Statistica Sinica Preprint No: SS-2024-0339	
Title	Nonparametric Spatial Modeling towards the Mode
Manuscript ID	SS-2024-0339
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0339
Complete List of Authors	Tao Wang and Weixin Yao
Corresponding Authors	Tao Wang
E-mails	taow@uvic.ca
Notice: Accepted author version.	

Statistica Sinica

Nonparametric Spatial Modeling towards the Mode

Tao Wang[a] Weixin Yao[b]

a. University of Victoria b. University of California Riverside

Abstract: Existing models for spatial data analysis typically rely on mean or quantile regression to model the association between a dependent variable and covariates. We in this paper propose a novel spatial modal regression by assuming that the conditional mode of the response Y given covariates $\mathbf{X}$ follows a nonparametric regression structure, defined as $m : \mathbf{X} \mapsto m(\mathbf{X}) := Mode(Y_{\mathbf{i}} \mid \mathbf{X}_{\mathbf{i}})$, $\mathbf{X}_{\mathbf{i}} \in \mathbb{R}^d$ and $\mathbf{i} \in \mathbb{Z}^N$. The suggested spatial modal regression can be utilized to capture the "most likely" effect and may reveal new interesting data structures that are possibly missed by the conditional mean or quantiles, especially in cases of asymmetric data distributions. We derive the asymptotic distributions for the resulting modal estimators with appropriate choices of bandwidths. To numerically estimate the developed model, we recommend a modified modal expectation-maximization (MEM) algorithm with the assistance of a Gaussian kernel. Numerical examples are presented to demonstrate the favorable finite sample performance of the estimators. We also generalize the propounded spatial modal regression to an additive sum form to offer a versatile solution to handle high-dimensional datasets.

Key words and phrases: Additive model, Local linear, MEM algorithm, Modal regression, Spatial process.

1. Introduction

Because of its ability to capture the "most likely" value of a distribution, the mode offers a compelling alternative to traditional measures of central tendency such as the mean or median. This motivates the development of *modal regression*, defined as $Mode(Y \mid \mathbf{X}) = \arg\max_Y f(Y \mid \mathbf{X})$, where $f(Y \mid \mathbf{X})$ represents the conditional distribution of Y given $\mathbf{X}$; see Yao and Li (2014), Chen et al. (2016), Chen (2018), Kemp et al. (2020), Xiang and Yao (2022), Ullah et al. (2021, 2022, 2023), and Wang (2025). Modal regression is particularly advantageous in the presence of skewness, heavy tails, or multimodality, where mean or median regression may provide misleading summaries. In such cases, the conditional mean may fall in regions of low probability density, masking dominant patterns and rendering interpretation difficult. For instance, in agricultural applications, outcomes such as chemical concentrations often exhibit non-Gaussian structures with multiple local peaks, driven by site-specific heterogeneity or diverse land-use practices. By targeting the most probable outcome, modal regression can provide a more robust and interpretable insights of the underlying process.

However, existing modal regression literature has largely focused on time series, cross-sectional, or panel data, with limited development in spatial contexts. This is due, in part, to the lack of natural ordering in space and the complex dependence structures that make spatial modal estimation especially challenging. In addition, spatial data analysis has traditionally relied on conditional mean or quantile regression; see Lu and Chen (2004), Robinson (2008, 2011), Hallin et al. (2004), Hallin et al. (2009), among others. While these methods are effective in capturing average or percentile behavior, they may obscure key features of the underlying distribution in the presence of heterogeneity or multiple modes. For instance, as demonstrated in our soil chemistry analysis in Section 3, the conditional distribution of cation exchange capacity exhibits multiple peaks, reflecting variation in soil treatment regimes and land use history. In such cases, the conditional mean may misrepresent the central structure of the data and lie in regions of low density. Motivated by these considerations, we propose a novel *spatial modal regression model*, which estimates the conditional mode function to provide more informative and distribution-aware summaries in spatial settings.

We in this paper consider data observed over a space of general dimension N, represented by $\mathbb{Z}^N$, where $N \geq 1$ and $\mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$, denoting the set of integer lattice points in an N-dimensional Euclidean space. A point $\mathbf{i} = (i_1, \cdots, i_N)$ in $\mathbb{Z}^N$ is referred to as a *site* that typically reflects economic characteristics or geographical positions, and may contain a time component. Spatial data are conceptualized as finite realizations of vector stochastic processes indexed by $\mathbf{i} \in \mathbb{Z}^N$, commonly known as *random fields*. We focus on strictly stationary $(d+1)$-dimensional real random fields of the form $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}); \mathbf{i} \in \mathbb{Z}^N\}$, where $Y_{\mathbf{i}} \in \mathbb{R}$, a scalar dependent variable, and $\mathbf{X}_{\mathbf{i}} \in \mathbb{R}^d$, covariates with compact support $\Gamma \subseteq \mathbb{R}^d$, are defined over some probability space $(\Omega, \mathcal{F}, P)$. Our objective is to characterize the spatial dependence between $Y_{\mathbf{i}}$ and $\mathbf{X}_{\mathbf{i}}$ by estimating the *spatial modal regression function*

$$m : \mathbf{X} \mapsto m(\mathbf{X}) := Mode(Y_{\mathbf{i}} \mid \mathbf{X}_{\mathbf{i}}) \text{ almost surely (a.s.),} \tag{1.1}$$

assuming that given a fixed $\mathbf{X}_{\mathbf{i}}$, the dependent variable $Y_{\mathbf{i}}$ has a unique global mode, where $m(\cdot)$ is regarded as a well-defined real-valued $\mathbf{X}$-measurable function, defined almost everywhere except on a P-null set of $\mathbf{X}$-values.

To estimate the proposed model, we extend the local linear approximation method to the spatial setting due to its several desirable properties (Fan and Gijbels, 1996), contrasting with spatial smoothing methods that smooth over site $\mathbf{i}$ based on proximity or dependence structures. We establish the asymptotic properties of the resulting estimators and show that their convergence rate is $(\tilde{\mathbf{n}} h_2^d h_1^3)^{1/2}$ ($\tilde{\mathbf{n}}$, h_1, and h_2 are defined in Section 2), which is slower than the rate $(\tilde{\mathbf{n}} h_2^d)^{1/2}$ associated with spatial mean regression. Our framework imposes no restrictions on the configuration of the sample region and enables the sample to expand in various directions at different rates. We derive theoretically optimal bandwidths under our proposed modal criterion and suggest a data-driven bandwidth selection procedure for empirical implementation. We further develop a modified MEM algorithm with the assistance of a Gaussian kernel to numerically estimate the targeted model. To illustrate the flexibility of our approach, we provide several extended modal regression models in the Supplementary Material-S6.

Despite these advantages, the proposed spatial modal regression remains susceptible to the curse of dimensionality when the covariate dimension d becomes large. This issue is especially pronounced in lattice-based applications, such as when analyzing spatial grid data $\{Y_{i,j}, (i,j) \in \mathbb{Z}^2\}$, where estimating the conditional mode of $Y_{i,j}$ based on neighboring values $\mathbf{X}_{i,j} = \{Y_{i-1,j}, Y_{i,j-1}, Y_{i+1,j}, Y_{i,j+1}\}$ requires a four-dimensional nonparametric regression. Although several dimension-reduction techniques have been developed in the literature of nonparametric regression, such as additive modeling and sparse regularization (Hastie and Tibshirani, 1990; Gao et al., 2006; Lu et al., 2007, 2014; Nandy et al., 2017), these approaches have not been extended to modal regression in spatial settings. To address this challenge, we in the end extend the proposed model to develop an *additive spatial modal regression* framework, which decomposes the multivariate regression surface into additive components while accounting for spatially dependent errors. Due to space limitations, the full specification of this model, along with its asymptotic theory, is provided in the Supplementary Material-S5.

The rest of this paper is organized as follows. Section 2 presents the local linear modal estimation procedure and establishes its asymptotic properties under stationary spatial dependence. Section 3 evaluates the finite sample performance of the proposed estimators through simulations and real data analysis. Section 4 concludes the paper. Additional simulation studies, technical comments and proofs, and generalizations to extended spatial modal regression models are provided in the Supplementary Material.

2. Local Linear Spatial Modal Regression

We begin this section by formulating the modeling framework and introducing the local linear modal estimation procedure, accompanied by a practical numerical algorithm. We then delve into examining the consistency and asymptotic properties of the resulting estimators across various scenarios.

2.1 Model Framework

We suppose that the random field is observed over a rectangular region of the form $I_{\mathbf{n}} = \left\{\mathbf{i} = (i_1, \cdots, i_N) \in \mathbb{Z}^N : 1 \le i_l \le n_l, l = 1, \cdots, N\right\}$ with $\mathbf{n} = (n_1, \cdots, n_N) \to \infty$ if $\min_{1\le l\le N}\{n_l\} \to \infty$ and $|n_k/n_l| < C$, $k = 1, \cdots, N$, for a constant C such that $0 < C < \infty$. The defined $I_{\mathbf{n}}$ form implies that n_l,

$1 \leq l \leq N$, tend to infinity at the same rate, known as *isotropic divergence*. As discussed later in this section, our asymptotic results remain valid with suitable conditions if the rates of expansion are not the same along all directions (i.e., only $\min_{1\leq l\leq N}\{n_l\} \to \infty$ holds), which is termed as *nonisotropic divergence*. The total sample size is thus $\tilde{\mathbf{n}} = \prod_{l=1}^{N} n_l$. To formulate the spatial modal regression, we assume that $\{Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}\}$ observed on $I_{\mathbf{n}}$ satisfies

$$Y_{\mathbf{i}} = Mode(Y_{\mathbf{i}} \mid \mathbf{X}_{\mathbf{i}}) + \varepsilon_{\mathbf{i}} = m(\mathbf{X}_{\mathbf{i}}) + \varepsilon_{\mathbf{i}}, \tag{2.1}$$

where $\{\varepsilon_{\mathbf{i}}, \mathbf{i} \in I_{\mathbf{n}}\}$ are the random disturbances with zero conditional mode and have identical marginal distributions but may exhibit dependence among each other such that the strong spatial mixing property remains valid. Before illustrating the developed estimation procedure, we first formalize the concept of the modal estimator with the following definition.

Definition 2.1. If the data $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})\}_{\mathbf{i}\in I_{\mathbf{n}}}$ are independent and identically distributed (i.i.d.), given a kernel function $K(\cdot)$ satisfying condition C2 stated in Subsection 2.3 and a shrinking bandwidth h, the modal estimator of $\boldsymbol{\theta}$ with respect to a function $m(\mathbf{X}_{\mathbf{i}}, \boldsymbol{\theta})$ is defined as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \lim_{h\to 0} \mathbb{E}[L_{\boldsymbol{\theta}}(Y, \mathbf{X})], \text{ where } L_{\boldsymbol{\theta}}(Y, \mathbf{X}) = \frac{1}{h} K\left(\frac{Y - m(\mathbf{X}, \boldsymbol{\theta})}{h}\right).$$

Definition 2.1 is consistent with the principles of kernel density estimation as discussed in Chen et al. (2016). Let $g(\varepsilon)$ denote the continuous density function of ε, and let $K(\cdot)$ be a bounded and integrable probability density function with compact support, satisfying condition C2 in Subsection 2.3. Using standard properties of kernel smoothing, we obtain

$$\begin{aligned}\sup_{\varepsilon\in\mathbb{R}}\left|g(\varepsilon)-\int K(w)g(\varepsilon+wh)dw\right| &\leq \sup_{\varepsilon\in\mathbb{R}}\int |g(\varepsilon)-g(\varepsilon+wh)|K(w)dw \\ &\leq \sup_{\varepsilon\in\mathbb{R}}\int |g^{(1)}(\varepsilon)wh|K(w)dw \to 0\end{aligned}$$

as $h\to 0$, where $g^{(1)}(\varepsilon)$ represents the first derivative of $g(\varepsilon)$. Therefore, there exists a modal parameter $\boldsymbol{\theta}$ that can maximize the density of ε, forming the basis of the modal estimation procedure.

Building on this foundation, we propose a spatial modal estimator as a natural extension of its i.i.d. counterpart. Specifically, the objective function $L_{\boldsymbol{\theta}}(Y,\mathbf{X})$ defined in Definition 2.1 remains valid in the presence of spatial dependence, with its asymptotic properties rigorously justified via a spatial mixing framework. To this end, we assume that $\{(Y_{\mathbf{i}},\mathbf{X}_{\mathbf{i}})\}_{\mathbf{i}\in I_{\mathbf{n}}}$ satisfies the following spatial mixing condition, extending the strong mixing assumptions used for continuous-time stochastic processes and time series models; see Hallin et al. (2004), Gao et al. (2006), and Hallin et al. (2009).

Definition 2.2. Suppose that S and S' are two sets of sites. The Borel fields $\mathcal{B}(S)=\mathcal{B}((Y_{\mathbf{i}},\mathbf{X}_{\mathbf{i}}):\mathbf{i}\in S)$ and $\mathcal{B}(S')=\mathcal{B}((Y_{\mathbf{i}'},\mathbf{X}_{\mathbf{i}'}):\mathbf{i}'\in S')$ are the σ-fields generated by $(Y,\mathbf{X})$. Let $d(S,S')=\min\{\|\mathbf{i}-\mathbf{i}'\|\mid \mathbf{i}\in S,\mathbf{i}'\in S'\}$ denote the distance between S and S', where $\|\mathbf{i}\| = (i_1^2 + \cdots + i_N^2)^{1/2}$ stands for the Euclidean norm. Then, spatial mixing is defined such that there exists a function $\varphi(t) \downarrow 0$ as $t \to \infty$ and whenever $S, S' \subset Z^N$,

$$\alpha\left(\mathcal{B}(S), \mathcal{B}\left(S'\right)\right) = \sup\left\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(S), B \in \mathcal{B}\left(S'\right)\right\}$$
$$\leq \chi\left(\mathrm{Card}(S), \mathrm{Card}\left(S'\right)\right) \varphi\left(d\left(S, S'\right)\right),$$

where Card (S) (respectively Card (S')) indicates the cardinality of S (respectively S') and $\chi(\cdot)$ is a symmetric positive function nondecreasing in each of its two arguments. If $\chi(\infty, \infty) = C$ for some positive constant C, the sequence $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})\}_{\mathbf{i} \in I_{\mathbf{n}}}$ is called α-mixing (or strong mixing).

The α-mixing dependence is a mild restriction towards achieving asymptotic properties among a variety of mixing conditions. It was discussed in Hallin et al. (2004) that a spatial process of the form $\mathbf{X}_{\mathbf{n}} = \sum_{\mathbf{i} \in \mathbb{Z}^N} a_{\mathbf{i}} Z_{\mathbf{n}-\mathbf{i}}$ can satisfy $\alpha\left(\mathcal{B}(S), \mathcal{B}\left(S'\right)\right)$ if $Z_{\mathbf{i}}$'s are independent random variables, $a_{\mathbf{i}} \to 0$ grows exponentially fast, and the probability density function of $Z_{\mathbf{i}}$ exists.

2.2 Local Linear Modal Estimation

We generalize the local linear framework of Fan and Gijbels (1996) to accommodate the structure of the proposed spatial modal regression. Under the smoothness assumptions specified in condition C4 (Subsection 2.3), the regression function $m(\mathbf{X})$ admits a first-order Taylor expansion in a neighborhood of the target point $\mathbf{x}$, given by $m(\mathbf{X}) = m(\mathbf{x})+(m^{(1)}(\mathbf{x}))^T(\mathbf{X}-\mathbf{x})+R(\mathbf{X})$, where $m^{(1)}(\mathbf{x})$ denotes the gradient vector at $\mathbf{x}$, and $R(\mathbf{X})$ is a second-order remainder term. For observations $\mathbf{X}$ sufficiently close to $\mathbf{x}$, the remainder term becomes asymptotically negligible, yielding the approximation

$$m(\mathbf{X}) \approx m(\mathbf{x}) + (m^{(1)}(\mathbf{x}))^T(\mathbf{X} - \mathbf{x}) := a + \mathbf{b}^T(\mathbf{X} - \mathbf{x}), \qquad (2.2)$$

where the notation "$\approx$" signifies equality up to a higher-order error, with $R(\mathbf{X}) = o(\|\mathbf{X}-\mathbf{x}\|)$ as $\mathbf{X} \to \mathbf{x}$. Therefore, estimating $m(\mathbf{x})$ and its gradient $m^{(1)}(\mathbf{x})$ is locally equivalent to estimating $(a, \mathbf{b}^T)^T = (a(\mathbf{x}), \mathbf{b}^T(\mathbf{x}))^T$.

Based on Definition 2.1 and associated arguments, we can obtain the following kernel-based objective function for achieving spatial modal estimates

$$Q_{\tilde{\mathbf{n}}}(\boldsymbol{\theta}) = \frac{1}{\tilde{\mathbf{n}} h_1 h_2^d} \sum_{\mathbf{i}\in I_{\mathbf{n}}} \phi\left(\frac{Y_{\mathbf{i}} - a - \mathbf{b}^T(\mathbf{X_i} - \mathbf{x})}{h_1}\right) K\left(\frac{\mathbf{X_i} - \mathbf{x}}{h_2}\right), \qquad (2.3)$$

where $\boldsymbol{\theta} = (a, \mathbf{b}^T)^T$, the kernel $\phi(\cdot) : \mathbb{R} \to \mathbb{R}$ is defined on $\mathbb{R}$ with bandwidth $h_1 = h_1(\mathbf{n}) > 0$ tending to 0 as $\mathbf{n} \to 0$, and the kernel $K(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is a nonnegative weight function defined on $\mathbb{R}^d$ with bandwidth $h_2 = h_2(\mathbf{n}) > 0$ such that $\lim_{\mathbf{n}\to 0} h_2(\mathbf{n}) = 0$. Note that $\phi(\cdot)$ is utilized to capture the mode value according to Definition 2.1, while $K(\cdot)$, in line with nonparametric mean estimation, represents the weight assigned locally to observations $\{(Y_{\mathbf{i}}, \mathbf{X_i})\}_{\mathbf{i}\in I_{\mathbf{n}}}$. According to Yao and Li (2014) and Ullah et al. (2021, 2022, 2023), the choice of kernels is not particularly essential for modal regression models. For computational simplicity, we select the Gaussian kernel for $\phi(\cdot)$ in this paper to employ a so-called MEM Algorithm 1. We denote the corresponding estimators from (2.3) as $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{\mathbf{b}}^T)^T = (\hat{m}(\mathbf{x}), \hat{m}^{(1)}(\mathbf{x})^T)^T$.

Algorithm 1: MEM Algorithm for Spatial Modal Regression

Data: Sample observations $\{(Y_{\mathbf{i}}, X_{\mathbf{i}})\}_{\mathbf{i}\in I_{\mathbf{n}}}$ and bandwidths h_1, h_2.
Result: Final modal estimates $(\hat{m}(\mathbf{x}), \hat{m}^{(1)}(\mathbf{x})^T)^T$.
while *two consecutive solutions are not close enough, i.e.,* $\|\hat{m}(\mathbf{x})^{(g)} - \hat{m}(\mathbf{x})^{(g-1)}\| > 10^{-4}$*, or a pre-specified maximum number of iterations (i.e., g =100) is not reached* **do**
  if *current estimate* $\hat{m}(\mathbf{x})^{(g)}$ *with iterative indicator* $g \geq 1$ **then**
    E-Step: Calculate the weight $\pi\big(\mathbf{i} \mid \boldsymbol{\theta}^{(g)}\big)$, $\mathbf{i} \in I_{\mathbf{n}}$, with the preliminary estimates of the modal parameters as

$$\pi\big(\mathbf{i} \mid \boldsymbol{\theta}^{(g)}\big) = \frac{\phi\Big(\frac{Y_{\mathbf{i}} - a^{(g)} - \mathbf{b}^{(g)T}(\mathbf{X}_{\mathbf{i}} - \mathbf{x})}{h_1}\Big) K\Big(\frac{\mathbf{X}_{\mathbf{i}} - \mathbf{x}}{h_2}\Big)}{\sum_{\mathbf{i}\in I_{\mathbf{n}}} \phi\Big(\frac{Y_{\mathbf{i}} - a^{(g)} - \mathbf{b}^{(g)T}(\mathbf{X}_{\mathbf{i}} - \mathbf{x})}{h_1}\Big) K\Big(\frac{\mathbf{X}_{\mathbf{i}} - \mathbf{x}}{h_2}\Big)}$$
$$\propto \phi\Big(\frac{Y_{\mathbf{i}} - a^{(g)} - \mathbf{b}^{(g)T}(\mathbf{X}_{\mathbf{i}} - \mathbf{x})}{h_1}\Big) K\Big(\frac{\mathbf{X}_{\mathbf{i}} - \mathbf{x}}{h_2}\Big),$$

    which is nonnegative and sums to one.
    M-Step: Update the estimates with the weight computed in the E-Step by log-maximization

$$\boldsymbol{\theta}^{(g+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{\mathbf{i}\in I_{\mathbf{n}}} \Big\{\pi\big(\mathbf{i} \mid \boldsymbol{\theta}^{(g)}\big) \log \frac{1}{h_1}\phi\Big(\frac{Y_{\mathbf{i}} - a - \mathbf{b}^T(\mathbf{X}_{\mathbf{i}} - \mathbf{x})}{h_1}\Big)\Big\}$$
$$= (\mathbf{X}^{*T} W_{\mathbf{X}} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} W_{\mathbf{X}} \mathbf{Y},$$

    where $\mathbf{X}^* = (\mathbf{X}^*_{i_1}, \cdots, \mathbf{X}^*_{i_N})^T$ with $\mathbf{X}^*_{i_j} = (1, \mathbf{X}_{i_j} - \mathbf{x})$, $j = 1, \cdots, N$, $\mathbf{Y} = (\mathbf{Y}_{i_1}, \cdots, \mathbf{Y}_{i_N})^T$, and $W_{\mathbf{X}}$ is an $\tilde{\mathbf{n}} \times \tilde{\mathbf{n}}$ diagonal matrix with diagonal elements $\{\pi(\mathbf{i} \mid \boldsymbol{\theta}^{(g)}), \mathbf{i} \in I_{\mathbf{n}}\}$.
  end
end

In contrast to spatial mean regression, the maximizer $\boldsymbol{\theta}$ of the kernel-based objective function (2.3) does not admit a closed-form solution. To address this, we adopt a modified efficient MEM algorithm inspired by Yao et al. (2012) and Yao (2013), leveraging the Gaussian kernel for $\phi(\cdot)$. The MEM algorithm iteratively maximizes a surrogate lower bound of the objective function $Q_{\tilde{\mathbf{n}}}(\boldsymbol{\theta})$, consistent with standard EM theory. Specifically, by Jensen's inequality, we obtain $\log(Q_{\tilde{\mathbf{n}}}(\boldsymbol{\theta})) \geq \frac{1}{\tilde{\mathbf{n}} h_1 h_2^d} \sum_{\mathbf{i} \in I_{\mathbf{n}}} \pi(\mathbf{i} \mid \boldsymbol{\theta}^{(g)}) \log [\phi(\frac{Y_{\mathbf{i}} - a - \mathbf{b}^T(\mathbf{X}_{\mathbf{i}} - \mathbf{x})}{h_1}) K(\frac{\mathbf{X}_{\mathbf{i}} - \mathbf{x}}{h_2})]$, ensuring monotonic ascent towards a local optimum of the log-kernel likelihood. We emphasize, however, that the MEM algorithm does not guarantee convergence to the global mode. Since the objective surface is generally non-convex and may exhibit multiple local optima, the convergence behavior is inherently sensitive to initialization. We provide an extensive discussion in the Supplementary Material-S1 detailing practical heuristics to mitigate local convergence issues. These include (i) initializing from multiple starting points such as local linear mean or quantile estimates; (ii) employing tempered EM variants or incorporating controlled stochastic perturbations to escape shallow modes; and (iii) selecting the final estimate by comparing kernel likelihoods across candidate solutions. In addition, we theoretically establish in Theorem S1 (Supplementary Material-S1) that the MEM algorithm achieves local quadratic convergence when initialized sufficiently close to a mode under standard regularity conditions. This theoretical guarantee parallels the Newton–Kantorovich theorem (Ortega and Rheinboldt, 1970), which ensures exponential convergence within a local neighborhood of the true maximizer. Empirical results in Section 3 reinforce this behavior as the modal estimator consistently achieves higher kernel likelihood than local mean regression, demonstrating robust performance across a range of noise levels and initialization strategies. We also note a potential numerical instability in the M-step of Algorithm 1, which involves inversion of the weighted local design matrix $(\mathbf{X}^{*T} W_{\mathbf{X}} \mathbf{X}^*)$. As $h_2 \to 0$, the Gaussian kernel induces exponential decay in the weights $\pi(\mathbf{i} \mid \boldsymbol{\theta}^{(g)})$, which may result in near-singularity or high condition numbers due to insufficient local effective sample sizes. This ill-conditioning arises from over-localization of the kernel. Nevertheless, the bandwidth selection procedure in Subsection 2.4 inherently guards against such issues by discouraging excessive localization. The optimal bandwidth rate $\tilde{\mathbf{n}}^{-\frac{1}{d+7}}$ ensures a stable effective sample size in each neighborhood, preventing degeneracy of the design matrix. In practice, to further guard against numerical instability, we recommend implementing a ridge-regularized version of the M-step to stabilize matrix inversion; see Supplementary Material-S1 for details.

2.3 Asymptotic Properties

To simplify the exposition, we introduce some notations that will be utilized subsequently. The letter C denotes a constant, the value of which may vary for convenience. The symbol "$\xrightarrow{d}$" signifies the convergence in distribution. For clarity, we define $H_2 = \text{diag}(h_2, \cdots, h_2)_{d\times d}$, $\phi_{h_1}(\varepsilon_{\mathbf{i}}) = h_1^{-1}\phi(\varepsilon_{\mathbf{i}}/h_1)$, and $K_{\mathbf{i}} = K_{h_2}(\mathbf{X}_{\mathbf{i}} - \mathbf{x}) = h_2^{-d}K((\mathbf{X}_{\mathbf{i}} - \mathbf{x})/h_2)$. For a sequence of random variables X_n and numbers a_n, we define $X_n = o_p(a_n)$ if X_n/a_n converges to zero in probability and $X_n = O_p(a_n)$ if for every $c > 0$, there exists a finite C such that $P(|X_n/a_n| \geq C) \leq c$. We let a function $f(n) = O(1)$ if there exist some nonzero constant C and N such that $f(n)/C \to 1$ for $n \geq N$, and $f(n) = o(1)$ if $f(n)/C \to 0$ for any constant C. Recursively, $g(n) = O(f(n))$ implies $g(n)/f(n) = O(1)$ and $g(n) = o(f(n))$ indicates $g(n)/f(n) = o(1)$. We then impose the following regularity conditions C1-C7, which are commonly employed in the literature.

C1 The random errors $\{\varepsilon_{\mathbf{i}}, \mathbf{i} \in I_{\mathbf{n}}\}$ have zero mode and are permitted to be dependent on $\{\mathbf{X}_{\mathbf{i}}, \mathbf{i} \in I_{\mathbf{n}}\}$. The density function $g(\cdot)$ of $\{\varepsilon_{\mathbf{i}}, \mathbf{i} \in I_{\mathbf{n}}\}$ is continuous in a neighborhood of the point 0, and the conditional density $g(\varepsilon \mid \mathbf{X}) < g(0 \mid \mathbf{X})$ for all $\varepsilon \neq 0$. It is also assumed that $g(\varepsilon_{\mathbf{i}}, \varepsilon_{\mathbf{j}} \mid \mathbf{X}_{\mathbf{i}}, \mathbf{X}_{\mathbf{j}})$ exists and is uniformly bounded for all $\mathbf{i}, \mathbf{j} \in \mathbb{Z}^N$.

C2 For the kernel function $K(\cdot) : \mathbb{R}^d \to \mathbb{R}$, it is bounded with a compact support $[-M, M]^d$ for some constant $M > 0$ and satisfies $\int \mathbf{u}\mathbf{u}^T K(\mathbf{u})d\mathbf{u} < \infty$ and $\int \mathbf{u}\mathbf{u}^T K^2(\mathbf{u})d\mathbf{u} < \infty$. It is also integrable and continuous almost everywhere, with $\int K(\mathbf{u})d\mathbf{u} = 1$ and $\int |K(\mathbf{u})|d\mathbf{u} < \infty$.

C3 The random field is strictly stationary. For all district $\mathbf{i}$ and $\mathbf{j}$ in $\mathbb{Z}^N$, the vectors $\mathbf{X_i}$ and $\mathbf{X_j}$ admit a joint density $f_{\mathbf{i},\mathbf{j}}(\cdot)$. Moreover, $|f_{\mathbf{i},\mathbf{j}}(\mathbf{x}', \mathbf{x}'') - f(\mathbf{x}')f(\mathbf{x}'')| \leq C$ for all $\mathbf{i}, \mathbf{j} \in \mathbb{Z}^N$ and all $\mathbf{x}', \mathbf{x}'' \in \mathbb{R}^d$, where $0 < C < \infty$ and $f(\cdot)$ denotes the marginal density of $\mathbf{X}$. Also, the function $\mathbf{x} \mapsto f(\mathbf{x})$ is strictly positive and continuous for all $\mathbf{x}$.

C4 The spatial modal regression function $m(\cdot)$ is at least twice differentiable in an open neighborhood of $\mathbf{x}$. Denoting $m^{(1)}(\mathbf{x})$ and $m^{(2)}(\mathbf{x})$ as the gradient vector and the Hessian matrix of $m(\cdot)$ evaluated at $\mathbf{x}$, respectively, the function $\mathbf{x} \to m^{(2)}(\mathbf{x})$ is continuous at all $\mathbf{x}$.

C5 The function $\chi(\cdot)$ satisfies $\chi(n', n'') \leq C \min(n', n'')$, $\forall\ n', n'' \in \mathbb{N}$ for some $C > 0$. The mixing process fulfills a polynomial mixing condition $\varphi(t) \leq Ct^{-\mu}$ for some $\mu > (2+\delta)\ (N+a)/\delta$. In addition, the following equation is satisfied: $\lim_{m\to\infty} m^a \sum_{i=m}^{\infty} i^{N-1}\ \{\varphi(i)\}^{\delta/(2+\delta)} = 0$ for some constant $a > \delta N/(2+\delta)$.

C6 There exist two sequences $\mathbf{p_n} := (p_1, \cdots, p_N) \in \mathbb{Z}^N$ and $\mathbf{q_n} := (q, \cdots, q) \in \mathbb{Z}^N$, with $q \to \infty$ such that $p = \prod_{k=1}^N p_k = o((\tilde{\mathbf{n}} h_1 h_2^d)^{1/2})$, $q/p_k \to 0$, $n_k/p_k \to \infty$ for all $k = 1, \cdots, N$, and $\tilde{\mathbf{n}}\varphi(q) \to 0$ as $\mathbf{n} \to \infty$.

C7 The bandwidths h_1 and h_2 tend to zero in a manner that $q(h_1 h_2^d)^{\delta/a(2+\delta)}$ > 1 with integer q defined in condition C6, and $(h_1 h_2^d)^{-\delta/(2+\delta)} \sum_{i=q}^{\infty} i^{N-1}$

$\{\varphi(i)\}^{\delta/(2+\delta)} \to 0$ as $\mathbf{n} \to \infty$ with function $\varphi(\cdot)$ listed in condition C5.

The conditions itemized above are standard in the setting of local smoothers and modal regression models required for asymptotics, and their justification can be detailed. Due to space limitations, we have included the comprehensive explanations of the aforementioned conditions in the Supplementary Material-S2. We now state the consistency result for the developed estimators in the complex spatial dependence setting.

Theorem 2.1. *Under the regularity conditions C1-C7, with probability approaching one, as* $\mathbf{n} \to \infty$*,* $h_1 \to 0$*,* $h_2 \to 0$*, and* $\tilde{\mathbf{n}} h_2^d h_1^5 \to \infty$*, there exist consistent maximizers* $(\hat{m}(\mathbf{x}), \hat{m}^{(1)}(\mathbf{x}))$ *of* (2.3) *such that*

i. $|\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p\Big(\big(\tilde{\mathbf{n}} h_2^d h_1^3\big)^{-1/2} + h_1^2 + h_2^2\Big)$*;*

ii. $\|H_2(\hat{m}^{(1)}(\mathbf{x}) - m^{(1)}(\mathbf{x}))\| = O_p\Big(\big(\tilde{\mathbf{n}} h_2^d h_1^3\big)^{-1/2} + h_1^2 + h_2^2\Big)$.

Theorem 2.1 establishes that the convergence rates of the proposed spatial modal estimators remain unaffected by the presence of spatial dependence, provided that the degree of dependence satisfies the strong mixing condition specified in condition C5. In the special case where $d = 1$, our estimators reduce to classical ones for independent data, recovering the convergence behavior known from the i.i.d. setting; see Ullah et al. (2022).

Remark 2.1. The convergence rates of the resulting modal estimators match those in the independent case under the spatial strong mixing condition (condition C5), which ensures summable dependence across lattice sites. This condition enables the application of empirical process theory and central limit theorems for strongly mixing random fields; see Rio (2017). Conceptually, the mixing coefficients control the effective dependence range so that the cumulative dependence contributes bounded higher-order terms to the asymptotic variance. As a result, the leading stochastic order remains driven by the local kernel smoothing behavior rather than long-range dependence. However, if the decay of mixing coefficients is too slow, these bounds may no longer hold, and the convergence rates may degrade accordingly.

To establish asymptotic normality, we begin by obtaining a representation for $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ represents the true value. Let $R(\mathbf{X_i}) = S(\mathbf{X_i})-(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\mathbf{X_i}$ and $S(\mathbf{X_i}) = m(\mathbf{X_i}) - m(\mathbf{x}) - m^{(1)}(\mathbf{x})(\mathbf{X_i} - \mathbf{x})$. By taking the first derivative of (2.3), the solution $\hat{\boldsymbol{\theta}}$ satisfies $\sum_{\mathbf{i}\in I_\mathbf{n}} K_\mathbf{i}\phi_{h_1}^{(1)}\left(\varepsilon_\mathbf{i} + R(\mathbf{X_i})\right)\mathbf{X}_\mathbf{i}^* = \sum_{\mathbf{i}\in I_\mathbf{n}} K_\mathbf{i}\left\{\phi_{h_1}^{(1)}\left(\varepsilon_\mathbf{i}\right) + \phi_{h_1}^{(2)}\left(\varepsilon_\mathbf{i}\right)R(\mathbf{X_i}) + \frac{1}{2}\phi_{h_1}^{(3)}\left(\varepsilon_\mathbf{i}^*\right)R^2(\mathbf{X_i})\right\}\mathbf{X}_\mathbf{i}^* = 0$, where $\mathbf{X}_\mathbf{i}^* = [1\ (\mathbf{X_i} - \mathbf{x})h_2^{-1}]^T$, $\varepsilon_\mathbf{i}^*$ lies between $\varepsilon_\mathbf{i}$ and $\varepsilon_\mathbf{i} + R(\mathbf{X_i})$, and $\phi_{h_1}^{(c)}(\varepsilon_\mathbf{i})$ represents the cth derivative of $\phi_{h_1}(\varepsilon_\mathbf{i})$. The following two lemmas establish a Bahadur-type expansion in the context of modal regression.

Lemma 2.1. *Define* $(S_{\tilde{\mathbf{n}}})_{ij} := \left(\tilde{\mathbf{n}}h_2^d\right)^{-1}\sum_{\mathbf{i}\in I_\mathbf{n}}\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_2}\right)_i\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_2}\right)_j K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_2}\right)$, $i, j = 0, 1, \cdots, d$. *Assume that conditions C2, C3, and C5 hold, and the* *bandwidth* h_2 *tends to zero in such a way that* $\tilde{\mathbf{n}}h_2^d \to \infty$ *and* $n_k h_2^{\delta d/[a(2+\delta)]} > 1$ *as* $\mathbf{n} \to \infty$. *Then, for all* $\mathbf{x}$ *and* $\mathbf{u} = (u_1, \cdots, u_d)^T \in \mathbb{R}^d$,

$$S_{\tilde{\mathbf{n}}} \xrightarrow{p} S := \begin{pmatrix} f(\mathbf{x}) \int_{\mathbb{R}^d} K(\mathbf{u})d\mathbf{u} & f(\mathbf{x}) \int_{\mathbb{R}^d} \mathbf{u}^T K(\mathbf{u})d\mathbf{u} \\ f(\mathbf{x}) \int_{\mathbb{R}^d} \mathbf{u}K(\mathbf{u})d\mathbf{u} & f(\mathbf{x}) \int_{\mathbb{R}^d} \mathbf{u}\mathbf{u}^T K(\mathbf{u})d\mathbf{u} \end{pmatrix},$$

where "$\xrightarrow{p}$" *denotes the convergence in probability.*

Lemma 2.2. *Define* $W_{\tilde{\mathbf{n}}} = \sum_{\mathbf{i}\in I_{\mathbf{n}}} \mathbf{X}_{\mathbf{i}}^* K_{\mathbf{i}} \phi_{h_1}^{(1)}(\varepsilon_{\mathbf{i}})$ *and* $m_{ij}(\mathbf{x}) = \partial^2 m(\mathbf{x})/\partial x_i \partial x_j$. *Assume that the same conditions in Theorem 2.1 hold. We can then obtain*

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \frac{h_2^2}{2} S^{-1}\Lambda(1 + o_p(1)) + \frac{S^{-1}W_{\tilde{\mathbf{n}}}}{\tilde{\mathbf{n}}g^{(2)}(0 \mid \mathbf{X} = \mathbf{x})}(1 + o_p(1)),$$

where $\Lambda = \begin{pmatrix} f(\mathbf{x}) \sum_{i=1}^d \sum_{j=1}^d m_{ij}(\mathbf{x}) \int_{\mathbb{R}^d} u_i u_j K(\mathbf{u})d\mathbf{u} \\ f(\mathbf{x}) \sum_{i=1}^d \sum_{j=1}^d m_{ij}(\mathbf{x}) \int_{\mathbb{R}^d} u_i u_j \mathbf{u}K(\mathbf{u})d\mathbf{u} \end{pmatrix}$.

The proof of Lemma 2.1 involves straightforward calculations and is omitted in this paper. Built on the preceding lemmas, the asymptotic normality of the local linear spatial modal estimators under weak conditions can be demonstrated by establishing the result $W_{\tilde{\mathbf{n}}}^* = \frac{1}{\sqrt{\tilde{\mathbf{n}}h_1 h_2^d}} \sum_{\mathbf{i}\in I_{\mathbf{n}}} \phi^{(1)} \left(\frac{\varepsilon_{\mathbf{i}}}{h_1}\right) K\left(\frac{\mathbf{X}_{\mathbf{i}}-\mathbf{x}}{h_2}\right) \mathbf{X}_{\mathbf{i}}^* \xrightarrow{d} \mathcal{N}\left(0, \int \phi^2(t)t^2 dt g(0 \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x})\Sigma\right)$; see technical proofs in the Supplementary Material-S7. The expression of Σ is in Theorem 2.2.

Theorem 2.2. *With* $\tilde{\mathbf{n}}h_2^{d+4}h_1^3 = O(1)$ *and* $\tilde{\mathbf{n}}h_2^d h_1^7 = O(1)$, *under the same conditions as Theorem 2.1, if* $n_k \left(h_1 h_2^d\right)^{\delta/(2+\delta)a} > 1$ *for all* $k = 1, \cdots, N$ *as* $\mathbf{n} \to \infty$, *the estimators satisfying the consistency results in Theorem 2.1* *have the following asymptotic result*

$$\sqrt{\tilde{\mathbf{n}}h_2^d h_1^3}\left[\begin{pmatrix}\hat{m}(\mathbf{x})-m(\mathbf{x})\\ h_2(\hat{m}^{(1)}(\mathbf{x})-m^{(1)}(\mathbf{x}))\end{pmatrix}-S^{-1}\left(\frac{h_2^2}{2}\Lambda-\frac{h_1^2}{2}\frac{g^{(3)}(0\mid\mathbf{X}=\mathbf{x})}{g^{(2)}(0\mid\mathbf{X}=\mathbf{x})}\Gamma\right)\right]$$

$$\xrightarrow{d}\mathcal{N}\Big(0,\int\phi^2(t)t^2dt\frac{g(0\mid\mathbf{X}=\mathbf{x})}{(g^{(2)}(0\mid\mathbf{X}=\mathbf{x}))^2}f(\mathbf{x})S^{-1}\Sigma S^{-1}\Big),$$

where $\Gamma=\begin{pmatrix}f(\mathbf{x})\int_{\mathbb{R}^d}K(\mathbf{u})d\mathbf{u}\\ f(\mathbf{x})\int_{\mathbb{R}^d}\mathbf{u}K(\mathbf{u})d\mathbf{u}\end{pmatrix}$ *and* $\Sigma=\begin{pmatrix}\int_{\mathbb{R}^d}K^2(\mathbf{u})d\mathbf{u} & \int_{\mathbb{R}^d}\mathbf{u}^TK^2(\mathbf{u})d\mathbf{u}\\ \int_{\mathbb{R}^d}\mathbf{u}K^2(\mathbf{u})d\mathbf{u} & \int_{\mathbb{R}^d}\mathbf{u}\mathbf{u}^TK^2(\mathbf{u})d\mathbf{u}\end{pmatrix}.$

Corollary 2.1. *With* $\tilde{\mathbf{n}}h_2^{d+4}h_1^3=O(1)$, $\tilde{\mathbf{n}}h_2^dh_1^7=O(1)$, *and* $n_k\left(h_1h_2^d\right)^{\delta/(2+\delta)a}$ >1 *for all* $k=1,\cdots,N$ *as* $\mathbf{n}\to\infty$, *if furthermore, the kernel* $K(\cdot)$ *is a symmetric function, the result in Theorem 2.2 can be reinforced into*

$$\sqrt{\tilde{\mathbf{n}}h_2^dh_1^3}\begin{pmatrix}\hat{m}(\mathbf{x})-m(\mathbf{x})-B_0(\mathbf{x})\\ h_2(\hat{m}^{(1)}(\mathbf{x})-m^{(1)}(\mathbf{x})-B_1(\mathbf{x}))\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\begin{pmatrix}\sigma_0^2(\mathbf{x}) & 0\\ 0 & \sigma_1^2(\mathbf{x})\end{pmatrix}\right),$$

where $B_1(\mathbf{x})=0$, $B_0(\mathbf{x})=\left[\int_{\mathbb{R}^d}K(\mathbf{u})d\mathbf{u}\right]^{-1}\frac{h_2^2}{2}\sum_{i=1}^d m_{ii}(\mathbf{x})\int_{\mathbb{R}^d}u_i^2K(\mathbf{u})d\mathbf{u}-$ $\frac{h_1^2}{2}\frac{g^{(3)}(0\mid\mathbf{X}=\mathbf{x})}{g^{(2)}(0\mid\mathbf{X}=\mathbf{x})}$, $\sigma_0^2(\mathbf{x})=\frac{\int\phi^2(t)t^2dt}{f(\mathbf{x})}\frac{g(0\mid\mathbf{X}=\mathbf{x})}{(g^{(2)}(0\mid\mathbf{X}=\mathbf{x}))^2}\left[\int_{\mathbb{R}^d}K(\mathbf{u})d\mathbf{u}\right]^{-1}$ $\int_{\mathbb{R}^d}K^2(\mathbf{u})d\mathbf{u}\left[\int_{\mathbb{R}^d}K(\mathbf{u})d\mathbf{u}\right]^{-1}$, *and* $\sigma_1^2(\mathbf{x})=\frac{\int\phi^2(t)t^2dt}{f(\mathbf{x})}\frac{g(0\mid\mathbf{X}=\mathbf{x})}{(g^{(2)}(0\mid\mathbf{X}=\mathbf{x}))^2}$ $\left[\int_{\mathbb{R}^d}\mathbf{u}\mathbf{u}^TK(\mathbf{u})d\mathbf{u}\right]^{-1}\int_{\mathbb{R}^d}\mathbf{u}\mathbf{u}^TK^2(\mathbf{u})d\mathbf{u}\left[\int_{\mathbb{R}^d}\mathbf{u}\mathbf{u}^TK(\mathbf{u})d\mathbf{u}\right]^{-1}$.

The asymptotic normality results in Theorem 2.2 are stated for $m(\mathbf{x})$ and $m^{(1)}(\mathbf{x})$ at a specific site $\mathbf{x}$ that is not too near the boundary of the support of the covariates. In fact, by applying the traditional Cramér-Wold device, these results can be conveniently generalized to a joint asymptotic normality result for multiple sites. With optimal bandwidths obtained by minimizing asymptotic MSE, $\hat{m}(\mathbf{x})$ converges to $m(\mathbf{x})$ at a rate of $O_p(\tilde{\mathbf{n}}^{-2/(7+d)})$, which is smaller than the convergence rare $(O_p(\tilde{\mathbf{n}}^{-1/(4+d)}))$ achieved by local linear mean estimation; see Hallin et al. (2004). This slower convergence rate is the trade-off for estimating the mode. With undersmoothing ($\lim_{\tilde{\mathbf{n}}\to\infty} \tilde{\mathbf{n}} h_2^{d+4} h_1^3 \to 0$ and $\lim_{\tilde{\mathbf{n}}\to\infty} \tilde{\mathbf{n}} h_2^d h_1^7 \to 0$), the asymptotic bias can be disregarded, and the estimators are centered at true values.

The spatial dependence complicates the proof of Theorem 2.2. We utilize Bernstein's technique by decomposing the finite summation into smaller pieces involving "large" and "small" blocks. Particularly, we need to verify whether condition C6 holds for the preceding theorem. Suppose there exists a sequence of positive integers $q = q_{\mathbf{n}} \to \infty$ such that $q_{\mathbf{n}} = o(\left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/(2N)})$. We can choose sequence $s_{\mathbf{n}} \to 0$ in a manner that $q = \left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/2N} s_{\mathbf{n}}$. Taking $p_k := \left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/2N} s_{\mathbf{n}}^{1/2}$, $k = 1, \cdots, N$, it follows that $q/p_k = s_{\mathbf{n}}^{1/2} \to 0$, $p = \left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/2} s_{\mathbf{n}}^{N/2} = o\big(\left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/2}\big)$, and $\tilde{\mathbf{n}}\varphi(q) = \tilde{\mathbf{n}} q^{-\mu} \to 0$. As $\mathbf{n} \to \infty$, we obtain $p < \left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/2}$ for large $\tilde{\mathbf{n}}$. Because of $\tilde{\mathbf{n}}/p > (\tilde{\mathbf{n}} h_1^{-1} h_2^{-d})^{1/2} \to \infty$, it is apparent that $n_k/p_k \to \infty$ for all k. As a result, condition C6 is satisfied. The details of the proof are listed in the Supplementary Material-S7.

Remark 2.2. The asymptotic results also hold for the nonisotropic divergence case under appropriate conditions. Assume that conditions C1-C4 hold with $\varphi(t) = O\left(t^{-\mu}\right)$ for some $\mu > (2+\delta)\ (N+a)/\delta$. In addition, let the sequence of positive integers $q = q_{\mathrm{n}} \to \infty$, the bandwidth h_1 factor into $h_1 := \prod_{i=1}^{N} h_{1_i}$, and the bandwidth h_2 factor into $h_2 := \prod_{i=1}^{N} h_{2_i}$, such that $\tilde{\mathbf{n}} q^{-\mu} \to 0$, $q = o\big(\min_{1\le k\le N} \left(n_k h_{1_k} h_{2_k}^d\right)^{1/2}\big)$, and $q(h_1 h_2^d)^{\delta/a(2+\delta)} > 1$ for some constant $a > \delta N/(2+\delta)$. Under these conditions, Theorem 2.2 remains valid as $\mathbf{n} \to \infty$ with $\min_{1\le l\le N}\{n_l\} \to \infty$ (*nonisotropic divergence*). To demonstrate this result, we can follow the steps for the proof of Theorem 2.2 in the Supplementary Material-S7, ensuring that condition C6 is satisfied. Suppose that there exists a sequence $s_{n_k} \to 0$ such that $q = o\big(\min_{1\le k\le N} \left(n_k h_{1_k} h_{2_k}^d\right)^{1/2} s_{n_k}\big)$ as $\mathbf{n} \to \infty$. By taking $p_k = \left(n_k h_{1_k} h_{2_k}^d\right)^{1/2} s_{n_k}^{1/2}$, we have $q/p_k \le s_{n_k}^{1/2} \to 0$, $p = \left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/2} \prod_{k=1}^{N} s_{n_k}^{1/2} = o\big(\left(\tilde{\mathbf{n}} h_1 h_2^d\right)^{1/2}\big)$, and $\tilde{\mathbf{n}}\varphi(q) = \tilde{\mathbf{n}} q^{-\mu} \to 0$. As $\mathbf{n} \to \infty$, we obtain $p_k < \left(n_k h_{1_k} h_{2_k}^d\right)^{1/2}$. It is evident that $n_k/p_k > \left(n_k h_{1_k} h_{2_k}^{-d}\right)^{1/2} \to \infty$. As a result, condition C6 is fulfilled.

Remark 2.3. If we impose $\chi\left(n', n''\right) \le C\left(n' + n'' + 1\right)^{\kappa}$ for some $C > 0$ and $\kappa > 1$ in condition C5 and replace the last requirement in condition C6 with $\left(\tilde{\mathbf{n}}^{\kappa+1}/p\right)\varphi(q) \to 0$, Theorem 2.2 still holds as $\mathbf{n} \to \infty$. To establish this result, it suffices to verify that the bound on the term Q_1 in the Supplementary Material-S7 remains valid, which is true since $Q_1 \le C \sum_{i=1}^{M}$

$[p + (M - i)p + 1]^{\kappa} \varphi(q) \leq C p^{\kappa} M^{\kappa+1} \varphi(q) \leq C \left(\tilde{\mathbf{n}}^{(\kappa+1)}/p\right) \varphi(q) \to 0.$

The local linear approximation offers significantly improved boundary behavior compared to the local constant approach. For simplicity, we assume that there is a univariate regressor X ($d = 1$) with a bounded support, i.e., $[-M, M]$. By employing an argument similar to the one developed in the proof of Theorem 2.2, it can be shown that asymptotic normality still holds near the boundary point $x = -M + ch_2$ with $c > 0$. However, there are adjustments in the asymptotic biases and variances; see Supplementary Material-S3 for the details. Indeed, this boundary advantage would likely become more pronounced as N grows. Consequently, local linear modal estimation exhibits automatic good behavior at boundaries without the need for boundary correction. This holds true for both the left boundary point $x = -M + ch_2$ and the right boundary point $x = M - ch_2$. Even if point M were an interior point, the same results would still apply with $c = M$.

2.4 Optimal Bandwidths

The shape and smoothness of the spatial modal function, like other nonparametric regression models, depends to a large extent on the values of the bandwidths, which typically involve a trade-off between bias and variance. Nevertheless, there is no data-driven rule that permits automatic and optimal selection of bandwidth values in the context of spatial modal regression. The asymptotic results in Theorem 2.2 enable finding asymptotically optimal bandwidths by minimizing the asymptotic MSE (AsyMSE) of $\hat{m}(\mathbf{x})$

$$\begin{aligned}
&\text{AsyMSE}(\hat{m}(\mathbf{x})) = \text{Bias}(\hat{m}(\mathbf{x}))^2 + \text{Var}(\hat{m}(\mathbf{x}))\\
\approx&\left\{e^T S^{-1}\left(\frac{h_2^2}{2}\Lambda - \frac{h_1^2}{2}\frac{g^{(3)}(0 \mid \mathbf{X}=\mathbf{x})}{g^{(2)}(0 \mid \mathbf{X}=\mathbf{x})}\Gamma\right)\right\}^2 + \frac{\int \phi^2(t)t^2dt}{\tilde{\mathbf{n}}h_2^d h_1^3}\frac{g(0 \mid \mathbf{X}=\mathbf{x})}{(g^{(2)}(0 \mid \mathbf{X}=\mathbf{x}))^2}\\
&f(\mathbf{x})e^T S^{-1}\Sigma S^{-1}e,
\end{aligned}$$

where $e = (\underbrace{1,\cdots,1}_{d},\underbrace{0,\cdots,0}_{d})^T$. Accordingly, the optimal bandwidths h_1 and h_2 satisfy $\hat{h}_1 = \hat{h}_2\Delta_1^{1/2}$ with $\Delta_1 = \frac{A\sqrt{(3+d)^2+12d}-(3+d)A}{2dB}$, $A = e^T S^{-1}\Lambda$, and $B = e^T S^{-1}\frac{g^{(3)}(0|\mathbf{X}=\mathbf{x})}{g^{(2)}(0|\mathbf{X}=\mathbf{x})}\Gamma$, where

$$\hat{h}_2 = \left(\frac{g(0 \mid \mathbf{X}=\mathbf{x})}{(g^{(2)}(0 \mid \mathbf{X}=\mathbf{x}))^2}\frac{d\int \phi^2(t)t^2dtf(\mathbf{x})e^T S^{-1}\Sigma S^{-1}e}{\Delta_1^{3/2}(A^2 - AB\Delta_1)}\right)^{\frac{1}{d+7}}\tilde{\mathbf{n}}^{-\frac{1}{d+7}}.$$

As evident, the optimal rate for bandwidths is larger than that for spatial mean regression. However, the exact values of bandwidths depend on unknown quantities, i.e., the derivatives of the density of the error components, making direct application of the provided bandwidth expression challenging.

Practically selecting bandwidths involves adopting the plug-in method to approximate unknown terms in the expressions of the asymptotically optimal bandwidths, as suggested by Ullah et al. (2021). Nonetheless, this can be challenging for nonparametric estimators in reality, leading to potential inaccuracies and computational expenses due to additional tuning parameters for estimating the density function. To facilitate the selection process, we employ the approach outlined by Ullah et al. (2023) for bandwidth choice in this paper. Particularly, we let $\hat{h}_2 = 1.05\sigma_m \tilde{\mathbf{n}}^{-1/(d+6)}$ and $\hat{h}_1 = 1.6\text{MAD}\tilde{\mathbf{n}}^{-1/(d+6)}$, where σ_m represents the standard deviation of $\varepsilon_{\mathbf{i},m} = Y_{\mathbf{i}} - \hat{m}_m(\mathbf{X}_{\mathbf{i}})$, $MAD = med_{\mathbf{j}}\{|\varepsilon_{\mathbf{j},m} - med_{\mathbf{i}}(\varepsilon_{\mathbf{i},m})|\}$, $\hat{m}_m(\cdot)$ denotes the corresponding mean estimate, and *med* is the median value. These choices take into account the AsyMSE-optimal rate and the requirement for undersmoothing. While this method for choosing bandwidths may not produce globally optimal estimates (i.e., obtained by minimizing the integrated MSE), it provides a practical data-based "rule of thumb" for real-world applications that has been demonstrated to work well in simulations listed in Section 3.

In addition to the previous bandwidth selection procedure, we can also employ the density-based cross-validation method to choose bandwidths. Specifically, we minimize the integrated squared error (ISE) of the general conditional density estimate ISE $= \int\int(\hat{f}_{\mathbf{X},Y}(\mathbf{X}_{\mathbf{i}}, Y_{\mathbf{i}}) - f_{\mathbf{X},Y}(\mathbf{X}_{\mathbf{i}}, Y_{\mathbf{i}}))^2 f_{\mathbf{X}}(\mathbf{X}_{\mathbf{i}})d\mathbf{X}_{\mathbf{i}}dY_{\mathbf{i}}$, which gives the cross-validation criterion listed below

$$\min CV(h_1, h_2) = \frac{1}{\tilde{\mathbf{n}}} \sum_{\mathbf{i}\in I_{\mathbf{n}}} \frac{\int \hat{f}^2_{\mathbf{X},Y,-\mathbf{i}}(\mathbf{X}_{\mathbf{i}}, Y)dY}{\hat{f}^2_{\mathbf{X},-\mathbf{i}}(\mathbf{X}_{\mathbf{i}})} - \frac{2}{\tilde{\mathbf{n}}} \sum_{\mathbf{i}\in I_{\mathbf{n}}} \frac{\hat{f}_{\mathbf{X},Y,-\mathbf{i}}(\mathbf{X}_{\mathbf{i}}, Y_{\mathbf{i}})}{\hat{f}_{\mathbf{X},-\mathbf{i}}(\mathbf{X}_{\mathbf{i}})},$$

under which $-\mathbf{i}$ indicates the absence of observations $\mathbf{X}_{\mathbf{i}}$ and $Y_{\mathbf{i}}$. We utilize this method in the simulation examples to assess the effectiveness of the previously suggested "rule of thumb" selection procedure; see Section 3 and Supplementary Material-S4. However, the detailed theoretical investigation of bandwidth selection using modal cross-validation is outside the reach of this paper due to the absence of a natural order for time series with spatial data. Such a technical analysis will be pursued in future research.

Remark 2.4. In practice, the assumption of stationarity may often be violated. An alternative approach is to assume that nonstationarity arises from the presence of a spatial trend. Instead of the stationary process $\{Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}\}$, for example, we actually observe $\{\tilde{Y}_{\mathbf{i}}, \tilde{\mathbf{X}}_{\mathbf{i}}\}$ with $\tilde{Y}_{\mathbf{i}} = \mu_Y(\mathbf{s}_{\mathbf{i}}) + Y_{\mathbf{i}}$, $\tilde{\mathbf{X}}_{\mathbf{i}} = \mu_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}) + \mathbf{X}_{\mathbf{i}}$, $\mathbf{i} \in I_{\mathbf{n}}$, in which $\mathbf{s}_{\mathbf{i}} = (s_{i_1}, \cdots, s_{i_N}) := (i_1/n_1, \cdots, i_N/n_n)$ and $\mathbf{s} \in [0,1]^N \mapsto (\mu_Y(\mathbf{s}), \mu_{\mathbf{X}}(\mathbf{s}))$ is some deterministic but unknown trend functions. Since $\{Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}\}$ is unobservable, we need to estimate $\{\mu_Y(\mathbf{s}_{\mathbf{i}}), \mu_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}})\}$ first using local constant spatial mean regression such that

$$\hat{\mu}_Y(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \tilde{Y}_{\mathbf{i}} w(\mathbf{s}_{\mathbf{i}}, \mathbf{s}) \text{ and } \hat{\mu}_{\mathbf{X}}(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \tilde{\mathbf{X}}_{\mathbf{i}} w(\mathbf{s}_{\mathbf{i}}, \mathbf{s}),$$

where $w(\mathbf{s}_{\mathbf{i}}, \mathbf{s}) := W((\mathbf{s}_{\mathbf{i}} - \mathbf{s})/g_h) / \left[\sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{n}}} W((\mathbf{s}_{\mathbf{j}} - \mathbf{s})/g_h)\right]$, $W(\cdot)$ is a kernel function, and g_h is the bandwidth that approaches zero. The estimation procedure described in the preceding parts is subsequently performed to the residuals $\{(\hat{Y}_{\mathbf{i}}, \hat{\mathbf{X}}_{\mathbf{i}}) := (\tilde{Y}_{\mathbf{i}} - \hat{\mu}_Y(\mathbf{s}_{\mathbf{i}}), \tilde{\mathbf{X}}_{\mathbf{i}} - \hat{\mu}_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}))\}$, which is supposed to satisfy the stationarity assumption. The local linear estimators of the spatial modal regression are then specified as

$$(\hat{a}, \hat{\mathbf{b}}) = \arg\max_{a,\mathbf{b}} \frac{1}{\tilde{\mathbf{n}} h_1 h_2^d} \sum_{\mathbf{i}\in I_{\mathbf{n}}} \phi\left(\frac{\hat{Y}_{\mathbf{i}} - a - \mathbf{b}^T(\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{x})}{h_1}\right) K\left(\frac{\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{x}}{h_2}\right).$$

As the preliminary spatial smoothing $(\hat{\mu}_Y(\mathbf{s}), \hat{\mu}_{\mathbf{X}}(\mathbf{s}))$ of the original data $\{\tilde{Y}_{\mathbf{i}}, \tilde{\mathbf{X}}_{\mathbf{i}}\}$ is derived via mean regression, which converges faster than the spatial modal estimators, we can achieve the same asymptotic normality results as stated in Theorem 2.2 under certain suitable conditions.

3. Numerical Examples

We in this section present numerical studies to gain insights into the established estimation procedure, where the bandwidth selection procedures outlined in Subsection 2.3 are implemented. Due to space constraints, we include additional simulation studies in the Supplementary Material-S4.

3.1 Monte Carlo Experiments

We consider a model in a two-dimensional space (N=2), where we denote the sites $\mathbf{i}$ in $\mathbb{Z}^2$ as (i, j) instead of (i_1, i_2). For simplicity, we assume X to be a scalar random variable. In what follows, we use DGP to represent the data generating process. We generate two hundred simulated spatial datasets independently. To measure the performance of estimators, we utilize the average MSE (AMSE) indicator, i.e., AMSE $= \frac{1}{200 n_1 n_2} \sum_{l=1}^{200} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left(\hat{m}^{(l)}(X_{i,j}) - m_0(X_{i,j})\right)^2$, where $m_0(X_{i,j})$ denotes the true spatial regression function, and $\hat{m}^{(l)}(X_{i,j})$ represents the estimate in the lth replication.

DGP 1 (Asymmetric Data) We generate data from the following model

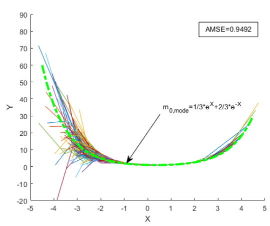
$$Y_{i,j} = m(X_{i,j}) + \sigma(X_{i,j})\varepsilon_{i,j} \text{ with } m(x) = \frac{1}{3}e^{x} + \frac{1}{3}e^{-x}, \tag{3.1}$$

where $\sigma(X_{i,j}) = \frac{1}{3}e^{-X_{i,j}}$, $\{\varepsilon_{i,j}, (i,j) \in \mathbb{Z}^2\}$ are originated from $0.5\mathcal{N}(-1, 2.5^2)$ $+0.5\mathcal{N}(1, 0.5^2)$ with mean zero and mode one, and $\{X_{i,j}, (i,j) \in \mathbb{Z}^2\}$, according to Definition 2.1, are produced by the spatial autoregression $X_{i,j} = \sin(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}) + e_{i,j}$ with $\{e_{i,j}, (i,j) \in \mathbb{Z}^2\} \sim \mathcal{N}(0,1)$. We then have the following spatial modal regression function

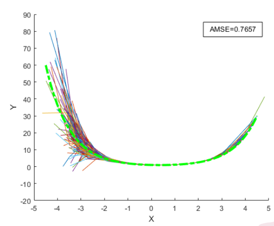
$$Mode(Y_{i,j} \mid X_{i,j}) = m_{0,mode}(X_{i,j}) = \frac{1}{3}e^{X_{i,j}} + \frac{2}{3}e^{-X_{i,j}}, \tag{3.2}$$

which is different from mean regression $\mathbb{E}(Y_{i,j} \mid X_{i,j}) = \frac{1}{3}e^{X_{i,j}} + \frac{1}{3}e^{-X_{i,j}}$.

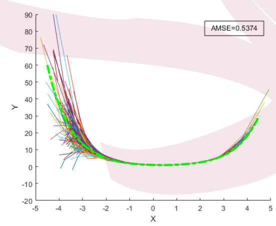
The above model is utilized to simulate data across a rectangular region of $n_1 \times n_2$ sites, represented as a grid $\{(i,j) : 76 \le i \le 75 + n_1, 76 \le j \le 75 + n_2\}$ for various values of n_1 and n_2. Each replication is obtained iteratively through the following procedures: initially simulating i.i.d. random variables $e_{i,j}$ over the grid $\{(i,j) : 1 \le i \le 150 + n_1, 1 \le j \le 150 + n_2\}$; subsequently, setting all initial values of $X_{i,j}$ to zero and generating $X_{i,j}$'s over $\{(i,j) : 1 \le i \le 150 + n_1, 1 \le j \le 150 + n_2\}$ recursively using spatial autoregression models. Starting with these generated values, the process is iterated 20 times. To achieve stationarity, the results at the final iteration step for (i, j) inside $\{(i, j) : 76 \leq i \leq 75+n_1, 76 \leq j \leq 75+n_2\}$ are considered as the simulated $n_1 \times n_2$ sample. For comprehensive illustration, we conduct simulation studies using four different combinations of sample sizes, namely (n_1, n_2)=(10, 10), (15, 15), (20, 20), and (20, 30), respectively.



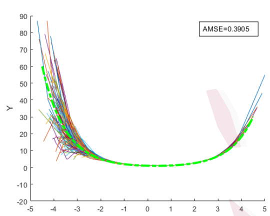
(a) $(n_1, n_2) = (10, 10)$



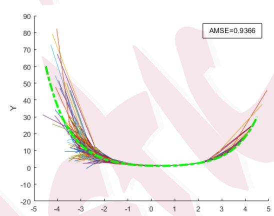
(b) $(n_1, n_2) = (15, 15)$



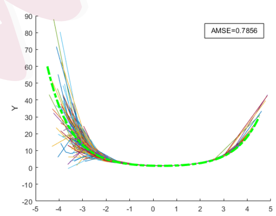
(c) $(n_1, n_2) = (20, 20)$



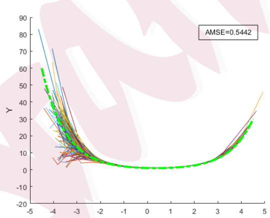
(d) $(n_1, n_2) = (20, 30)$



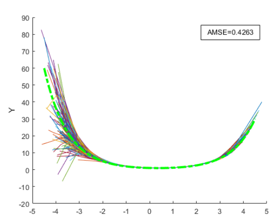
(e) $(n_1, n_2) = (10, 10)$-CV



(f) $(n_1, n_2) = (15, 15)$-CV



(g) $(n_1, n_2) = (20, 20)$-CV



(h) $(n_1, n_2) = (20, 30)$-CV

Figure 1: Local Linear Modal Estimator of Function $m(\cdot)$

The simulation results based on 200 independent replications are displayed in Figure 1, where the true spatial modal regression function is indicated by the green dotted line, and the estimated functions $m(\cdot)$ across all replications are shown as solid curves. For each scenario, we report the AMSE to quantify estimation accuracy. Although spatial mean regression estimates are not shown in the figure due to their poor comparability under asymmetric error structures, we note that the proposed modal estimator consistently yields substantially higher conditional kernel likelihood values when compared to the mean estimator, reinforcing its appropriateness for modal estimation. These comparative results are available upon request. It can be observed in Figure 1-(a)-(d) that the proposed estimation procedure behaves effectively and can capture the spatial regression line even for sample sizes as small as $(n_1, n_2) = (10, 10)$. As expected, the simulation results align well with the asymptotic theory presented in Subsection 2.3, where the AMSE decreases as the sample size increases. To demonstrate the applicability of the suggested bandwidth selection method, we also report results with bandwidth values obtained from the modal cross-validation (CV) method. It is evident from Figure 1 that there is no significant difference in estimates between these two bandwidth choices, indicating the good performance of the suggested bandwidth selection procedure. Note that while the MEM algorithm used for estimation does not guarantee global convergence, our simulation experiments demonstrate that with carefully chosen initial values, such as those based on local linear mean or quantile regression, the algorithm consistently converges to a dominant mode of the conditional distribution. Across all 200 replications, we observe no evidence of convergence to spurious or suboptimal local modes. Moreover, we implement a post-estimation validation step by comparing conditional kernel likelihood values across multiple candidate solutions, selecting the maximizer with the highest likelihood. These empirical observations, together with the robustness heuristics described in the Supplementary Material-S1, offer strong support for the reliability of the proposed modal estimation framework.

3.2 Empirical Analysis of Soil Data

The spatial modeling of soil nutrient levels play a critical role in agronomy, environmental monitoring, and precision agriculture. Accurate spatial estimation of key chemical properties, such as calcium, phosphorus, and potassium, is essential for informing site-specific soil management, optimizing fertilizer application, and minimizing environmental impacts; see White (2005). Among these properties, cation exchange capacity (CTC) is particularly important, as it reflects a soil's ability to retain essential nutrients and thus directly influences crop productivity. Understanding how CTC varies across space in response to underlying soil chemistry is therefore not only of scientific interest but also of direct relevance to agricultural planning and policy. However, the conditional distribution of soil properties often deviates substantially from normality due to heterogeneity in environmental factors such as micro-climate, historical land use, and contamination; see Figure 2. These complexities frequently give rise to skewed, heavy-tailed, or multimodal patterns that render mean regression inadequate. Modal regression, by targeting the most frequent value, provides a robust alternative that better characterizes the dominant spatial behavior of soil properties.

To illustrate the utility of our proposed spatial modal regression, we analyze the *soil*250 dataset available from the R package *GeoR*, which originates from a uniformity trial involving 250 undisturbed soil samples collected at a depth of 25cm and spaced regularly over a 25×10 grid. This dataset comprises 22 variables representing key soil chemistry properties relevant to agronomy and environmental monitoring. For our analysis, we focus on modeling the relationship between calcium (Ca) and CTC ($N = 2$ and $d = 1$). Ca is a key determinant of soil fertility, playing a critical role in improving soil structure, enhancing root development, and facilitating nutrient uptake. Moreover, Ca deficiency or imbalance can significantly alter the soil's ion-exchange properties, which directly impacts CTC. Therefore, understanding how Ca affects the modal (i.e., most typical) level of CTC is crucial for site-specific soil management and sustainable agricultural practices. To estimate this relationship, we consider the model

$$CTC_{i,j} = m(Ca_{i,j}) + \varepsilon_{i,j},\ 1 \le i \le 25,\ 1 \le j \le 10, \tag{3.3}$$

where the x-coordinate corresponds to Linha, the y-coordinate is Coluna, and the conditional mode satisfies $Mode(\varepsilon_{i,j} \mid Ca_{i,j}) = 0$. As noted in Remark 2.4, we first apply the `sm.regression` function from the `sm` R package to remove large-scale spatial trends in both Ca and CTC, ensuring approximate stationarity in the residual spatial processes prior to modal estimation.

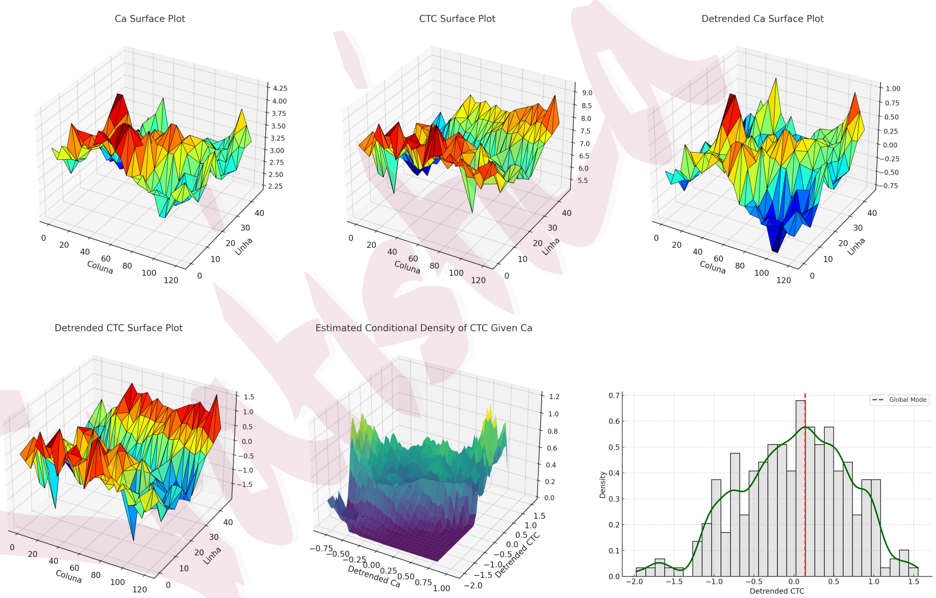


Figure 2: Spatial Perspectives of Soil Dataset

Figure 2 provides a spatial and distributional overview of the soil dataset and motivates the use of spatial modal regression. The top row displays the raw surfaces of Ca and CTC, both of which exhibit spatial trends and large-scale variation, indicating the presence of nonstationarity in the original data. Following the detrending procedure, the resulting surfaces (bottom-left panels) reveal localized irregularities and fine-scale spatial heterogeneity, suggesting complex spatial dynamics that are unlikely to be captured by global smoothing techniques or mean regression models. Notably, the empirical density of the detrended CTC values (bottom-right panel) shows marked asymmetry and possible multimodality, with heavy tails and deviations from Gaussian behavior. These distributional departures are further substantiated by the estimated conditional density of CTC given Ca (bottom-middle panel), which displays complex, ridged patterns and non-elliptical contours. Such structures violate key assumptions of mean regression, potentially leading to estimates that fall in regions of low probability density and obscure the dominant spatial response behavior. As modal regression is robust to distributional asymmetry and tail behavior, it provides a more interpretable model for conditional spatial structure of the soil data.

Figure 3 presents a comparative analysis of spatial estimation results derived from mean regression and the proposed modal regression, highlighting substantive differences in both magnitude and spatial structure. The top-left panel displays the estimated mean regression surface, which shows a

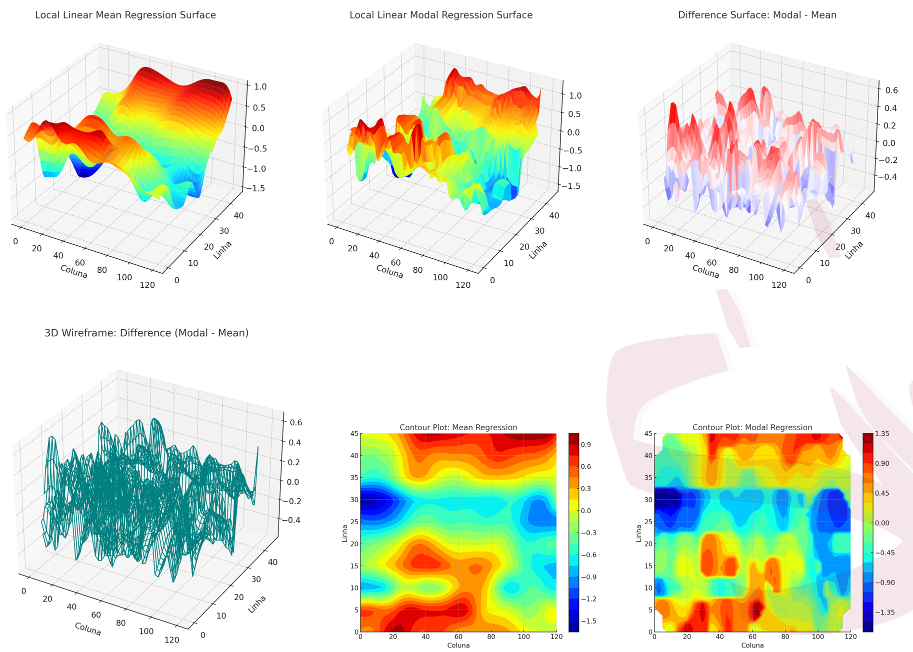


Figure 3: Mean and Modal Estimation Results

generally smooth gradient, with an initial upward trend in CTC followed by a noticeable decline across certain regions of the spatial domain. The modal regression surface (top-middle panel), by contrast, reveals more localized variability and maintains a plateau-like structure in high-Ca regions. This suggests that modal regression better preserves spatial features that may be flattened or misrepresented by mean regression. The top-right panel depicts the pointwise difference between the two regression surfaces (modal minus mean). Notably, the modal estimates are consistently higher in regions with elevated Ca content, suggesting that the mean regression surface may be biased downward by sparse or extreme low-CTC values. The bottom-left wireframe plot offers a complementary 3D view of these differences, emphasizing the high-frequency spatial structure and spatially clustered deviations that occur primarily in agronomically meaningful areas. The lower row of contour plots reinforces this interpretation. The mean regression contours (bottom-middle) appear overly smooth and may miss important local features, whereas the modal regression contours (bottom-right) exhibit sharper boundaries and more localized plateaus. These localized features are agriculturally interpretable, as they may reflect real spatial heterogeneity driven by underlying soil composition, management zones, or drainage patterns. Importantly, the plateauing behavior observed in the modal surface aligns with established agronomic findings, which suggest that increasing Ca concentration enhances CTC only up to a certain point, beyond which additional Ca does not yield further gains; see White (2005) and Brady and Weil (2016). The modal regression surface accurately captures this saturation pattern by maintaining high but stable CTC values in regions with high Ca. By contrast, the mean regression surface suggests a decline in CTC in these regions, a result that is not only inconsistent with agronomic findings but also likely driven by the influence of heavy-tailed residuals or skewed error structures. These findings underscore the utility of modal regression as a robust and informative tool for spatial modeling in agricultural contexts.

4. Concluding Remarks

We in this paper propose a nonparametric spatial modal regression model designed to uncover relationships between the dependent variable and spatial covariates that may be overlooked by conventional spatial mean or quantile regression approaches. Under mild regularity conditions, the asymptotic normality of the estimators of $m(\mathbf{x})$ and its gradient is established by relying on a Bahadur representation. By virtue of a Gaussian kernel, we construct a modified MEM algorithm to numerically estimate the developed model and provide rules for bandwidth selection in practice. The proposed estimation procedure is supported by both asymptotic theory and favorable finite sample properties through simulation studies and empirical analysis. Additionally, we discuss several extensions, including scenarios where the size of the rectangular domain tends to infinity at different rates, and provide insights into various other model perspectives to offer a comprehensive understanding of the suggested regression. Finally, to avoid the curse of dimensionality and relax the linearity assumption, we extend the developed model to propose an additive spatial modal regression model; see Supplementary Material-S5. All presented findings can serve as a foundation for future research on mode-based analysis in spatial settings, with broad potential for applications in environmental science, agriculture, and beyond.

Supplementary Material

The Supplementary Material contains comments for MEM algorithm and theoretical conditions, boundary analysis, additional simulations, generalizations to additive and extended spatial modal regression models, as well as technical proofs of the main theorems and supporting lemmas.

Acknowledgements

We are deeply grateful to the Co-Editor Yi-Hau Chen, Associate Editor, and two anonymous referees for their constructive comments, leading to the substantial improvement of the paper. This research is supported by SSHRC-IDG (430-2023-00149), NSERC Discovery Grant (RGPIN-2025-04185 and DGECR-2025-00343), and NSF Grant (DMS-2210272).

References

Brady, N. C. and Weil, R. R. (2016). *The Nature and Properties of Soils*. 15th ed. Pearson.

Chen, Y. C. (2018). Modal Regression using Kernel Density Estimation: A Review. *Wiley Interdisciplinary Reviewers: Computational Statistics*, 10:e1431.

Chen, Y. C., Genovese, C. R., Tibshirani, R. J., and Wasserman, L. (2016). Nonparametric Modal Regression. *The Annals of Statistics*, 44, 489-514.

Dabo-Niang, S., Ould-Abdi, S., Ould-Abdi, A., and Diop, A. (2014). Consistency of A Nonparametric Conditional Mode Estimator for Random Fields. *Statistical Methods & Applications*, 23, 1-39.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Application*. London: Chapman and Hall.

Feng, Y., Fan, J., and Suykens, J. A. K. (2020). A Statistical Learning Approach to Modal Regression. *Journal of Machine Learning Research*, 21, 1-35.

Gao, J., Lu, Z., and Tjøstheim, D. (2006). Estimation in Semiparametric Spatial Regression. *The Annals of Statistics*, 34, 1395-1435.

Hallin, M., Lu, Z., and Tran, L. T. (2004). Local Linear Spatial Regression. *The Annals of Statistics*, 32, 2469-2500.

Hallin, M., Lu, Z., and Yu, K. (2009). Local Linear Spatial Quantile Regression. *Bernoulli*, 15, 659-686.

Hastie, T. and Tibshirani, R. (1990). *Generalised Additive Models*. Chapman & Hall, London.

Kemp, G. C. R., Parente, P. M. D. C., and Santos Silva, J. M. C. (2020). Dynamic Vector Mode Regression. *Journal of Business & Economic Statistics*, 647-661.

Lu, Z. and Chen, X. (2004). Spatial Kernel Regression Estimation: Weak Consistency. *Statistics & Probability Letters*, 68, 125-136.

Lu, Z., Lundervold, A., Tjøstheim, D., and Yao, Q. (2007). Exploring Spatial Nonlinearity using Additive Approximation. *Bernoulli*, 13, 447-472.

Lu, Z., Tang, Q., and Cheng, L. (2014). Estimating Spatial Quantile Regression with Functional Coefficients: A Robust Semiparametric Framework. *Bernoulli*, 20, 164-189.

Nandy, S., Lim, C., and Maiti, T. (2017). Additive Model Building for Spatial Regression. *Journal of the Royal Statistical Society Series B*, 79, 779-800.

Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press.

Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes.* Springer Berlin, Heidelberg.

Robinson, P. M. (2008). Developments in the Analysis of Spatial Data. *Journal of the Japan Statistical Society*, 38, 87-96.

Robinson, P. M. (2011). Asymptotic Theory for Nonparametric Regression with Spatial Data. *Journal of Econometrics*, 165, 5-19.

Ullah, A., Wang, T., and Yao, W. (2021). Modal Regression for Fixed Effects Panel Data. *Empirical Economics*, 60, 261-308.

Ullah, A., Wang, T., and Yao, W. (2022). Nonlinear Modal Regression for Dependent Data with Application for Predicting COVID-19. *Journal of the Royal Statistical Society Series A*, 185, 1424-1453.

Ullah, A., Wang, T., and Yao, W. (2023). Semiparametric Partially Linear Varying Coefficient Modal Regression. *Journal of Econometrics*, 1001-1026.

Wang, T. (2025). Parametric Modal Regression with Autocorrelated Error Process. *Statistica Sinica*, 35, 457-478.

White, R. E. (2005). *Principles and Practice of Soil Science: The Soil as a Natural Resource*. 4th ed. Wiley-Blackwell.

Xiang, S. and Yao, W. (2022). Nonparametric Statistical Learning Based on Modal Regression. *Journal of Computational and Applied Mathematics*, 409, 114-130.

Yao, W. (2013). A Note on EM Algorithm for Mixture Models. *Statistics & Probability Letters*, 83, 519-526.

Yao, W. and Li, L. (2014). A New Regression Model: Modal Linear Regression. *Scandinavian Journal of Statistics*, 41, 656-671.

Yao, W., Lindsay, B. G., and Li, R. (2012). Local Modal Regression. *Journal of Nonparametric Statistics*, 24, 647-663.

a. Department of Economics and Department of Mathematics and Statistics (by courtesy), University of Victoria, BC V8W 2Y2, Canada. E-mail: taow@uvic.ca

b. Department of Statistics, University of California Riverside, CA 92521, USA. E-mail: weixin.yao@ucr.edu