendix.

2 Error control for excursion sets - CoPE sets

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. We call the topological boundary \( \partial A_c \) of the excursion set \( A_c \), the contour of \( \mu \) at the level \( c \).

Assumption 1. We assume that

- (a) the target function \( \mu \) is continuous and the level set \( \{ s : \mu(s) = c \} \) is equal to \( \partial A_c \).
- (b) the estimator \( \hat{\mu}_n(s) \) is continuous in \( s \) (for all \( n \in \mathbb{N} \)).
- (c) there is a sequence of numbers \( \tau_n \) and a continuous function \( \sigma : S \to \mathbb{R}^+ \) such that

\[
\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \to G(s),
\]
We will obtain nested estimates by thresholding the surface \( \hat{\mu}_n(s) \) as follows:

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

where \( a^+ \) and \( a^- \) are appropriate non-negative constants to be determined. Note that in this notation \( A_c = \hat{A}_c(0) \) 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 (3) are in fact nested. 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 (3) can be chosen such that \( \hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \) with a predefined probability.

**Theorem 1.** If the Assumptions \( \square \) hold, then

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

A direct consequence of Theorem \( \square \) 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 \( P \left[ \sup_{s \in \partial A_c} |G(s)| \geq a \right] \leq \alpha \). The determination of \( a \) poses a computational challenge since the distribution of the supremum of \( |G(s)| \) and the set \( \partial A_c \) are unknown. In Section \( 3.2 \) we propose an easy and fast way to approximate this distribution by a multiplier bootstrap.

As mentioned in the Introduction, confidence sets for the excursion set \( A_c \) yield confidence sets for the contour \( \partial A_c \). More precisely, we have the following

**Corollary 1.** Under the assumptions of Theorem \( \square \), we have

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

where \( \text{cl} \) denotes the topological closure.

Note that, conversely, confidence sets for the contour do not automatically give confidence sets 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 \( \square \). The key is to divide the region \( S \) into a close-range zone where \( \mu(s) \) is close to \( c \) and a long-range zone. More precisely, the close-range zone is given by the inflated boundary \( \hat{A}_c^0 = \{ s \in S : c - \eta \sigma(s) \leq \mu(s) \leq c + \eta \sigma(s) \} \). Then, the strategy of the proof is to let the parameter \( \eta \) go to zero at an appropriate rate as \( n \to \infty \) such that, eventually, the probability of a part of \( \hat{A}_c^+ \) falsely appearing in the long-range zone \( S \setminus \hat{A}_c^0 \) (as shown in Figure \( 2 \)) vanishes. The probability of making an error remains in the close-range zone, and is asymptotically given by \( P \left[ \sup_{\partial A_c} |G(s)| > a \right] \).

We want to emphasize that Theorem 1 and its Corollary are valid for any estimator \( \hat{\mu}_n \) satisfying Assumption 1. Thus, they hold generally whether the estimator is 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). However, for concreteness, we focus on the former situation, which we develop in detail in the following section.
3 CoPE sets for general linear models

For concreteness and 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 begin by proving an analog of Theorem 1 for the parameters of a general linear model. The most difficult part is to establish a Central Limit Theorem as in (2). This will require conditions on the error field as well as on the design. We consider the model

\[
Y(s) = Xb(s) + \epsilon(s), \quad s \in S \subset \mathbb{R}^N
\]

(4)

where \( Y(s) \) is a 1 \( \times \) 1 vector of observations, \( X \) is a known \( n \times p \) design matrix, \( b(s) = (b_1(s), \ldots, b_p(s)) \) is an unknown \( p \times 1 \) vector of parameters and \( \epsilon(s) = (\epsilon_1(s), \ldots, \epsilon_n(s)) \) with \( \epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} \) \( \epsilon \) an unknown stochastic process. Results of the kind presented here are well-known (see e.g. Eicker (1963)). We show and prove versions tailored for our specific purpose for coherence and convenience.

The least squares regression estimator for \( b(s) \) in the model (4) is

\[
\hat{b}(s) = (X^T X)^{-1} X^T Y(s).
\]

In the notation of Section (2), the target function \( \mu \) is now one of the parameter functions of the model (4), \( b_1 \), say. Of course, the choice of \( b_1 \) is arbitrary and any other coefficient of \( b \) may be considered, with the obvious modifications of the assumptions and theorems. Naturally, \( b_1(s) \) now plays the role of the estimator \( \hat{\mu}_n(s) \).

Further, define \( \sigma : S \rightarrow \mathbb{R}_{\geq 0} \) via \( \sigma^2(s) = \text{var}[\epsilon(s)] \) and the correlation function \( \varphi : S \times S \rightarrow (-1, 1) \) by

\[
\varphi(s_1, s_2) = \frac{\text{cov}[\epsilon(s_1), \epsilon(s_2)]}{\sigma(s_1)\sigma(s_2)}.
\]

Recall that for \( 1 \leq p \leq \infty \), 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 \( x \). In the special case \( p = \infty \) the matrix norm \( ||A||_\infty \) is the maximum absolute row sum of \( A \), i.e.

\[
||A||_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|.
\]

Definition 1. (a) For vectors \( s, t \in \mathbb{R}^N \) define the block \( [s, t] = (s_1, t_1) \times \cdots \times (s_N, t_N) \subset \mathbb{R}^N \) and for a stochastic process \( \epsilon(s) \) with index set containing \( [s, t] \) define the increment of \( \epsilon(s) \) around \( [s, t] \) (cf. Bickel and Wichura (1971)) as

\[
\epsilon([s, t]) = \sum_{\kappa_1=0,1} \cdots \sum_{\kappa_N=0,1} (-1)^{N-\sum_{i=1}^{N} \kappa_i} \epsilon(s_1 + \kappa_1(t_1 - s_1), \ldots, s_N + \kappa_N(t_N - s_N)).
\]

(b) We denote the Lebesgue measure of a set \( A \subset S \) by \( |A| \). For non-negative numbers \( \delta, \gamma, \beta \) we say that the error field \( \epsilon(s) \) has the properties

- N1-\( \delta \), if \( \sup_{s \in S} \sigma(s)^{-2+\delta} E||\epsilon(s)||^{2+\delta} < \infty; \)
- N2-(\( \gamma, \beta \)), if there exists a constant \( C > 0 \) such that \( E[(\sigma^{-1}\epsilon)(B)]^{2+\gamma} \leq C |B|^{1+\beta} \) for all blocks \( B \subset S \).

In dimension \( N = 1 \) the definition of an increment yields \( \epsilon([s_1, t_1]) = \epsilon(t_1) - \epsilon(s_1) \), the usual increment. In Dimension \( N = 2 \) we get \( \epsilon([s, t]) = \epsilon(t_1, t_2) - \epsilon(s_1, t_2) - \epsilon(t_1, s_2) + \epsilon(s_1, s_2) \).

Assumption 2. Assume that
(a) the parameter functions \( b(s) = (b_1(s), \ldots, b_p(s)) \) are continuous and the level set \( \{ s : b_1(s) = c \} \) is equal to \( \partial A_c(b_1) \).

(b) the noise field \( \epsilon(s) \) has continuous sample paths with probability one. Moreover, a centered unit variance Gaussian field with correlation function \( c(s_1, s_2) \) also has continuous sample paths with probability one.

(c) the variance function \( \sigma(s) \) is continuous.

(d) there exists a \( \delta > 0 \) such that \( \epsilon(s) \) has the property \( N1-\delta \) and \( n \left\| X(X^T X)^{-1/2} \right\|^{2+\delta}_\infty \to 0 \) as \( n \to \infty \).

(e) there exist \( \gamma \geq 0 \) and \( \beta > 0 \) such that \( \epsilon(s) \) has the property \( N2-(\gamma, \beta) \) and \( \max_{n \in \mathbb{N}} n \left\| X(X^T X)^{-1/2} \right\|^{2+\gamma}_\infty < \infty \).

Part (a) and (b) of Assumption 2 are tantamount to the first two conditions in Assumption 1. Parts (c), (d) and (e) will ensure that the parameter function \( b_1 \) enjoys a Central Limit Theorem. Note that the assumptions on the increments of the error field and on the design matrix \( X \) are coupled. The following Theorem 2 gives convergence results and explains how we can obtain CoPE sets for \( b_1 \).

**Theorem 2.** Under Assumption 2 the following is true.

(a) the weak convergence

\[
\sqrt{n} \left( b(s) - b(s) \right) / \sigma(s) \to G^\otimes p(s),
\]

holds, where \( G^\otimes p \) is an \( \mathbb{R}^p \)-valued mean zero, unit variance Gaussian random field with correlation function

\[
\text{cov} \left[ G^\otimes p(s_1), G^\otimes p(s_2) \right] = c(s_1, s_2) I_p.
\]

(b) if additionally the top-left entry \( \pi_n = \left( (X^T X)^{-1} \right)_{11} \) of the matrix \( X^T X \) is not zero and with \( e_1^T = (1, 0, \ldots, 0) \), the first standard basis vector,

\[
\pi_n^{-1/2} e_1^T (X^T X)^{-1/2} \to v^T \in \mathbb{R}^p, \quad \text{as} \ n \to \infty,
\]

then we have

\[
\left\| v \right\|^2_2 \pi_n^{-1/2} \sigma(s) \to \left( \hat{b}_1(s) - b_1(s) \right) \to G(s),
\]

weakly, where \( G \) is a mean zero, unit variance Gaussian field on \( S \) with correlation function \( \text{cov} [G(s_1), G(s_2)] = c(s_1, s_2) \).

(c) under the additional assumptions of part (b), and if we define

\[
\hat{A}_c^+(b_1) := \left\{ s : \left( \hat{b}_1(s) - c \right) / \left\| v \right\|^2_2 \sigma(s) \geq \pm a \right\},
\]

then

\[
\lim_{n \to \infty} P \left[ \hat{A}_c^+(b_1) \subset A_c(b_1) \subset \hat{A}_c^-(b_1) \right] = P \left[ \sup_{\partial A_c(b_1)} |G(s)| \leq a \right].
\]

### 3.2 Approximating the tail probabilities of \( G \)

#### 3.2.1 Multiplier bootstrap

In order to obtain CoPE sets from Theorem 2 we need to know the tail distribution of the supremum of the limiting (non-stationary) Gaussian field \( G \). In applications, as for example our climate data, the distribution of \( G \) (and hence of its supremum) 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 residuals \( (R_1(s), \ldots, R_n(s)) = R(s) = Y(s) - Xb(s) \) of the linear regression.

A way of approximating the distribution of the limiting Gaussian field \( G \) in this situation is given by the multiplier or wild bootstrap first introduced by Wu [1986] and later studied by Mammen [1992, 1993],
It is based on the following idea. Let \(g_1, \ldots, g_n\) be i.i.d standard Gaussian random variables independent of the data. Consider the random field

\[
\tilde{G}(s) = n^{-\frac{1}{2}} \sum_{j=1}^{n} g_j R_j(s).
\]

Then, conditional on the residuals \(\{R_j(s)\}_{j=1}^{n}\), the field \(\tilde{G}(s)\) is Gaussian and has covariance

\[
\text{cov} \left[ \tilde{G}(s_1), \tilde{G}(s_2) \right] = \frac{1}{n} \sum_{i,j=1}^{n} R_i(s_1) R_j(s_2) \text{cov} [g_i, g_j] = \frac{1}{n} \sum_{j=1}^{n} R_j(s_1) R_j(s_2),
\]

the sample covariance. For large \(n\), we expect the sample covariance to resemble the true covariance. The idea is to take the distribution of \(\tilde{G}\) as an approximation of the distribution of \(G\). In particular, we can approximate \(P \left[ \sup_{s \in \partial A} |G(s)| \leq a \right]\), needed in Theorem 1, by \(P \left[ \sup_{s \in \partial A} |\tilde{G}(s)| \leq a \mid \{R_j(s)\}_{j=1}^{n} \right]\). In practice, the latter can be efficiently computed by generating a large number \(M\) of i.i.d. copies \(\tilde{G}_1(s), \ldots, \tilde{G}_M(s)\) of \(\tilde{G}(s)\), conditional on \(\{R_j(s)\}_{j=1}^{n}\), via (6) and evaluating \(M^{-1} \sum_{j=1}^{M} \left\{ \sup_{s \in S} |\tilde{G}_j(s)| \leq a \right\}\).

**Theoretical considerations:** The above approximation of the distribution of the supremum requires some justification because the sample covariance (7) 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 is about the distribution of the supremum of the process instead. For a discrete set of locations, the distribution of the supremum \(\sup_{s \in \partial A} |G(s)|\) is similar to the distribution of the maximum of a high-dimensional Gaussian random vector, recently considered by Chernozhukov et al. (2013). Substantially 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.

**Computational considerations:** Besides these theoretical considerations, the multiplier bootstrap is also computationally attractive. In comparison with the direct simulation of the limiting field. While it is theoretically possible to simulate a number of realizations of a non-stationary Gaussian field with a given covariance as in (7) and to obtain tail probabilities from these, in practice this is computationally infeasible. Assuming that all fields are observed at \(L\) locations in \(S\) the direct method first requires computing the \(L \times L\) covariance matrix which is of complexity \(O(nL^2)\). Then, a Cholesky decomposition of the covariance matrix must be computed in time \(O(L^3)\). Finally, for each realization a matrix-vector product with the triangular matrix from the Cholesky decomposition is required and hence \(N\) realizations can be obtained in \(O(L^2 N)\) time. This yields a total complexity of \(O(L^3 + L^2(n + N))\) which is prohibitively large (in our data \(L = 9051\)). This problem has also been encountered by Adler et al. (2012) in a setting where information about the field is available through the true covariance function instead of realizations of it.

In contrast, creating one multiplier bootstrap realizations of the field \(\tilde{G}\) requires computing a linear combination of \(n\) vectors of dimension \(L\) which is of complexity \(O(nL)\). Hence, \(N\) multiplier bootstrap realizations can be generated in \(O(nNL)\) time. Further, note that the simulation of \(N\) bootstrap realizations can be written as a matrix multiplication. Let \(E\) be a \(L \times n\) matrix such that each column is one residual \(R_j\) and let \(V\) be a \(n \times N\) matrix with i.i.d. standard Gaussian entries. Then, the columns of \(EV\) correspond to realizations of \(\tilde{G}\). This makes the multiplier bootstrap very efficient because very fast implementations of matrix multiplication are available and it is an operation that can easily be parallelized.

A comparison of the computational burden to existing methods is not easy, since all work that is known to the authors considers different settings. However, to put the above considerations in perspective, we briefly discuss computational requirements of the method proposed by French (2014). This method requires (French 2014, Sec. 3.7) the computation of a kriging estimate in time \(O(n_l^3)\), where \(n_l\) is the number of observed locations, and a Cholesky decomposition which is of complexity \(O((n_l + n_g)^3)\), where \(n_g\) is the number of
grid locations (corresponding to \( L \) in our notation). This increased complexity compared to the multiplier bootstrap is reflected in actual empirical computation times. According to his own experiments, the method of French (2014) applied to \( n = 100 \) observed locations and a grid size of \( 100 \times 100 \), can be computed, on average, in just under five minutes. In contrast, the entire data analysis for the mean summer temperature, including linear regression at each grid point and computation of CoPE sets with the multiplier bootstrap, with a grid size of \( L \approx 10,000 \) and \( n = 58 \) is performed in under five seconds on a machine comparable to the one used by French (2014).

### 3.2.2 An alternative method for smooth fields

If one is willing to assume that the limiting field is twice differentiable and that the error field \( \epsilon(s) \) in (4) is Gaussian itself then the Gaussian Kinematic Formula (GKF) (e.g. Taylor (2006), Taylor et al. (2005), Taylor and Worsley (2007)) offers another way to approximate tail probabilities. Our motivation for presenting it here is twofold: First, it offers, under the additional assumptions made above, an elegant and accurate way of computing tail probabilities of Gaussian fields that does not require simulations and that has been successfully applied (Taylor and Worsley, 2007). Second, it shows that, at least for smooth fields, the tail probability of the supremum is intrinsically low-dimensional. More precisely, for a field on \( \mathbb{R}^N \) it is given (up to an exponential error term) by \( N + 1 \) numbers. 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.

The GKF is based on two properties of smooth Gaussian fields. The first is that for such fields the expected Euler characteristic in (8) has a closed formula. Indeed, with \( \rho(d) \) the Euler characteristic (see e.g. Adler and Taylor (2007)). The second property is that the probability of the supremum is precisely, for any set \( B \subseteq S \) with smooth boundary,

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

where \( \chi \) is the Euler characteristic (see e.g. Adler and Taylor (2007)). The second property is that the expected Euler characteristic in (8) has a closed formula. Indeed, with \( \Lambda(s) = \text{var} \left( \hat{G}(s) \right) \), where \( \hat{G}(s) \) is the vector of partial derivatives of \( G \), we can write (cf. e.g. Taylor (2006))

\[
E \left[ \chi(A_n(G) \cap B) \right] = \sum_{d=0}^{N} \mathcal{L}_d(B, \Lambda) \rho_d(a),
\]

with \( \mathcal{L}_d \) the \( d \)-th order Lipschitz-Killing curvature (LKC) (see e.g. Taylor (2006) for details on these quantities) and known functions \( \rho_d(a) \). We can use this to obtain CoPE sets: If the Assumptions (1) are satisfied then Theorem (11) implies in conjunction with (8) and (9) that

\[
\lim_{n \to \infty} P \left[ \hat{A}^+_c \subseteq A_c \subseteq \hat{A}_c^- \right] = 1 - P \left[ \sup_{s \in \hat{A}_c^-} |G(s)| \geq a \right]
\]

\[
= 1 - \sum_{d=0}^{N-1} \mathcal{L}_d(\partial A_c, \Lambda) \rho_d(a) + \mathcal{O}(\exp(-a^2/2)).
\]

To use (10), the problem amounts to estimating the LKCs of the boundary \( \partial A_c \). Taylor and Worsley (2007) propose a method to estimate the LKCs based on a finite number of realizations of \( G \). Applying this method requires a triangulation of the plug-in estimate \( \hat{A}_c \) of the boundary \( \partial A_c \).

While the triangulation is challenging, yet feasible, Taylor and Worsley (2007) prove the validity of their method only when applied to realizations of the Gaussian field \( G \). 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 \( Y(s) = Xb(s) + \epsilon(s) \).
Algorithm 1. Given a design matrix $X$ and observations $Y(s)$ following the linear model (4). If Assumptions hold, the following yields CoPE sets for $b_1(s)$.

(a) Compute the estimate $\hat{b}(s) = (X^TX)^{-1}X^TY(s)$ and the corresponding residuals $R(s) = Y(s) - Xb(s)$. With the empirical variance $\hat{\sigma}^2(s) = n^{-1}\sum_{j=1}^n R_j^2(s)$ compute the normalized residuals $\tilde{R}(s) = \hat{\sigma}(s)^{-1}R(s)$.

(b) Determine $a$ such that approximately $P[\sup_{s \in S}|G(s)| \geq a] \leq \alpha$. For example, use the multiplier bootstrap procedure presented in Section 3.2.1 with the residuals $\tilde{R}(s)$ to generate i.i.d. copies of a Gaussian $\tilde{G}$ field with covariance structure given by the sample covariance. With these, determine $a$ such that $P[\sup_{s \in S}|\tilde{G}(s)| \geq a] \leq \alpha$.

(c) Obtain the nested CoPE sets defined in equation (5).

4 Simulations

This section includes some artificial simulations to show that the proposed methods provide approximately the right coverage in practical non-asymptotic situations with non-smooth, non-stationary and non-Gaussian noise. We will describe ways to obtain error fields with these properties. Our objective in the design of the error fields described below 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

As a simple instance of the general linear model (4) we consider the signal plus noise model

$$y_j(s) = \mu(s) + \epsilon_j(s), \quad j = 1, \ldots, n$$

with non-stationary and non-Gaussian noise $\epsilon(s)$ over a square region $S$ of size $10 \times 10$ consisting of $64 \times 64$ square pixels. We will consider three different noise fields and we describe in the following how to obtain a realization of each.

Noise 1 In the upper half of $S$, each pixel is assigned the value of a standard normal random variable, all of which are independent. In the lower half, the pixels are grouped together in blocks of 4 by 4 pixels and each block is assigned the value of a standard normal, again all independent (cf. Figure 3b). Finally, the entire picture is convolved with a Gaussian kernel with bandwidth one and all values are multiplied by a scaling factor of 50.

Noise 2 Identical to Noise 1 except the image is smoothed by a Laplace kernel with bandwidth one instead of a Gaussian and the scaling factor is 100.

Noise 3 Each pixel in the upper half is assigned the value of a Laplace distributed random variable with mean zero and variance two. In the lower half, pixels are assigned the values of independent Student $t$-distributed random variables with 10 degrees of freedom. The entire picture is convolved with a Gaussian kernel of bandwidth one and multiplied by a scaling factor of 25.

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

The signal $\mu$ is a linear combination of three Gaussians. Figure 3a shows the signal $\mu$. Each one realization of the three noise fields is shown in Figure 3b.

We controlled the probability of coverage at the level $1 - \alpha = 0.9$ using Theorem 1. The estimator for $\mu$ here is the mean $\hat{\mu}_n(s) = n^{-1}\sum_{j=1}^n y_j(s)$ and the thresholds for the CoPE sets are obtained using Algorithm 1 with $||\nu||_2^{-1/2}\alpha^{-1/2} = \sqrt{n}$. The threshold $\alpha$ was computed using the multiplier bootstrap procedure proposed...
in Section 3.2.1 using either the true boundary $\partial A_c$ or the plug-in estimate $\partial \hat{A}_c$. The results of our method using $\partial \hat{A}_c$ for each one run with the three noise fields and sample sizes $n = 60$, $n = 120$ and $n = 240$ are shown in Figure 4.

4.2 Performance of CoPE sets

We analyzed the performance of our method on each 5000 runs of the toy examples shown in Section 4.1 with sample sizes $n = 60$, $n = 120$ and $n = 240$. Table 1 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 $\partial A_c$ or the plug-in estimator $\partial \hat{A}_c$ was used to determine the threshold.

We see that the empirical coverage is smaller than the nominal level in all experiments but approaches the nominal level reasonably fast as the sample size increases. In fact, when $n = 240$, the simulation confidence interval cover the nominal level of 90%, suggesting asymptotic unbiasedness. Comparing the two columns, 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.

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 $n = 240$, each on a grid of $64 \times 64 = 4096$ locations could be computed in less than two seconds on a standard laptop.

4.3 Comparison with Taylor’s Method

In this Section we compare the multiplier bootstrap with the method proposed by Taylor and Worsley (2007), as described in Section 3.2.2. We use both methods to approximate the distribution of $\sup_{s \in \partial A_c} |\epsilon(s)/\sigma(s)|$, where $\epsilon(s)$ is distributed according to Noise 1 (see Section 4.1 above), $\sigma^2(s) = \text{var}\{\epsilon(s)\}$ and $\partial A_c$ is the contour $A_c(\mu)$ of the function $\mu$ shown in Figure 3a at level $c = 4/3$. The true cumulative density function
Fig. 4: The output of our method for the three noise fields described above (corresponding to rows) and for sample sizes \( n = 60, 120, 240 \) (corresponding to columns) with the target function \( \mu(s) \) shown in Figure 3a. In all pictures we show a heat map of the estimator \( \hat{\mu}_n(s) \), the boundary of \( A_c(\mu) \) in purple as well as the boundaries of \( \hat{A}_c^+ \) and \( \hat{A}_c^- \) in red and green, respectively. The threshold \( a \) was obtained according to Theorem 1 to guarantee inclusion \( \hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \) with confidence \( 1 - \alpha = 0.9 \).

| Noise field 1  | \( n = 60 \) | \( n = 120 \) | \( n = 240 \) |
|----------------|-------------|-------------|-------------|
|                | 86.20% ± 0.49% | 86.16% ± 0.49% | 86.16% ± 0.49% |
|                | 88.62% ± 0.45% | 88.74% ± 0.45% | 88.74% ± 0.45% |
|                | 88.94% ± 0.44% | 88.9% ± 0.44%  | 88.9% ± 0.44%  |
| Noise field 2  | \( n = 60 \) | \( n = 120 \) | \( n = 240 \) |
|                | 87.22% ± 0.47% | 88.74% ± 0.45% | 88.74% ± 0.45% |
|                | 89.22% ± 0.44% | 89.26% ± 0.44% | 89.26% ± 0.44% |
|                | 89.76% ± 0.43% | 89.70% ± 0.43% | 89.70% ± 0.43% |
| Noise field 3  | \( n = 60 \) | \( n = 120 \) | \( n = 240 \) |
|                | 86.44% ± 0.48% | 86.62% ± 0.48% | 86.62% ± 0.48% |
|                | 88.60% ± 0.44% | 88.78% ± 0.45% | 88.78% ± 0.45% |
|                | 89.76% ± 0.43% | 89.94% ± 0.43% | 89.94% ± 0.43% |

**Tab. 1:** Percentage of trials in which \( \hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \). The nominal coverage probability is 90%.
for $\sup_{s \in \partial A_c} |\epsilon(s)/\sigma(s)|$ and its empirical approximations based on the multiplier bootstrap and Taylor’s method are shown in Figure 5. The empirical cdfs are each based on a single i.i.d. sample $\epsilon_1(s), \ldots, \epsilon_n(s)$ 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 $\epsilon(s)$.

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.

5 Application to the climate data

5.1 Data setup

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. More precisely we have

$$
Y_j(s) = T^{(a)}(s) + m^{(a)}(s) t_j^{(a)} + \epsilon_j(s), \quad j = 1, \ldots, n^{(a)}
$$

$$
Y_j(s) = T^{(b)}(s) + m^{(b)}(s) t_j^{(b)} + \epsilon_j(s), \quad j = n^{(a)} + 1, \ldots, n^{(a)} + n^{(b)}.
$$

Without loss of generality, we may assume that $\sum_{j=1}^{n^{(a)}} t_j^{(a)} = 0$ and $\sum_{j=n^{(a)}+1}^{n^{(a)}+n^{(b)}} t_j^{(b)} = 0$. We will denote the covariance of the error field $\epsilon(s)$ by $\text{cov}[\epsilon(s_1), \epsilon(s_2)]$. Our goal is to give CoPE sets for the excursion sets of the difference $T^{(b)}(s) - T^{(a)}(s)$. Therefore, we define the parameter vector and the design matrix

$$
\begin{pmatrix}
  b_1(s) \\
  b_2(s) \\
  b_3(s) \\
  b_4(s)
\end{pmatrix}
= 
\begin{pmatrix}
  T^{(b)}(s) - T^{(a)}(s) \\
  T^{(a)}(s) \\
  m^{(a)}(s) \\
  m^{(b)}(s)
\end{pmatrix},
\quad
X = 
\begin{pmatrix}
  0 & 1 & t_1^{(a)} & 0 \\
  : & : & : & : \\
  0 & 1 & t_n^{(a)} & 0 \\
  1 & 1 & 0 & t_1^{(b)} \cdot n^{(a)} + 1 \\
  : & : & : & : \\
  1 & 1 & 0 & t_n^{(b)} \cdot n^{(a)} + n^{(b)}
\end{pmatrix},
$$

to be able to rewrite (11) as a general linear model $Y(s) = X b(s) + \epsilon(s)$. Our objective is now to formulate Assumptions on the design and the noise under which we can apply Algorithm 1 to the data. This is done in the following.
Assumption 3. Assume that

(a) the parameter functions $b$ are continuous and and the level set $\{s : b_1(s) = c\}$ is equal to $\partial A_c(b_1)$.
(b) the noise field $\epsilon(s)$ has continuous sample paths with probability one and a centered unit variance Gaussian field with correlation function $c(s_1, s_2)$ also has continuous sample paths with probability one.
(c) the variance function $\sigma(s)$ is continuous.
(d) there exist numbers $\delta, \beta > 0$ and $\gamma \geq 0$ such that the error field $\epsilon(s)$ has the properties $N1-\delta$ and $N2-(\gamma, \beta)$.
(e) $n(a) = n(b) = n/2$ and that both sets of design points $t_j^{(a)}$ and $t_j^{(b)}$ are equally spaced (possibly with different spacing for the periods $(a)$ and $(b)$).

The next final and central statement now asserts that these Assumptions are indeed sufficient for Algorithm 1 to be valid. Its proof is a direct application of Theorem 2.

Proposition 1. Under Model (1) Assumptions 3 imply that Assumptions 1 hold for the target function $b_1(s)$ and the estimator $\hat{b}_1(s)$ with $\tau_n = 2n^{-1/2}$. In particular,

$$\sqrt{n} \left( \hat{b}_1(s) - b_1(s) \right) \to G(s),$$

weakly, where $G$ is a mean zero unit variance Gaussian field with correlation function $\text{cov} \{G(s_1), G(s_2)\} = c(s_1, s_2)$. Consequently, Algorithm 1 can be used to obtain CoPE sets for the excursion sets of $b_1(s)$.

5.2 Data analysis

The results for the climate data described in the Introduction, shown in Figure 1, correspond to CoPE sets for $b_1(s) = T^{(b)}(s) - T^{(a)}(s)$ obtained via Algorithm 1 with $||v||_2 \pi n^{1/2} \sigma(s)^{-1} = 2/\sqrt{n}$ and $n = 29 + 29 = 58$. The target level is $c = 2^\circ C$ 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 $2^\circ C$ 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.

For the computation time we remark that the entire analysis of one season, including the pointwise linear regression and the multiplier bootstrap to obtain the CoPE sets was performed in under five seconds on a regular laptop.

Acknowledgment

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).
References

Robert J. Adler. On Excursion Sets, Tube Formulas and Maxima of Random Fields. *The Annals of Applied Probability*, 10(1):1–74, 2000.

Robert J. Adler and Jonathan E Taylor. *Random fields and geometry*. Springer, New York, 2007.

Robert J. Adler, Jose H. Blanchet, and Jingchen Liu. Efficient Monte Carlo for high excursions of Gaussian random fields. *The Annals of Applied Probability*, 22(3):1167–1214, 2012.

Kevin Anderson and Alice Bows. 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, 2011.

Jr. Bassett, Gilbert and Roger Koenker. Asymptotic Theory of Least Absolute Error Regression. *Journal of the American Statistical Association*, 73(363):618–622, 1978.

Simeon M. Berman. Sojourns and Extremes of Stationary Processes. *The Annals of Probability*, 10(1):1–46, 1982.

P. J. Bickel and M. J. Wichura. Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *The Annals of Mathematical Statistics*, 42(5):1656–1670, 1971.

David Bolin and Finn Lindgren. Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. To appear.

Benoît Cadre. Kernel estimation of density level sets. *Journal of Multivariate Analysis*, 97(4):999–1023, 2006.

Laurent Cavalier. Nonparametric Estimation of Regression Level Sets. *Statistics*, 29(2):131–160, 1997.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 2013.

Antonio Cuevas, Wenceslao González-Manteiga, and Alberto Rodríguez-Casal. Plug-in Estimation of General Level Sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19, 2006.

F. Eicker. Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions. *The Annals of Mathematical Statistics*, 34(2):447–456, 1963.

G. M. Flato. The third generation coupled global climate model (CGCM3). *Available on line at [http://www.cccma.bc.ec.gc.ca/models/cgcm3.shtml]* 2005.

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

Joshua P. French and Stephan R. Sain. Spatio-temporal exceedance locations and confidence regions. *The Annals of Applied Statistics*, 7(3):1421–1449, 2013.

Christopher R. Genovese, Nicole A. Lazar, and Thomas Nichols. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, 15(4):870–878, 2002.

W. Hardle and E. Mammen. Comparing Nonparametric Versus Parametric Regression Fits. *The Annals of Statistics*, 21(4):1926–1947, 1993.

Davar Khoshnevisan. *Multiparameter Processes: an introduction to random fields*. Springer, 2002.

Georg Lindgren and Igor Rychlik. How reliable are contour curves? Confidence sets for level contours. *Bernoulli*, 1(4):301–319, 1995.
Enno Mammen. Bootstrap, wild bootstrap, and asymptotic normality. *Probability Theory and Related Fields*, 93(4):439–455, 1992.

Enno Mammen. Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *The Annals of Statistics*, 21(1):255–285, 1993.

Enno Mammen and Wolfgang Polonik. Confidence regions for level sets. *Journal of Multivariate Analysis*, 122:202–214, 2013.

David M. Mason and Wolfgang Polonik. Asymptotic normality of plug-in level set estimates. *The Annals of Applied Probability*, 19(3):1108–1142, 2009.

L. O. Mearns, S. Sain, L. R. Leung, M. S. Bukovsky, S. McGinnis, S. Biner, D. Caya, R. W. Arritt, W. Gutowski, E. Takle, M. Snyder, R. G. Jones, A. M. B. Nunes, S. Tucker, D. Herzmann, L. McDaniel, and L. Sloan. Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Climatic Change*, 120(4):965–975, 2013.

Linda O. Mearns, William Gutowski, Richard Jones, Ruby Leung, Seth McGinnis, Ana Nunes, and Yun Qian. A Regional Climate Change Assessment Program for North America. *Eos, Transactions American Geophysical Union*, 90(36):311–311, 2009.

Linda O. Mearns, Ray Arritt, Sébastien Biner, Melissa S. Bukovsky, Seth McGinnis, Stephan Sain, Daniel Caya, James Correia, Dave Flory, William Gutowski, Eugene S. Takle, Richard Jones, Ruby Leung, Wilfran Moufouma-Okia, Larry McDaniel, Ana M. B. Nunes, Yun Qian, John Roads, Lisa Sloan, and Mark Snyder. The North American Regional Climate Change Assessment Program: Overview of Phase I Results. *Bulletin of the American Meteorological Society*, 93(9):1337–1362, 2012.

J. Michalakes, J. Dudhia, D. Gill, T. Henderson, J. Klemp, W. Skamarock, and W. Wang. 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. World Scientific, 2004.

Viet-Hung Pham. On the rate of convergence for central limit theorems of sojourn times of Gaussian fields. *Stochastic Processes and their Applications*, 123(6):2158–2174, 2013.

Thomas Polfeldt. On the quality of contour maps. *Environmetrics*, 10(6):785–790, 1999.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL [http://www.R-project.org/](http://www.R-project.org/).

Calyampudi Radhakrishna Rao and Helge Toutenburg. *Linear models*. Springer, 1995.

Philippe Rigollet and Régis Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15(4):1154–1178, 2009.

Joeri Rogelj, Bill Hare, Julia Nabel, Kirsten Macey, Michiel Schaeffer, Kathleen Markmann, and Malte Meinshausen. Halfway to Copenhagen, no way to 2 °C. *Nature Reports Climate Change*, (0907):81–83, 2009.

Armin Schwartzman and Xihong Lin. The effect of correlation in false discovery rate estimation. *Biometrika*, 98(1):199–214, 2011.

Armin Schwartzman, Robert F. Dougherty, and Jonathan E. Taylor. Group Comparison of Eigenvalues and Eigenvectors of Diffusion Tensors. *Journal of the American Statistical Association*, 105(490):588–599, 2010.

Katsuhiro Shiohama and Hong-Wei Xu. An Integral Formula for Lipschitz-Killing Curvature and the Critical Points of Height Functions. *Journal of Geometric Analysis*, 21(2):241–251, 2011.

Aarti Singh, Clayton Scott, and Robert Nowak. Adaptive Hausdorff estimation of density level sets. *The Annals of Statistics*, 37(5B):2760–2782, 2009.
Max Sommerfeld. *cope: Coverage Probability Excursion (CoPE) sets.*, 2015. URL http://www.cran.r-project.org/package=cope, R package version 0.1.

Jonathan Taylor, Akimichi Takemura, and Robert J. Adler. Validity of the expected Euler characteristic heuristic. *The Annals of Probability*, 33(4):1362–1396, 2005.

Jonathan E. Taylor. A Gaussian kinematic formula. *The Annals of Probability*, 34(1):122–158, 2006.

Jonathan E. Taylor and Robert J. Adler. Euler Characteristics for Gaussian Fields on Manifolds. *The Annals of Probability*, 31(2):533–563, 2003.

Jonathan E. Taylor and Robert J. Adler. Gaussian processes, kinematic formulae and Poincaré’s limit. *The Annals of Probability*, 37(4):1459–1482, 2009.

Jonathan E. Taylor and Keith J. Worsley. Detecting Sparse Signals in Random Fields, with an Application to Brain Mapping. *Journal of the American Statistical Association*, 102(479):913–928, 2007.

A. B. Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969, 1997.

Almuth Wameling. Accuracy of geostatistical prediction of yearly precipitation in Lower Saxony. *Environmetrics*, 14(7):699–709, 2003.

R.M. Willett and R.D. Nowak. Minimax Optimal Level-Set Estimation. *IEEE Transactions on Image Processing*, 16(12):2965–2979, 2007.

Keith J. Worsley, Sean Marrett, Peter Neelin, Alain C. Vandal, Karl J. Friston, Alan C. Evans, et al. A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73, 1996.

C. F. J. Wu. Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*, 14(4):1261–1295, 1986.

## A Proofs

**Proof of Theorem [1]** We start by showing that

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

For \( \eta > 0\) define the inflated boundary \( A_c^\eta = \{ s \in S : c - \eta \sigma(s) \leq \mu(s) \leq c + \eta \sigma(s) \} \). The idea of the proof is that, loosely speaking, points outside of \( A_c^\eta \) become irrelevant in the limit \( n \to \infty \) since their values are far from \( c \) and, if we let \( \eta \) go to zero at an appropriate rate, we finally end up with the boundary \( \partial A_c \). More precisely, we note that

\[
\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \geq -a \text{ for all } s \in A_c \cap A_c^\eta \quad \text{and} \quad \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \geq -\eta \tau_n^{-1} - a \text{ for all } s \in A_c \setminus A_c^\eta,
\]

implies that \( \hat{\mu}_n(s) \geq c - \tau_n \sigma(s)a \) for all \( s \in A_c \) and hence \( A_c \subset \hat{A}_c^- \). Similarly,

\[
\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < a \text{ for all } s \in (S \setminus A_c) \cap A_c^\eta \quad \text{and} \quad \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < \eta \tau_n^{-1} + a \text{ for all } s \in (S \setminus A_c) \setminus A_c^\eta
\]

implies \( A_c^+ \subset A_c \). Combining these observations, we see that \( \hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \) holds, provided that \( \sup_{s \in A_c^\eta} \left| \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \right| < a \) and \( \sup_{s \in S \setminus A_c^\eta} \left| \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} \right| < a + \eta \tau_n^{-1} \). Now, let \( \{ \eta_n \}_{n \in \mathbb{N}} \) be a sequence of
positive numbers such that $\eta_n \to 0$ and $\eta_n \tau_n^{-1} \to \infty$. We can then write

\[
P\left[\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-\right] \geq P\left[\sup_{s \in A_n^+} \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < a \right. \quad \text{and} \quad \sup_{s \in S \setminus A_n^+} \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < a + \eta_n \tau_n^{-1}\] 

\[
\geq P\left[\sup_{s \in A_n^+} \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < a\right] + P\left[\sup_{s \in S \setminus A_n^+} \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < a + \eta_n \tau_n^{-1}\right] - 1. \tag{13}
\]

We first show that the term (II) goes to zero. To this end let $\delta > 0$ arbitrary. Let $b \in \mathbb{R}$ such that $P[\sup_{s \in S} |G(s)| < b] \geq 1 - \delta$ and $n_0 \in \mathbb{N}$ such that $a + \eta_n \tau_n^{-1} \geq b$ for all $n \geq n_0$. Also, let $n_1$ large enough such that

\[
\left|P\left[\sup_{s \in S} \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < b\right] - P\left[\sup_{s \in S} |G(s)| < b\right]\right| < \delta,
\]

for all $n \geq n_1$. In consequence, for all $n \geq \max\{n_0, n_1\}$

\[
P\left[\sup_{s \in S \setminus A_n^+} \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < a + \eta_n \tau_n^{-1}\right] \geq P\left[\sup_{s \in S} \frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)} < b\right] \geq 1 - 2\delta.
\]

Since $\delta > 0$ was arbitrary it follows that (II) converges to zero as $n \to \infty$.

To prove convergence of (I) we need the following

**Lemma 1.** Under Assumptions I part (a) if $\eta_n \to 0$ then the Hausdorff distance $\delta_n := d_H(A^n_L, \partial A_c) \to 0$.

**Proof.** Let us define the set $(\partial A_c)_\varepsilon := \{s \in S : d(s, \partial A_c) < \varepsilon\}$. We prove the assertion by showing that for any $\varepsilon > 0$ there exists an $\eta > 0$ such that $A^n_L \subset (\partial A_c)_\varepsilon$. To this end, assume the contrary. Then, there exists $\varepsilon > 0$ such that for any $\eta > 0$ we find $s_n \in A^n_L$ with $d(s_n, \partial A_c) \geq \varepsilon$. The sequence $(s_n)_{n \in \mathbb{N}}$ is contained in the compact set $S$ and hence has a convergent subsequence with limit $s^*$, say. By construction, we have $s^* \in \cap_{\eta > 0} A^n_L = \partial A_c$. On the other hand, $0 = d(s^*, \partial A_c) = \lim_{\eta \to 0} d(s_\eta, \partial A_c) \geq \varepsilon$, a contradiction. □

Recall that for a function $f : S \to \mathbb{R}$ and some number $\delta > 0$ the modulus of continuity is defined as $w(f, \delta) = \sup_{|s_1 - s_2| \leq \delta} |f(s_1) - f(s_2)|$. Since $\left\{\tau_n^{-1} \sigma(s)^{-1} (\hat{\mu}_n(s) - \mu(s))\right\}_{n \in \mathbb{N}}$ is weakly convergent, we have

\[
\lim_{\delta \to 0} \limsup_{n \to \infty} P\left[w\left(\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}, \delta\right) \geq \zeta\right] = 0 \tag{14}
\]

for all positive $\zeta$. [Khoshnevisan 2002 Prop. 2.4.1 and Exc. 3.3.1]. Together with Lemma 1 this implies

\[
\left|\sup_{s \in A^n_L} \left|\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}\right| - \sup_{s \in \partial A_c} \left|\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}\right|\right| \leq w\left(\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}, \delta_n\right) \to 0,
\]

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

\[
\sup_{s \in A^n_L} \left|\frac{\hat{\mu}_n(s) - \mu(s)}{\tau_n \sigma(s)}\right| \to \sup_{s \in \partial A_c} |G(s)|
\]

in distribution. In view of (13) this completes the proof of (12).

It remains to prove the opposite inequality, i.e.

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

If for some arbitrary $\delta > 0$ we have $\tau_n^{-1} \sigma(s)^{-1} (\hat{\mu}_n(s) - c) \geq a + \delta$ for some $s \in \partial A_c$ then by continuity there is a $s \in S \setminus A_c$ for which $\tau_n^{-1} \sigma(s)^{-1} (\hat{\mu}_n(s) - c) \geq a$ and hence the inclusion $\hat{A}_c^+ \subset A_c$ does not hold. Since
an analogous argument works for the inclusion $A_c \subset \hat{A}_c$, we have
\[ P \left[ \hat{A}_c^+ \subset A_c \subset \hat{A}_c \right] \leq 1 - P \left[ \exists s \in \partial A_c : \frac{\hat{\mu}_n(s) - c}{\tau_n \sigma(s)} \geq a + \delta \text{ or } \frac{\hat{\mu}_n(s) - c}{\tau_n \sigma(s)} \leq -a - \delta \right] \leq 1 - P \left[ \sup_{s \in \partial A_c} \left| \frac{\hat{\mu}_n(s) - c}{\tau_n \sigma(s)} \right| \geq a + \delta \right] \to P \left[ \sup_{s \in \partial A_c} |G(s)| \leq a + \delta \right]. \]
Since $\delta > 0$ was arbitrary and $\sup_{s \in \partial A_c} |G(s)|$ has a continuous distribution the bound (15) follows. \qed

Proof of Corollary 7. For any pair of nested sets $\hat{A}_c^+ \subset A_c \subset \hat{A}_c^-$ 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 $\sup_{s \in \partial A_c} \left| \frac{\hat{\mu}_n(s) - c}{\tau_n \sigma(s)} \right| > a$. Combining these two observations yields
\[ P \left[ \hat{A}_c^+ \subset A_c \subset \hat{A}_c^- \right] \leq P \left[ \partial A_c \subset \text{cl}(\hat{A}_c^- \setminus \hat{A}_c^+) \right] \leq P \left[ \sup_{s \in \partial A_c} \left| \frac{\hat{\mu}_n(s) - c}{\tau_n \sigma(s)} \right| \leq a \right]. \]
Taking the limit $n \to \infty$ of this inequality and using Theorem 1 gives the assertion. \qed

Proof of Theorem 2. We begin by proving part (a). Let us define $A = (X^T X)^{-1} X^T$, giving $\hat{b}(s) - b(s) = A \epsilon(s)$. In order 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 $s_1, \ldots, s_K \in S$ be arbitrary. We need to show that with
\[ U = \left( \sqrt{X^T X} \sigma(s_1)^{-1} \left( \hat{b}(s_1) - b(s_1) \right), \ldots, \sqrt{X^T X} \sigma(s_K)^{-1} \left( \hat{b}(s_K) - b(s_K) \right) \right) = I_K \otimes \left[ \sqrt{X^T X} \right] \left( \sigma(s_1)^{-1} A \epsilon(s_1), \ldots, \sigma(s_K)^{-1} A \epsilon(s_K) \right), \]
(here, '$\otimes$' denotes the Kronecker product of two matrices) we have convergence $U \to \mathcal{N}(0, [\epsilon(s_i, s_j)]_{i,j=1}^K \otimes I_p)$ in distribution. We readily see that $E[U] = 0$ and for the covariance we compute
\[ \text{cov} [U] = I_K \otimes \left[ \sqrt{X^T X} A \right] E \left\{ (\sigma(s_1)^{-1} \epsilon(s_1), \ldots, \sigma(s_K)^{-1} \epsilon(s_K))(\sigma(s_1)^{-1} \epsilon(s_1), \ldots, \sigma(s_K)^{-1} \epsilon(s_K))^T \right\} \]
\[ = I_K \otimes \left[ \sqrt{X^T X} A \right] \left( [\epsilon(s_i, s_j)]_{i,j=1}^K \right) \otimes I_p \left( I_K \otimes \left[ \sqrt{X^T X} A \right] \right) \]
\[ = [\epsilon(s_i, s_j)]_{i,j=1}^K \otimes \sqrt{X^T X} A A^T \sqrt{X^T X} = [\epsilon(s_i, s_j)]_{i,j=1}^K \otimes I_p. \]
We employ the Cramér-Wold device to show convergence of $U$. Indeed, let and $(\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^{K \times p}$ be some fixed arbitrary vector and compute
\[ \langle U, \alpha \rangle = \langle (\sigma(s_1)^{-1} \epsilon(s_1), \ldots, \sigma(s_K)^{-1} \epsilon(s_K)), I_K \otimes \left[ A^T \sqrt{X^T X} \right] \rangle \]
\[ = \sum_{i=1}^K \sigma(s_i)^{-1} \langle \epsilon(s_i), A^T \sqrt{X^T X} \alpha_i \rangle = \sum_{i=1}^K \sum_{j=1}^n \sigma(s_i)^{-1} \epsilon_j(s_i) \left( A^T \sqrt{X^T X} \alpha_i \right). \]
By interchanging the sums and defining $W_j = \sum_{i=1}^K \sigma(s_i)^{-1} \epsilon_j(s_i) \left( A^T \sqrt{X^T X} \alpha_i \right)$, we have 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 $\text{var} \left( \sum_{j=1}^n W_j \right) = \text{var} \left( \langle U, \alpha \rangle \right) = \alpha^T \text{cov} [U] \alpha$ 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 $\delta$ be as in Assumption 2 to give
\[ \sum_{j=1}^n E|W_j|^{2+\delta} \leq K \sum_{j=1}^n \sum_{i=1}^K \sigma(s_i)^{-2(2+\delta)} E|\epsilon_j(s_i)|^{2+\delta} \left( A^T \sqrt{X^T X} \alpha_i \right)^{2+\delta} \]
\[ \leq n \sum_{j=1}^K CK \left\| \left( A^T \sqrt{X^T X} \alpha_i \right) \right\|_\infty^{2+\delta} \leq CK^2 \| \alpha \|_\infty n \left\| A^T \sqrt{X^T X} \right\|_\infty^{2+\delta} \to 0, \]
max

Now we note that since the design points are as in Assumption 2

$$\pi_n^{-1/2} \sigma(s)^{-1} \left( \hat{b}_1(s) - b_1(s) \right) = \pi_n^{-1/2} \epsilon_1^T (X^T X)^{-1/2} \sigma(s)^{-1} \sqrt{X^T X} \left( b(s) - b(s) \right)$$

$$\rightarrow v G^p(s) \overset{D}{=} \|v\|_2 G(s),$$

where the weak convergence from the previous part is used.

The last part (c) of the Theorem is a direct application of Theorem 1 where parts (a) and (b) guarantee that the assumptions are satisfied.

Proof of Proposition 1 In order to be able to apply Theorem 2 we compute

$$X^T X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{\sigma(s)} \end{pmatrix} \cdot \omega(a) = \sum_{j=1}^{n(s)} (t_j^{(a)})^2, \quad \omega(b) = \sum_{j=n(s)+1}^{n(s)+n(b)} (t_j^{(b)})^2.$$

It follows that

$$(X^T X)^{-1} = \begin{pmatrix} 4/n & 2/n & 0 \\ -2/n & 2/n & 0 \\ 0 & 0 & 2/\sqrt{\sigma(s)} \end{pmatrix}, \quad (X^T X)^{-1/2} = \sqrt{n/10} \begin{pmatrix} 6/n & -2/n & 0 \\ -2/n & 4/n & 0 \\ 0 & 0 & \sqrt{10}\omega(a) \end{pmatrix}.$$

With this we obtain

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

Now we note that since the design points $t_j^{(a)}$ and $t_j^{(b)}$ are equally spaced by Assumption 3 we have $\max_{1 \leq j \leq n} |t_j^{(a)}| = O(n)$ and $\omega(a) = O(n^3)$, and the same is true for the (b)-counterparts. This shows that $\|X(X^T X)^{-1/2}\|_\infty = O(n^{-1/2})$ and therefore Assumptions 3 imply Assumptions 2. Now, in the notation of Theorem 2 we have $\pi_n^{-1/2} = \sqrt{n/2}$ and

$$\pi_n^{-1/2} \epsilon_1^T (X^T X)^{-1/2} = \sqrt{n/2} \sqrt{n/10} \left( \frac{6}{n} - \frac{2}{n} \right) = \frac{1}{2\sqrt{10}} (6 - 2) = v^T,$$

so that $\|v\|_2 = 1$. This finally gives $\sqrt{\frac{2}{n}} \sigma(s)^{-1} (\hat{b}_1(s) - b_1(s)) \rightarrow G(s).$