



Robust Maximum L_q -Likelihood Covariance Estimation for Replicated Spatial Data

Sihan Chen
King Abdullah University
of Science and Technology

Joydeep Chowdhury
King Abdullah University
of Science and Technology

Marc G. Genton
King Abdullah University
of Science and Technology

Abstract

Parameter estimation with the maximum L_q -likelihood estimator (ML q E) is an alternative to the maximum likelihood estimator (MLE) that considers the q -th power of the likelihood values for some $0 < q < 1$. In this method, extreme values are down-weighted because of their lower likelihood values, which yields robust estimates. In this work, we study the properties of the ML q E for spatial data with replicates. We investigate the asymptotic properties of the ML q E for Gaussian random fields with a Matérn covariance function, and carry out simulation studies to investigate the numerical performance of the ML q E. We show that it can provide more robust and stable estimation results when some of the replicates in the spatial data contain outliers. In addition, we develop a mechanism to find the optimal choice of the hyper-parameter q for the ML q E. The robustness of our approach is further verified on a United States precipitation dataset. Compared with other robust methods for spatial data, our proposal is more intuitive and easier to understand, yet it performs well when dealing with datasets containing outliers.

Keywords: Gaussian random fields, large-scale computation, maximum L_q -likelihood estimator, robust statistics, spatial statistics.

1. Introduction

The Maximum Likelihood Estimator (MLE) has been one of the most powerful statistical methods widely used in various domains (Aldrich 1997). Its applications have been extended across various fields, including econometrics (Greene 1980; Cramer 1989), genetics (Shaw 1987; Beerli 2006) and geography (Elhorst 2005; Ree and Smith 2008). The MLE has gained popularity for its asymptotic properties, such as consistency and efficiency (see, e.g., Casella and Berger 2024). It also plays an important role in the analysis of spatial and spatio-temporal data (Mardia and Marshall 1984; Gelfand et al. 2010), and in recent years, some tools have been developed to efficiently compute the MLE for large-scale spatial data, for example **ExaGeoStat** (Abdulah et al. 2018) and **GpGp** (Katzfuss and Guinness 2021).

However, when using the MLE with datasets containing outliers, the likelihood evaluation might be largely distorted by these extreme values, and this can lead to inaccurate estimation results (Chen et al. 2014). The maximum L_q -Likelihood estimator (ML q E) proposed by Ferrari and Yang (2010) provides an effective way to address this problem. Unlike the MLE which calculates the direct summation of the log-likelihood values at each data point, the ML q E sums the q -th power of the likelihood values for some $q < 1$, which down-weights the extreme values and leads to robust estimation. Ferrari and Yang (2010) and Ferrari and La Vecchia (2012) established the asymptotic normality of the ML q E under some regularity conditions and provided the expression of its asymptotic variance.

In this work, our aim is to employ the ML q E method to estimate the covariance parameters for replicated spatial data. We consider the replications of spatial data as independent realisations of a random field over the spatial domain, and estimate the parameters of the covariance function of the random field using the ML q E. The outliers in this work are defined in the following way. Outliers are considered to be replicates of the random field whose dependence structures are largely different from the rest of the sample. The fraction of outliers in the data depends highly on the application and can vary significantly; therefore, in the simulation experiments of this work, we consider several different settings, where the fraction of contaminated data in the datasets varies from 1% to 20%. We focus on Gaussian random fields with the Matérn covariance function, which is one of the most influential models for spatial statistics. The Matérn covariance function was proposed by Matérn (1960), and became popular later after the work of Handcock and Stein (1993). It is used in many well-known tools in R, such as **RandomFields** (Schlather et al. 2013) and **fields** (Nychka et al. 2021). A comprehensive review of the application of the Matérn covariance function in spatial statistics and many other related research fields can be found in Porcu et al. (2024). There are several different widely accepted ways to parameterise the Matérn covariance function; see Wang et al. (2023) for a comprehensive comparison of the three most popular parameterisations. In this paper, we adopt the following formulation of the Matérn covariance function, which is the first type of parametrisation in Wang et al. (2023):

$$\mathcal{M}(h; \boldsymbol{\theta}) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{h}{\beta}\right)^\nu \mathcal{K}_\nu\left(\frac{h}{\beta}\right). \quad (1)$$

Here, h is the distance between a pair of spatial locations, and $\boldsymbol{\theta} = (\sigma^2, \beta, \nu)^\top$ is the parameter vector that we aim to estimate in a robust way with our proposed method. Its

components σ^2, β, ν are the variance, range and smoothness parameters, respectively, and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν .

In the literature, there have been various attempts to apply robust statistical methods on spatial data. For example, [Cressie and Hawkins \(1980\)](#) proposed a robust procedure for estimating the variogram of spatial data, and [Hawkins and Cressie \(1984\)](#) also proposed a robust kriging method based on their previous work. [Genton \(1998\)](#) proposed a highly robust variogram estimator and compared it with the one in [Cressie and Hawkins \(1980\)](#), while showing that the latter is actually not strictly robust but only less non-robust than the empirical variogram estimator. A comprehensive review of classical robust methods for spatial data can be found in [Lark \(2000\)](#). [Marchant and Lark \(2007\)](#) used the residual maximum likelihood for robust estimation of the variogram, which was based on the robustified MLE proposed in [Richardson and Welsh \(1995\)](#) for linear mixed models. The latest approach proposed by [Künsch et al. \(2013\)](#) is a robust method to estimate the external drift and the variogram of spatial data. Their main idea is to replace the residual function with another bounded function of the observations, so that the influence of outliers in the data can be bounded. Unfortunately, their method is not very practical because it is not able to deal with large-scale datasets. Indeed, in their numerical studies they only used datasets limited to a few hundreds of locations, which is often not sufficient because the scale of modern spatial datasets is getting very large. Hence, a robust likelihood-based method for estimating the parameters of the spatial covariance function for large-scale datasets is lacking.

We first demonstrate how the MLqE diminishes the influence of outliers in the dataset on parameter estimation results using a motivating example of precipitation data from the United States retrieved from <https://www.image.ucar.edu/Data/US.monthly.met/USmonthlyMet.shtml>. The detailed definition of the MLqE can be found in [Section 2](#). The dataset that we consider here contains hourly precipitation data from 1895 to 1997 recorded at thousands of observation stations throughout the country. In this example, we consider the average hourly precipitation in the month of January from 1928 to 1997 recorded at 621 different monitoring stations located in the contiguous United States. The coordinates of these locations are normalised to be within the 2-dimensional unit square. We first conduct the parameter estimation using the MLE and the MLqE with several different values of the hyper-parameter q , and then try to remove the years in the data with the most outliers and redo the parameter estimation using the MLE, to see whether the MLE after removing the outliers is closer to the MLqE without removing the outliers for some $q < 1$. For the average January data for each year, we first estimate the empirical variogram of the data, which is an empirical function that describes the spatial dependence of the random field, using the R package [geoR \(Ribeiro Jr and Diggle 2007\)](#); next, we find out the years for which the January data behave differently from other years via a functional boxplot of the corresponding empirical variogram, using the R command `fbplot` ([Sun and Genton 2011](#)). The resulting functional boxplot for the January data of all the 70 years is shown in the first sub-figure of [Figure 1](#), where the dashed-curves represent the outliers identified among the 70 empirical variograms. Here we identify 5 outliers in total. Additionally, in the same plot, the black curve is the median of the variograms, and the purple area represents the 50% central region. In the last three sub-figures of [Figure 1](#), the blue horizontal lines represent the MLE results with the original data, and the red dashed

lines are the MLE results after removing the 5 outliers from the dataset, while the black curves are the $MLqE$ results without removing the outliers. The blue dots in those sub-figures are the $MLqE$ values for $q = 1$, which coincide with the MLE values. We can see that the red dashed line intersects the black curve at a $q < 1$, which means that the existence of outliers significantly affects the parameter estimation results, and the $MLqE$ is able to reduce their effect with a $q < 1$. It suggests that applying the $MLqE$ for some $q < 1$ is able to provide parameter estimation results that automatically diminish the effect of outliers, without the need to identify and remove outliers manually in the dataset beforehand.

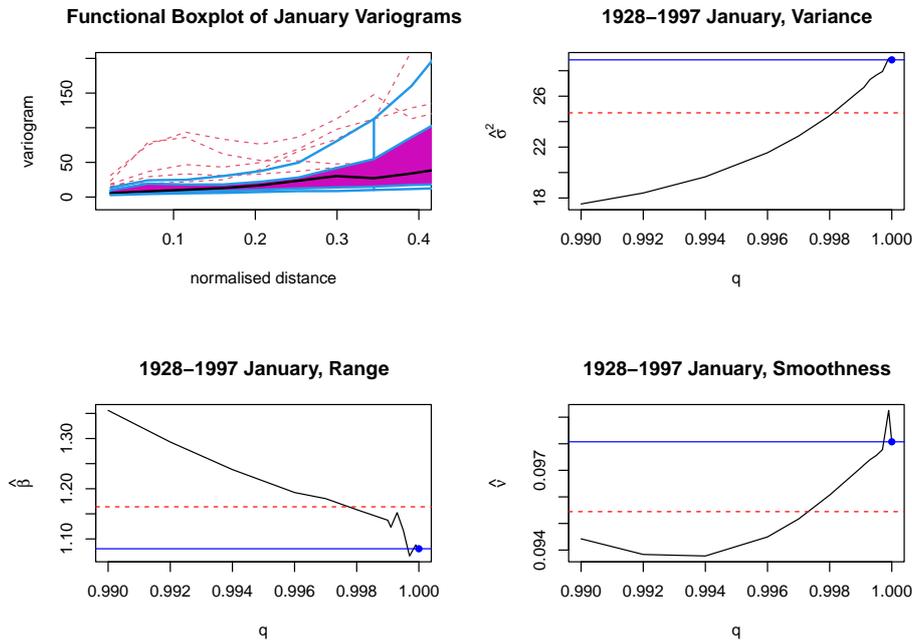


Figure 1: Plots from the motivating example of US precipitation data. In the first sub-figure we present the functional boxplot of the variograms of the January precipitation data, which we use to identify the outliers in the dataset. The remaining three sub-figures present MLE and $MLqE$ estimation results of the three parameters, variance (σ^2), range (β) and smoothness (ν), with the January data from 1928 to 1997. The black curves represent the $MLqE$ results with different values of q , and when $q = 1$ (indicated by the blue dot on the curve), the $MLqE$ is the same as the MLE. The blue horizontal lines and the red dashed lines denote the MLE results with and without the outliers in the dataset.

In this paper, we first investigate the asymptotic properties of the $MLqE$ in the scenario of Gaussian random fields with Matérn covariance function. Then, we carry out simulation studies to investigate the numerical performance of the $MLqE$ with synthetic spatial data, which demonstrate that the $MLqE$ is able to provide a more robust and stable parameter estimation in the presence of outliers. This observation is also corroborated by experiments with real data. Next, we develop a methodology to choose the optimal value of the hyper-parameter q for the $MLqE$. Since spatial data are often large-scale in practice, we implement our method using the software **ExaGeoStat**, which provides a high-performance computing framework for spatial and spatio-temporal data, making

our implementation capable of dealing with large-scale data efficiently.

The remainder of this paper is organised as follows. Section 2 describes the proposed methodology and provides some of its theoretical properties on Gaussian random fields, with some of the corresponding proofs placed in Appendix A. We also present our procedure of tuning the hyper-parameter q for the ML q E, as well as our algorithm for computing the L_q -likelihood. In Section 3, we present the data generation method and numerical results from simulation studies for synthetic data, and compare the behaviour of the ML q E and the MLE on both clean data and contaminated data, while some additional experimental results can be found in the supplementary materials. The experiments on our mechanism for choosing the optimal value of q are included in Section 3 as well. Section 4 expands the application of the ML q E on the precipitation data in the United States. Conclusions and discussions are presented in Section 5.

2. Method

2.1. Maximum L_q -likelihood estimator

Let f_1 and f_2 be two probability density functions and X be a random variable whose density exists. The q -entropy (Tsallis 1988) of f_2 with respect to f_1 is defined as

$$\mathcal{H}_q(f_1, f_2) = -\mathbb{E}_{f_1}\{L_q[f_2(X)]\}, \quad q > 0, \quad (2)$$

where

$$L_q(u) = \begin{cases} \log u, & \text{if } q = 1; \\ (u^{1-q} - 1)/(1 - q), & \text{otherwise.} \end{cases} \quad (3)$$

The Maximum L_q -Likelihood Estimator (ML q E) is defined through the empirical q -entropy, where we replace f_1 and f_2 in (2) by the empirical density and the density of the parametric model, respectively. Consider an i.i.d. sample X_1, \dots, X_n from the distribution $g(x; \boldsymbol{\theta}_0)$ with parameter $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ for some integer $p \geq 1$. The ML q E for the true parameter $\boldsymbol{\theta}_0$ is formulated as

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i=1}^n L_q\{g(X_i; \boldsymbol{\theta})\}, \quad 0 < q \leq 1. \quad (4)$$

From this definition we can see that, when $q = 1$, the ML q E is exactly the same as the MLE, because in this case the function L_q is nothing but the logarithmic function. In addition, it can be shown that, when $q \rightarrow 1$, if the ML q E $\hat{\boldsymbol{\theta}}$ exists, it approaches the MLE.

2.2. ML q E for Gaussian random fields

The main objective of this work is to apply the ML q E to spatial data with replicates. We consider m sets of independent realisations of an identical zero-mean Gaussian random field, with $m > 1$, denoted by $\{Z_i(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, where $i = 1, \dots, m$ and $d \in \mathbb{Z}^+$. Here, d denotes the dimension of the locations. For most geographical and

environmental datasets, we have $d = 2$, and this will be the setting for all the datasets that we use in the numerical experiments of this work as well. The spatial covariance function of each of the Z_i s is assumed to be the Matérn covariance function shown in (1). We consider the case in which all the m realisations of the random fields, denoted by $\{\mathbf{Z}_i, i = 1, \dots, m\}$, are recorded on a fixed set of n locations, where $n > 1$. Therefore, each \mathbf{Z}_i is an n -dimensional vector with identical zero-mean multivariate Gaussian distribution with common probability density

$$f(\mathbf{z}; \boldsymbol{\theta}) = \frac{1}{|2\pi\boldsymbol{\Sigma}_{\mathcal{M}}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1}\mathbf{z}\right), \quad \mathbf{z} \in \mathbb{R}^n, \quad (5)$$

where $\boldsymbol{\Sigma}_{\mathcal{M}}$ is the $n \times n$ covariance matrix obtained from the Matérn covariance function $\mathcal{M}(h; \boldsymbol{\theta})$ in (1) based on the n locations.

It follows that the MLqE $\hat{\boldsymbol{\theta}}$ defined in (4) for $\{\mathbf{Z}_i, i = 1, \dots, m\}$ can be expressed as

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^m L_q\{f(\mathbf{Z}_i; \boldsymbol{\theta})\}, \quad 0 < q \leq 1,$$

where the function f is as in (5) and L_q is as in (3).

Define

$$\mathbf{U}(\mathbf{z}; \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{z}; \boldsymbol{\theta}).$$

Then for $\{\mathbf{Z}_i, i = 1, \dots, m\}$, the MLqE $\hat{\boldsymbol{\theta}}$ can alternatively be represented by the solution of the following equation:

$$\sum_{i=1}^m \mathbf{U}(\mathbf{Z}_i; \hat{\boldsymbol{\theta}}) f(\mathbf{Z}_i; \hat{\boldsymbol{\theta}})^{1-q} = \mathbf{0}, \quad 0 < q \leq 1. \quad (6)$$

2.3. Asymptotic properties

In this subsection, we prove the consistency and derive the asymptotic variance of the MLqE for Gaussian random fields with Matérn covariance.

Note that the density function (5) can be rewritten as

$$f(\mathbf{z}; \boldsymbol{\theta}) = \exp\left[\{\boldsymbol{\eta}(\boldsymbol{\theta})\}^\top \mathbf{b}(\mathbf{z}) - A(\boldsymbol{\theta})\right],$$

where we define

$$\mathbf{b}(\mathbf{z}) = \begin{bmatrix} \mathbf{z} \\ \text{vec}(\mathbf{z}\mathbf{z}^\top) \end{bmatrix}, \quad \boldsymbol{\eta}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0}_n \\ \text{vec}\left(-\frac{1}{2}\boldsymbol{\Sigma}_{\mathcal{M}}^{-1}\right) \end{bmatrix},$$

and $A(\boldsymbol{\theta})$ is a normalising constant. Recall that according to (1), the dimension of $\boldsymbol{\theta}$ is 3, which is smaller than the dimension of the vector $\boldsymbol{\eta}(\boldsymbol{\theta})$ for any integer $n > 1$. Therefore, $f(\mathbf{z}; \boldsymbol{\theta})$ belongs to a curved exponential family, see p. 25 and Note 10.6 on p. 79 in Lehmann and Casella (1998).

In Section 3 of Ferrari and Yang (2010), the asymptotic distribution theory for MLqE was derived for full-rank exponential families. However, the parameters for curved exponential families estimated via the MLqE also follow similar asymptotic normality as derived in Theorem 3.2 in Ferrari and Yang (2010), and further generalised in Theorem 4.2 in the same article. Using these results, we have the following lemma and theorem.

Lemma 1. *Let $q_m > 0$ satisfy $q_m \rightarrow 1$ as $m \rightarrow \infty$, and let the value of the underlying parameter $\boldsymbol{\theta}_0$ of the sample be an interior point of the parameter space Θ , which is compact. Then, the probability that Equation (6) has a unique solution $\hat{\boldsymbol{\theta}}$ converges to 1 as $m \rightarrow \infty$ and $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ as $m \rightarrow \infty$.*

Proof. This lemma follows directly from Theorem 3.2 in Ferrari and Yang (2010). \square

Next, the following theorem states asymptotic Gaussianity of the MLqE along with the form of the asymptotic dispersion matrix. Before stating the theorem, we first define some necessary quantities for it. Let

$$\begin{aligned} \mathbf{U}_n^*(\mathbf{Z}; \boldsymbol{\theta}, q) &= \left(\frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{\mathcal{M}}|^{1/2}} \right)^{1-q} \exp \left\{ -\frac{1-q}{2} (\mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z}) \right\} \\ &\quad \times \left[\frac{1}{2} \mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z} - \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) \right], \end{aligned} \quad (7)$$

$$\begin{aligned} \mathbf{V}_n^*(\mathbf{Z}; \boldsymbol{\theta}, q) &= (1-q) \left(\frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{\mathcal{M}}|^{1/2}} \right)^{1-q} \exp \left(-\frac{1-q}{2} (\mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z}) \right) \\ &\quad \times \left[\frac{1}{2} \mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z} - \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) \right]^2 \\ &\quad + \left(\frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{\mathcal{M}}|^{1/2}} \right)^{1-q} \exp \left(-\frac{1-q}{2} (\mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z}) \right) \\ &\quad \times \left\{ \left[\frac{1}{2} \mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \left[\frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} - 2 \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \right] \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z} \right] \right. \\ &\quad \left. - \frac{1}{2} \text{tr} \left(-\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} \right) \right\}, \end{aligned} \quad (8)$$

where $\boldsymbol{\Sigma}_{\mathcal{M}}$, $\partial \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}$ and $\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}^2$ are functions of $\boldsymbol{\theta}$, but we drop the argument $\boldsymbol{\theta}$ for brevity. Here for any vector \mathbf{a} , we denote the operation $\mathbf{a}^{\otimes 2} = \mathbf{a} \mathbf{a}^\top$; in addition, $\partial \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}$ is a p -dimensional vector of matrices, and $\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}^2$ is a $p \times p$ matrix of matrices, where p is the dimension of $\boldsymbol{\theta}$. Every element of $\partial \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}$ and $\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}^2$ is an $n \times n$ matrix, and $\partial \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}$ and $\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}^2$ are defined as follows:

$$\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} = \left(\frac{\partial}{\partial \theta_1} \boldsymbol{\Sigma}_{\mathcal{M}}, \dots, \frac{\partial}{\partial \theta_p} \boldsymbol{\Sigma}_{\mathcal{M}} \right)^\top; \quad \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{\Sigma}_{\mathcal{M}} = \begin{bmatrix} \frac{\partial^2}{\partial \theta_1^2} \boldsymbol{\Sigma}_{\mathcal{M}} & \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \boldsymbol{\Sigma}_{\mathcal{M}} & \dots & \frac{\partial^2}{\partial \theta_1 \partial \theta_p} \boldsymbol{\Sigma}_{\mathcal{M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \theta_p \partial \theta_1} \boldsymbol{\Sigma}_{\mathcal{M}} & \frac{\partial^2}{\partial \theta_p \partial \theta_2} \boldsymbol{\Sigma}_{\mathcal{M}} & \dots & \frac{\partial^2}{\partial \theta_p^2} \boldsymbol{\Sigma}_{\mathcal{M}} \end{bmatrix}.$$

The matrix operations involving $\partial \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}$ and $\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}^2$ are carried out element-wise over the individual $n \times n$ matrix elements, so that $\mathbf{U}_n^*(\mathbf{Z}; \boldsymbol{\theta}, q)$ is a p -dimensional vector and $\mathbf{V}_n^*(\mathbf{Z}; \boldsymbol{\theta}, q)$ is a $p \times p$ matrix. Define $\boldsymbol{\theta}_m^*$ such that it satisfies $\mathbb{E}_{\boldsymbol{\theta}_0} [\mathbf{U}_n^*(\mathbf{Z}; \boldsymbol{\theta}_m^*, q_m)] = \mathbf{0}$. Also define $\mathbf{J}_m = \mathbb{E}_{\boldsymbol{\theta}_0} [\mathbf{V}_n^*(\mathbf{Z}; \boldsymbol{\theta}_m^*, q_m)]$, $\mathbf{K}_m = \mathbb{E}_{\boldsymbol{\theta}_0} [\mathbf{U}_n^*(\mathbf{Z}; \boldsymbol{\theta}_m^*, q_m) \mathbf{U}_n^*(\mathbf{Z}; \boldsymbol{\theta}_m^*, q_m)^\top]$. We have the following theorem:

Theorem 1. *Let $q_m \rightarrow 1$ as $m \rightarrow \infty$, and the parameter value $\boldsymbol{\theta}_0$ of the sample be an interior point of the compact parameter space Θ . Then,*

$$\sqrt{m} \left(\mathbf{J}_m^{-1} \mathbf{K}_m \mathbf{J}_m^{-1} \right)^{-1/2} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_m^* \right) \xrightarrow{d} N_p(\mathbf{0}_p, \mathbf{I}_p) \text{ as } m \rightarrow \infty.$$

If $\sqrt{m}(q_m - 1) \rightarrow 0$ as $m \rightarrow \infty$, then we have

$$\sqrt{m} \left(\mathbf{J}_m^{-1} \mathbf{K}_m \mathbf{J}_m^{-1} \right)^{-1/2} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) \xrightarrow{d} N_p(\mathbf{0}_p, \mathbf{I}_p) \text{ as } m \rightarrow \infty.$$

Here, $\mathbf{0}_p$ and \mathbf{I}_p represent the zero vector of dimension p and the identity matrix of dimension $p \times p$, respectively. The proof of this theorem, as well as the expressions of the derivatives $\partial \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}$ and $\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}^2$, are deferred to Appendix A.

2.4. Influence function

The robustness of the MLqE method was explored in Ferrari and La Vecchia (2012) for multivariate observations using the influence function. Similar results can be obtained for replicated spatial data in our setup, which we present here.

Lemma 2. *Let the influence function for $\hat{\boldsymbol{\theta}}$ be denoted by $\mathbf{IF}_q(\mathbf{z}, \boldsymbol{\theta})$. Then,*

$$\mathbf{IF}_q(\mathbf{z}, \boldsymbol{\theta}) = -q^{-1} \mathbf{J}_m^{-1} \mathbf{U}_n^*(\mathbf{z}; \boldsymbol{\theta}, q).$$

The proof of the lemma follows directly from the discussion in Section 3.2 in Ferrari and La Vecchia (2012). Note that

$$\min\{|\lambda_i|\} \|\mathbf{z}\|^2 \leq \left| \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z} \right| \leq \max\{|\lambda_i|\} \|\mathbf{z}\|^2,$$

where $\{\lambda_i\}$ are the eigenvalues of $\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} (\partial \boldsymbol{\Sigma}_{\mathcal{M}} / \partial \boldsymbol{\theta}) \boldsymbol{\Sigma}_{\mathcal{M}}^{-1}$. Since for any $q < 1$,

$$\int \|\mathbf{z}\|^2 \exp \left\{ -\frac{1-q}{2} \left(\mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z} \right) \right\} d\mathbf{z} < \infty,$$

as $\exp[-\{(1-q)/2\}(\mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z})]$ corresponds to the density of a Gaussian random vector, we have

$$\|\mathbf{z}\|^2 \exp \left\{ -\frac{1-q}{2} \left(\mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z} \right) \right\} \rightarrow 0$$

as $\|\mathbf{z}\| \rightarrow \infty$. Therefore, it follows that the influence function $\mathbf{IF}_q(\mathbf{z}, \boldsymbol{\theta})$ is bounded when $q < 1$. On the other hand, when $q = 1$, which is the case of the MLE, the exponential term in $\mathbf{U}_n^*(\mathbf{z}; \boldsymbol{\theta}, q)$ vanishes, and hence $\mathbf{IF}_q(\mathbf{z}, \boldsymbol{\theta})$ is unbounded. The boundedness can be further verified by the sensitivity curves for the MLqE, which are presented in Figures S1 and S2 in the supplementary materials.

2.5. Choice of q

Although smaller values of the hyper-parameter q can reduce the influence of outliers in the data to a larger extent, it also results in higher variance and numerical instability, as we will see from the numerical results in Sections 3.1 and 3.2. Therefore, a mechanism for finding the optimal value of q is desired. The method we apply to tune the hyper-parameter q is inspired by the one introduced in Ribeiro and Ferrari (2022), with the idea of performing an adequate grid search. The goal is to find a sub-interval of $(0, 1]$ such that all values of q taken from this sub-interval lead to similar estimation results.

For this purpose, we need to find a metric that jointly evaluates the three parameters that we estimate.

The first option is to follow the idea in [Ribeiro and Ferrari \(2022\)](#), which is to standardise the parameter estimates with their corresponding asymptotic variance, and then sum them up. Define an ordered grid of values for q , denoted by $q_0 = 1 > q_1 > q_2 > \dots > q_K > 0$, and for each q_k with $0 \leq k \leq K$, denote the corresponding p -dimensional MLqE by

$$\hat{\boldsymbol{\theta}}_{q_k} = \left(\hat{\theta}_{q_k}^1, \dots, \hat{\theta}_{q_k}^p \right)^\top,$$

and define the corresponding vector of standardised estimates $\boldsymbol{\zeta}_{q_k}$ as

$$\boldsymbol{\zeta}_{q_k} = \left(\frac{\hat{\theta}_{q_k}^1}{\sqrt{m} \text{se}(\hat{\theta}_{q_k}^1)}, \dots, \frac{\hat{\theta}_{q_k}^p}{\sqrt{m} \text{se}(\hat{\theta}_{q_k}^p)} \right)^\top,$$

where m is the number of replicates of the spatial data, and $\text{se}(\hat{\boldsymbol{\theta}})$ denotes the asymptotic standard deviation of $\hat{\boldsymbol{\theta}}$ derived in [Section 2.3](#). For $1 \leq r \leq p$, $\text{se}(\hat{\theta}_{q_k}^r)$ is the r -th diagonal entry of the matrix

$$\left(\mathbf{J}_m(\hat{\boldsymbol{\theta}}_{q_k})^{-1} \mathbf{K}_m(\hat{\boldsymbol{\theta}}_{q_k}) \mathbf{J}_m(\hat{\boldsymbol{\theta}}_{q_k})^{-1} \right)^{1/2}.$$

Furthermore, we define the standardised quadratic variation (SQV) for each q_k with $1 \leq k \leq K$ as

$$\text{SQV}_{q_k} = \frac{\|\boldsymbol{\zeta}_{q_{k-1}} - \boldsymbol{\zeta}_{q_k}\|}{p},$$

and choose the value of q based on this quantity.

The corresponding algorithm to find the optimal value of q , denoted by q^* , through SQV, is shown in [Algorithm 1](#). This algorithm basically follows what is proposed in [Ribeiro and Ferrari \(2022\)](#) and aims to find an optimal q^* such that the MLqE is stabilised in a neighborhood of q^* . However, if no optimal value of q^* can be found in this way, the algorithm still returns $q^* = 1$: we may not be able to obtain robustness in this case, so we still choose the MLE to ensure stability.

However, directly summing up the standardised values of the parameters may not be the best way to select q . Another option we consider is to make use of the function κ defined as

$$\kappa(\sigma^2, \beta, \nu) = \sigma^2 \beta^{-2\nu}, \quad (9)$$

which was proved in [Zhang \(2004\)](#) to be able to identify Matérn covariance parameters under infill asymptotics. Therefore, compared to the previous option using the SQV, the function κ is capable of providing an evaluation of the fitness of the model with the estimated parameters. Moreover, the computation time using the function κ is much faster, since computing the SQV requires computing the asymptotic standard errors of the parameters, which is extremely time-consuming, as it involves a lot of complicated large matrix operations.

Similarly to the previous option, for the parameter estimation results using each q_k with $0 \leq k \leq K$, we calculate the corresponding value of the function κ , denoted by κ_{q_k} . The algorithm for finding the optimal value of q , denoted by q^* , through the function κ , is shown in [Algorithm 2](#).

Algorithm 1: Tuning the hyper-parameter q with SQV.

Input : An ordered grid for q : $q_0 = 1 > q_1 > \dots > q_K = q_{min} > 0$, a threshold for the SQV: $L > 0$, a threshold for the difference between values of q : $\epsilon > 0$.

while $q_0 - q_{min} > \epsilon$, **do**
 Initialise $q^* \leftarrow 1$;
 Calculate SQV_{q_k} for each $k = 1, \dots, K$;
 if $\forall k, SQV_{q_k} < L$ **then**
 $q^* \leftarrow q_0$;
 break;
 end
 Let k^* be the largest integer among $1, \dots, K$ such that $SQV_{q_{k^*}} \geq L$;
 Define a new equally spaced grid for q : $q_0 = q_{k^*} > q_1 > \dots > q_K = q_{min} > 0$;
end
Output: q^*

Algorithm 2: Tuning the hyper-parameter q with the function κ .

Input : An ordered grid for q : $q_0 = 1 > q_1 > \dots > q_K = q_{min} > 0$, a threshold coefficient: $L > 0$, a threshold for the difference between values of q : $\epsilon > 0$.

while $q_0 - q_{min} > \epsilon$, **do**
 Initialise $q^* \leftarrow 1$;
 Calculate κ_{q_k} for each $k = 0, \dots, K$;
 Calculate $d\kappa_{q_k} = |\kappa_{q_{k-1}}/\kappa_{q_k} - 1|$ for each $k = 1, \dots, K$;
 if $\max_k \{d\kappa_{q_k}\} < L \cdot \min_k \{d\kappa_{q_k}\}$ **then**
 $q^* \leftarrow q_0$;
 break;
 end
 Let k^* be the largest integer among $1, \dots, K$ such that $d\kappa_{q_{k^*}} \geq L \cdot \min_k \{d\kappa_{q_k}\}$;
 Define a new equally spaced grid for q : $q_0 = q_{k^*} > q_1 > \dots > q_K = q_{min} > 0$;
end
Output: q^*

In addition to the evaluation metric, another change we make here is that instead of the fixed threshold in Algorithm 1, we apply a floating threshold based on the values of κ_{q_k} , with a threshold coefficient L . In our simulation studies in Section 3, we set the value of the coefficient L equal to 4.

In general, the results given by Algorithms 1 and 2 are quite similar, while the results from Algorithm 2 are more stable than those of Algorithm 1, as will be shown in Section 3 through a comparison. Moreover, the computation using Algorithm 2 is much faster than Algorithm 1 due to its simplicity. Therefore, in Sections 3 and 4, in which we conduct numerical experiments on both simulated and real world datasets, we tune the hyper-parameter q using Algorithm 2.

2.6. Computations

The computation of the ML q E is implemented in the software **ExaGeoStat** (Abdulah et al. 2018). To make computation easier and more compatible with the environment in **ExaGeoStat**, while evaluating the L_q -likelihood, we first evaluate the log-likelihood and perform an additional transformation on the log-likelihood if $q \neq 1$.

From (5), the log-likelihood for a realisation \mathbf{Z} of length n from a zero-mean Gaussian random field with covariance matrix $\Sigma_{\mathcal{M}}$ parametrised by $\boldsymbol{\theta}$ can be expressed as

$$l(\mathbf{Z}; \boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \mathbf{Z}^\top \Sigma_{\mathcal{M}}^{-1} \mathbf{Z} - \frac{1}{2} \log |\Sigma_{\mathcal{M}}|.$$

Furthermore, from (3), the expression of the L_q -likelihood in this case is

$$L_q(\mathbf{Z}; \boldsymbol{\theta}) = \begin{cases} l(\mathbf{Z}; \boldsymbol{\theta}), & \text{if } q = 1; \\ \{\exp[l(\mathbf{Z}; \boldsymbol{\theta}) \times (1 - q)] - 1\} / (1 - q), & \text{otherwise.} \end{cases}$$

The detailed Algorithm 3 for computing the L_q likelihood can be found in Appendix B. In addition, as Ferrari and Yang (2010) pointed out, for a fixed q , the ML q E $\hat{\boldsymbol{\theta}}$ would converge to $\boldsymbol{\theta}_0/q$ in probability; therefore, a correction of the final estimation result could be considered. However, these authors also mentioned that the numerical results after the correction are not promising, a claim that was further confirmed by our own experiments. Therefore, in all the numerical studies in this work, we directly consider the estimation results without correction to be the ML q E results.

3. Simulation Study

In this section, we conduct simulation studies on the proposed method, in which we use the software **ExaGeoStat** to both generate the data and estimate the parameters. The optimisation algorithm we use for parameter estimation is BOBYQA (Powell 2009), which is a bound-constrained algorithm without using derivatives embedded in **ExaGeoStat**.

In each of the experiments, we generate synthetic data from a zero-mean Gaussian random field with Matérn covariance matrix $\Sigma_{\mathcal{M}}$ parametrised by $\boldsymbol{\theta}$, with n locations and m replicates, and the level of contamination r . If $r = 0$, then we are generating clean data without outliers, which is used in Section 3.1; if $r > 0$, which means that the synthetic data are contaminated by outliers and is used in Section 3.2, then we generate the clean data first and add noises to parts of the data afterward. The detailed Algorithm 4 is shown in Appendix B. In the rest of this section, we show the estimation results of the variance, range and smoothness parameters in each setting. In addition, we plot the estimated value of the function κ in (9), to evaluate the fitness of the model with the estimated parameters. Throughout this section and the supplementary materials, in the figures in which we show our experimental results, the “function κ ” refers to the function κ in (9). Moreover, for all the simulation experiments in this section, we tune the value of q using Algorithm 2.

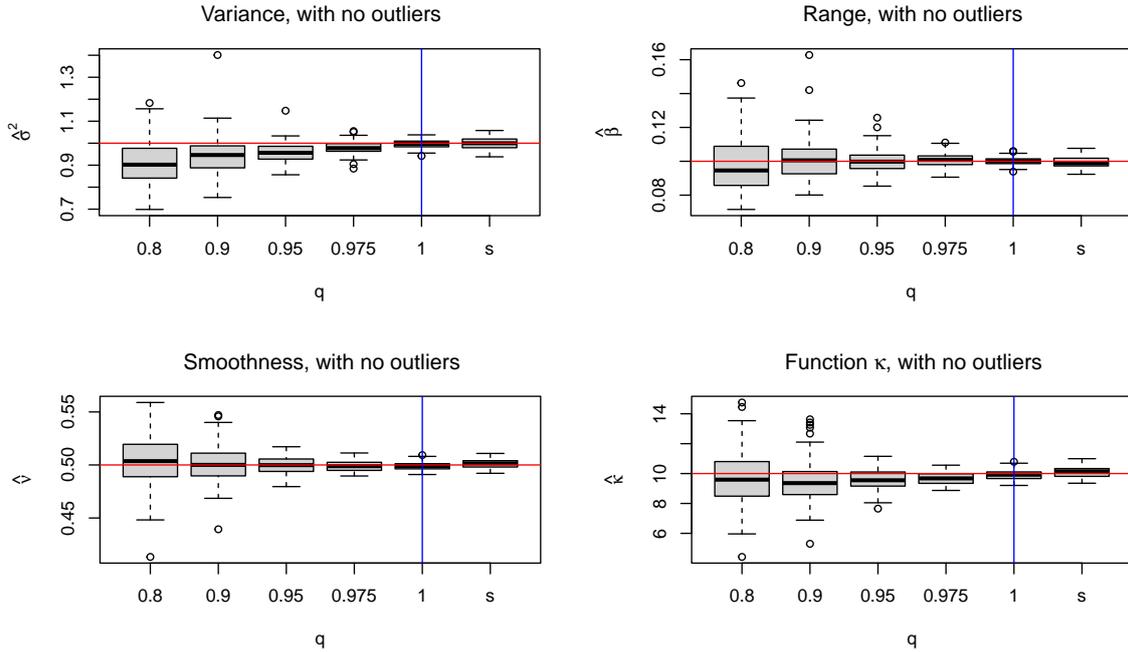


Figure 2: The MLE and $MLqE$ estimation results with no outlier in the data. The red horizontal lines correspond to the true values of the parameters or the function κ in (9). The blue vertical lines correspond to the value of q leading to the smallest mean squared error (MSE), which in this case is $q = 1$ (MLE).

3.1. Clean data

First of all, we performed some experiments to compare the performance of $MLqE$ with MLE when there are no outliers in the simulated data.

In Figure 2, we show the experimental results with simulated data from the Matérn covariance function with $\sigma^2 = 1, \beta = 0.1, \nu = 0.5$, which is essentially the exponential covariance function with a medium spatial dependence strength. Here we simulate the data with $n = 1,600$ locations and $m = 100$ replicates, and the experiment is repeated on 100 different datasets generated from the same random field to make the boxplots. We use red horizontal lines to indicate the true values of the parameters and the function κ in (9), and blue vertical lines correspond to the value of q leading to the smallest mean squared error (MSE). We can see that in this case the MLE is giving the best performance in terms of both consistency and efficiency. We also notice that when q is close to 1, the $MLqE$ is converging to the MLE as expected.

However, in real applications, we do not have the luxury of performing parameter estimation on different datasets from the same random field to find the value of q with the smallest MSE. Therefore, we also conduct choice-of- q experiments using Algorithm 2 for each of the 100 datasets individually, to examine whether the optimal value of q can also be found in this way. In the last column of each panel of Figure 2, denoted by “s”, we present the boxplot of the parameter estimation results using the values of q chosen by Algorithm 2, which have slightly larger bias than the MLE results with $q = 1$. This is also what we expect as the compromise of using a robust method, i.e., that it results in slightly larger variability than the MLE when applied on clean data.

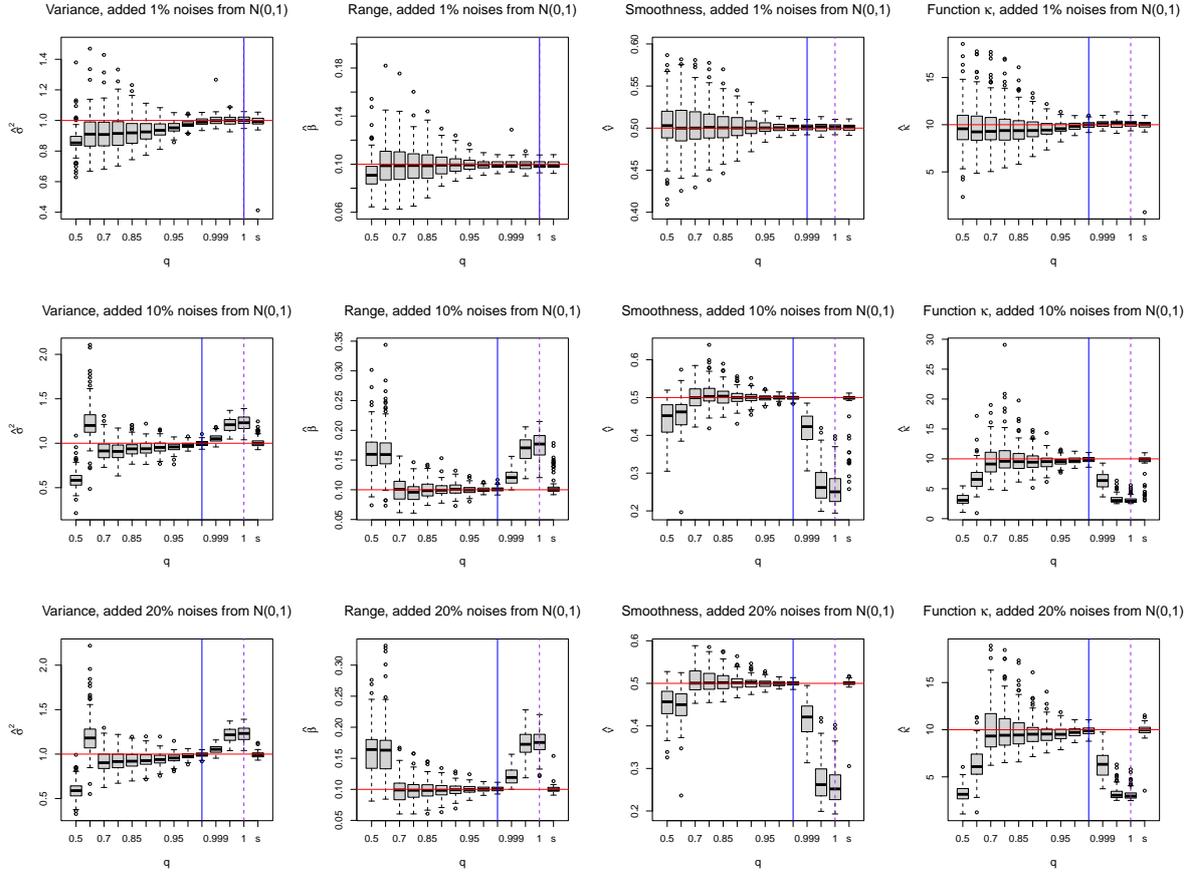


Figure 3: The MLE and $MLqE$ estimation results with the data contaminated by noises generated from $N(0,1)$, with 1%, 10% and 20% probability respectively from the first row to the third. The last column of each panel with the label “s” indicates the estimation results using the q values selected using Algorithm 2. The values of q we use here are 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.925, 0.95, 0.975, 0.99, 0.999, 0.9999, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in (9), and the blue vertical lines correspond to the value of q leading to the smallest MSE. The MLE ($q = 1$) results are indicated by the purple vertical dashed lines. In the first row, the q values leading to the smallest MSE for variance and range are 1, and in the corresponding figures the blue and purple lines overlap.

3.2. Contaminated data

Next, we test the performance of the $MLqE$ and compare it with the MLE when there are different types of outliers in the simulated data. For all the results shown here in Figures 3 and 4, the data are generated from a Gaussian random field with the Matérn covariance function with $\sigma^2 = 1, \beta = 0.1, \nu = 0.5$. Here also, we simulate the data with $n = 1,600$ locations and $m = 100$ replicates, and the experiment is repeated on 100 different datasets generated from the same random field to make the boxplots. The mechanism of contaminating the data is as described in Algorithm 4. However, since we only have 100 replicates for each dataset, in the experiments in which the data are contaminated with probability 1% (see the first rows of Figures 3 and 4) we actually

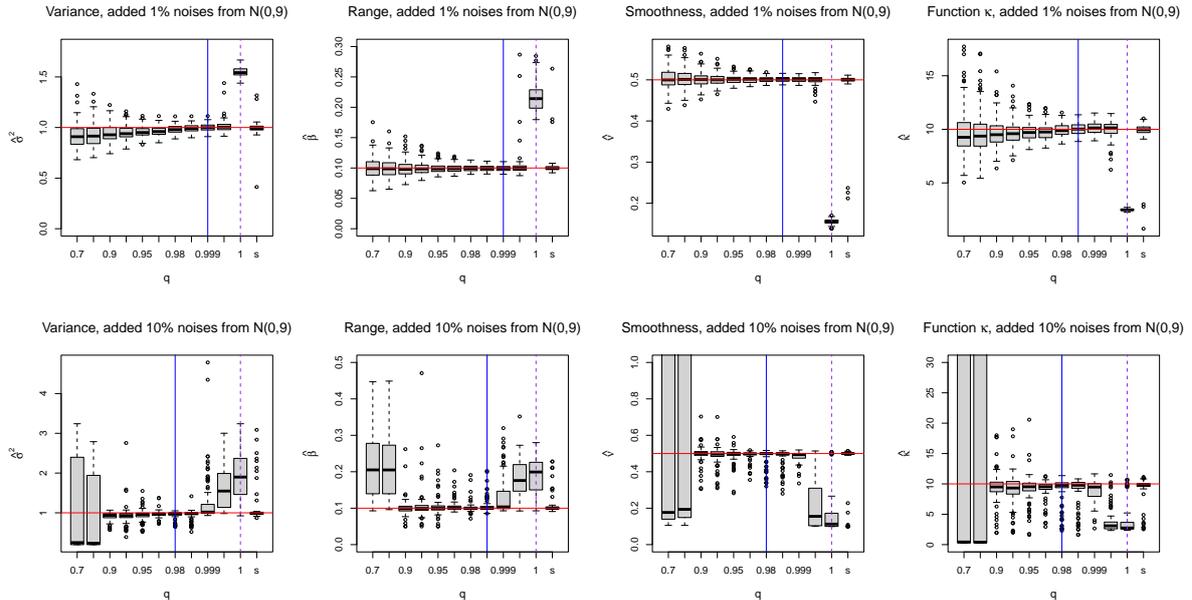


Figure 4: The MLE and $MLqE$ estimation results with the data contaminated by noises generated from $N(0, 9)$, with 1% probability in the first row and 10% in the second. The last column of each panel with the label “s” indicates the estimation results using the q values selected using Algorithm 2. The values of q we use here are 0.7, 0.8, 0.9, 0.925, 0.95, 0.96, 0.98, 0.99, 0.999, 0.9999, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in (9), and the blue vertical lines correspond to the value of q leading to the smallest MSE. The MLE ($q = 1$) results are indicated by the purple vertical dashed lines.

randomly select exactly 1 of the 100 replicates to be contaminated.

In the first row of Figure 3, where we only add noises from $N(0, 1)$ to 1% of the data, the MLE still performs well. For the variance and the range parameters, our hyperparameter tuning method suggests that $q = 1$ is the best choice. However, as we can see from the rest of Figures 3 and 4, when we increase the proportion of data with noises from $N(0, 1)$ to 10% or 20%, the MLE results deviate significantly from the true values. Nevertheless, even if q is only slightly smaller than 1, the $MLqE$ results are much closer to the true values, and $q = 0.99$ is suggested to be the best choice for these two cases. For data with outliers generated by adding noises from $N(0, 9)$ with 1% or 10% probability, as we show in Figure 4, since the magnitude of outliers is much larger, smaller values of q are needed to provide better $MLqE$ results, compared with the data with added noises from $N(0, 1)$ with the same probability. In addition, as before, we use red horizontal lines to indicate the true values of the parameters and the function κ in (9), and blue vertical lines correspond to the value of q leading to the smallest MSE. The MLE ($q = 1$) results are indicated by the vertical purple dashed lines.

Furthermore, similar to Figure 2, in Figures 3 and 4, we still add one column to the right of each panel labeled with “s”, to present the boxplot of the parameter estimation results using the values of q chosen by Algorithm 2. The choice-of- q experiments are conducted on each of the 100 datasets individually. We notice that in all three of these

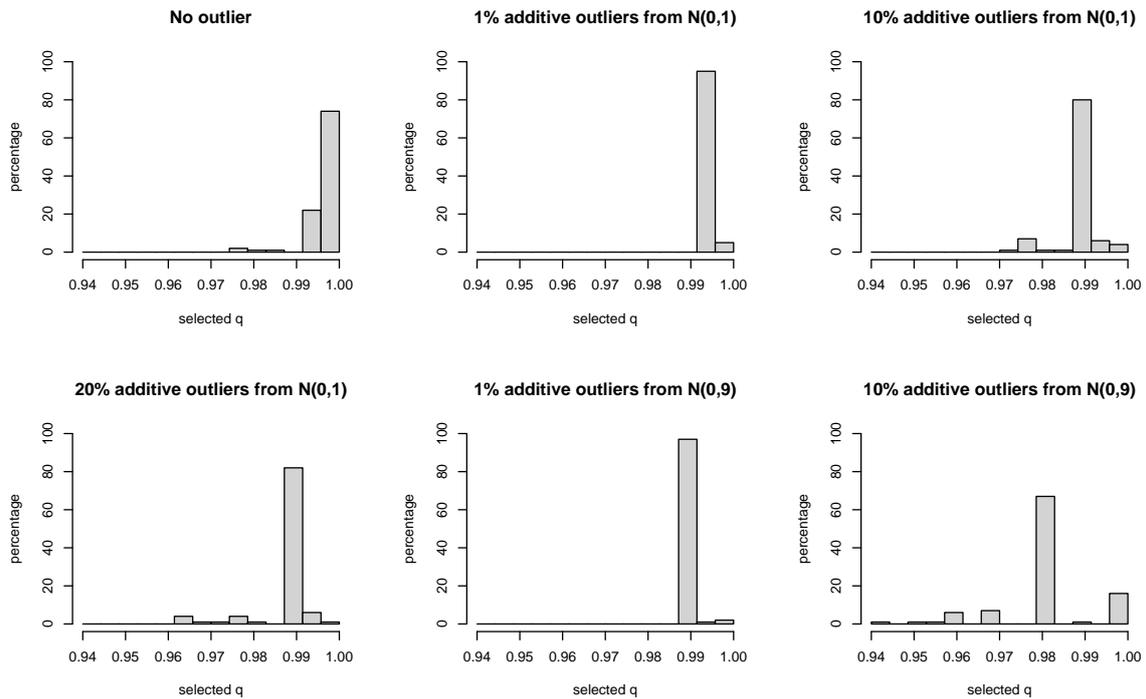


Figure 5: The histograms of the values of q chosen by Algorithm 2, with the same synthetic data as in the experiments shown in Figures 2 to 4. The y -axes indicate the frequency (in percentage) of the selected q .

cases, for most simulated datasets, the estimation results using the chosen q are similar to the results using the value of q with the smallest values of MSE indicated by the blue vertical lines. In Figure 4, since outliers in the datasets are generated in a relatively extreme way, it is more often that the optimal value of q is not chosen, compared to Figure 3. Nevertheless, despite those several extreme cases, Algorithm 2 is working well and the resulting parameter estimation results are very close to the true values.

Experimental results with synthetic datasets from some other values of the true parameters (i.e., other strengths of dependence and smoothness) are presented in Figures S3 to S14 in the supplementary materials. In fact, the behaviour of the ML q E in those experiments is similar to the results shown here in Figures 2 to 4. For clean data, the MLE has the lowest bias and variability; the bias of the ML q E is also quite low as long as q is not too small (generally not less than 0.95), while it is not as good as the MLE in terms of efficiency. For contaminated data, the MLE is always biased for all cases, while the ML q E with q between 0.97 and 0.99 generally gives fairly unbiased estimation results.

3.3. Chosen values of q

In addition to the boxplots, we also present histograms of the frequency (in percentage) of the chosen values of q using Algorithm 2 in Figure 5, with the same datasets that we used to plot Figures 2 to 4.

For clean data, as expected, $q = 1$ is chosen for most simulated datasets, and among the

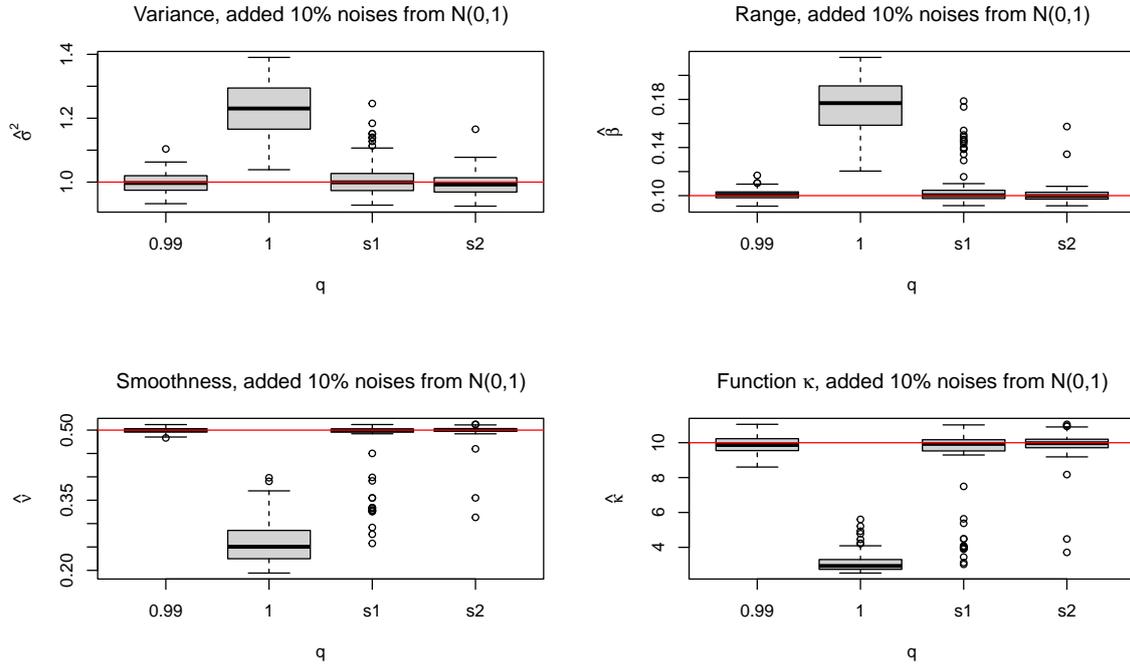


Figure 6: Comparison of the MLE and ML q E of several different cases. In each of the four sub-figures, from the left to the right, the four boxplots represent the ML q E of the smallest MSE in the second row of Figure 3, the MLE ($q = 1$), the ML q E with q selected by Algorithm 1 (“s1”), and the ML q E with q selected by Algorithm 2 (“s2”), respectively. The red horizontal lines correspond to the true values of the parameters or the function κ in (9).

rest, $q > 0.99$ is chosen for most cases, with $q < 0.99$ chosen only a very small number of times. For contaminated data, in each of the five cases, the most frequently chosen value of q is less than 1, which means the ML q E outperforms the MLE. When the dataset is contaminated by noises from $N(0, 9)$, the most frequently chosen value of q would be smaller than for the dataset contaminated by the same fraction of noises from $N(0, 1)$. Moreover, the algorithm to tune q is designed in such a way that it chooses 1 if the ML q E results are too unstable, which explains why in the most extreme case where the data contain 10% outliers generated by adding noises from $N(0, 9)$, $q = 1$ is chosen more frequently than the other cases of contaminated data. Those datasets for which $q = 1$ is chosen also correspond to most of the outlying points in the columns labeled “s” in the second row of Figure 4.

In addition, we present a comparison between the two methods of choosing q , namely Algorithms 1 and 2, in Figure 6, using the same data as the experiment shown in the second row of Figure 3. In Figure 6, we show the parameter estimation results with the values of q chosen by Algorithm 1 (in the columns named “s1”) and Algorithm 2 (in the columns named “s2”), together with the MLE results ($q = 1$) and ML q E results with $q = 0.99$, which is the q leading to the smallest MSE in the second row of Figure 3. The red horizontal lines correspond to the true values of the parameters or the function κ in (9). Both Algorithms 1 and 2 are able to give quite unbiased estimation results, while the results in “s2” are more stable than in “s1”, since there are many fewer outlying

points. Further considering the massive advantage on computation speed of Algorithm 2 over Algorithm 1, in practice, it is recommended to apply Algorithm 2 to tune the hyper-parameter q .

4. Precipitation Data Analysis

We return to the US precipitation data we used as a motivating example in Section 1. To eliminate the influence of missing data on our experiment, we select the data from 1928 to 1997 on $n = 621$ observation stations whose data do not contain any missing values for those $m = 70$ years. In addition, we extract the data of each month in the dataset and study them separately, making 12 sub-datasets. For example, for the experiment labeled “Jan”, we used only the precipitation data of January from 1928 to 1997 at all observation stations. In this way, it is reasonable to assume that all the replicates (i.e., data from different years) in each experiment are from the same distribution. Moreover, to remove the trend in precipitation, the data for each year are centralised separately.

In Figure 7, we present the experimental results using the data. In each of the subplots, the results corresponding to three consecutive months are presented for compactness. The triplets of the months are formed based on seasons: December, January and February correspond to Winter, March, April and May correspond to Spring, June, July and August correspond to Summer, and finally September, October and November correspond to Autumn. In the first 3 columns, we show the estimation results using ML q E with different values of q , while it degenerates to the MLE when $q = 1$; in the last column, we show the estimated value of the function κ in (9). The values of q that we use here are 0.9, 0.925, 0.95, 0.96, 0.97, 0.98, 0.99, 0.999, 0.9999, and 1. The dot on each curve shows the optimal value of q selected for the data using Algorithm 2.

We can see from Figure 7 that, for all 12 months, the ML q E results with most values of $q < 1$ are different from the corresponding MLE results. This observation indicates that outliers exist in the 12 sub-datasets of the 12 months; otherwise the ML q E results with q slightly smaller than 1 are expected to be similar to the MLE results. To examine whether there are really outliers in all the 12 sub-datasets, we apply the same method that we used to detect the outliers in the January data that we described in Section 1, which is to first find the empirical variogram of the data from each year, and then find out the years for which the data behave differently from other years via functional boxplots using the R command `fbplot`. It turns out that outliers exist in all of the 12 sub-datasets, which explains why the MLE and ML q E results with $q < 1$ are different for all the 12 months. In addition, we note that for the October data, the optimal selected value of q is 1, because the ML q E results, especially for the range parameter, cannot stabilise as q decreases, and the MLE is considered as the optimal choice to ensure that we still have a reasonable estimation result. For the data for each of the other months, the optimal values of q are all selected as a point such that the curve of the estimation results becomes relatively flat in a neighborhood of it, which means that stable ML q E results can be achieved using the values q close to this value. We also present the ML q E results using the selected value of q , together with the MLE results, in Table 1. Except for October, where the ML q E and MLE are identical because $q = 1$ is chosen, the MLE overestimates the variance for all other months due

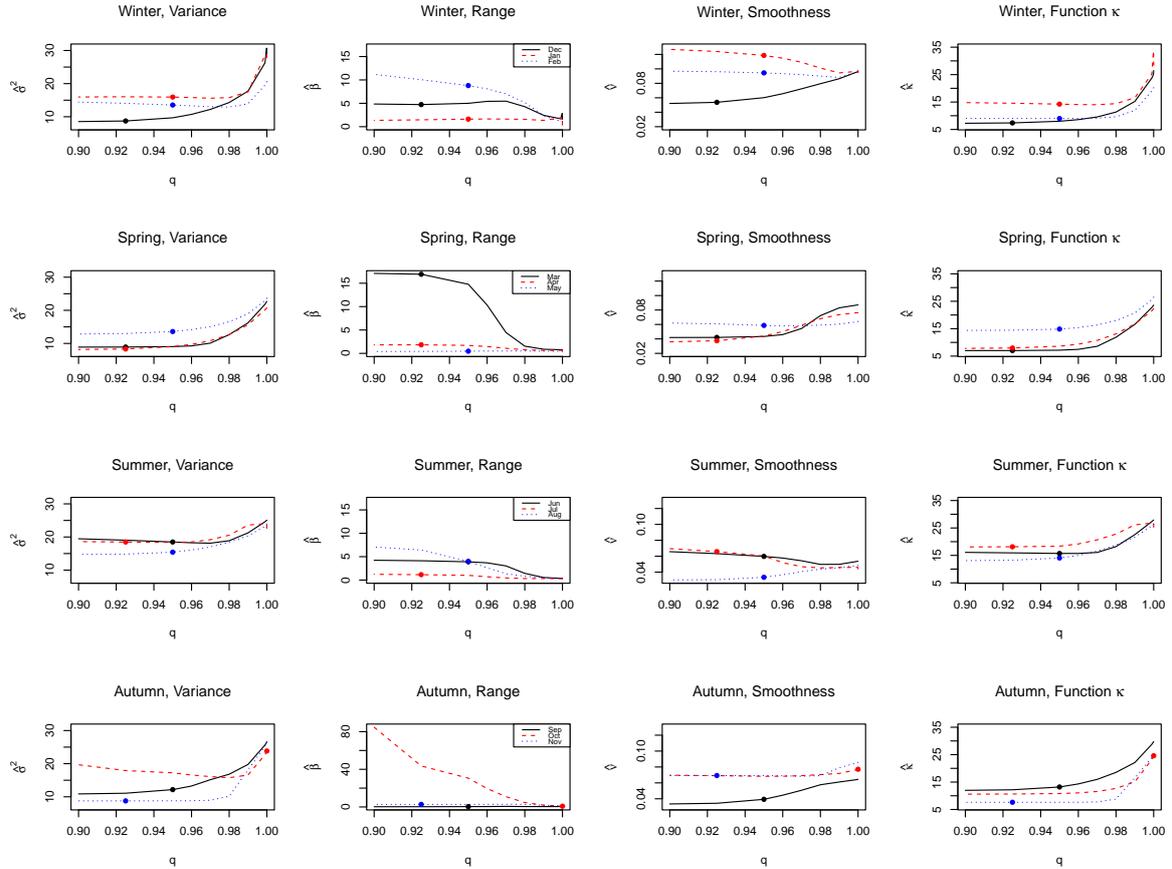


Figure 7: The MLE and $MLqE$ estimation results with different values of q , for the US precipitation data of different months. Each column shows the plot of the estimation of the same parameter or the function κ in (9), and each row shows the results for the same months. The legends in the second column of each row apply to all sub-figures in the same row. The dots on the curves indicate the values of q selected with Algorithm 2 and their corresponding estimation results.

to the presence of outliers in the datasets. Meanwhile, the MLE often underestimates both the range and smoothness of the Matérn covariance function. For the 11 months where the $MLqE$ differs from the MLE, the range parameter is underestimated for 10 months and the smoothness is underestimated for 8 months. Outliers generally have higher variance compared to non-outliers, and the noise in the data reduces the range and smoothness of the random field. Consequently, the non-robust MLE tends to overestimate variance while underestimating range and smoothness. In contrast, the robustness of the $MLqE$ provides estimates that more accurately reflect the true spatial dependence of the datasets.

5. Conclusion

To conclude, both simulation and real data experiments have shown that the $MLqE$ can indeed outperform the MLE with the existence of outliers in the data. With a suitable

Table 1: The ML q E results of the three parameters of Matérn covariance function (1), σ^2 , β and ν , for the US precipitation data of all the 12 months, using the values of q selected via Algorithm 2, followed by the MLE results of the same data.

	q selected	$\hat{\sigma}^2$	MLE	$\hat{\beta}$	MLE	$\hat{\nu}$	MLE
Jan	0.95	15.94	28.85	1.614	1.080	0.119	0.098
Feb	0.95	13.54	20.57	8.797	1.099	0.094	0.096
Mar	0.925	8.908	22.69	16.92	0.799	0.042	0.087
Apr	0.925	8.329	20.73	1.835	0.621	0.037	0.076
May	0.95	13.59	23.62	0.467	0.408	0.059	0.064
Jun	0.95	18.49	25.09	3.906	0.371	0.060	0.053
Jul	0.925	18.61	22.62	1.230	0.283	0.069	0.044
Aug	0.95	15.42	23.92	3.998	0.360	0.034	0.048
Sep	0.95	12.16	26.46	0.354	0.401	0.039	0.064
Oct	1	23.82	23.82	0.810	0.810	0.077	0.077
Nov	0.925	8.765	26.70	2.712	1.395	0.069	0.085
Dec	0.925	8.725	28.93	4.730	1.580	0.054	0.095

choice of q , the parameter estimation results via the ML q E exhibit very low bias and high efficiency, compared to the MLE, when the data are contaminated. On the theory side, the ML q E holds desirable asymptotic properties as well. The mechanism of tuning the hyper-parameter q for the ML q E that we developed was also proved to work very well in our experiments. Moreover, our method can efficiently deal with data on a much larger scale than in the literature, with the help of the high-performance computing framework **ExaGeoStat**, which is extremely important for spatial data analysis.

A limitation of the current work is that the current method is only able to provide robust estimation for spatial data with multiple replicates, since it can only treat some of the replicates as outliers as a whole. An interesting follow-up work would be investigating how to adjust this method so that it can be applied on spatial data with only one replicate as well. Another possible extension of this work would be integrating the proposed method with high-performance approximation methods for numerical linear algebra, such as the tile low-rank algorithm and the mixed-precision method, so that the ML q E can be further applied on even larger-scale datasets efficiently.

There are also several other possible extensions of the method presented in this work. First, for convenience, here we only considered the case in which all the replicates of the spatial data are from the same fixed set of locations. However, it is also possible to apply the proposed method on replicated spatial datasets where the locations for each of the replicates are not exactly the same, given that they are still from the same region so that we are able to assume that those replicated data are essentially from the same Gaussian random field. Second, we did not include the nugget effect in the model. It is also possible to estimate the nugget effect together with the three parameters in the Matérn covariance function that we considered, with some minor modifications of the code. However, for the choice-of- q mechanism, since the nugget effect does not appear in the function κ in (9), Algorithm 2 may not work that well in this case. Furthermore, the methodology presented here can also be potentially

adapted to estimate the parameters of non-Gaussian random fields, such as t -random fields (Røislien and Omre 2006) or Tukey g -and- h random fields (Xu and Genton 2017). These processes are useful to model non-Gaussian heavy-tailed or skewed spatial data. In the present work, however, we concentrated on the more popular Gaussian process with Matérn covariance function. In future work, the performance of this methodology for non-Gaussian processes can be explored.

Acknowledgements

This publication is based upon work supported by King Abdullah University of Science and Technology Research Funding (KRF) under Award No. ORFS-2022-CRG11-5069.

References

- Abdulah, S., Ltaief, H., Sun, Y., Genton, M. G., and Keyes, D. E. (2018). **ExaGeoStat**: A high performance unified software for geostatistics on manycore systems. *IEEE Transactions on Parallel and Distributed Systems*, 29(12):2771–2784, DOI: [10.1109/tpds.2018.2850749](https://doi.org/10.1109/tpds.2018.2850749).
- Aldrich, J. (1997). R.A. Fisher and the making of maximum likelihood 1912–1922. *Statistical Science*, 12(3):162–176, DOI: [10.1214/ss/1030037906](https://doi.org/10.1214/ss/1030037906).
- Berli, P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, 22(3):341–345, DOI: [10.1093/bioinformatics/bti803](https://doi.org/10.1093/bioinformatics/bti803).
- Casella, G. and Berger, R. L. (2024). *Statistical Inference*. Chapman and Hall/CRC, 2nd edition, DOI: [10.1201/9781003456285](https://doi.org/10.1201/9781003456285).
- Chen, W., Shi, J., Qian, L., and Azen, S. P. (2014). Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: A simulation study. *BMC Medical Research Methodology*, 14:1–8, DOI: [10.1186/1471-2288-14-82](https://doi.org/10.1186/1471-2288-14-82).
- Cramer, J. S. (1989). *Econometric Applications of Maximum Likelihood Methods*. CUP Archive, DOI: [10.1017/cbo9780511572050](https://doi.org/10.1017/cbo9780511572050).
- Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, 12:115–125, DOI: [10.1007/bf01035243](https://doi.org/10.1007/bf01035243).
- Elhorst, J. P. (2005). Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis*, 37(1):85–106, DOI: [10.1111/j.1538-4632.2005.00577.x](https://doi.org/10.1111/j.1538-4632.2005.00577.x).
- Ferrari, D. and La Vecchia, D. (2012). On robust estimation via pseudo-additive information. *Biometrika*, 99(1):238–244, DOI: [10.1093/biomet/asr061](https://doi.org/10.1093/biomet/asr061).

- Ferrari, D. and Yang, Y. (2010). Maximum L_q -likelihood estimation. *The Annals of Statistics*, 38(2):753–783, DOI: [10.1214/09-AOS687](https://doi.org/10.1214/09-AOS687).
- Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of Spatial Statistics*. CRC press, DOI: [10.1201/9781420072884](https://doi.org/10.1201/9781420072884).
- Genton, M. G. (1998). Highly robust variogram estimation. *Mathematical Geology*, 30:213–221, DOI: [10.1023/A:1021728614555](https://doi.org/10.1023/A:1021728614555).
- Greene, W. H. (1980). Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics*, 13(1):27–56, DOI: [10.1016/0304-4076\(80\)90041-x](https://doi.org/10.1016/0304-4076(80)90041-x).
- Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35(4):403–410, DOI: [10.2307/1270273](https://doi.org/10.2307/1270273).
- Hawkins, D. M. and Cressie, N. (1984). Robust kriging — a proposal. *Journal of the International Association for Mathematical Geology*, 16:3–18, DOI: [10.1007/bf01036237](https://doi.org/10.1007/bf01036237).
- Katzfuss, M. and Guinness, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statistical Science*, 36(1):124–141, DOI: [10.1214/19-sts755](https://doi.org/10.1214/19-sts755).
- Künsch, H., Papritz, A. J., Schwierz, C., and Stahel, W. A. (2013). Robust estimation of the external drift and the variogram of spatial data. In *ISI 58th World Statistics Congress of the International Statistical Institute*. Eidgenössische Technische Hochschule Zürich. <https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/154005/eth-7080-01.pdf>
- Lark, R. (2000). A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*, 51(1):137–157, DOI: [10.1046/j.1365-2389.2000.00280.x](https://doi.org/10.1046/j.1365-2389.2000.00280.x).
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer Texts in Statistics, 1st edition, DOI: [10.1007/b98854](https://doi.org/10.1007/b98854).
- Marchant, B. and Lark, R. (2007). Robust estimation of the variogram by residual maximum likelihood. *Geoderma*, 140(1-2):62–72, DOI: [10.1016/j.geoderma.2007.03.005](https://doi.org/10.1016/j.geoderma.2007.03.005).
- Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1):135–146, DOI: [10.2307/2336405](https://doi.org/10.2307/2336405).
- Matérn, B. (1960). Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden fran Statens Skogsforskningsinstitut*, 49(5):144. <https://www.cabidigitallibrary.org/doi/full/10.5555/19600602215>.
- Nychka, D., Furrer, R., Paige, J., Sain, S., Gerber, F., Iverson, M., Johnson, R. (2021). **fields**: Tools for spatial data. DOI: [10.32614/CRAN.package.fields](https://doi.org/10.32614/CRAN.package.fields).

- Porcu, E., Bevilacqua, M., Schaback, R., and Oates, C. J. (2024). The Matérn model: A journey through statistics, numerical analysis and machine learning. *Statistical Science*, 39(3):469–492, DOI: [10.1214/24-sts923](https://doi.org/10.1214/24-sts923).
- Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06*, University of Cambridge, Cambridge, 26. <https://optimization-online.org/?p=11137>.
- Ree, R. H. and Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57(1):4–14, DOI: [10.1080/10635150701883881](https://doi.org/10.1080/10635150701883881).
- Ribeiro, T. K. and Ferrari, S. L. (2022). Robust estimation in beta regression via maximum L_q -likelihood. *Statistical Papers*, pages 321–353, DOI: [10.1007/s00362-022-01320-0](https://doi.org/10.1007/s00362-022-01320-0).
- Ribeiro Jr, P. J. and Diggle, P. J. (2007). The **geoR** package. *R news*, 1(2):14–18, DOI: [10.32614/CRAN.package.geoR](https://doi.org/10.32614/CRAN.package.geoR).
- Richardson, A. M. and Welsh, A. H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics*, pages 1429–1439, DOI: [10.2307/2533273](https://doi.org/10.2307/2533273).
- Røislien, J. and Omre, H. (2006). T-distributed random fields: A parametric model for heavy-tailed well-log data 1. *Mathematical Geology*, 38:821–849, DOI: [10.1007/s11004-006-9050-z](https://doi.org/10.1007/s11004-006-9050-z).
- Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O. (2013). **RandomFields**: Simulation and analysis of random fields. *R Package Version*, 2:66. <https://rdrr.io/cran/RandomFields/>
- Shaw, R. G. (1987). Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution*, 41(4):812–826, DOI: [10.2307/2408890](https://doi.org/10.2307/2408890).
- Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334, DOI: [10.1198/jcgs.2011.09224](https://doi.org/10.1198/jcgs.2011.09224).
- Tsallis, C. (1988). Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 52:479–487, DOI: [10.1007/bf01016429](https://doi.org/10.1007/bf01016429).
- Wang, K., Abdulah, S., Sun, Y., and Genton, M. G. (2023). Which parameterization of the Matérn covariance function? *Spatial Statistics*, 58:100787, DOI: [10.1016/j.spasta.2023.100787](https://doi.org/10.1016/j.spasta.2023.100787).
- Xu, G. and Genton, M. G. (2017). Tukey g -and- h random fields. *Journal of the American Statistical Association*, 112(519):1236–1249, DOI: [10.1080/01621459.2016.1205501](https://doi.org/10.1080/01621459.2016.1205501).
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, DOI: [10.1198/016214504000000241](https://doi.org/10.1198/016214504000000241).

A. Proofs

A.1. Proof of Theorem 1

Note that

$$\begin{aligned}\log f(\mathbf{z}; \boldsymbol{\theta}) &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_{\mathcal{M}}| - \frac{1}{2} \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z}, \\ \frac{\partial}{\partial \boldsymbol{\theta}} \log |\boldsymbol{\Sigma}_{\mathcal{M}}| &= \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right), \\ \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z} &= \mathbf{z}^\top \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \mathbf{z} = -\mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z}.\end{aligned}$$

Hence, we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{z}; \boldsymbol{\theta}) = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) + \frac{1}{2} \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z}.$$

Therefore, we let

$$\begin{aligned}\mathbf{U}_n^*(\mathbf{Z}; \boldsymbol{\theta}, q) &= \frac{\partial}{\partial \boldsymbol{\theta}} L_q[f(\mathbf{Z}; \boldsymbol{\theta})] = \frac{\partial}{\partial \boldsymbol{\theta}} \frac{f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) - 1}{1-q} = f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{Z}; \boldsymbol{\theta}) \\ &= \left(\frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{\mathcal{M}}|^{1/2}} \right)^{1-q} \exp \left(-\frac{1-q}{2} (\mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z}) \right) \\ &\quad \left[\frac{1}{2} \mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z} - \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) \right].\end{aligned}\quad (10)$$

From (10) we get that the expression of $\mathbf{U}_n^*(\mathbf{Z}; \boldsymbol{\theta}, q)$, which is the same as (7) in the main text of this article, and it matches the quantity $\mathbf{U}^*(X; \theta, q)$ defined in (2.6) in Ferrari and Yang (2010). Therefore, $\boldsymbol{\theta}_m^*$ defined here corresponds to θ_n^* defined in (3.2) in Ferrari and Yang (2010).

Next,

$$\begin{aligned}\frac{\partial^2}{\partial \boldsymbol{\theta}^2} L_q[f(\mathbf{Z}; \boldsymbol{\theta})] &= \left(\frac{\partial}{\partial \boldsymbol{\theta}} f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) \right) \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{Z}; \boldsymbol{\theta}) + f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log f(\mathbf{Z}; \boldsymbol{\theta}) \\ &= (1-q) f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{Z}; \boldsymbol{\theta}) \right)^2 + f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log f(\mathbf{Z}; \boldsymbol{\theta}).\end{aligned}$$

We proceed to derive the expression of $\partial^2 \log f(\mathbf{Z}; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}^2$. Note that

$$\begin{aligned}\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log |\boldsymbol{\Sigma}_{\mathcal{M}}| &= \frac{\partial}{\partial \boldsymbol{\theta}} \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) = \text{tr} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) \right) \\ &= \text{tr} \left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{\Sigma}_{\mathcal{M}} \right), \\ \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}} \mathbf{z} &= \mathbf{z}^\top \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \mathbf{z} = -\mathbf{z}^\top \left(\frac{\partial}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \right) \mathbf{z},\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) &= \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \\
&= \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \\
&= -2 \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \\
&= \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \left[\frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} - 2 \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \right] \boldsymbol{\Sigma}_{\mathcal{M}}^{-1}, \\
\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}} \mathbf{z} &= -\mathbf{z}^\top \left(\frac{\partial}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \right) \mathbf{z} \\
&= \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \left[-\frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} + 2 \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \right] \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log f(\mathbf{z}; \boldsymbol{\theta}) &= -\frac{1}{2} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log |\boldsymbol{\Sigma}_{\mathcal{M}}| - \frac{1}{2} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}} \mathbf{z} \\
&= -\frac{1}{2} \text{tr} \left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{\Sigma}_{\mathcal{M}} \right) \\
&\quad - \frac{1}{2} \mathbf{z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \left[-\frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} + 2 \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \right] \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{z},
\end{aligned}$$

and it follows that

$$\begin{aligned}
&\frac{\partial^2}{\partial \boldsymbol{\theta}^2} L_q[f(\mathbf{Z}; \boldsymbol{\theta})] \\
&= (1-q) f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{Z}; \boldsymbol{\theta}) \right)^2 + f^{1-q}(\mathbf{Z}; \boldsymbol{\theta}) \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log f(\mathbf{Z}; \boldsymbol{\theta}) \\
&= (1-q) \left(\frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{\mathcal{M}}|^{1/2}} \right)^{1-q} \exp \left(-\frac{1-q}{2} (\mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z}) \right) \\
&\quad \times \left[\frac{1}{2} \mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z} - \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}} \right) \right]^2 \\
&\quad + \left(\frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{\mathcal{M}}|^{1/2}} \right)^{1-q} \exp \left(-\frac{1-q}{2} (\mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z}) \right) \\
&\quad \times \left\{ \left[\frac{1}{2} \mathbf{Z}^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \left[\frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} - 2 \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \right] \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \mathbf{Z} \right] \right. \\
&\quad \left. - \frac{1}{2} \text{tr} \left(-\boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}} + \boldsymbol{\Sigma}_{\mathcal{M}}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}_{\mathcal{M}}}{\partial \boldsymbol{\theta}^2} \right) \right\}.
\end{aligned}$$

Therefore, the expressions of \mathbf{K}_m and \mathbf{J}_m defined in the statement of the theorem match with K_n and J_n defined in (3.4) and (3.5) in [Ferrari and Yang \(2010\)](#), respectively, which completes the proof with reference to Theorems 3.2 and 4.3 in [Ferrari and Yang \(2010\)](#).

A.2. Expressions of derivatives

For practical use of Theorem 1, we need to evaluate $\partial \Sigma_{\mathcal{M}} / \partial \theta$ and $\partial^2 \Sigma_{\mathcal{M}} / \partial \theta^2$. Note that $\partial \Sigma_{\mathcal{M}} / \partial \theta$ and $\partial^2 \Sigma_{\mathcal{M}} / \partial \theta^2$ are arrays consisting of the second-order derivatives of the terms of $\Sigma_{\mathcal{M}}$ with respect to θ . The terms of $\Sigma_{\mathcal{M}}$ are of the form as in (1).

Then we have

$$\begin{aligned} \frac{\partial \mathcal{M}(h; \theta)}{\partial \sigma^2} &= \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right); \\ \frac{\partial^2 \mathcal{M}(h; \theta)}{\partial (\sigma^2)^2} &= 0; \\ \frac{\partial \mathcal{M}(h; \theta)}{\partial \beta} &= \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left\{ -\frac{\nu}{\beta} \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) - \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}'_{\nu} \left(\frac{h}{\beta}\right) \frac{h}{\beta^2} \right\}; \\ \frac{\partial^2 \mathcal{M}(h; \theta)}{\partial \beta^2} &= \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left\{ \frac{h}{\beta^2} \nu \left(\frac{h}{\beta}\right)^{\nu-1} \left[\frac{\nu}{\beta} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) + \mathcal{K}'_{\nu} \left(\frac{h}{\beta}\right) \frac{h}{\beta} \right] \right. \\ &\quad \left. - \left(\frac{h}{\beta}\right)^{\nu} \left[-\frac{\nu}{\beta^2} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) - \frac{\nu}{\beta} \frac{h}{\beta^2} \mathcal{K}'_{\nu} \left(\frac{h}{\beta}\right) - \frac{h}{\beta^2} \mathcal{K}'_{\nu} \left(\frac{h}{\beta}\right) - \frac{h^2}{\beta^3} \mathcal{K}''_{\nu} \left(\frac{h}{\beta}\right) \right] \right\}; \\ \frac{\partial \mathcal{M}(h; \theta)}{\partial \nu} &= \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left\{ -(\log(2) + \Psi(\nu)) \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) + \frac{\partial}{\partial \nu} \left[\left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) \right] \right\}; \\ \frac{\partial^2 \mathcal{M}(h; \theta)}{\partial \nu^2} &= -(\log(2) + \Psi(\nu)) \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \\ &\quad \times \left\{ -(\log(2) + \Psi(\nu)) \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) + \frac{\partial}{\partial \nu} \left[\left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) \right] \right\} \\ &\quad - \Psi'(\nu) \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) + \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \\ &\quad \times \left\{ -(\log(2) + \Psi(\nu)) \frac{\partial}{\partial \nu} \left[\left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) \right] + \frac{\partial^2}{\partial \nu^2} \left[\left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) \right] \right\}. \end{aligned}$$

Here, $\Psi(\cdot)$ represents the digamma function.

For the cross terms:

$$\begin{aligned} \frac{\partial^2 \mathcal{M}(h; \theta)}{\partial (\sigma^2) \partial \beta} &= \frac{1}{\Gamma(\nu)2^{\nu-1}} \left\{ -\frac{\nu}{\beta} \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) - \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}'_{\nu} \left(\frac{h}{\beta}\right) \frac{h}{\beta^2} \right\}; \\ \frac{\partial^2 \mathcal{M}(h; \theta)}{\partial (\sigma^2) \partial \nu} &= \frac{1}{\Gamma(\nu)2^{\nu-1}} \left\{ -(\log(2) + \Psi(\nu)) \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) + \frac{\partial}{\partial \nu} \left[\left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) \right] \right\}; \\ \frac{\partial^2 \mathcal{M}(h; \theta)}{\partial \beta \partial \nu} &= \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left\{ (\log(2) + \Psi(\nu)) \left[\frac{\nu}{\beta} \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) + \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}'_{\nu} \left(\frac{h}{\beta}\right) \frac{h}{\beta^2} \right] \right. \\ &\quad \left. - \frac{1}{\beta} \left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) - \frac{\nu}{\beta} \frac{\partial}{\partial \nu} \left[\left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}_{\nu} \left(\frac{h}{\beta}\right) \right] - \frac{h}{\beta^2} \frac{\partial}{\partial \nu} \left[\left(\frac{h}{\beta}\right)^{\nu} \mathcal{K}'_{\nu} \left(\frac{h}{\beta}\right) \right] \right\}. \end{aligned}$$

B. Algorithms

Here in Algorithm 3, we describe how we evaluate the L_q -likelihood for an observation vector \mathbf{Z} from a zero-mean Gaussian random field with Matérn covariance parametrised by $\boldsymbol{\theta}$. The Cholesky factorisation, solving of the linear system and dot product are all parallelised in **ExaGeoStat**.

Algorithm 3: Calculation of the L_q -likelihood.

Input : Observation vector \mathbf{Z} of length n , parameter $\boldsymbol{\theta}$ (initial value or the value from the previous iteration of optimisation), the n locations.

Calculate the Matérn covariance matrix $\boldsymbol{\Sigma}_{\mathcal{M}}$ using $\boldsymbol{\theta}$ and the locations based on (1);

$\mathbf{L}\mathbf{L}^\top = \boldsymbol{\Sigma}_{\mathcal{M}}$: Cholesky factorisation;

Calculate \mathbf{Z}_{new} by solving the linear system $\mathbf{L}\mathbf{L}^\top \mathbf{Z}_{new} = \mathbf{Z}$;

$logdet \leftarrow \log |\boldsymbol{\Sigma}_{\mathcal{M}}|$;

$prod \leftarrow \mathbf{Z}_{new}^\top \mathbf{Z}_{new}$;

$lq \leftarrow -0.5 \cdot prod - 0.5 \cdot logdet - 0.5 \cdot n \cdot \log(2\pi)$;

if $q \neq 1$ **then**

$lq \leftarrow \{\exp[lq \times (1 - q)] - 1\} / (1 - q)$;

end

Output: The L_q likelihood value lq .

Algorithm 4: Generating synthetic data for simulation experiments.

Input : Fixed locations, true parameter $\boldsymbol{\theta}$, number of locations n , number of replicates m , level of contamination r with $0 \leq r < 1$.

Calculate the Matérn covariance matrix $\boldsymbol{\Sigma}_{\mathcal{M}}$ using $\boldsymbol{\theta}$ and the locations, based on (1);

$\mathbf{L}\mathbf{L}^\top = \boldsymbol{\Sigma}_{\mathcal{M}}$: Cholesky factorisation;

for i in $1 : m$ **do**

 Generate vector \mathbf{e}_i of length n from i.i.d. standard normal distribution;

$\mathbf{Z}_i \leftarrow \mathbf{L} \cdot \mathbf{e}_i$;

if $r > 0$ **then**

 Generate r_i from continuous uniform distribution between 0 and 1;

if $r_i < r$ **then**

 Generate random noise vector \mathbf{z}_i of length n from i.i.d. normal (or other reasonable distributions);

$\mathbf{Z}_i \leftarrow \mathbf{Z}_i + \mathbf{z}_i$;

end

end

end

Output: Realisations $\{\mathbf{Z}_i, i = 1, \dots, m\}$.

In practice, since q is fixed throughout one single optimisation process, when $q \neq 1$, to simplify the calculation in the final step of the algorithm, we only need to let $lq \leftarrow \exp[lq \times (1 - q)]$ instead of evaluating the full expression. Moreover, since the value of the log-likelihood can be very small, the term $\exp[lq \times (1 - q)]$ is likely to be exactly 0 in most computing systems when the number of locations n gets large. To solve this issue, again in the final step of the algorithm, when $q \neq 1$, we let $lq \leftarrow \exp[(lq + n) \times (1 - q)]$, which means that we multiply the L_q -likelihood by the constant $\exp[n \times (1 - q)]$, to make the computation possible.

In Algorithm 4, we describe how we generate synthetic data from a zero-mean Gaussian random field with Matérn covariance matrix $\Sigma_{\mathcal{M}}$ parametrised by $\boldsymbol{\theta}$, with n locations and m replicates, and level of contamination r .

Affiliation:

Sihan Chen

Statistics Program

King Abdullah University of Science and Technology

Thuwal, Saudi Arabia, 23955

E-mail: sihan.chen@kaust.edu.sa

URL: <https://cemse.kaust.edu.sa/stsds/people/person/sihan-chen>

Joydeep Chowdhury

Statistics Program

King Abdullah University of Science and Technology

Thuwal, Saudi Arabia, 23955

E-mail: joydeep.chowdhury@kaust.edu.sa

URL: <https://cemse.kaust.edu.sa/stsds/people/person/joydeep-chowdhury>

Marc G. Genton

Statistics Program

King Abdullah University of Science and Technology

Thuwal, Saudi Arabia, 23955

E-mail: marc.genton@kaust.edu.sa

URL: <https://cemse.kaust.edu.sa/stsds/people/person/marc-g-genton>