

SPATIAL STATISTICS  
RESEARCH ARTICLE

# Maximum composite $L_q$ -likelihood for robust spatial covariance estimation

SIHAN CHEN<sup>1</sup>  and MARC G. GENTON<sup>1,\*</sup> 

<sup>1</sup>Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

(Received: 2 November 2025 · Accepted: 04 December 2025)

## Abstract

We develop a new method based on the  $L_q$ -likelihood to perform robust covariance estimation on spatial data. We propose the maximum composite  $L_q$ -likelihood estimator (MCL $q$ E), which is obtained from a combination of the maximum  $L_q$ -likelihood and the composite likelihood. In order to down-weight a single value at a spatial location, we make use of the composite likelihood of the spatial data, which is in the form of the product of the pairwise likelihood of all pairs of the observations. We apply the maximum  $L_q$ -likelihood on each pair to form the MCL $q$ E, which effectively down-weights abnormal observations in a spatial dataset and gives more robust covariance parameter estimation results when the data are contaminated by outliers. We adapt the hyperparameter tuning procedure developed for the ML $q$ E to the new MCL $q$ E framework. The robustness of the MCL $q$ E is further verified by both simulation studies and experiments on a precipitation dataset from the US.

**Keywords:** composite likelihood · Gaussian random fields · maximum composite  $L_q$ -likelihood estimator · robust statistics · spatial statistics

**Mathematics Subject Classification:** Primary 62M30 · Secondary 62F35

## 1. INTRODUCTION

Statistical analysis of spatial data plays a crucial role in numerous scientific fields, including environmental monitoring [1], climate modeling [2], epidemiology [3], geosciences [4, 5], and remote sensing [6]. Spatial data are often characterized by spatial dependence, irregular sampling, and increasingly large volumes of observations due to advances in sensing technologies and simulation models. Although many classical methods have proven effective for spatial interpolation and prediction, their application to large datasets is hindered by substantial computational costs and sensitivity to data contamination such as outliers or measurement errors.

A primary computational challenge in spatial statistics arises from the dense covariance matrices associated with Gaussian models, whose inversion requires  $\mathcal{O}(n^3)$  operations and  $\mathcal{O}(n^2)$  memory, where  $n$  is the number of spatial locations. To address this, various approximation methods have been developed, such as the tile low-rank approximation method used in [7] and the mixed-precision method in [8].

---

\*Corresponding author. Email: [marc.genton@kaust.edu.sa](mailto:marc.genton@kaust.edu.sa) (M.G. Genton)

In addition to computational methods that focus on accelerating the computation itself without changing the mathematical formulas, statistical methods that aim to find alternatives to the large matrix operations involved have also been developed. Among them, the composite likelihood approach [9] offers a particularly attractive solution by decomposing the full likelihood into products of lower-dimensional marginal or conditional likelihoods, such as pairwise likelihoods. This dramatically reduces the computational burden and makes inference feasible for datasets with a large number of spatial locations. Since then, composite likelihood has inspired new developments in statistics, for instance, in multivariate data analysis [10] and in model selection [11] in high-dimensional data analysis. An overview of the developments of composite likelihood methods can be found in [12].

In the context of spatial models, composite likelihood methods have been successfully applied to Gaussian spatial fields, generalized linear spatial models, and spatial point processes. For example, in [13, 14], it were applied the composite likelihood on binary spatial data, while in [15, 16], it were developed new methods based on the composite likelihood to perform parameter estimation and address spatial inhomogeneity for spatial point processes. In [17], proposed a composite likelihood-based method to model spatially correlated survival data and applied their method to e-Commerce data. In [18], it was combined composite likelihood and Bayesian inference, and applied their method to model spatial extremes. In [19, 20], it was demonstrated the effectiveness of pairwise and blockwise composite likelihoods for large-scale spatial prediction and parameter estimation, especially when combined with parallel or distributed computing frameworks. However, a known limitation of composite likelihood estimators is their sensitivity to outliers, which can distort pairwise relationships and bias parameter estimation. This is especially problematic in environmental and remote sensing applications, where anomalies are frequent.

In addition to the efficiency of the computation, the composite likelihood method in [19] also allows more flexibility when analysing spatial data. Instead of directly modeling spatial data using a multivariate distribution, it decomposes the full likelihood into products of pairwise likelihoods, making it possible to treat potential outliers in the dataset separately, thus allowing more robustness. To enhance robustness in spatial data analysis when there are outliers, the maximum  $L_q$ -likelihood estimation (ML $q$ E) method [21, 22] can be used. The ML $q$ E modifies the conventional likelihood by re-weighting each contribution to a power  $q \in (0, 1]$ , which systematically down-weights the influence of observations that have low likelihood under the model, such as potential outliers. The resulting estimator retains many desirable properties of the maximum likelihood estimator (MLE) under the assumed model but is significantly more resistant to data contaminations.

This article focuses on integrating the pairwise composite likelihood framework with the ML $q$ E to develop a new type of robust estimator, named the maximum composite  $L_q$ -likelihood estimator (MCL $q$ E). This estimator is designed to be both computationally scalable and robust to atypical observations in spatial datasets. In the MCL $q$ E, the composite likelihood is modified by applying the  $L_q$ -transformation to each pairwise component, therefore introducing a flexible robustness mechanism within a computationally efficient framework. The key distinction between the use of the MCL $q$ E in this work and the ML $q$ E in [23] lies in the type of outliers being addressed: cell-wise and row-wise outliers. The differences between these two types of contamination have been extensively discussed in recent works, such as [24, 25]. To be specific, in spatial statistics, a row-wise outlier refers to an anomalous data vector observed across all spatial locations at a given time point, within replicated realizations of the same spatial field. In contrast, a cell-wise outlier corresponds to a single abnormal measurement at one spatial location and time point. In [23], it were mainly considered row-wise outliers, while in this work, we focus on datasets contaminated by cell-wise outliers. Consequently, we do not require replicated spatial datasets; instead, we focus on handling individual abnormal values occurring within a single dataset.

Before introducing the MCLqE for spatial data, we begin with a motivating example that illustrates the handling of cell-wise outliers using the MLqE on time series data generated from an autoregressive AR(1) model. We independently simulate 100 Gaussian time series of length  $n = 400$  from an AR(1) process with parameter  $\rho = 0.5$ , and estimate  $\rho$  for each of the 100 time series using both the MLE and the MLqE (with the number of replicates equal to one), where in this time series illustration the MLqE is applied to the Gaussian conditional likelihood  $\prod_{t=2}^n f(X_t | X_{t-1}; \rho)$  of the AR(1) model. As shown in Figure 1, when there are no outliers, the MLE (that is,  $q = 1$ ) produces more accurate estimates than the MLqE for any  $q < 1$ .

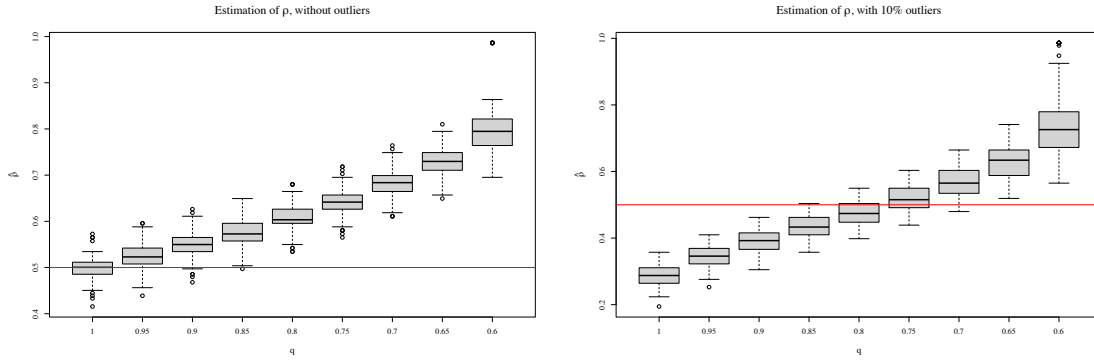


Figure 1. Boxplots (based on 100 simulations) of the MLE and MLqE estimation results with different values of  $q$ , for Gaussian AR(1) data on  $n = 400$  time points without outliers (top) and with 10% of outliers (bottom).

Next, to assess robustness, we introduce contamination by randomly selecting 10% of the time points and multiplying their values by 3, thus creating cell-wise outliers. We then repeat the estimation using both the MLE and the MLqE. The results presented in Figure 1 show that the MLE becomes sensitive to these outliers and no longer provides the most accurate estimates. In contrast, the MLqE with  $q = 0.75$  demonstrates improved robustness and delivers the most accurate estimation under contamination.

For spatial data, handling cell-wise outliers and performing robust estimation of covariance is more complicated than for time series, thus we have to introduce the MCLqE because the MLqE cannot be directly applied. Although the MCLqE cannot explicitly target individual observations due to its composite (pairwise) structure, it still effectively down-weights outliers by reducing the influence of anomalous pairs, those in which one or both observations deviate strongly from the model assumptions. This makes it particularly well-suited for large spatial datasets with irregular structures and potential contamination, such as gridded climate records, atmospheric monitoring data, or health surveillance data subject to reporting errors.

The remainder of this article is organized as follows. In Section 2, we first introduce the MLqE and the composite likelihood, and then define the MCLqE based on them. We also investigate how to choose the hyperparameters in this section. In Section 3, simulation studies are performed and compared the performance of the MCLqE with the MLE in different settings, including clean and contaminated data. Some additional experimental results can be found in Appendix A. Section 4 presents the application of the MCLqE on a precipitation data set from the US, also compared to the MLE. Conclusions and discussions about the use of the MCLqE on spatial data are presented in Section 5.

## 2. METHODOLOGY

In this section, we present the MLqE, the composite likelihood framework, and the resulting MCLqE, along with the procedure for selecting their hyperparameters.

### 2.1 Context

In this article, we use the definition of the MLqE in [23]. Let  $X_1, \dots, X_n$  be an independent identically distributed sample from a density  $f(x; \boldsymbol{\theta}_0)$ , where  $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$  for some integer  $p \geq 1$ . The MLqE of  $\boldsymbol{\theta}_0$  is expressed as

$$\tilde{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\{ \sum_{i=1}^n \lambda_q(f(X_i; \boldsymbol{\theta})) \right\}, \quad 0 < q \leq 1,$$

with

$$\lambda_q(u) = \begin{cases} \log(u), & \text{if } q = 1; \\ \frac{u^{1-q} - 1}{1 - q}, & \text{otherwise.} \end{cases}$$

Suppose  $\mathbf{Z}$  is a realization of a zero-mean Gaussian random field with a Matérn covariance function stated as

$$\mathcal{M}(h; \boldsymbol{\theta}) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left( \frac{h}{\beta} \right)^\nu \mathcal{K}_\nu \left( \frac{h}{\beta} \right), \quad (2.1)$$

where  $h$  is the distance between spatial locations,  $\boldsymbol{\theta} = (\sigma^2, \beta, \nu)^\top$ , and  $\sigma^2, \beta > 0$ , and  $\nu > 0$  represent the variance, range, and smoothness parameters, respectively. Here,  $\mathcal{K}_\nu(\cdot)$  denotes the modified Bessel function of the second kind of order  $\nu$ . The joint density of  $\mathbf{Z}$  is then expressed as

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

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

In [23], the setting of replicated spatial data from a Gaussian random field was considered. Specifically, let  $\{\mathbf{Z}_i, i = 1, \dots, m\}$  be  $m$  independent realizations of an identical zero-mean Gaussian random field, observed at the same set of spatial locations  $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathbb{R}^d$ . Then, the Lq-likelihood for estimating the parameter vector  $\boldsymbol{\theta}$  associated with the Matérn covariance function presented in (2.1) is given by

$$L_q(\boldsymbol{\theta}) = \begin{cases} \sum_{i=1}^m \left\{ -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_{\mathcal{M}}(\boldsymbol{\theta})|) - \frac{1}{2} \mathbf{Z}_i^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1}(\boldsymbol{\theta}) \mathbf{Z}_i \right\}, & \text{if } q = 1; \\ \sum_{i=1}^m \left\{ \left( \frac{1}{(|2\pi \boldsymbol{\Sigma}_{\mathcal{M}}(\boldsymbol{\theta})|)^{\frac{1-q}{2}}} \exp \left( -\frac{(1-q)}{2} \mathbf{Z}_i^\top \boldsymbol{\Sigma}_{\mathcal{M}}^{-1}(\boldsymbol{\theta}) \mathbf{Z}_i \right) - 1 \right) / (1-q) \right\}, & \text{otherwise.} \end{cases} \quad (2.2)$$

This  $Lq$ -likelihood expression aggregates the contribution of each realization of the Gaussian random field. When certain realizations exhibit behavior that deviates from the majority, the ML $q$ E can effectively down-weight their influence, resulting in a more robust parameter estimation. However, this robustness mechanism is only effective when multiple realizations are available in the dataset, and the outliers are entire realizations, rather than individual observations.

In scenarios where only a single realization is available, and contamination occurs at specific spatial locations, such as in the case of cell-wise outliers, the  $Lq$ -likelihood formulation stated in (2.2) fails to down-weight these anomalous observations. This is because the joint likelihood does not decompose across individual locations. Consequently, the ML $q$ E cannot distinguish and reduce the influence of cell-wise outliers.

To overcome this limitation, we seek an alternative decomposition of the likelihood that expresses it as a sum over individual or pairwise contributions, which allows us to apply the ML $q$ E at a finer resolution. This motivates the development of a composite likelihood framework that enables robust estimation by targeting outliers at the level of individual spatial observations.

## 2.2 Composite likelihood

Our approach in this work is to adapt the composite likelihood framework for spatial data introduced in [19], and combine it with the ML $q$ E method to achieve robust and efficient inference.

Let  $\mathbf{Z} \equiv \{Z(\mathbf{s}_i), i = 1, \dots, n\}$  denote observations from a real-valued, stationary Gaussian random field with constant and finite variance, where each  $\mathbf{s}_i \in \mathbb{R}^2$  represents a spatial location. Thus, we assume  $n$  observations collected at locations  $\mathbf{s}_1, \dots, \mathbf{s}_n$ .

For each distinct pair of locations indexed by  $1 \leq i < j \leq n$ , define the difference  $U_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ . Since the underlying field is Gaussian, each  $U_{ij}$  is normally distributed with mean zero and variance  $2\gamma_{ij}$ , where  $\gamma_{ij}$  is the semivariogram between the two locations. Then, the likelihood contribution from this pair is given by

$$\ell_{ij}(\boldsymbol{\theta}) = \frac{1}{2\sqrt{\pi\gamma_{ij}(\boldsymbol{\theta})}} \exp\left(-\frac{U_{ij}^2}{4\gamma_{ij}(\boldsymbol{\theta})}\right), \quad (2.3)$$

where  $\boldsymbol{\theta}$  denotes the parameters of the covariance function. In this work, we specifically use the Matérn covariance function as described in (2.1). Let  $\gamma_{ij}(\boldsymbol{\theta}) = \gamma(h; \boldsymbol{\theta})$  denote the semivariogram. The associated variogram is  $2\gamma(h; \boldsymbol{\theta})$ , which is related to the covariance structure through

$$2\gamma(h; \boldsymbol{\theta}) = \text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)), \quad h = \|\mathbf{s}_i - \mathbf{s}_j\|.$$

We define the corresponding pairwise log-likelihood contribution (up to an additive constant) as

$$\begin{aligned} l_{ij}(\boldsymbol{\theta}) &= \log(\ell_{ij}(\boldsymbol{\theta})) + \frac{1}{2} \log(4\pi) \\ &= -\frac{1}{2} \log(\gamma_{ij}(\boldsymbol{\theta})) - \frac{U_{ij}^2}{4\gamma_{ij}(\boldsymbol{\theta})}. \end{aligned} \quad (2.4)$$

This formulation allows us to approximate the full likelihood of the spatial observations in the vector  $\mathbf{Z} \equiv \{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}^\top$  using the sum of the marginal log-likelihoods from all distinct pairs. This leads to the composite log-likelihood expression formulated as

$$cl(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j>i}^n l_{ij}(\boldsymbol{\theta}) = - \sum_{i=1}^n \sum_{j>i}^n \left( \frac{1}{2} \log(\gamma_{ij}(\boldsymbol{\theta})) + \frac{U_{ij}^2}{4\gamma_{ij}(\boldsymbol{\theta})} \right). \quad (2.5)$$

Though the expression stated in (2.5) does not retain all the information contained in the full likelihood, it offers the flexibility to re-weight individual data points. This ability to adjust the influence of specific observations during likelihood evaluation enables robust parameter estimation. Another major advantage of this pairwise composite likelihood is that it avoids computationally intensive operations, such as inversion or multiplication of the full covariance matrix of the spatial data, which are often required for the full Gaussian likelihood and become infeasible for large spatial datasets. Additionally, computational efficiency can be further enhanced by ignoring location pairs that are far apart. Specifically, we introduce a weighted composite likelihood as

$$wcl(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j>i}^n l_{ij}(\boldsymbol{\theta}) w_{ij}, \quad (2.6)$$

where  $w_{ij} = 1$  if  $\|\mathbf{s}_i - \mathbf{s}_j\| \leq d$ , and  $w_{ij} = 0$  otherwise. Here,  $d$  is a user-specified distance threshold that controls the inclusion of pairwise terms. This localized approximation retains the spatial structure while significantly reducing the computational burden, making it particularly suitable for large-scale spatial data analysis.

### 2.3 MCLqE for Gaussian random fields

Based on the MLqE and the composite likelihood framework, we now introduce the MCLqE.

**DEFINITION 2.1** Let  $\mathbf{Z} = \{Z(\mathbf{s}_i), i = 1, \dots, n\}$ , where  $\mathbf{s}_i \in \mathbb{R}^2$ , be a realization from a stationary Gaussian random field with finite and constant variance, and a covariance function parameterized by  $\boldsymbol{\theta}$ . The MCLqE is defined as

$$\tilde{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \left\{ \sum_{i=1}^n \sum_{j>i}^n L_{q,ij}(\boldsymbol{\theta}) w_{ij} \right\}, \quad 0 < q \leq 1,$$

where

$$L_{q,ij}(\boldsymbol{\theta}) = \begin{cases} -\frac{1}{2} \log(4\pi\gamma_{ij}(\boldsymbol{\theta})) - \frac{U_{ij}^2}{4\gamma_{ij}(\boldsymbol{\theta})}, & \text{if } q = 1; \\ \frac{\left( \frac{1}{\sqrt{2\pi}\sqrt{2\gamma_{ij}(\boldsymbol{\theta})}} \exp\left(-\frac{U_{ij}^2}{4\gamma_{ij}(\boldsymbol{\theta})}\right) \right)^{1-q} - 1}{1-q}, & \text{otherwise;} \end{cases}$$

and  $U_{ij}$ ,  $\gamma_{ij}$ , and  $w_{ij}$  for all pairs  $1 \leq i < j \leq n$  are defined as in (2.3) and (2.6).

Note that for  $q = 1$ ,  $L_{1,ij}(\boldsymbol{\theta})$  coincides with the usual Gaussian pairwise log-likelihood contribution, and the MCLqE with  $q = 1$  yields the same estimator as the standard maximum composite likelihood method—since  $l_{ij}(\boldsymbol{\theta})$  in (2.4) differs from  $L_{1,ij}(\boldsymbol{\theta})$  only by an additive constant that does not depend on  $\boldsymbol{\theta}$ .

Unlike the MLqE, the MCLqE does not operate directly on individual observations, as it is formulated based on the sum of likelihood contributions from all observation pairs. Nevertheless, it still has robustness to outliers: pairwise differences involving outlying observations tend to stand out among all pairwise combinations, thereby contributing less to the overall likelihood. As a result, the MCLqE can reduce the influence of such outliers and is expected to produce more robust parameter estimates compared to the standard composite likelihood approach.

#### 2.4 Choice of hyperparameters

Compared to [23], tuning the hyperparameters is a more challenging task in this work, since we now have an additional hyperparameter,  $d$ , to select. As we will see from the numerical results in Section 3, different values of  $q$  may lead to very different estimation results when there are outliers in the dataset. In addition, the choice of the threshold  $d$  has a significant impact on computational efficiency. Therefore, it is essential to develop a mechanism to choose both  $q$  and  $d$  appropriately.

For the threshold  $d$ , since smaller values of  $d$  lead to faster computation but may also result in less accurate estimates, the goal is to find the smallest possible value of  $d$  without incurring a substantial loss of accuracy. Here, we adopt the method in [19]. This method makes use of the Godambe information  $\mathbf{G}_n(\boldsymbol{\theta}; d)$  for the weighted composite likelihood given in (2.6), which is stated as

$$\mathbf{G}_n(\boldsymbol{\theta}; d) = \mathbf{H}_n(\boldsymbol{\theta}; d) \mathbf{J}_n(\boldsymbol{\theta}; d)^{-1} \mathbf{H}_n(\boldsymbol{\theta}; d)^\top,$$

where

$$\mathbf{H}_n(\boldsymbol{\theta}; d) = -\mathbb{E} \left[ \frac{\partial^2 wcl(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right] = \frac{1}{2} \sum_{i=1}^n \sum_{j>i}^n \frac{\gamma'_{ij}(\boldsymbol{\theta}) \gamma'_{ij}(\boldsymbol{\theta})^\top}{\gamma_{ij}^2(\boldsymbol{\theta})} w_{ij},$$

and

$$\begin{aligned} \mathbf{J}_n(\boldsymbol{\theta}; d) &= \mathbb{E} \left[ \frac{\partial wcl(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial wcl(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right] \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j>i}^n \sum_{l=1}^n \sum_{k>l}^n \frac{\gamma'_{ij}(\boldsymbol{\theta}) \gamma'_{lk}(\boldsymbol{\theta})^\top}{\gamma_{ij}(\boldsymbol{\theta}) \gamma_{lk}(\boldsymbol{\theta})} \text{Corr} [U_{ij}^2, U_{lk}^2] w_{ij} w_{lk}. \end{aligned}$$

Here,  $\gamma'_{ij}(\boldsymbol{\theta})$  denotes the derivative of  $\gamma_{ij}(\boldsymbol{\theta})$  with respect to the parameter vector  $\boldsymbol{\theta}$ . In practice, directly evaluating the Godambe information is not feasible for large datasets, as the computational complexity of  $\mathbf{J}_n(\boldsymbol{\theta}; d)$  is  $\mathcal{O}(n^4)$ . In [26], a window-based method was proposed to approximate  $\mathbf{J}_n(\boldsymbol{\theta}; d)$  by subsampling  $m_n$  overlapping subsets of  $\mathbf{s}_1, \dots, \mathbf{s}_n$ , denoted by  $S_1, \dots, S_{m_n}$ . For each subset  $S_r$ , the corresponding composite score is defined as

$$\mathbf{s}_r(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{(i,j) \in S_r} l_{ij}(\boldsymbol{\theta}) w_{ij} = -\frac{1}{2} \sum_{(i,j) \in S_r} \frac{\gamma'_{ij}(\boldsymbol{\theta})}{\gamma_{ij}(\boldsymbol{\theta})} \left( 1 - \frac{U_{ij}^2}{2\gamma_{ij}(\boldsymbol{\theta})} \right) w_{ij}.$$



Then, an estimator of the variability matrix is given by

$$\hat{\mathbf{J}}_n(\boldsymbol{\theta}; d) = \frac{1}{m_n} \sum_{r=1}^{m_n} \mathbf{s}_r(\boldsymbol{\theta}) \mathbf{s}_r(\boldsymbol{\theta})^\top.$$

Using this, we can estimate the Godambe information as

$$\hat{\mathbf{G}}_n(\boldsymbol{\theta}; d) = \mathbf{H}_n(\boldsymbol{\theta}; d) \hat{\mathbf{J}}_n(\boldsymbol{\theta}; d)^{-1} \mathbf{H}_n(\boldsymbol{\theta}; d)^\top,$$

and select the optimal value of  $d$ , denoted by  $d^*$ , by

$$d^* = \operatorname{argmin}_{d \in \mathcal{D}} \left\{ \operatorname{tr} \left( \hat{\mathbf{G}}_n^{-1}(\boldsymbol{\theta}^*; d) \right) \right\},$$

where  $\mathcal{D}$  is the set of all feasible values of  $d$  and  $\boldsymbol{\theta}^*$  is an initial estimator of  $\boldsymbol{\theta}$  (for example, the MLE). The standard errors of the parameter estimates, or associated confidence intervals for the MCLqE method, can also be computed by making use of the Godambe information in a similar manner to [19], or by bootstrap-based methods.

For the hyperparameter  $q$ , we adapt the method in [23], which was inspired from [27]. The target is to find a sub-interval of  $(0, 1]$ , within which all values of  $q$  lead to similar parameter estimation results. We define an ordered grid  $q_0 = 1 > q_1 > \dots > q_K > 0$  for  $q$ , and for each  $q_k$ ,  $0 \leq k \leq K$ , denote the corresponding MCLqE by

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

where  $p$  is the dimension of  $\boldsymbol{\theta}$ .

Also, define the corresponding vector of standardized estimates  $\boldsymbol{\zeta}_{q_k}$  as

$$\boldsymbol{\zeta}_{q_k} = \left( \frac{\tilde{\theta}_{q_k}^1}{C_1}, \dots, \frac{\tilde{\theta}_{q_k}^p}{C_p} \right)^\top.$$

Unlike [23] who used the asymptotic standard error to normalize the estimation results, here we instead use standardizing constants  $C_1, \dots, C_p$ , because of computational feasibility. Afterwards, we use the standardized quadratic variation (SQV) defines as  $\text{SQV}_{q_k} = \|\boldsymbol{\zeta}_{q_{k-1}} - \boldsymbol{\zeta}_{q_k}\|/p$ , for each  $q_k$ , with  $1 \leq k \leq K$ , and choose the optimal value  $q^*$  based on this quantity. The corresponding algorithm to choose both  $d^*$  and  $q^*$  is shown in Algorithm 1. During the tuning of the hyperparameter  $q$ , the algorithm iteratively shrinks the target interval until the interval becomes sufficiently small or the estimation results obtained using the  $q$  values within the interval differ by less than a prescribed tolerance. The optimal value  $q^*$  is then returned as the largest value of  $q$  in the final interval. It would be ideal to tune both hyperparameters simultaneously so that they are optimally aligned with each other. However, performing a joint search is too computationally expensive. For this reason, we first select  $d$ , and then determine the best value of  $q$  conditional on the chosen  $d$ . This sequential strategy places priority on ensuring the robustness of the MCLqE while keeping the overall computational burden manageable. The constants  $\varepsilon, L, C_1, C_2, C_3$  can be pragmatically selected, depending on how sensitive the estimation results are to the choice of  $q$ . In the numerical experiments of this work, we set  $\varepsilon = 0.01$  and  $L = 0.06$ , while the constants  $C_1, C_2$ , and  $C_3$  were chosen based on the MLE estimates of the data.



---

**Algorithm 1:** Tuning the hyperparameters  $d$  and  $q$ 


---

**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  $q$ :  $\varepsilon > 0$ .

Find an initial estimator  $\theta^*$  of  $\theta$ , e.g., the MLE;

Determine  $d^*$  by finding the minimizer of  $\text{tr}\{\hat{G}_n^{-1}(\theta^*; d)\}$ ;

Set  $q^* \leftarrow 1$ ;

**while**  $q_0 - q_{\min} > \varepsilon$  **do**

    Compute  $SQV_{q_k}$  for each  $k = 1, \dots, K$ , using the optimal  $d^*$ ;

**if**  $SQV_{q_k} < L$  for all  $k = 1, \dots, K$  **then**

$q^* \leftarrow q_0$ ;

**break**;

**end**

    Let  $k^*$  be the largest integer in  $\{1, \dots, K\}$  such that  $SQV_{q_{k^*}} \geq L$ ;

    Redefine an equally spaced grid for  $q$ :  $q_0 \leftarrow q_{k^*} > q_1 > \dots > q_K \leftarrow q_{\min} > 0$ ;

**end**

**if**  $q_0 - q_{\min} \leq \varepsilon$  **then**

$q^* \leftarrow q_0$ ;

**end**

**Output:**  $d^*, q^*$

---

### 3. SIMULATION EXPERIMENTS

In this section, we perform simulation studies on the MCLqE. We use the R [28] package `geoR` [29] to generate the spatial data, then tune the hyperparameters and estimate the parameters using our implementation in R. The optimization algorithm we apply for parameter estimation is BOBYQA [30], which is the same as the algorithm embedded in `ExaGeoStat` [31]. The experiments are run on a Intel Xeon CPU with 48 cores.

#### 3.1 Context

In the experiments, we generate single-replicate synthetic spatial data from a zero-mean Gaussian random field with Matérn covariance matrix  $\Sigma_{\mathcal{M}}$  parametrized by  $\theta$ , with  $n$  locations in the unit square. In this section, we focus mainly on simulated datasets from the Matérn covariance function with true parameters  $\sigma^2 = 1, \beta = 0.1, \nu = 0.5$  (medium dependence, rough realizations). Experiments using datasets generated from other settings are shown in Appendix A. For the data used in Section 3.3, we generate the clean data first and add noise to parts of the data afterwards. In the rest of this section, we show the estimation results of the variance, range and smoothness parameters in each setting. For all simulation experiments in this section, we tune the hyperparameters  $d$  and  $q$  using Algorithm 1.

#### 3.2 Clean data

We start with experiments that compare the performance of MCLqE with the ordinary composite likelihood when there are no outliers in the simulated data.

In Figure 2, we present the experimental results for clean data. Each experiment is repeated on 90 independently generated datasets, with the results summarized in boxplots. In the second and third rows of this figure, we fix  $q = 1$  and  $q = 0.9$ , respectively, and vary  $d$ . In the third and fourth rows, we vary  $q$  with  $d = 0.3$  and with  $d$  selected using Algorithm 1, and compare these estimates with MLE results computed on the same datasets using the same optimization method. The red horizontal lines indicate the true values of the parameters, and the blue vertical lines mark the  $q$  or  $d$  values that yield the smallest mean squared error (MSE).

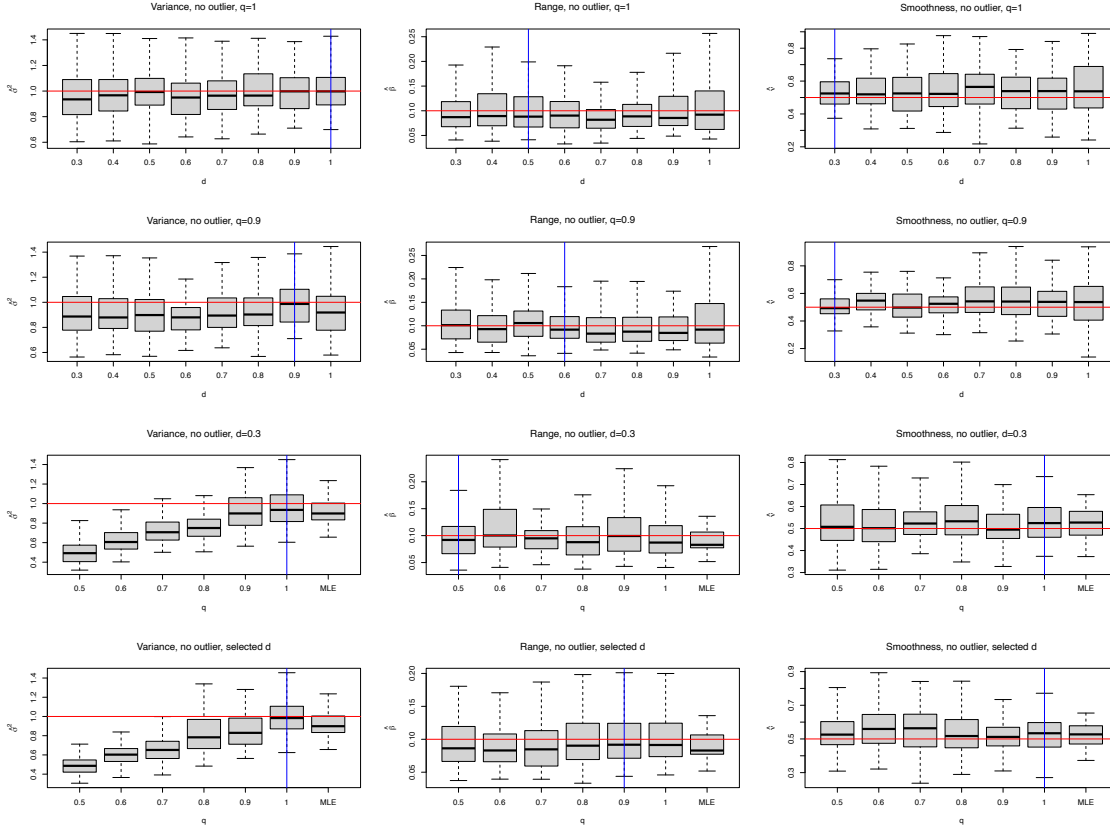


Figure 2. Boxplots (based on 90 simulations) of the MCLqE results with no outlier in the spatial data, generated using the medium-rough setting of the Matérn covariance function with parameters  $\sigma^2 = 1, \beta = 0.1, \nu = 0.5$  on  $n = 400$  locations. The red horizontal lines correspond to the true values of the parameters. The blue vertical lines correspond to the value of  $q$  or  $d$  leading to the smallest MSE. The cases  $q = 1$  correspond to ordinary composite likelihood estimation results.

Because the composite likelihood does not fully capture the covariance structure and is based on a reduced problem size, the estimation results are less stable here than in [23]. Nevertheless, as long as the value of  $q$  is not too small, the estimation results using the MCLqE do not deviate much from the MLE results. In the first and second rows, where  $q$  is fixed and  $d$  is varied, there is no clear pattern in how  $d$  affects accuracy and the impact appears minor. In the third row, where  $d = 0.3$  and  $q$  is varied, the MCLqE does not replicate the behavior of the MLqE, where  $q = 1$  clearly gave the most unbiased and stable results. Nevertheless, the MSE for both the variance and smoothness parameters is still the smallest at  $q = 1$ , and for the range parameter  $q = 1$  remains close to optimal, which is consistent with our expectations for clean data. In the last row, where  $d$  is first selected using Algorithm 1 and then  $q$  is varied, the estimates are slightly more accurate than in the previous row, and it becomes clearer that  $q$  values equal to or near 1 produce the best results.

### 3.3 Contaminated data

Next, we evaluate the performance of the MCLqE in the presence of outliers in the simulated data. For all results shown in Figure 3, the data are generated from a zero-mean Gaussian random field with the Matérn covariance function, using parameters  $\sigma^2 = 1, \beta = 0.1$ , and  $\nu = 0.5$ , over  $n = 400$  spatial locations. Each experiment is repeated on 90 independently simulated datasets under the same settings, and the resulting estimates are summarized in boxplots.

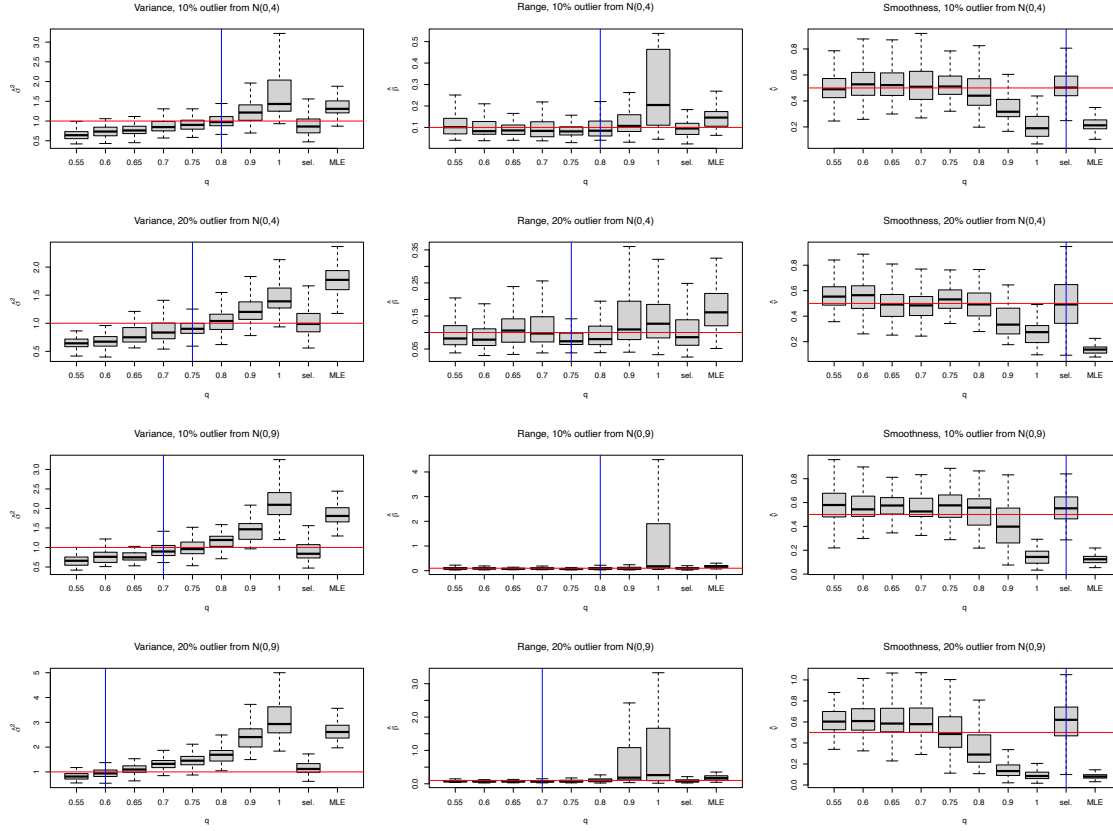


Figure 3. Boxplots (based on 90 simulations) of the MCLqE results under varying levels of contamination in the spatial data, generated using the medium-rough setting of the Matérn covariance function with parameters  $\sigma^2 = 1, \beta = 0.1, \nu = 0.5$  on  $n = 400$  locations. Red horizontal lines indicate the true parameter values. Blue vertical lines correspond to the value of  $q$  (fixed or selected) that yields the smallest MSE. The cases  $q = 1$  correspond to ordinary composite likelihood estimates. The label “sel.” indicates results obtained using  $q$  values selected by Algorithm 1.

In each row of Figure 3, the same uncontaminated datasets are altered to produce varying contamination levels, by randomly adding Gaussian noise from  $N(0, 4)$  or  $N(0, 9)$  to 10% or 20% of the spatial locations. For all experiments, the distance cutoff parameter  $d$  is selected using Algorithm 1, and in the final column of each plot,  $q$  is also chosen via the same algorithm. For reference, the MLE results obtained using the same datasets and optimization method are included for comparison.

From the plots, it is evident that when  $q$  is equal to or close to 1, the MCLqE produces estimates that deviate considerably from the true parameter values. In contrast, the most accurate estimates are generally obtained when  $q$  is between 0.6 and 0.8. These findings demonstrate that the MCLqE can yield more robust parameter estimates in the presence of data contamination, compared to both the ordinary MLE and composite likelihood approaches. In addition, Figure 4 presents experimental results based on datasets simulated under the same settings as in Figures 2 and 3, except that we set  $\nu = 1$  instead of  $\nu = 0.5$ , so that the realizations are smoother. From the first to the fifth row, the datasets correspond respectively to clean data, data with 10% and 20% of locations contaminated by noise from  $N(0, 4)$ , and data with 10% and 20% of locations contaminated by noise from  $N(0, 9)$ . In all experiments, the parameter  $d$  is selected using Algorithm 1. The general patterns are consistent with those observed in Figures 2 and 3, that is, for clean datasets, the MSE of the range parameter is the smallest when  $q = 1$ , while the MSEs of the other two parameters at  $q = 1$  remain close to their minimum values.

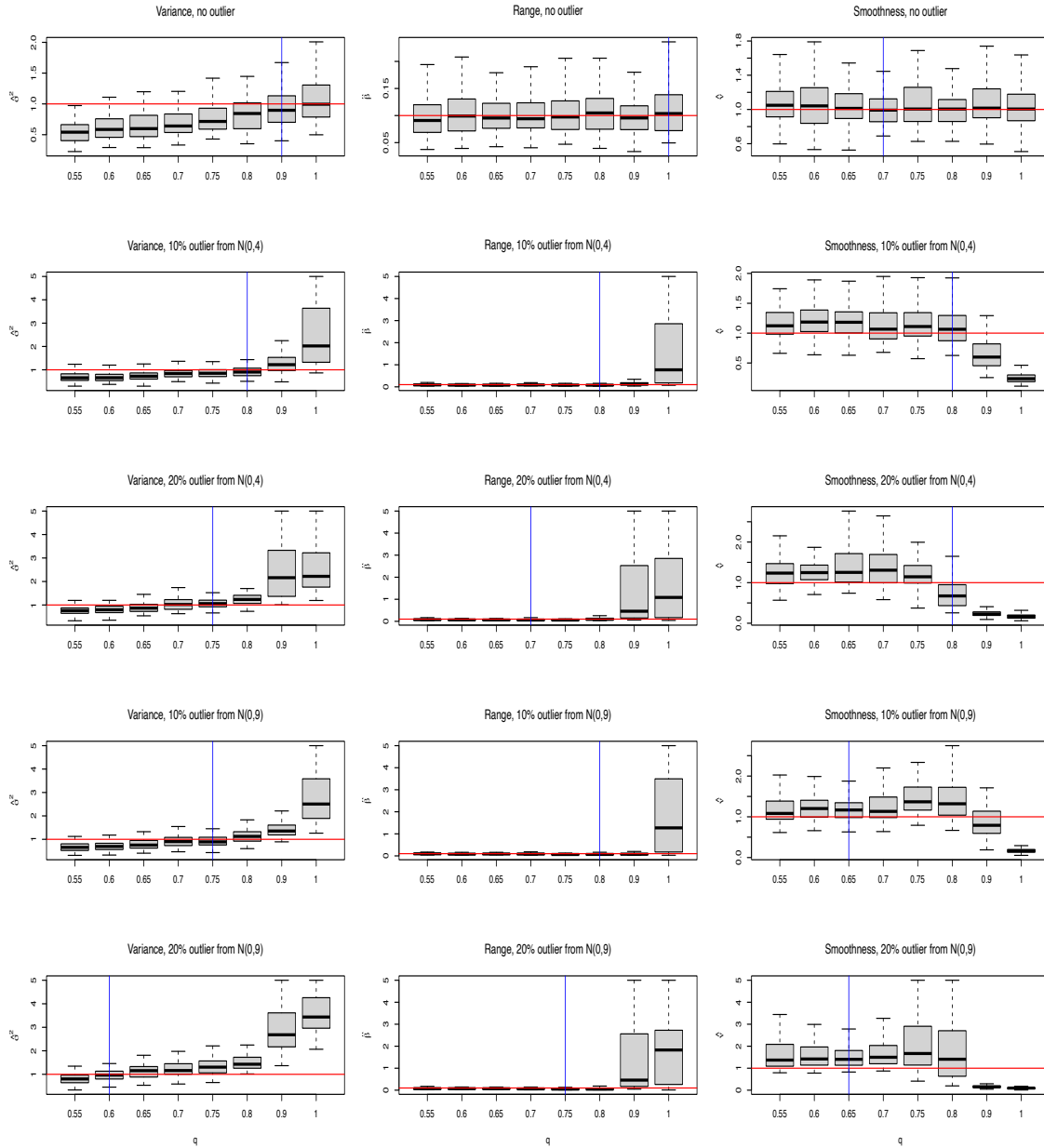


Figure 4. Boxplots (based on 90 simulations) of the MCLqE estimation results for spatial datasets generated under the medium-smooth setting of the Matérn covariance function with parameters  $\sigma^2 = 1, \beta = 0.1, \nu = 1$  on  $n = 400$  locations. The values of  $q$  that we use here are 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1. The red horizontal lines correspond to the true values of the parameters, and the blue vertical lines indicate the smallest MSE.

However, under data contamination, the use of  $q = 1$  leads to substantial bias across all experiments, while the most accurate estimates are obtained for  $q$  values between 0.6 and 0.8. These results confirm once again that the MCLqE provides robust parameter estimation in the presence of contaminated spatial data.

Unlike the results in [23], the estimates based on selected values of  $q$  do not always coincide with the smallest observed MSE, reflecting the greater variability of the composite likelihood relative to the full likelihood. Nevertheless, the selected- $q$  estimates remain substantially more accurate than those with  $q = 1$ , and are typically close to the optimal MSE in each scenario.

## 4. PRECIPITATION DATA ANALYSIS

We revisit the US precipitation dataset introduced in [23], obtained from

<https://www.image.ucar.edu/Data/US.monthly.met/USmonthlyMet.shtml>.

We extract the data for each month separately, producing 12 sub-datasets, and this time we focus on detecting and handling cell-wise outliers within single-month data. We select observations from 1928 to 1997 at  $n = 621$  stations that do not contain missing values over the 70-year period to eliminate the influence of missing data on the analysis. The monthly data are obtained by averaging the original hourly observations and then centered throughout the whole spatial domain, which produces a smoothed dataset that is more appropriately modeled using stationary Gaussian random fields. The spatial locations are also normalized into the unit square. Similarly to the simulations, we estimate the variance, range and smoothness parameters in the Matérn covariance function for the data, excluding a nugget effect.

In Table 1, we present several representative estimation results using the MLE and MCLqE, as well as the original composite likelihood, applied to the May datasets from the years 1942, 1948, 1953, and 1957. For the MCLqE, the hyperparameters  $d$  and  $q$  are selected using Algorithm 1, with the initial grid for selecting  $q$  being set to 1, 0.8, 0.7, 0.6, 0.5,  $L = 0.1$  and  $\varepsilon = 0.01$ . The results show that, for 1942 and 1948, the differences between the MLE and MCLqE estimates of the three parameters are relatively small, while for 1953 and 1957, the differences are much more significant. This behavior might be explained by Figure 5, where we present boxplots of these datasets and calculate the number of outlying data points among each of these datasets, though these univariate boxplots for spatial data are only indicative.

Table 1. The MLE and MCLqE results, as well as the estimation results using the original composite likelihood (indicated by “MCLE” for maximum composite likelihood estimation in the table) of the US precipitation data of May in several years. The MCLqE results are obtained with hyperparameters  $d$  and  $q$  selected using Algorithm 1, and the selected values of  $q$  are listed in the last column of the table named “Selected  $q$ ”.

Method	MLE			MCLE			MCLqE			Selected $q$
Parameter	$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\nu}$	$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\nu}$	$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\nu}$	
Year										
1942	22.272	0.016	0.057	23.969	0.025	0.029	18.838	0.034	0.037	0.80
1948	30.735	0.332	0.222	27.932	0.072	0.043	24.606	0.061	0.023	0.80
1953	41.998	0.055	0.210	38.069	0.044	0.239	21.755	0.056	0.022	0.75
1957	36.912	0.048	0.413	34.010	0.113	0.027	22.996	0.093	0.021	0.75

In the first row of Figure 5, the 1942 and 1948 datasets contain only 15 and 14 detected outliers, respectively, and their magnitudes are not very large. In contrast, while the 1953 dataset in the second row does not contain many more outliers than the previous two examples, we can see that several of these outliers are extreme deviations. The 1957 dataset shows the most severe contamination, with more than twice as many outliers as in 1942 and 1948. Therefore, it is reasonable that the robust MCLqE is closer to the non-robust MLE on the 1942 and 1948 datasets, as these two datasets are less contaminated, and the MCLqE behaves more differently from the MLE on more contaminated datasets.

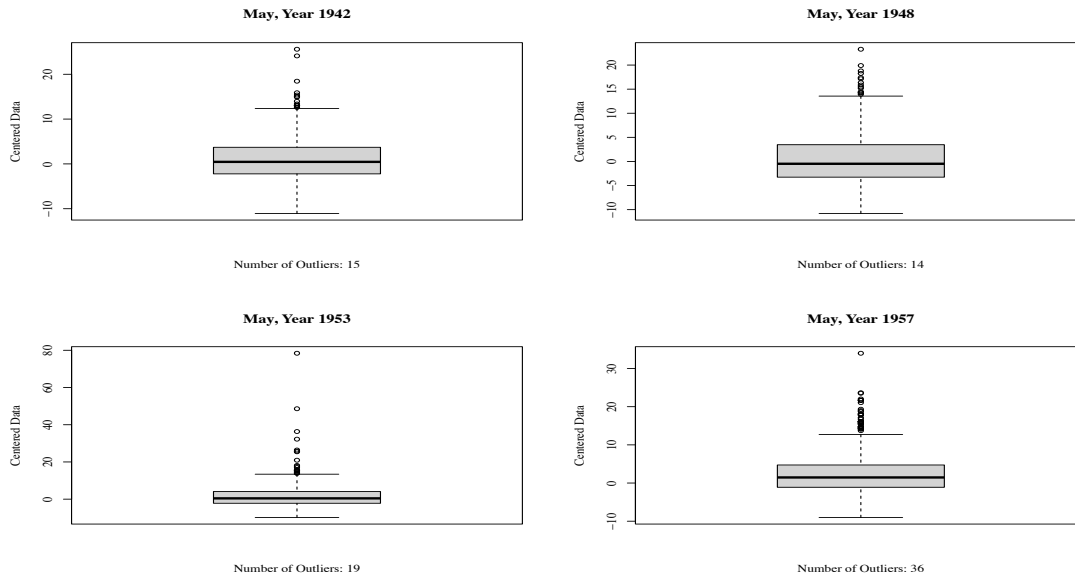


Figure 5. Boxplots of the US precipitation data in May of the years 1942, 1948, 1953 and 1957. For each year, the number of outliers detected among the 621 locations in total is shown below the corresponding boxplot.

## 5. CONCLUSIONS

To conclude, this work extends the application of the  $Lq$ -likelihood framework to spatial data without replicates. As discussed in [23], a key limitation of the  $MLqE$  is that it is only applicable for robust estimation when multiple replicates of spatial data are available, treating entire replicates as potential outliers. To address this issue, we introduced the  $MCLqE$ , which integrates  $MLqE$  with the composite likelihood approach. This combination allows the  $Lq$ -likelihood of the spatial random field to be decomposed into the sum of the  $Lq$ -likelihoods of pairs of the spatial locations, allowing robust estimation in the presence of cell-wise outliers, even when no replicates are available. As a result, the proposed method substantially broadens the applicability of robust estimation techniques in spatial statistics.

Through both simulation studies and real-world data experiments, we have shown that  $MCLqE$  consistently outperforms both the MLE and the original composite likelihood approach when outliers are present in spatial data without replicates. On clean data (Figure 2), the  $MCLqE$  estimates with not-too-small values of  $q$  remain not too far from those of the MLE, indicating no substantial loss of efficiency. Under contamination (Figure 3), and with suitable choices of  $q$ , the  $MCLqE$  delivers parameter estimates with relatively low bias and high efficiency compared to the MLE and composite likelihood. Furthermore, the method is computationally more efficient as it avoids the costly matrix operations typically associated with likelihood-based spatial estimation, which is a crucial advantage for large-scale spatial analyses.

There are several possible extensions to the methodology presented in this work. For instance, we have not yet incorporated the nugget effect into the model. With minor modifications to the existing code, it would be feasible to estimate the nugget effect alongside the three parameters of the Matérn covariance function considered here.

Additionally, the hyperparameter tuning strategy employed in this work may not be optimal. Although it is already capable of suggesting suitable values of  $q$  that lead to much less biased estimation results than the MLE, it is neither as computationally efficient nor as accurate compared to the approach in [23], partly due to the increased complexity of simultaneously tuning both  $d$  and  $q$ . Developing more effective tuning procedures is a worthwhile direction for future research.

## STATEMENTS

*Acknowledgements*

We dedicate this article to the memory of Joydeep Chowdhury who started this research with us but sadly passed away in February 2025.

*Author contributions*

Conceptualization, S. Chen, M.G. Genton; methodology, S. Chen, M.G. Genton; writing of original draft, S. Chen, M.G. Genton; final writing and editing, S. Chen, M.G. Genton; All authors have read and agreed to the published version of the article.

*Conflicts of interest*

The authors declare no conflict of interest.

*Data and code availability*

The data and code are available from <https://github.com/Sihan96/MCLqE>

*Declaration on the use of artificial intelligence (AI) technologies*

The authors declare that no generative AI was used in the preparation of this article.

*Funding*

This research was supported by the King Abdullah University of Science and Technology.

*Open access statement*

The Chilean Journal of Statistics (ChJS) is an open-access journal. Articles published in the ChJS are distributed under the terms of the Creative Commons AttributionNonCommercial-ShareAlike 4.0 License, which permits others to remix, adapt, and build upon the material for non-commercial purposes, provided that appropriate credit is given and derivative works are licensed under identical terms

## APPENDIX A. SUPPLEMENTARY EXPERIMENTAL RESULTS

In the main text of this article, we mainly presented simulation experiments using spatial datasets under the medium-rough and medium-smooth settings, where the true parameters are  $\sigma^2 = 1$ ,  $\beta = 0.1$ , and  $\nu = 0.5$  (rough) or  $\nu = 1$  (smooth). In Figure A1 to A4, we present additional experimental results using simulated spatial datasets in other settings. We consider the cases with parameter  $\beta = 0.03, 0.1, 0.3$ , which we refer to as weak, medium, and strong correlation, respectively, and  $\nu = 0.5, 1$ , which we refer to as rough and smooth field, respectively. For all the experimental results shown here, we set the number of locations to  $n = 400$ , and we repeat the experiment 90 times to make the box-plots. The values of  $d$  used in the experiments are selected using Algorithm 1, and various values of  $q$  are used for comparison. The findings from these additional experiments are consistent with those presented in the main text. For clean datasets, the optimal choice of  $q$  is typically 1 or close to 1, while for contaminated datasets, setting  $q = 1$  consistently results in large bias, and the most accurate estimates are generally obtained with smaller values of  $q$ . These results further demonstrate that the robustness of the MCLqE is not confined to a specific experimental configuration, but rather reflects a general and reliable property of the proposed estimation framework across various scenarios.



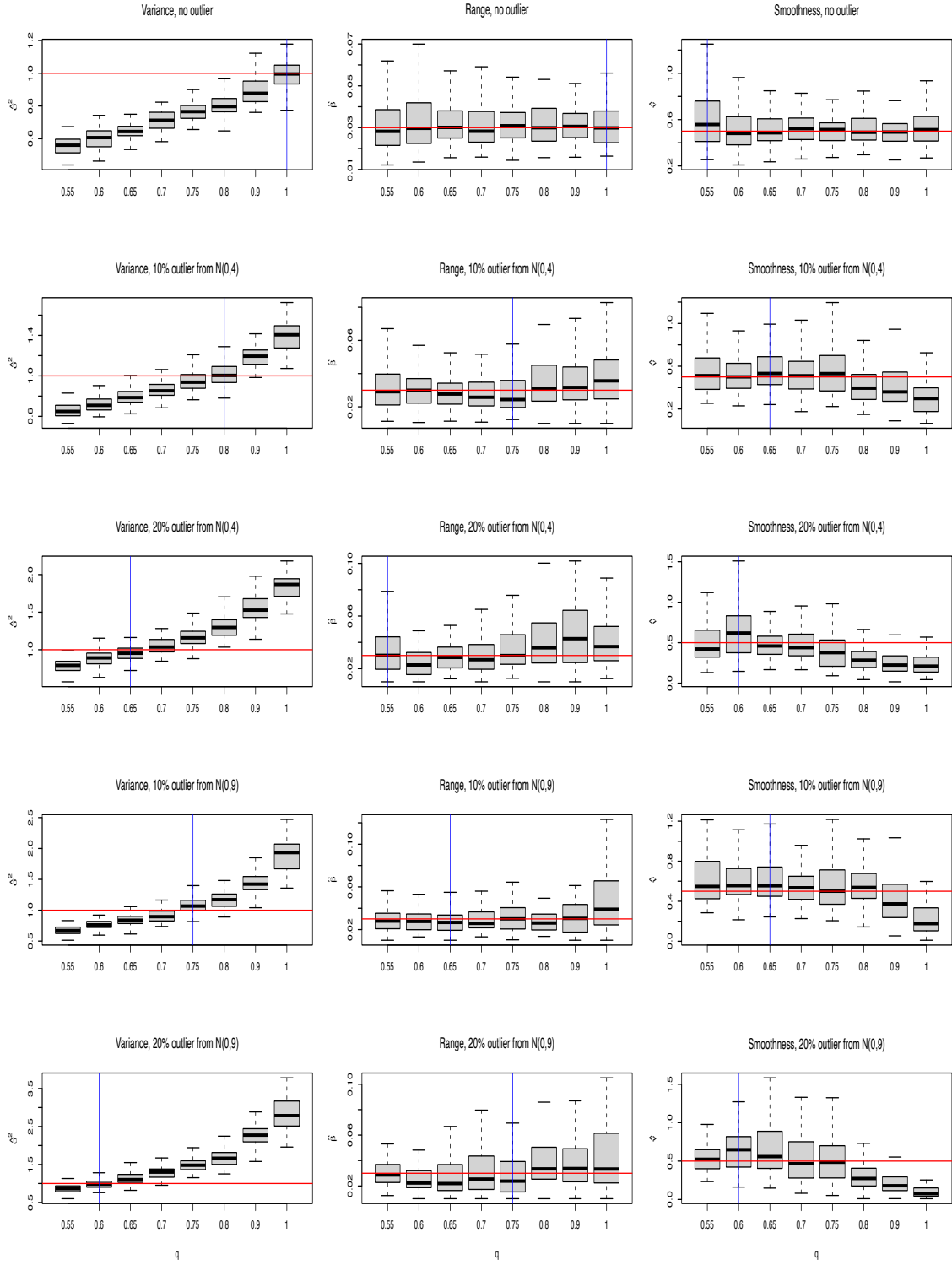


Figure A1. Boxplots (based on 90 simulations) of the MCLqE estimation results for spatial datasets generated under the weak-rough setting of the Matérn covariance function with parameters  $\sigma^2 = 1, \beta = 0.03, \nu = 0.5$ . The values of  $q$  that we use here are 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1. The red horizontal lines correspond to the true values of the parameters, and the blue vertical lines indicate the smallest MSE.

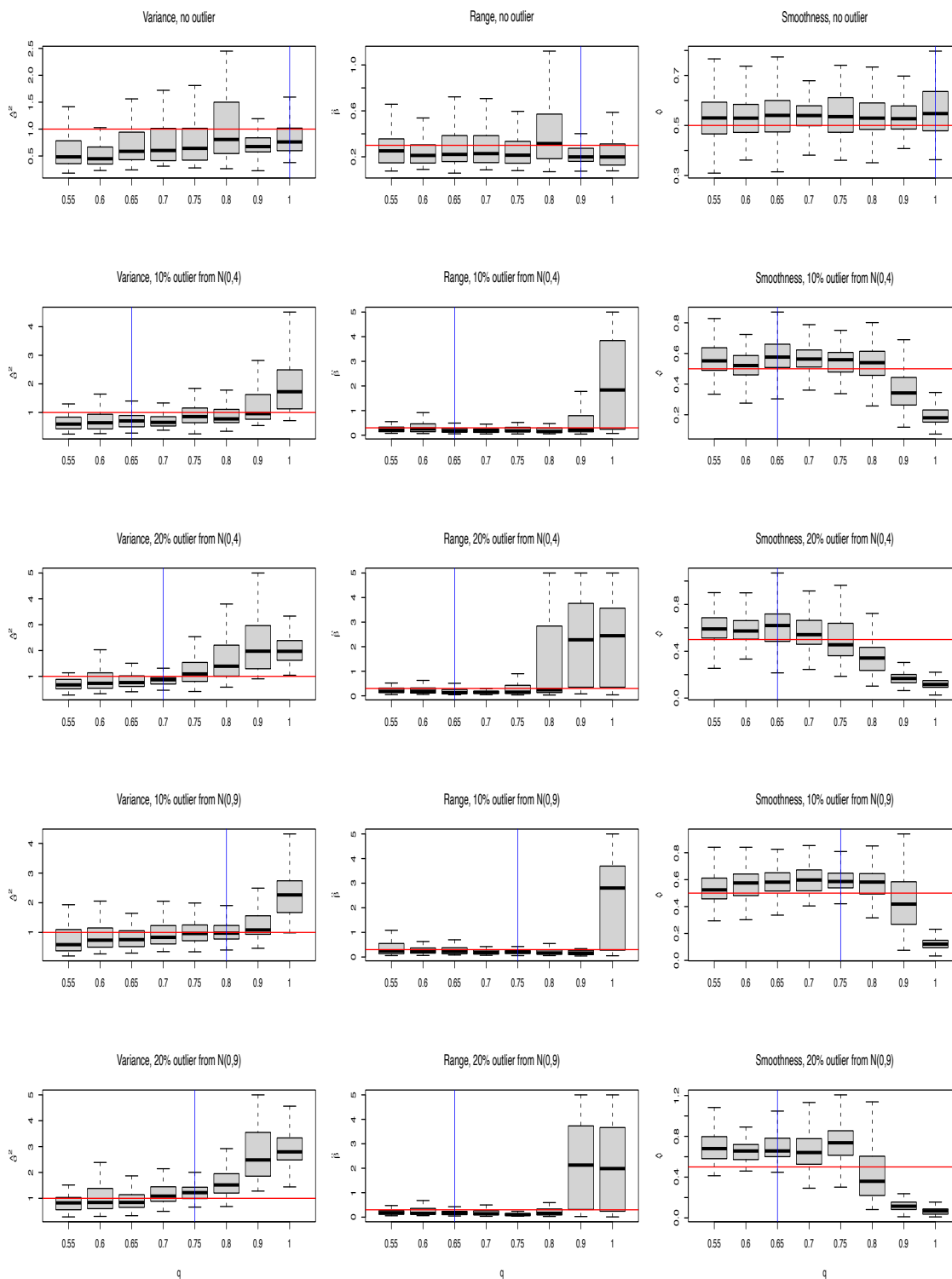


Figure A2. Boxplots (based on 90 simulations) of the MCLqE estimation results for spatial datasets generated under the strong-rough setting of the Matérn covariance function with parameters  $\sigma^2 = 1$ ,  $\beta = 0.3$ ,  $\nu = 0.5$ . The values of  $q$  that we use here are 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1. The red horizontal lines correspond to the true values of the parameters, and the blue vertical lines indicate the smallest MSE.

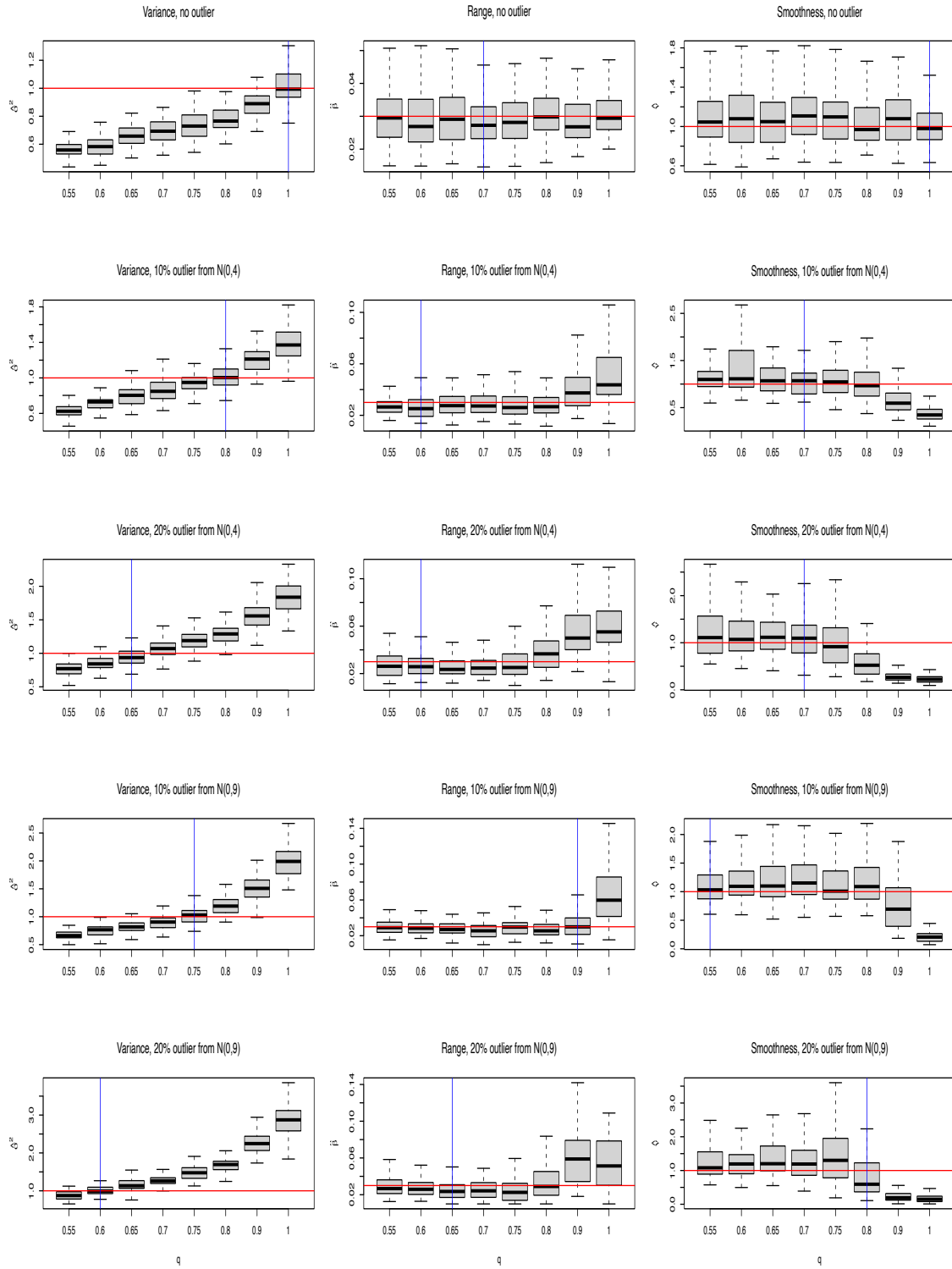


Figure A3. Boxplots (based on 90 simulations) of the MCLqE estimation results for spatial datasets generated under the weak-smooth setting of the Matérn covariance function with parameters  $\sigma^2 = 1, \beta = 0.03, \nu = 1$ . The values of  $q$  that we use here are 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1. The red horizontal lines correspond to the true values of the parameters, and the blue vertical lines indicate the smallest MSE.

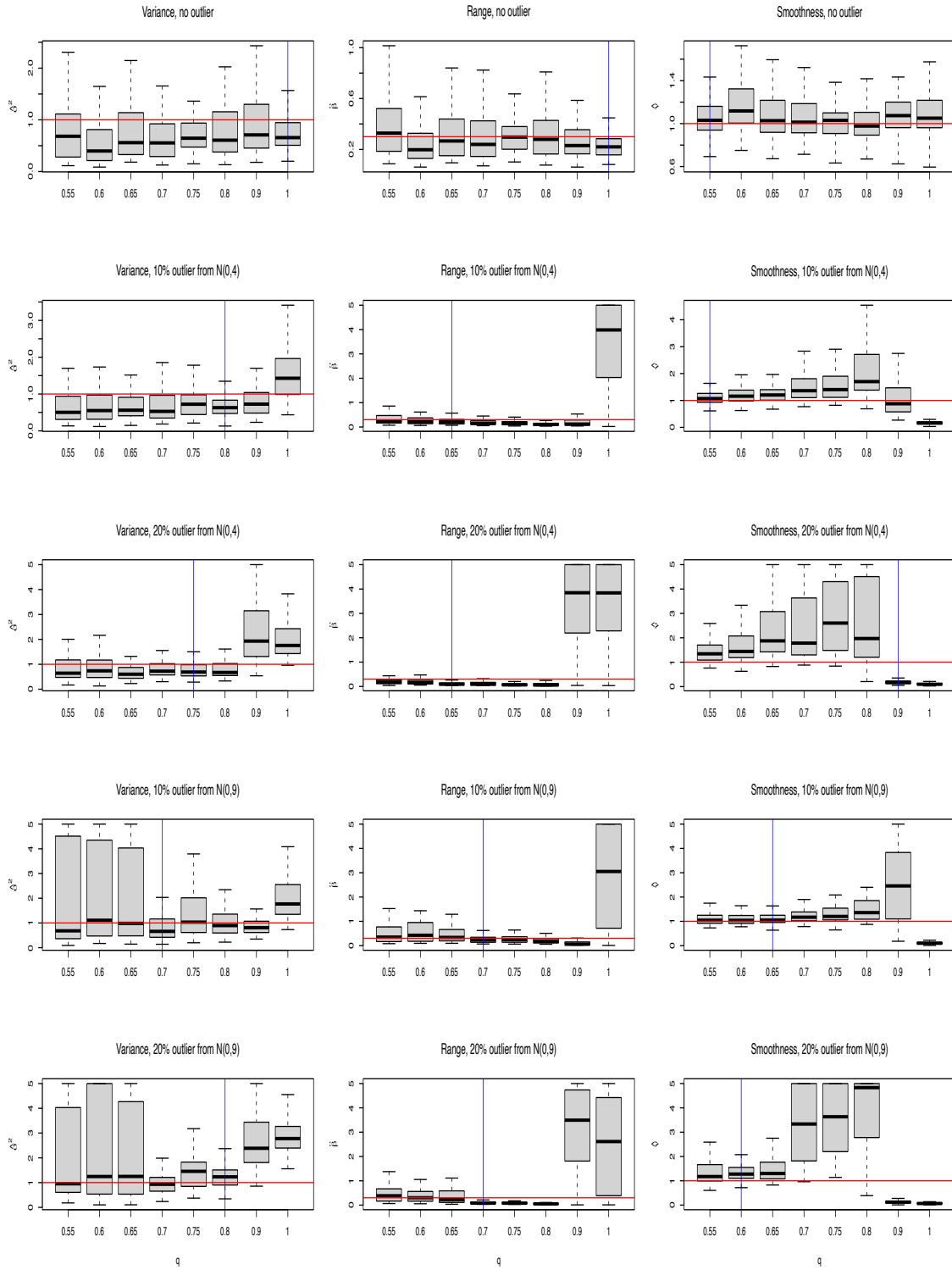


Figure A4. Boxplots (based on 90 simulations) of the MCLqE estimation results for spatial datasets generated under the strong-smooth setting of the Matérn covariance function with parameters  $\sigma^2 = 1$ ,  $\beta = 0.3$ ,  $\nu = 1$ . The values of  $q$  that we use here are 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1. The red horizontal lines correspond to the true values of the parameters, and the blue vertical lines indicate the smallest MSE.

## REFERENCES

- [1] Smith, R.L., 2000. Spatial statistics in environmental science. Fitzgerald, W.J., Smith, R.L., Walden, A.T., and Young, P.C. (Eds.). *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press, Cambridge, UK, pp. 152–183. URL: [rsl.sites.oasis.unc.edu/postscript/rs/ss.pdf](https://rsl.sites.oasis.unc.edu/postscript/rs/ss.pdf).
- [2] Sain, S.R., Furrer, R., and Cressie, N., 2011. A spatial analysis of multivariate output from regional climate models. *The Annals of Applied Statistics*, 22, 150–175. DOI: [10.1214/10-AOAS369](https://doi.org/10.1214/10-AOAS369).
- [3] Lawson, A.B., 2018. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman and Hall/CRC, Boca Raton, FL, USA. DOI: [10.1201/9781351271769](https://doi.org/10.1201/9781351271769).
- [4] Chiles, J.P. and Delfiner, P., 2012. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, NY, USA. URL: [onlinelibrary.wiley.com/doi/book/10.1002/9781118136188](https://onlinelibrary.wiley.com/doi/book/10.1002/9781118136188).
- [5] Gräler, B., Pebesma, E., and Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. *The R Journal*, 8, 203–218. DOI: [10.32614/RJ-2016-014](https://doi.org/10.32614/RJ-2016-014).
- [6] Cressie, N. and Kornak, J., 2003. Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statistical Science*, 18, 436–456. DOI: [10.1214/ss/1081443228](https://doi.org/10.1214/ss/1081443228).
- [7] Chen, S., Abdulah, S., Sun, Y., and Genton, M.G., 2024. On the impact of spatial covariance matrix ordering on tile low-rank estimation of Matérn parameters. *Environmetrics*, 35, e2868. DOI: [10.1002/env.2868](https://doi.org/10.1002/env.2868).
- [8] Baboulin, M., Buttari, A., Dongarra, J., Kurzak, J., Langou, J., Langou, J., Luszczek, P., and Tomov, S., 2009. Accelerating scientific computations with mixed precision algorithms. *Computer Physics Communications*, 180, 2526–2533. DOI: [10.1016/j.cpc.2008.11.005](https://doi.org/10.1016/j.cpc.2008.11.005).
- [9] Lindsay, B.G., 1988. Composite likelihood methods. *Statistical Inference from Stochastic Processes: Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference Held August 9–15, 1987, with support from the National Science Foundation and the Army Research Office*, American Mathematical Society, Ithaca, NY, USA, p. 221. DOI: [10.1090/conm/080/999014](https://doi.org/10.1090/conm/080/999014).
- [10] Zhao, Y. and Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33, 335–356. DOI: [10.1002/cjs.5540330303](https://doi.org/10.1002/cjs.5540330303).
- [11] Gao, X. and Song, P. X.K., 2010. Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association*, 105, 1531–1540. DOI: [10.1198/jasa.2010.tm09414](https://doi.org/10.1198/jasa.2010.tm09414).
- [12] Varin, C., Reid, N., and Firth, D., 2011. An overview of composite likelihood methods. *Statistica Sinica*, 21, 5–42. URL: [www3.stat.sinica.edu.tw/sstest/oldpdf/A21n11.pdf](http://www3.stat.sinica.edu.tw/sstest/oldpdf/A21n11.pdf).
- [13] Heagerty, P.J. and Lele, S.R., 1998. A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 93, 1099–1111. DOI: [10.2307/2669853](https://doi.org/10.2307/2669853).
- [14] Feng, X., Zhu, J., Lin, P.S., and Steen-Adams, M.M., 2014. Composite likelihood estimation for models of spatial ordinal data and spatial proportional data with zero/one values. *Environmetrics*, 25, 571–583. DOI: [10.1002/env.2306](https://doi.org/10.1002/env.2306).
- [15] Guan, Y., 2006. A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101, 1502–1512. DOI: [10.1198/016214506000000500](https://doi.org/10.1198/016214506000000500).
- [16] Baddeley, A., 2017. Local composite likelihood for spatial point processes. *Spatial Statistics*, 22, 261–295. DOI: [10.1016/j.spasta.2017.03.001](https://doi.org/10.1016/j.spasta.2017.03.001).

- [17] Paik, J. and Ying, Z., 2012. A composite likelihood approach for spatially correlated survival data. *Computational Statistics and Data Analysis*, 56, 209–216. DOI: [10.1016/j.csda.2011.07.004](https://doi.org/10.1016/j.csda.2011.07.004).
- [18] Ribatet, M., Cooley, D., and Davison, A.C., 2012. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22, 813–845. URL: [www3.stat.sinica.edu.tw/ssctest/oldpdf/A22n217.pdf](http://www3.stat.sinica.edu.tw/ssctest/oldpdf/A22n217.pdf).
- [19] Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E., 2012. Estimating space and space-time covariance functions for large data sets: A weighted composite likelihood approach. *Journal of the American Statistical Association*, 107, 268–280. DOI: [10.1080/01621459.2011.646928](https://doi.org/10.1080/01621459.2011.646928)
- [20] Eidsvik, J., Shaby, B.A., Reich, B.J., Wheeler, M., and Niemi, J., 2014. Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, 23, 295–315. DOI: [10.1080/10618600.2012.760460](https://doi.org/10.1080/10618600.2012.760460).
- [21] Ferrari, D. and Yang, Y., 2010. Maximum  $L_q$ -likelihood estimation. *The Annals of Statistics*, 38, 753–783. DOI: [10.1214/09-AOS687](https://doi.org/10.1214/09-AOS687).
- [22] Ferrari, D. and La Vecchia, D., 2012. On robust estimation via pseudo-additive information. *Biometrika*, 99, 238–244. URL: [www.jstor.org/stable/41720686](http://www.jstor.org/stable/41720686).
- [23] Chen, S., Chowdhury, J., and Genton, M.G., 2025. Robust maximum  $L_q$ -likelihood covariance estimation for replicated spatial data. *Journal of Data Science, Statistics, and Visualisation*, 5, 4. DOI: [10.52933/jdssv.v5i4.126](https://doi.org/10.52933/jdssv.v5i4.126).
- [24] Raymaekers, J. and Rousseeuw, P.J., 2021. Handling cellwise outliers by sparse regression and robust covariance. *Journal of Data Science, Statistics, and Visualisation*, 1, 3. DOI: [10.52933/jdssv.v1i3.18](https://doi.org/10.52933/jdssv.v1i3.18).
- [25] Raymaekers, J. and Rousseeuw, P.J., 2024. Challenges of cellwise outliers. *Econometrics and Statistics*, in press. DOI: [10.1016/j.ecosta.2024.02.002](https://doi.org/10.1016/j.ecosta.2024.02.002).
- [26] Heagerty, P.J. and Lumley, T., 2000. Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association*, 95, 197–211. DOI: [10.1080/01621459.2000.10473914](https://doi.org/10.1080/01621459.2000.10473914).
- [27] Ribeiro, T.K.A. and Ferrari, S.L.P., 2022. Robust estimation in beta regression via maximum  $L_q$ -likelihood. *Statistical Papers*, 64, 321–353. DOI: [10.1007/s00362-022-01320-0](https://doi.org/10.1007/s00362-022-01320-0).
- [28] R Core Team, 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: [www.R-project.org](http://www.R-project.org).
- [29] Ribeiro Jr, P.J. and Diggle, P.J., 2007. The geoR package. *R Journal*, 1, 14–18. DOI: [10.32614/CRAN.package.geoR](https://doi.org/10.32614/CRAN.package.geoR).
- [30] Powell, M.J.D., 2009. The BOBYQA algorithm for bound constrained optimization without derivatives. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, UK, 26. URL: [api.semanticscholar.org/CorpusID:2488733](https://api.semanticscholar.org/CorpusID:2488733).
- [31] Abdulah, S., Ltaief, H., Sun, Y., Genton, M.G., and Keyes, D.E., 2018. Exa-GeoStat: A high performance unified software for geostatistics on manycore systems. *IEEE Transactions on Parallel and Distributed Systems*, 29, 2771–2784. DOI: [10.1109/TPDS.2018.2850749](https://doi.org/10.1109/TPDS.2018.2850749).

**Disclaimer/publisher’s note:** The views, opinions, data, and information presented in all publications of the ChJS are solely those of the individual authors and contributors, and do not necessarily reflect the views of the journal or its editors. The journal and its editors assume no responsibility or liability for any harm to people or property resulting from the use of ideas, methods, instructions, or products mentioned in the content.

