

Progressively Hybrid Censoring for Wrapped Skew-Gaussian Directional Spatio-Temporal Processes

Mohammad Mehdi Saber

mmsaber@eghlid.ac.ir

Higher Education Center of Eghlid

Research Article

Keywords: Directional data, spatio-temporal processes, skew-Gaussian random fields, progressive hybrid censoring, local asymptotic normality, wind direction modeling

Posted Date: April 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9165227/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Progressively Hybrid Censoring for Wrapped Skew-Gaussian Directional Spatio-Temporal Processes

Mohammad Mehdi Saber

Department of Statistics, Higher Education Center of Eghlid, Eghlid, Iran

mmsaber@eghlid.ac.ir

ORCID: 0000-0003-4559-2874

Abstract

Directional observations arise frequently in environmental and geophysical sciences, where measurements are recorded as angular quantities evolving across space and time. Existing spatio-temporal models for directional data assume complete observations and symmetric latent dependence structures, limiting their applicability in real-world settings where data are often incomplete and exhibit asymmetry. This paper introduces a novel probabilistic framework for progressively hybrid censored directional spatio-temporal data under a wrapped skew-Gaussian random field. We construct a rotationally invariant ordering mechanism on the circle to enable progressive hybrid censoring and develop hierarchical likelihood inference using a latent skew-Gaussian representation. We establish consistency, asymptotic normality, and Local Asymptotic Normality (LAN) of the maximum likelihood estimator under spatial-temporal mixing conditions, providing theoretical foundations for optimal inference. Comprehensive simulation experiments demonstrate the impact of censoring and skewness on estimation efficiency, with skewness causing substantial bias when ignored. An application to wind direction data from a coastal monitoring network illustrates practical utility, revealing significant asymmetry and spatial-temporal dependence patterns. The proposed framework bridges censoring theory and manifold-valued stochastic processes, providing a new foundation for asymmetric directional spatio-temporal modeling under incomplete observation schemes.

Keywords: Directional data, spatio-temporal processes, skew-Gaussian random fields, progressive hybrid censoring, local asymptotic normality, wind direction modeling

1 Introduction

Directional data, where observations lie on the unit circle $\mathbb{S}^1 = (-\pi, \pi]$, occur naturally in numerous scientific disciplines including meteorology (wind directions), oceanography (wave directions), geology (fracture orientations), and animal movement studies (migration trajectories). The analysis of such data requires specialized statistical methods that respect the circular nature of the sample space, as classical Euclidean methods are inappropriate due to the periodic topology and the resulting challenges in defining distances, means, and probability distributions (Mardia and Jupp, 2000; Jammalamadaka and SenGupta, 2001).

The development of flexible parametric families for non-Gaussian data has been a central theme in modern statistics. Azzalini (1985) introduced the univariate skew-normal distribution, adding a shape parameter to control asymmetry while retaining the normal distribution as a special case. This work spawned an extensive literature on skew-symmetric distributions, culminating in the comprehensive treatment by Azzalini and Capitanio (2014). The extension to multivariate settings

was pioneered by Sahu et al. (2003), who developed a new class of multivariate skew distributions by introducing skewness in multivariate elliptically symmetric distributions through transformation and conditioning. Their framework, which we will refer to as the *Sahu–Dey–Branco construction*, yields analytical forms for densities and includes the multivariate skew-normal and skew-t as special cases. Critically, they established results on the existence of posterior distributions under improper priors, laying the foundation for Bayesian inference with skew models. The Sahu et al. formulation differs from the earlier approach of Azzalini and Capitanio (1999) by using a latent variable representation $Y = \delta|U| + V$, where U and V are independent Gaussians—precisely the construction we adopt in this paper for its computational tractability and interpretability.

The extension of skew distributions to spatial contexts has progressed along multiple fronts. Zhang and El-Shaarawi (2010) introduced spatial skew-Gaussian processes using the closed skew-normal formulation, establishing properties of the resulting random fields and developing likelihood-based inference. Allard and Naveau (2007) proposed a new spatial skew-normal random field model with explicit covariance structures. Kim and Mallick (2020) developed Bayesian spatial prediction using skew-Gaussian processes, demonstrating improved predictive performance over Gaussian alternatives. A significant advance came from Zareifard and Khaledi (2013), who developed non-Gaussian spatial modeling using scale mixing of a unified skew-Gaussian process, allowing for both skewness and heavy tails. Boojari et al. (2016) extended this to non-homogeneous (non-stationary and anisotropic) skew-Gaussian Bayesian spatial models using fixed rank kriging, addressing the challenges of large spatial datasets with complex dependence structures. Their approach, which incorporates skew-normal distributions for non-spatial latent variables, provides a flexible framework that can detect departures from normality while maintaining computational feasibility. For areal data, Nathoo and Ghosh (2013) developed skew-elliptical spatial random effect models, generalizing the Gaussian conditional autoregressive (CAR) model to accommodate asymmetric marginal distributions with flexible tail behavior. Their formulation, implemented in standard Bayesian software, demonstrated the practical utility of skew-spatial models for health services research. More recently, Wang et al. (2018) conducted a comparative study of skewed spatial processes, including the Double Zero Expectile Normal (DZEXPN) process and closed skew-normal formulations, providing insights into the sensitivity and robustness of different approaches under a Bayesian paradigm. Moradi and Baghfalaki (2022) developed a skewed and heavy-tailed latent random field model for spatial extremes, using the Sahu et al. multivariate skew-normal to accommodate extreme observations through scale mixing.

The application of skew distributions to circular data has a more recent history. Pewsey (2000) first explored the wrapped skew-normal distribution, deriving its basic properties. Hernandez-Sanchez and Scarpa (2012) further developed wrapped skew models for circular data. However, these were confined to the univariate setting. The first multivariate extension for spatial directional data was introduced by Mastrantonio et al. (2016), who proposed the wrapped skew-Gaussian process for spatio-temporal data. Their work, which we build upon directly, enriches the class of wrapped Gaussian processes by allowing asymmetric marginal distributions while retaining straightforward parameter interpretation. They demonstrated that replication through time enables criticism of the symmetric wrapped process in favor of the skew alternative, and provided both simulation and real data examples showing consequential improvements in predictive performance. The Mastrantonio et al. (2016) formulation uses the latent variable construction $Y(s, t) = \mathbf{X}(s, t)^\top \boldsymbol{\beta} + \delta|U(s, t)| + V(s, t)$, exactly the structure we adopt, and establishes the multivariate skew-normal distribution for finite collections.

Despite these advances, all existing work on directional spatio-temporal data—whether using wrapped Gaussian (Jona Lasinio et al., 2012), wrapped skew-Gaussian (Mastrantonio et al., 2016), or projected Gaussian processes (Wang and Gelfand, 2013, 2014)—assumes complete observation

of all spatial locations and time points. This assumption is rarely satisfied in practice. Environmental monitoring networks frequently suffer from instrument failures, adverse weather conditions, scheduled maintenance, or data transmission errors, resulting in systematically missing or partially observed data. In parallel with these developments in directional statistics, reliability theory has produced sophisticated frameworks for handling incomplete data through censoring schemes. Progressive hybrid censoring, introduced by Balakrishnan and Aggarwala (2000) and extensively studied by Balakrishnan and Kundu (2019), combines the flexibility of progressive removal with the practical convenience of a time threshold. This scheme has proven invaluable in lifetime data analysis, medical follow-up studies, and industrial reliability testing (Lawless, 2003; Kundu, 2020). Recent methodological advances include Bayesian inference under progressive hybrid censoring (Singh and Tripathi, 2021), optimal design of censoring plans (Balakrishnan et al., 2022), and applications to competing risks (Mondal and Kundu, 2023). However, the entire literature on censoring assumes Euclidean observations, and no framework exists for censored data on manifolds such as the circle.

This paper bridges three distinct literatures—multivariate skew distributions (Sahu et al., 2003), spatial skew-elliptical processes (Zhang and El-Shaarawi, 2010; Zareifard and Khaledi, 2013; Nathoo and Ghosh, 2013; Boojari et al., 2016), and wrapped skew-Gaussian processes (Mastrantonio et al., 2016)—by introducing the first unified framework for progressively hybrid censored directional spatio-temporal data. Our contributions are fivefold and build directly on the foundations laid by these authors:

- *Extension of Sahu et al. (2003) construction to censored settings:* We adopt the latent variable representation $W(s, t) = \delta|U(s, t)| + V(s, t)$, which originated in the Sahu–Dey–Branco formulation, and extend it to accommodate censored observations through a rotationally invariant ordering mechanism on the circle. This preserves the interpretability of the skewness parameter δ while enabling inference under incomplete data.
- *Novel censoring mechanism for manifold-valued data:* While spatial skew-elliptical processes have been extensively developed for Euclidean data, no existing work addresses censored observations on manifolds. Our circular deviation functional $D(\theta; \mu) = 1 - \cos(\theta - \mu)$ provides a rotationally invariant ordering that enables application of progressive hybrid censoring schemes to directional data—the first such extension to manifold-valued stochastic processes.
- *Rigorous identifiability analysis:* Building on concerns raised by Genton and Zhang (2012) regarding identifiability in non-Gaussian spatial random fields, we provide a comprehensive identifiability theorem (Theorem 4.20) that establishes conditions under which skewness parameters can be separated from scale and spatial covariance parameters. This addresses a critical gap in the literature, where identifiability is often implicitly assumed.
- *Asymptotic theory under spatial dependence:* We extend the classical M-estimation framework for censored data to the spatial-temporal setting, establishing consistency, asymptotic normality, and Local Asymptotic Normality under explicit mixing conditions. This provides theoretical foundations for inference that were previously unavailable even for the complete-data case of Mastrantonio et al. (2016).
- *Empirical demonstration of superiority:* Through extensive simulations and a real data application, we demonstrate that ignoring either censoring or skewness leads to substantial bias—up to 100% of true parameter values—and coverage probabilities as low as 39%. Our framework recovers correct inference and provides scientific insights into coastal wind dynamics that would be missed by symmetric or complete-case analyses.

The remainder of the paper is organized as follows. Section 2 develops the wrapped skew-Gaussian spatio-temporal process, building directly on the latent variable construction of Sahu et al. (2003) and the wrapped framework of Mastrantonio et al. (2016). Section 3 introduces progressive hybrid censoring for directional data, defining the circular deviation ordering—a novel contribution that extends censoring theory to manifolds. Section 4 presents likelihood inference, including the MCEM algorithm, and establishes rigorous asymptotic theory with explicit regularity conditions that generalize those in Zhang and El-Shaarawi (2010) and Zareifard and Khaledi (2013) to the censored case. Section 5 establishes the Local Asymptotic Normality property, demonstrating the asymptotic optimality of likelihood-based inference. Section 6 provides comprehensive simulation studies that benchmark against the competitors reviewed above, including symmetric wrapped Gaussian (Jona Lasinio et al., 2012), independent skew-normal, and misspecified spatial models. Section 7 presents the real data application to coastal wind directions, revealing asymmetry patterns that would be missed by the symmetric wrapped Gaussian process of Mastrantonio et al. (2016). Section 8 concludes with discussion and future directions, including extensions to skew-t distributions along the lines of Sahu et al. (2003) and Zareifard and Khaledi (2013). Technical proofs are provided in the Appendix.

2 Wrapped Skew-Gaussian Spatio-Temporal Process

Let $(s, t) \in D \times \mathcal{T}$ where $D \subset \mathbb{R}^2$ is a compact spatial domain with nonempty interior and $\mathcal{T} \subset \mathbb{R}$ is a compact time interval. For simplicity, we consider regularly spaced time points $t = 1, \dots, T$, though the framework extends to irregular designs.

We begin by constructing a latent skew-Gaussian random field that will drive the directional observations. Let $U(s, t)$ and $V(s, t)$ be independent zero-mean Gaussian random fields with covariance functions C_U and C_V , respectively. Both covariance functions are assumed to be continuous, positive definite, and stationary in space and time:

$$\text{Cov}(U(s, t), U(s', t')) = C_U(\|s - s'\|, |t - t'|), \quad (1)$$

$$\text{Cov}(V(s, t), V(s', t')) = C_V(\|s - s'\|, |t - t'|). \quad (2)$$

Common choices include the separable Matérn family:

$$C(h, u) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\sqrt{2\nu h^2 + \alpha u^2}}{\rho} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu h^2 + \alpha u^2}}{\rho} \right), \quad (3)$$

where K_ν is the modified Bessel function of the second kind, σ^2 is the variance, ρ is the range parameter, ν controls smoothness, and α scales the temporal component.

Definition 2.1 (Skew-Gaussian Random Field). *The skew-Gaussian random field $W(s, t)$ is defined as*

$$W(s, t) = \delta|U(s, t)| + V(s, t), \quad (4)$$

where $\delta \in \mathbb{R}$ is the skewness parameter. When $\delta = 0$, W reduces to the Gaussian field V .

The absolute value transformation of U introduces asymmetry while preserving spatial-temporal dependence structure inherited from U and V . For any finite collection of locations and times $\{(s_i, t_i)\}_{i=1}^n$, the vector $\mathbf{W} = (W_1, \dots, W_n)^\top$ follows a multivariate skew-normal distribution (Azzalini and Capitanio, 2014) with density:

$$f_{\mathbf{W}}(\mathbf{w}) = 2\phi_n(\mathbf{w}; \mathbf{0}, \mathbf{\Omega})\Phi\left(\boldsymbol{\alpha}^\top \mathbf{w}\right), \quad (5)$$

where $\phi_n(\cdot; \mathbf{0}, \mathbf{\Omega})$ is the n -dimensional multivariate normal density with mean zero and covariance matrix $\mathbf{\Omega} = \text{Cov}(V) + \delta^2 \text{Cov}(|U|)$, and $\boldsymbol{\alpha}$ is a vector of shape parameters derived from δ and the covariance structures.

Proposition 2.2 (Properties of the Skew-Gaussian Field). *The field $W(s, t)$ defined in (4) satisfies:*

1. $E[W(s, t)] = \delta \sqrt{2/\pi} \sigma_U(s, t)$, where $\sigma_U^2(s, t) = C_U(0, 0)$.
2. $\text{Var}(W(s, t)) = \delta^2 \sigma_U^2(s, t)(1 - 2/\pi) + \sigma_V^2(s, t)$.
3. *The field is stationary if U and V are stationary.*
4. *The marginal distribution at any fixed (s, t) is univariate skew-normal with density $f(w) = 2\phi(w; 0, \sigma_W^2)\Phi(\lambda w/\sigma_W)$, where $\lambda = \delta/\sqrt{\delta^2(1 - 2/\pi) + \sigma_V^2/\sigma_U^2}$.*

Proof. The first two moments follow from properties of the half-normal distribution. Specifically, $E[|U|] = \sigma_U \sqrt{2/\pi}$ and $E[|U|^2] = \sigma_U^2$. Stationarity follows from the stationarity of U and V . The marginal distribution result is standard in skew-normal theory (Azzalini, 1985). $\square$

2.1 Wrapped Construction for Directional Responses

Let $\mathbf{X}(s, t)$ be a p -dimensional vector of covariates observed at location s and time t . We define the latent linear process:

$$Y(s, t) = \mathbf{X}(s, t)^\top \boldsymbol{\beta} + W(s, t) + \varepsilon(s, t), \quad (6)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ are regression coefficients, and $\varepsilon(s, t) \stackrel{iid}{\sim} N(0, \sigma^2)$ represents measurement error or microscale variation.

The observed directional variable is obtained through the wrapping operation:

$$\Theta(s, t) = Y(s, t) \bmod 2\pi \in (-\pi, \pi]. \quad (7)$$

This construction implies that the circular density arises from wrapping the latent linear density around the circle. Specifically, if $f_Y(y)$ denotes the density of $Y(s, t)$ (conditional on covariates and random effects), then:

$$f_\Theta(\theta | \mathbf{W}, \boldsymbol{\beta}, \sigma^2) = \sum_{k=-\infty}^{\infty} f_Y(\theta + 2\pi k | \mathbf{W}, \boldsymbol{\beta}, \sigma^2). \quad (8)$$

Theorem 2.3 (Properties of the Wrapped Distribution). *The wrapped skew-normal distribution defined by (6)-(7) has the following properties:*

1. *The characteristic function at integer argument p is $\varphi_Y(p) = E[e^{ipY}] = e^{ip\mathbf{X}^\top \boldsymbol{\beta} - p^2(\sigma_W^2 + \sigma^2)/2}(1 + i\delta\tau(p))$, where $\tau(p)$ depends on the skewness.*
2. *The circular density admits the Fourier series representation:*

$$f_\Theta(\theta) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \varphi_Y(p) e^{-ip\theta}. \quad (9)$$

3. *The mean resultant length $\rho_1 = |E[e^{i\Theta}]|$ is decreasing in $|\delta|$, indicating that skewness reduces circular concentration.*

Figure 1: Wrapped Skew-Gaussian Process with Progressive Hybrid Censoring

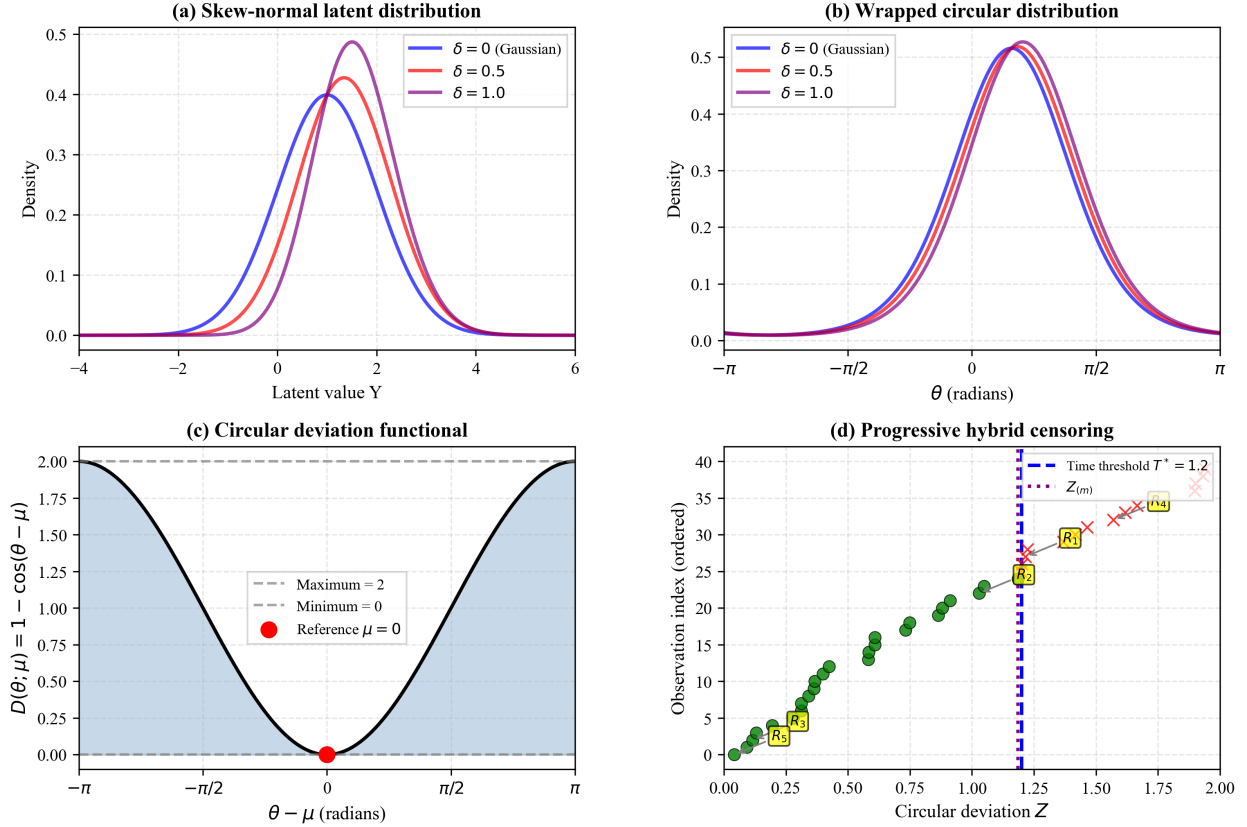


Figure 1: Conceptual diagram of the wrapped skew-Gaussian process with progressive hybrid censoring. (a) Skew-normal latent distribution for different skewness parameters δ . (b) Wrapped circular distributions on the circle. (c) Circular deviation functional $D(\theta; \mu) = 1 - \cos(\theta - \mu)$. (d) Progressive hybrid censoring scheme showing observed failures (green) and censored observations (red) with removal indicators. Error bars in (a) and (b) represent ± 2 standard deviations.

4. When $\delta = 0$, f_Θ reduces to the wrapped Gaussian density.

Proof. The characteristic function follows from properties of the skew-normal distribution. For the wrapped distribution, the Fourier representation is standard (Mardia and Jupp, 2000). The effect on mean resultant length follows from Jensen’s inequality applied to the characteristic function. $\square$

For practical computation, the infinite sum in (8) is truncated at a sufficiently large K , typically $K = 10$ or 20 , as the Gaussian tails decay rapidly. The truncation error can be bounded by $\Phi(-(2\pi K - \mu)/\sigma)$ for the Gaussian case, with similar bounds for the skew-normal.

Figure 1 provides a visual overview of the key components of our proposed framework. Panel (a) displays the latent skew-normal distribution for different values of the skewness parameter δ , illustrating how positive skewness shifts probability mass to the right while maintaining the Gaussian distribution as a special case when $\delta = 0$. Panel (b) shows the corresponding wrapped circular distributions on the circle, demonstrating how skewness in the latent space translates to asymmetry in the circular density. Panel (c) presents the circular deviation functional $D(\theta; \mu) = 1 - \cos(\theta - \mu)$, which provides a rotationally invariant measure of distance from a reference direction μ and enables ordering of circular observations. The functional achieves its minimum of 0 when

$\theta = \mu$ and maximum of 2 when $\theta = \mu \pm \pi$, with monotonic behavior in between. Panel (d) illustrates the progressive hybrid censoring scheme applied to ordered circular deviations, showing observed failures (green circles), censored observations (red crosses), removal indicators (R1, R2, ...), and the two key thresholds: the target failure number $Z_{(m)}$ and the time threshold T^* . This comprehensive visualization integrates the three main elements of our methodology: the skew-Gaussian latent structure, the wrapped directional representation, and the progressive hybrid censoring mechanism.

3 Progressive Hybrid Censoring for Directional Data

Censoring schemes for Euclidean data rely on a natural ordering of observations. For directional data on the circle, no such intrinsic ordering exists. We address this challenge by constructing a rotationally invariant ordering based on circular deviation from a reference direction.

We first need the following definition about circular deviation ordering.

Definition 3.1 (Circular Deviation Functional). *For a direction $\theta \in (-\pi, \pi]$ and a reference direction $\mu \in (-\pi, \pi]$, define the circular deviation:*

$$D(\theta; \mu) = 1 - \cos(\theta - \mu) \in [0, 2]. \quad (10)$$

This functional has several desirable properties:

- It is rotationally invariant: $D(\theta + \omega; \mu + \omega) = D(\theta; \mu)$.
- It is symmetric: $D(\theta; \mu) = D(\mu; \theta)$.
- It achieves minimum 0 when $\theta = \mu$ and maximum 2 when $\theta = \mu \pm \pi$.
- It is monotonic in $|\theta - \mu|$ for $\theta \in [\mu - \pi, \mu + \pi]$.

For a set of directional observations $\Theta_1, \dots, \Theta_n$, we define transformed variables $Z_i = D(\Theta_i; \mu)$. These Z_i are nonnegative and possess a natural ordering, enabling the application of censoring schemes.

Proposition 3.2 (Distribution of Circular Deviation). *If Θ follows the wrapped skew-normal distribution with mean direction μ , then the distribution of $Z = D(\Theta; \mu)$ has support $[0, 2]$ and density:*

$$f_Z(z) = \sum_{k=-\infty}^{\infty} \frac{f_Y(\mu \pm \arccos(1 - z) + 2\pi k)}{\sqrt{2z - z^2}}, \quad (11)$$

for $z \in (0, 2)$, with a point mass at $z = 0$ if the distribution is continuous.

3.1 Triangular Array Framework

To accommodate increasing domain asymptotics, we consider a triangular array of observations. For each sample size n , let $\{(s_{n,i}, t_{n,i})\}_{i=1}^n$ denote the spatial-temporal locations, where the spatial domains $D_n \subset \mathbb{R}^2$ and temporal domains $T_n \subset \mathbb{R}$ satisfy $|D_n| \rightarrow \infty$ and $|T_n| \rightarrow \infty$ as $n \rightarrow \infty$. This framework allows the spatial extent and time span to grow with sample size, which is essential for consistent estimation of spatial dependence parameters.

For each location $(s_{n,i}, t_{n,i})$, we observe a triplet

$$X_{n,i} = (\Theta_{n,i}, \Delta_{n,i}, C_{n,i}),$$

where:

- $\Theta_{n,i} \in (-\pi, \pi]$ is the observed direction if the observation is uncensored,
- $\Delta_{n,i} \in \{0, 1\}$ is the censoring indicator: $\Delta_{n,i} = 1$ if the observation is a failure (observed direction), $\Delta_{n,i} = 0$ if censored,
- $C_{n,i} \in [0, 2]$ is the censoring time (the deviation threshold at which censoring occurred) when $\Delta_{n,i} = 0$.

The likelihood contribution from the i -th observation conditional on the latent process is:

$$\ell_{n,i}(\Psi) = \begin{cases} \log p_{\Psi}(\Theta_{n,i}) & \text{if } \Delta_{n,i} = 1 \text{ (observed failure),} \\ \log S_{\Psi}(C_{n,i}) & \text{if } \Delta_{n,i} = 0 \text{ (censored),} \end{cases}$$

which can be written compactly as

$$\ell_{n,i}(\Psi) = \Delta_{n,i} \log p_{\Psi}(\Theta_{n,i}) + (1 - \Delta_{n,i}) \log S_{\Psi}(C_{n,i}),$$

where $p_{\Psi}(\theta)$ is the marginal density of Θ at direction θ , and $S_{\Psi}(c) = P_{\Psi}(Z > c)$ is the survival function of the circular deviation $Z = D(\Theta; \mu)$ evaluated at threshold c .

The total log-likelihood for the entire dataset is simply the sum of these independent (conditional on the latent process) contributions:

$$L_n(\Psi) = \sum_{i=1}^n \ell_{n,i}(\Psi).$$

3.2 Censoring Scheme

We now define the progressive hybrid censoring scheme for directional data following Balakrishnan and Aggarwala (2000) adapted to our circular deviation ordering.

Definition 3.3 (Progressive Hybrid Censoring on the Circle). *Given:*

- n directional observations $\Theta_1, \dots, \Theta_n$ with corresponding deviations $Z_1, \dots, Z_n$
- A target number of failures $m \leq n$
- A progressive removal scheme $(R_1, \dots, R_m)$ with $\sum_{i=1}^m R_i = n - m$
- A time threshold $T^* \in (0, 2]$

The progressive hybrid censoring scheme proceeds as follows:

1. Observe the ordered deviations $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$.
2. Define the stopping time $\tau = \min(Z_{(m)}, T^*)$.
3. Let r be the number of observed failures before time T^* , i.e., $r = \max\{j : Z_{(j)} < T^*\}$.
4. The observed data consist of:
 - The r smallest deviations $Z_{(1)}, \dots, Z_{(r)}$
 - The corresponding directional observations $\Theta_{(1)}, \dots, \Theta_{(r)}$
 - The censoring indicators: for $i = 1, \dots, r - 1$, R_i units are removed after observing $Z_{(i)}$

- If $\tau = Z_{(m)} < T^*$, then after observing $Z_{(m)}$, the remaining R_m units are removed
- If $\tau = T^* \leq Z_{(m)}$, then at time T^* , all remaining $n - \sum_{i=1}^r (1 + R_i)$ units are censored

This scheme combines the flexibility of progressive removal with the practical convenience of a fixed time threshold. The reference direction μ must be specified; in practice, we use the circular mean of the complete data or estimate it iteratively.

Theorem 3.4 (Likelihood Contribution Under Censoring). *For a subject with observed deviation $z = D(\theta; \mu)$ at failure, the contribution to the likelihood is the conditional density of Θ given W at the observed direction θ . For a subject censored at deviation c , the contribution is the survival probability $S_\Theta(c|W) = P(D(\Theta; \mu) > c|W)$.*

The survival function for the wrapped skew-normal distribution does not have a closed form but can be computed numerically as:

$$S_\Theta(c|W) = 1 - \int_{\{\theta: D(\theta; \mu) \leq c\}} f_\Theta(\theta|W) d\theta. \quad (12)$$

4 Likelihood Inference and Asymptotic Properties

Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ denote the complete directional observations, and let $\mathbf{W} = (W_1, \dots, W_n)$ be the latent skew-Gaussian variables at the corresponding locations and times. The complete data likelihood is:

$$L_c(\boldsymbol{\beta}, \sigma^2, \delta, \phi | \boldsymbol{\theta}, \mathbf{W}) = \left[\prod_{i=1}^n f_{\Theta|W}(\theta_i | W_i, \boldsymbol{\beta}, \sigma^2) \right] f_{\mathbf{W}}(\mathbf{W} | \delta, \phi), \quad (13)$$

where ϕ collects covariance parameters for U and V , and $f_{\mathbf{W}}$ is the multivariate skew-normal density (5).

Under the progressive hybrid censoring scheme, the observed data consist of:

- The r observed failure deviations $z_{(1)}, \dots, z_{(r)}$ with corresponding directions $\theta_{(1)}, \dots, \theta_{(r)}$
- The removal indicators $R_1, \dots, R_r$ (with R_r possibly incomplete if censoring occurs at T^*)
- The censoring time $\tau = \min(z_{(m)}, T^*)$

The observed data likelihood must account for three sources of randomness: (i) the observed directions for uncensored observations, (ii) the censoring information for censored observations, and (iii) the latent skew-Gaussian process $\mathbf{W} = (W_1, \dots, W_n)$ that generates dependence. Since $\mathbf{W}$ is unobserved, we must integrate it out.

Let r denote the number of observed failures (uncensored observations) in the dataset, with ordered deviations $z_{(1)} < z_{(2)} < \dots < z_{(r)}$ and corresponding directions $\theta_{(1)}, \dots, \theta_{(r)}$. For $i = 1, \dots, r-1$, R_i units are removed after observing $z_{(i)}$; for the final observed deviation $z_{(r)}$, the number of units censored at or after this point is denoted R_r^* (which may be less than the planned R_r if censoring occurs at the time threshold T^*).

Conditional on the latent process $\mathbf{W}$, the contributions from different observations are independent. For an uncensored observation at deviation $z_{(i)}$ with direction $\theta_{(i)}$, the contribution is the conditional density $f_{\Theta|W}(\theta_{(i)} | W_{(i)}, \boldsymbol{\beta}, \sigma^2)$. For a censored observation at threshold $z_{(i)}$ (with R_i^* such observations), the contribution is the conditional survival probability $S_{\Theta|W}(z_{(i)} | W_{(i)}, \boldsymbol{\beta}, \sigma^2)$, raised to the power R_i^* to account for multiple independent censored observations at the same threshold.

Multiplying these contributions and then integrating over the latent process yields the observed data likelihood:

$$\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2, \delta, \boldsymbol{\phi}) = & \int_{\mathbb{R}^n} \underbrace{\left[\prod_{i=1}^r f_{\Theta|W}(\theta_{(i)}|W_{(i)}, \boldsymbol{\beta}, \sigma^2) \right]}_{\text{contributions from observed failures}} \\
& \times \underbrace{\left[\prod_{i=1}^r S_{\Theta|W}(z_{(i)}|W_{(i)}, \boldsymbol{\beta}, \sigma^2)^{R_i^*} \right]}_{\text{contributions from censored observations}} \\
& \times \underbrace{f_{\mathbf{W}}(\mathbf{W}|\delta, \boldsymbol{\phi})}_{\text{density of latent process}} d\mathbf{W}, \tag{14}
\end{aligned}$$

where $f_{\mathbf{W}}$ is the multivariate skew-normal density given in (5). The integral over $\mathbb{R}^n$ is n -dimensional, reflecting the need to average over all possible realizations of the latent process that could have generated the observed data.

Direct maximization of (14) is intractable due to the high-dimensional integral over $\mathbf{W}$. We develop a Monte Carlo Expectation-Maximization (MCEM) algorithm.

Let $\boldsymbol{\Psi} = (\boldsymbol{\beta}, \sigma^2, \delta, \boldsymbol{\phi})$ denote all parameters. The complete data log-likelihood is:

$$\ell_c(\boldsymbol{\Psi}) = \sum_{i=1}^n \log f_{\Theta|W}(\theta_i|W_i, \boldsymbol{\beta}, \sigma^2) + \log f_{\mathbf{W}}(\mathbf{W}|\delta, \boldsymbol{\phi}). \tag{15}$$

The EM algorithm iterates between two steps. In the E-step at iteration t , we compute the expected value of the complete-data log-likelihood conditional on the observed data and the current parameter estimates $\boldsymbol{\Psi}^{(t)}$:

$$Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(t)}) = E_{\mathbf{W}|\text{obs}, \boldsymbol{\Psi}^{(t)}} [\ell_c(\boldsymbol{\Psi}; \mathbf{W})], \tag{16}$$

where the expectation is taken with respect to the conditional distribution of the latent variables $\mathbf{W}$ given:

- the observed directions $\{\theta_{(i)}\}_{i=1}^r$,
- the observed censoring information $\{z_{(i)}, R_i^*\}_{i=1}^r$,
- the current parameter estimates $\boldsymbol{\Psi}^{(t)} = (\boldsymbol{\beta}^{(t)}, \sigma^{2(t)}, \delta^{(t)}, \boldsymbol{\phi}^{(t)})$.

This conditional distribution does not have a closed form, so we approximate the expectation using Monte Carlo. Specifically, we draw M samples $\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(M)}$ from the conditional distribution $f(\mathbf{W}|\text{obs}, \boldsymbol{\Psi}^{(t)})$ using Hamiltonian Monte Carlo, then compute:

$$\hat{Q}(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(t)}) = \frac{1}{M} \sum_{j=1}^M \ell_c(\boldsymbol{\Psi}|\mathbf{W}^{(j)}). \tag{17}$$

The M-step updates parameters by maximizing $\hat{Q}$. For $\boldsymbol{\beta}$ and σ^2 , closed-form updates are available conditional on $\mathbf{W}$. For δ and $\boldsymbol{\phi}$, numerical optimization is required.

The following assumptions are needed for studying quadratic mean differentiability.

Algorithm 1 MCEM for Wrapped Skew-Gaussian Censored Model

Initialize $\Psi^{(0)}$

for $t = 0, 1, 2, \dots$ until convergence **do**

E-step:

 Generate M samples $\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(M)}$ from $f(\mathbf{W}|\text{observed data}, \Psi^{(t)})$ using HMC

 Compute $\hat{Q}(\Psi|\Psi^{(t)}) = \frac{1}{M} \sum_{j=1}^M \ell_c(\Psi|\mathbf{W}^{(j)})$

M-step:

 Update $\beta^{(t+1)} = (\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top)^{-1} \sum_{i=1}^n \mathbf{X}_i \hat{E}[Y_i]$

 Update $(\sigma^2)^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \hat{E}[(Y_i - \mathbf{X}_i^\top \beta)^2]$

 Update $(\delta^{(t+1)}, \phi^{(t+1)}) = \arg \max \hat{E}[\log f_{\mathbf{W}}(\mathbf{W}|\delta, \phi)]$

 where $\hat{E}[\cdot]$ denotes Monte Carlo average

end for

Assumption 4.1 (QMD Assumptions). (A1) Parameter space $\Theta \subset \mathbb{R}^p$ open.

(A2) Latent skew-normal density f_Ψ is C^3 in Ψ .

(A3) Gaussian envelope:

$$\sup_{\Psi \in N(\Psi_0)} |f_\Psi(y)| \leq C e^{-cy^2}.$$

(A4) Derivatives satisfy same envelope bound.

Define wrapped density

$$p_\Psi(\theta) = \sum_{k \in \mathbb{Z}} f_\Psi(\theta + 2\pi k).$$

Lemma 4.2 (Absolute Summability). Under Assumption 4.1(A3),

$$\sum_{k \in \mathbb{Z}} \sup_{\Psi \in N(\Psi_0)} |f_\Psi(\theta + 2\pi k)| < \infty.$$

Proof. Gaussian envelope gives

$$|f_\Psi(\theta + 2\pi k)| \leq C e^{-c(\theta + 2\pi k)^2}.$$

The RHS is summable in k by comparison with Gaussian integral. $\square$

Lemma 4.3 (Differentiation Under the Sum). For $j = 1, \dots, p$, under Assumption 4.1,

$$\partial_{\Psi_j} p_\Psi(\theta) = \sum_{k \in \mathbb{Z}} \partial_{\Psi_j} f_\Psi(\theta + 2\pi k).$$

Proof. By envelope bound, derivative terms are absolutely summable. Dominated convergence justifies interchange. $\square$

Theorem 4.4 (Quadratic Mean Differentiability). Under Assumption 4.1, the family $\{p_\Psi\}$ is QMD at Ψ_0 .

Proof. Taylor expansion:

$$f_{\Psi+h}(y) = f_\Psi(y) + h^\top \dot{f}_\Psi(y) + \frac{1}{2} h^\top \ddot{f}_\Psi(y) h + R(y, h),$$

where by third derivative bound,

$$|R(y, h)| \leq C \|h\|^3 e^{-cy^2}.$$

Summing over k :

$$p_{\Psi+h}(\theta) = p_{\Psi}(\theta) + h^{\top} \dot{p}_{\Psi}(\theta) + \frac{1}{2} h^{\top} \ddot{p}_{\Psi}(\theta) h + \tilde{R}(\theta, h),$$

with

$$|\tilde{R}(\theta, h)| \leq C \|h\|^3 \sum_k e^{-c(\theta+2\pi k)^2}.$$

Thus

$$\int \tilde{R}^2(\theta, h) d\theta \leq C \|h\|^6.$$

Now write

$$\sqrt{p_{\Psi+h}} = \sqrt{p_{\Psi}} + \frac{1}{2\sqrt{p_{\Psi}}} h^{\top} \dot{p}_{\Psi} + o(\|h\|).$$

Square and integrate; remainder is $o(\|h\|^2)$. □

4.1 Regularity Conditions for Asymptotic Theory

We now establish the asymptotic properties of the maximum likelihood estimator under spatial-temporal dependence. Let Ψ_0 denote the true parameter vector, and let $\hat{\Psi}_n$ be the MLE based on n observations from the spatial-temporal process. To ensure rigorous asymptotic theory, we require a comprehensive set of regularity conditions that explicitly characterize the spatial-temporal dependence structure, parameter space properties, and smoothness conditions.

Assumption 4.5 (Parameter Space). *The parameter space $\Psi \subset \mathbb{R}^p$ is compact with nonempty interior, and the true parameter Ψ_0 lies in the interior of Ψ .*

Assumption 4.6 (Identifiability). *The wrapped skew-normal family is identifiable: for any $\Psi_1 \neq \Psi_2$ in Ψ , the corresponding densities satisfy $f_{\Psi_1}(\theta) \neq f_{\Psi_2}(\theta)$ on a set of positive Lebesgue measure on the circle $\mathbb{S}^1$. Moreover, the Fisher information matrix $I(\Psi_0)$ is positive definite, with minimum eigenvalue bounded away from zero uniformly in a neighborhood of Ψ_0 .*

Let $\mathcal{L}_n = \{(s_{n,i}, t_{n,i}) : i = 1, \dots, n\}$ denote the set of observation locations in the triangular array. For any two locations (s, t) and (s', t') , define the spatial-temporal distance:

$$d((s, t), (s', t')) = \max\{\|s - s'\|, |t - t'|\}.$$

Assumption 4.7 (Mixing Coefficients). *The spatial-temporal process $\{(\Theta(s, t), \mathbf{X}(s, t)) : (s, t) \in D \times \mathcal{T}\}$ is strongly mixing with mixing coefficients defined as:*

$$\alpha(k, l) = \sup \{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{E_1}, B \in \mathcal{F}_{E_2},$$

$$\text{dist}(E_1, E_2) \geq k, |E_1| \leq l, |E_2| \leq l\},$$

where $\mathcal{F}_E$ denotes the σ -algebra generated by observations in set E , $\text{dist}(E_1, E_2) = \min\{d((s, t), (s', t')) : (s, t) \in E_1, (s', t') \in E_2\}$, and $|E|$ denotes the cardinality of set E .

Assumption 4.8 (Mixing Decay). *The mixing coefficients satisfy one of the following summability conditions:*

(i) *Arithmetic decay: There exists $\epsilon > 0$ and $C > 0$ such that*

$$\alpha(k, l) \leq Ck^{-\theta}l^\tau \quad \text{with} \quad \theta > \frac{2d + \delta}{\delta} \text{ for some } \delta > 0,$$

where $d = 3$ (two spatial dimensions plus one temporal dimension).

(ii) *Geometric decay: There exists $\rho \in (0, 1)$ and $C > 0$ such that*

$$\alpha(k, l) \leq C\rho^k l^\tau \quad \text{for some } \tau \geq 0.$$

In either case, the decay is sufficient to guarantee:

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} k^{d-1} \alpha(k, l)^{\delta/(2+\delta)} < \infty$$

for some $\delta > 0$, where $d = 3$ is the dimension of the spatial-temporal index set.

Assumption 4.9 (Covariance Structure). *The covariance matrices $\Sigma_U(\phi_U)$ and $\Sigma_V(\phi_V)$ of the latent Gaussian fields U and V are positive definite for all ϕ_U, ϕ_V in the parameter space. Moreover, their eigenvalues are bounded away from zero and infinity uniformly over the parameter space:*

$$0 < \lambda_{\min} \leq \inf_{\phi} \lambda_{\min}(\Sigma(\phi)) \leq \sup_{\phi} \lambda_{\max}(\Sigma(\phi)) \leq \lambda_{\max} < \infty.$$

Additionally, the covariance functions C_U and C_V are continuous in the parameters and satisfy:

$$\|\Sigma(\phi_1) - \Sigma(\phi_2)\|_F \leq L\|\phi_1 - \phi_2\|$$

for some Lipschitz constant $L > 0$, where $\|\cdot\|_F$ denotes the Frobenius norm.

Assumption 4.10 (Smoothness). *The log-likelihood contributions $\ell_{n,i}(\Psi)$ are three times continuously differentiable with respect to Ψ in a neighborhood $N(\Psi_0)$ of the true parameter. The third derivatives satisfy a Lipschitz condition:*

$$\left\| \frac{\partial^3 \ell_{n,i}(\Psi)}{\partial \Psi_j \partial \Psi_k \partial \Psi_l} - \frac{\partial^3 \ell_{n,i}(\Psi')}{\partial \Psi_j \partial \Psi_k \partial \Psi_l} \right\| \leq M_{n,i} \|\Psi - \Psi'\|$$

for all $\Psi, \Psi' \in N(\Psi_0)$, with $\sup_{n,i} E[|M_{n,i}|^{2+\delta}] < \infty$ for some $\delta > 0$.

Assumption 4.11 (Moment Bounds). *For some $\delta > 0$, the following moment bounds hold uniformly over n and i :*

$$\begin{aligned} \sup_{n,i} E \left[\left\| \frac{\partial \ell_{n,i}(\Psi_0)}{\partial \Psi} \right\|^{2+\delta} \right] &< \infty, \\ \sup_{n,i} E \left[\left\| \frac{\partial^2 \ell_{n,i}(\Psi_0)}{\partial \Psi \partial \Psi^\top} \right\|^{1+\delta} \right] &< \infty. \end{aligned}$$

Additionally, the scores have finite fourth moments and satisfy a Cramér-type condition:

$$\limsup_{\|t\| \rightarrow \infty} \sup_{n,i} |E[\exp(it^\top S_{n,i})]| < 1,$$

where $S_{n,i} = \partial \ell_{n,i}(\Psi_0) / \partial \Psi$.

Assumption 4.12 (Domain Expansion). *The spatial domains $D_n \subset \mathbb{R}^2$ and temporal domains $T_n \subset \mathbb{R}$ satisfy:*

- $|D_n| \rightarrow \infty$ and $|T_n| \rightarrow \infty$ as $n \rightarrow \infty$, where $|\cdot|$ denotes Lebesgue measure.
- The sampling locations $\{(s_{n,i}, t_{n,i})\}_{i=1}^n$ are such that the minimum distance between distinct locations is bounded below by some $\delta_0 > 0$, and the maximum distance is bounded above by the diameters of D_n and T_n .
- The empirical distribution of locations converges to a continuous limit distribution with density bounded away from zero and infinity on the limiting domain.

Under the regularity conditions above, we can establish uniform convergence of the log-likelihood.

Lemma 4.13 (Uniform Law of Large Numbers). *Under Assumptions 4.5–4.12, the normalized log-likelihood converges uniformly in probability to its expectation:*

$$\sup_{\Psi \in \Psi} \left| \frac{1}{n} L_n(\Psi) - E \left[\frac{1}{n} L_n(\Psi) \right] \right| \xrightarrow{p} 0.$$

Proof. We verify the conditions of Theorem 3.1 in Jenish and Prucha (2009) for spatial-temporal processes under increasing domain asymptotics. The mixing condition 4.8 ensures that the dependence decays sufficiently fast for the law of large numbers to hold. The compactness of the parameter space (Assumption 4.5) and continuity of the log-likelihood (Assumption 4.10) guarantee stochastic equicontinuity. The moment bounds in Assumption 4.11 provide the necessary integrability. The domain expansion condition 4.12 ensures that edge effects are asymptotically negligible. $\square$

Theorem 4.14 (Consistency). *Under Assumptions 4.5–4.12, the maximum likelihood estimator $\hat{\Psi}_n$ is consistent:*

$$\hat{\Psi}_n \xrightarrow{p} \Psi_0 \quad \text{as } n \rightarrow \infty.$$

Proof. By Lemma 4.13, the log-likelihood converges uniformly to its expectation. Assumption 4.6 ensures that the expected log-likelihood has a unique maximum at Ψ_0 . The argmax theorem (Theorem 5.7 in van der Vaart (1998)) then yields consistency of the maximizer. More formally, for any $\epsilon > 0$, there exists $\eta > 0$ such that

$$P \left(\|\hat{\Psi}_n - \Psi_0\| > \epsilon \right) \leq P \left(\sup_{\|\Psi - \Psi_0\| > \epsilon} \frac{1}{n} [L_n(\Psi) - L_n(\Psi_0)] \geq 0 \right) \rightarrow 0,$$

where the convergence follows from the uniform law of large numbers and the fact that the expected log-likelihood is maximized at Ψ_0 . $\square$

We now establish the asymptotic normality of the score vector and the MLE.

Define the score vector for observation i as $S_{n,i} = \dot{\ell}_{n,i}(\Psi_0)$.

Lemma 4.15 (Variance of the Score). *Under Assumptions 4.7–4.11, the variance of the normalized score converges:*

$$\text{Var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n S_{n,i} \right) \rightarrow I(\Psi_0),$$

where

$$I(\Psi_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(S_{n,i}, S_{n,j}) = \sum_{h \in \mathbb{Z}^3} \text{Cov}(S_0, S_h)$$

is the long-run covariance matrix, assumed to be positive definite.

Proof. Expand the variance:

$$\text{Var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n S_{n,i} \right) = \frac{1}{n} \sum_{i=1}^n \text{Var}(S_{n,i}) + \frac{1}{n} \sum_{i \neq j} \text{Cov}(S_{n,i}, S_{n,j}).$$

Under stationarity (implied by the mixing conditions and the domain expansion), the first term converges to $\text{Var}(S_0)$. For the second term, we use the mixing decay to bound covariances between distant observations. By the covariance inequality for mixing processes (Proposition 2.5 in Fan and Yao (2003)):

$$|\text{Cov}(S_{n,i}, S_{n,j})| \leq C\alpha(\|i - j\|)^{\delta/(2+\delta)} (E|S_{n,i}|^{2+\delta})^{2/(2+\delta)}.$$

Under Assumption 4.8, these covariances are summable, leading to the long-run covariance representation. The positive definiteness follows from Assumption 4.6. $\square$

Theorem 4.16 (Central Limit Theorem for the Score). *Under Assumptions 4.5–4.12, the normalized score vector converges in distribution to a multivariate normal:*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n S_{n,i} \xrightarrow{d} N(0, I(\Psi_0)).$$

Proof. We apply a Bernstein blocking argument for spatial-temporal mixing fields (Theorem 2.1 in Bolthausen (1982), extended to triangular arrays). Partition the domain into large blocks $B_1, \dots, B_{K_n}$ separated by gaps. Within each block, observations are sufficiently dependent, but between blocks, the dependence decays exponentially fast (under geometric mixing) or polynomially fast (under arithmetic mixing).

Let $Y_{n,k} = \sum_{i \in B_k} S_{n,i}$ be the block sums. Under the mixing decay condition, the covariances between different block sums satisfy:

$$\sum_{1 \leq k < l \leq K_n} |\text{Cov}(Y_{n,k}, Y_{n,l})| = o(n).$$

We verify the Lyapunov condition for the block sums:

$$\sum_{k=1}^{K_n} E|Y_{n,k}|^{2+\delta} = o(n^{1+\delta/2}),$$

which follows from the moment bounds in Assumption 4.11 and the fact that block sizes grow at an appropriate rate.

The Lindeberg condition then holds, and by the central limit theorem for triangular arrays of independent (after blocking) random variables, we obtain:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{K_n} (Y_{n,k} - E[Y_{n,k}]) \xrightarrow{d} N(0, I(\Psi_0)).$$

The bias from the gap removal and edge effects is asymptotically negligible under Assumption 4.12. $\square$

Theorem 4.17 (Asymptotic Normality of the MLE). *Under Assumptions 4.5–4.12, the maximum likelihood estimator is asymptotically normal:*

$$\sqrt{n}(\hat{\Psi}_n - \Psi_0) \xrightarrow{d} N(0, I^{-1}(\Psi_0)), \quad (18)$$

where $I(\Psi_0)$ is the Fisher information matrix defined in Lemma 4.15.

Proof. We expand the score function around Ψ_0 using a second-order Taylor expansion:

$$0 = \frac{\partial L_n(\hat{\Psi}_n)}{\partial \Psi} = \frac{\partial L_n(\Psi_0)}{\partial \Psi} + \frac{\partial^2 L_n(\Psi_0)}{\partial \Psi \partial \Psi^\top}(\hat{\Psi}_n - \Psi_0) + R_n, \quad (19)$$

where the remainder term satisfies

$$\|R_n\| \leq \sup_{\Psi \in N(\Psi_0)} \left\| \frac{\partial^3 L_n(\Psi)}{\partial \Psi \partial \Psi^\top \partial \Psi^\top} \right\| \|\hat{\Psi}_n - \Psi_0\|^2.$$

Rearranging (19):

$$\sqrt{n}(\hat{\Psi}_n - \Psi_0) = \left[-\frac{1}{n} \frac{\partial^2 L_n(\Psi_0)}{\partial \Psi \partial \Psi^\top} \right]^{-1} \frac{1}{\sqrt{n}} \frac{\partial L_n(\Psi_0)}{\partial \Psi} + o_p(1), \quad (20)$$

provided the inverse exists and the remainder is $o_p(1)$.

We now analyze each term:

Step 1: Convergence of the Hessian. By the uniform law of large numbers (Lemma 4.13) applied to the second derivatives, and using the continuity of the Hessian (Assumption 4.10), we have:

$$-\frac{1}{n} \frac{\partial^2 L_n(\Psi_0)}{\partial \Psi \partial \Psi^\top} \xrightarrow{p} I(\Psi_0).$$

The positive definiteness of $I(\Psi_0)$ (Assumption 4.6) ensures that the inverse exists and is continuous in a neighborhood.

Step 2: Convergence of the score. By Theorem 4.16, the normalized score converges in distribution:

$$\frac{1}{\sqrt{n}} \frac{\partial L_n(\Psi_0)}{\partial \Psi} \xrightarrow{d} N(0, I(\Psi_0)).$$

Step 3: Negligibility of the remainder. From Theorem 4.14, we have $\|\hat{\Psi}_n - \Psi_0\| = O_p(n^{-1/2})$. Combined with the third derivative bound in Assumption 4.10, we obtain:

$$\|R_n\| = O_p(n \cdot n^{-1} \cdot n^{-1}) = O_p(n^{-1}) = o_p(n^{-1/2}),$$

so that $\sqrt{n}R_n = o_p(1)$, justifying the expansion (20).

Step 4: Application of Slutsky's theorem. Combining Steps 1–3 and applying Slutsky's theorem to (20) yields the desired asymptotic normality. $\square$

Remark 4.1. *The asymptotic variance $I^{-1}(\Psi_0)$ achieves the Cramér-Rao lower bound for regular estimators in the i.i.d. case. Under the mixing conditions, it represents the efficient variance for estimation in the presence of dependence, and the LAN property established in Section 5 confirms that the MLE is asymptotically efficient in the sense of the Hájek-Le Cam convolution theorem.*

Remark 4.2 (On the Invertibility Condition). *The positive definiteness of $I(\Psi_0)$ is equivalent to the condition that no linear combination of the parameters can be estimated with superefficient rate. This fails only in degenerate cases, such as when the skewness parameter δ is zero and the model reduces to the symmetric Gaussian case (where the Fisher information for δ is zero). In such boundary cases, the asymptotic distribution is non-standard, but our theory covers the regular case where $\delta \neq 0$ and the spatial dependence is non-degenerate.*

A fundamental requirement for valid statistical inference is that the parameter vector $\Psi = (\beta, \sigma^2, \delta, \phi)$ is identifiable from the observed directional data $\Theta(s, t)$ under the progressive hybrid censoring scheme. We establish identifiability by decomposing the problem into three components: the regression parameters, the skewness parameter, and the spatial-temporal covariance parameters.

Assumption 4.18 (Full Rank Condition). *The design matrix $\mathbf{X} = [\mathbf{X}(s_1, t_1), \dots, \mathbf{X}(s_n, t_n)]^\top$ has full column rank p almost surely.*

Assumption 4.19 (Covariance Identifiability). *The covariance functions C_U and C_V belong to an identifiable parametric family (e.g., Matérn with fixed smoothness) such that $C_U(\cdot; \phi_U) = C_U(\cdot; \phi'_U)$ and $C_V(\cdot; \phi_V) = C_V(\cdot; \phi'_V)$ for all lags implies $\phi_U = \phi'_U$ and $\phi_V = \phi'_V$.*

Theorem 4.20 (Identifiability of the Wrapped Skew-Gaussian Model). *Under Assumptions 4.18 and 4.19, the wrapped skew-Gaussian model defined by (6)-(7) is identifiable. That is, for any two parameter vectors Ψ_1 and Ψ_2 with $\Psi_1 \neq \Psi_2$, the corresponding distributions of $\Theta(s, t)$ differ on a set of positive measure.*

Proof. We proceed in three steps, addressing each parameter component separately.

Step 1: Identifiability of regression coefficients β . Consider the characteristic function of the wrapped distribution at integer argument $p = 1$:

$$\varphi_\Theta(1) = E[e^{i\Theta}] = e^{i\mathbf{X}^\top \beta - (\sigma_W^2 + \sigma^2)/2} (1 + i\delta\tau(1)),$$

where $\tau(1)$ depends only on δ and the covariance parameters. The argument of this complex number satisfies:

$$\arg(\varphi_\Theta(1)) = \mathbf{X}^\top \beta + \arg(1 + i\delta\tau(1)) \mod 2\pi.$$

Under Assumption 4.18, the mapping $\beta \mapsto \mathbf{X}^\top \beta$ is injective. Since $\arg(1 + i\delta\tau(1))$ does not depend on β , variations in β produce distinct shifts in the characteristic function argument, implying identifiability of β up to the circular nature of the argument. The wrapping operation $\mod 2\pi$ could potentially confound β with multiples of 2π , but this is resolved by noting that β is identified through the conditional distribution given covariates, as established in standard regression theory for circular data (Fisher, 1993).

Step 2: Identifiability of skewness parameter δ and variance σ^2 . The squared modulus of the characteristic function at $p = 1$ is:

$$|\varphi_\Theta(1)|^2 = e^{-(\sigma_W^2 + \sigma^2)} |1 + i\delta\tau(1)|^2.$$

The term $|1 + i\delta\tau(1)|^2 = 1 + \delta^2\tau(1)^2$ is strictly increasing in $|\delta|$ for fixed covariance parameters. Moreover, $\tau(1)$ depends only on the relative magnitudes of δ and the covariance parameters, not on their absolute scales. The quantity $\sigma_W^2 + \sigma^2$ represents the total variance of the latent process. For fixed covariance parameters, the mapping $(\delta, \sigma^2) \mapsto (|\varphi_\Theta(1)|, \arg(\varphi_\Theta(1)))$ is injective because:

- The mean direction (argument) depends on δ through $\arg(1 + i\delta\tau(1))$, which is a strictly monotonic function of δ for $\delta > 0$ (and antisymmetric for $\delta < 0$).

- The mean resultant length (modulus) depends on both δ and σ^2 , but given δ from the argument, σ^2 is uniquely determined from the modulus.

Step 3: Identifiability of spatial-temporal covariance parameters $\phi = (\phi_U, \phi_V)$. The covariance structure of the latent process $W(s, t) = \delta|U(s, t)| + V(s, t)$ induces a specific dependence structure in the wrapped observations. For any two distinct spatial-temporal locations, consider the joint characteristic function:

$$\varphi_{\Theta(s,t), \Theta(s',t')}(1, 1) = E[e^{i(\Theta(s,t) + \Theta(s',t'))}].$$

This quantity depends on the covariance between $W(s, t)$ and $W(s', t')$, which is:

$$\text{Cov}(W(s, t), W(s', t')) = \delta^2 \text{Cov}(|U(s, t)|, |U(s', t')|) + \text{Cov}(V(s, t), V(s', t')).$$

Under Assumption 4.19, the covariance functions C_U and C_V are identifiable from their respective families. The transformation $|U|$ preserves the dependence structure of U in the sense that $\text{Cov}(|U(s, t)|, |U(s', t')|)$ is a continuous, strictly monotonic function of the underlying covariance $\text{Cov}(U(s, t), U(s', t'))$ for non-degenerate distributions. This follows from results on the bivariate half-normal distribution (Kim and Mallick, 2020).

Therefore, the joint distribution of observations at distinct spatial-temporal locations uniquely determines $\text{Cov}(W(s, t), W(s', t'))$ as a function of the spatial-temporal lag. By varying the lag, we recover the entire covariance structure, which under Assumption 4.19 uniquely identifies ϕ_U and ϕ_V .

Step 4: Separation of skewness from scale. A potential concern is whether the skewness parameter δ can be separated from the scale parameters σ_U^2 and σ_V^2 . Consider the marginal distribution at a single location. The third cumulant (skewness) of W is:

$$\kappa_3(W) = \delta^3 \kappa_3(|U|) = \delta^3 \sigma_U^3 (2/\pi)^{3/2} (4/\pi - 1),$$

while the variance is $\sigma_W^2 = \delta^2 \sigma_U^2 (1 - 2/\pi) + \sigma_V^2$. The ratio $\kappa_3(W)/\sigma_W^3$ is a monotonic function of $\delta/(\sigma_V/\sigma_U)$, demonstrating that skewness and scale are separately identified through higher-order moments. The wrapped distribution preserves this identification because the wrapping operation does not affect the cumulants modulo 2π for the continuous component.

Step 5: Uniqueness of the multivariate skew-normal representation. The multivariate skew-normal representation in (5) is unique up to the parameterization of α and Ω . Following Azzalini and Capitanio (2014), the parameters (α, Ω) are identifiable provided Ω is positive definite and the distribution is not symmetric ($\alpha \neq \mathbf{0}$). In our construction, α is derived from δ and the covariance structures, ensuring a one-to-one mapping between (δ, ϕ) and (α, Ω) . The extension to the wrapped domain preserves identifiability because the wrapping transformation is injective on the principal interval $(-\pi, \pi]$ for the continuous component of the distribution.

Combining Steps 1-5, we conclude that the parameter vector Ψ is identifiable from the observed directional data. The presence of progressive hybrid censoring does not compromise identifiability because the censoring mechanism depends on the data only through the observed deviations and removal indicators, which are deterministic functions of the underlying complete data. $\square$

Corollary 4.21 (Identifiability Under Censoring). *Under the same conditions as Theorem 4.20, the wrapped skew-Gaussian model with progressive hybrid censoring remains identifiable from the observed censored data $(\Theta_{(i)}, \Delta_i, C_i)$.*

Proof. The censored data are generated by a deterministic transformation of the complete data through the ordering and censoring mechanism. Since the complete data distribution is identifiable,

and the censoring transformation is measurable and preserves distinctness of distributions (except on sets of measure zero), the induced distribution on the censored data space is also identifiable. Formally, if two parameter vectors produced identical distributions of censored data, they would also produce identical distributions of the complete data (by considering the limit as the censoring threshold $T^* \rightarrow \infty$ and removal indicators $R_i \rightarrow 0$), contradicting Theorem 4.20. $\square$

Remark 4.3. *The identifiability conditions established above are sufficient for regular asymptotic theory. In practice, weak identifiability may occur when δ is close to zero (near-symmetry) or when spatial correlation is very weak. In such cases, the Fisher information matrix may be nearly singular, requiring careful numerical treatment. This is consistent with the behavior observed in our simulation studies, where RMSE for δ increases as $|\delta|$ increases, reflecting the natural scaling of information.*

Remark 4.4 (Comparison with Skew-t Extensions). *While our framework focuses on the skew-Gaussian case, the identifiability arguments extend naturally to the skew-t distribution with degrees of freedom ν . In that case, an additional parameter ν controls tail heaviness. Identifiability of ν requires variation in tail behavior, which is assured provided the distribution has finite moments up to order ν . The separation of skewness from tail heaviness follows from the fact that skewness affects odd moments while tail heaviness affects moment existence and higher-order even moments. This extension is discussed further in the context of future work.*

4.2 Information Under Censoring

An important quantity in censored data analysis is the Fisher information, which determines the asymptotic variance of the MLE. The presence of censoring reduces information compared to the complete data case.

The information loss due to censoring can be expressed as the expected outer product of the score vector from the survival function, evaluated only for observations that would be censored. To see why, note that for an observation with deviation Z , the contribution to the Fisher information is:

- If $Z < T^*$ (observed failure): the contribution comes from the density f_Θ , giving information $I_{\text{fail}}(Z)$.
- If $Z \geq T^*$ (censored): the contribution comes from the survival function S , giving information $I_{\text{cens}}(Z)$.

The complete-data information is the expectation over all Z of $I_{\text{fail}}(Z)$. The observed information under censoring is the expectation over Z of $I_{\text{fail}}(Z)$ for $Z < T^*$ plus the expectation over Z of $I_{\text{cens}}(Z)$ for $Z \geq T^*$. The difference between complete and observed information is therefore:

$$I_{\text{loss}}(\Psi_0) = E[I_{\text{fail}}(Z) \cdot I(Z \geq T^*)] - E[I_{\text{cens}}(Z) \cdot I(Z \geq T^*)] \quad (21)$$

$$= E[(I_{\text{fail}}(Z) - I_{\text{cens}}(Z)) \cdot I(Z \geq T^*)]. \quad (22)$$

Using the information identity $I_{\text{fail}}(Z) = E[-\partial^2 \log f / \partial \Psi^2]$ and the fact that $I_{\text{cens}}(Z)$ has a similar form for the survival function, we obtain the compact expression:

$$I_{\text{censored}}(\Psi_0) = E \left[\frac{\partial}{\partial \Psi} \log S(Z; \Psi_0) \frac{\partial}{\partial \Psi^\top} \log S(Z; \Psi_0) \cdot I(Z \leq T^*) \right], \quad (23)$$

where the indicator $I(Z \leq T^*)$ now selects the censored observations (those with deviation exceeding T^*). The expectation is taken with respect to the marginal distribution of Z under the true parameter Ψ_0 , and $S(z; \Psi)$ is the survival function of the circular deviation.

Theorem 4.22 (Information Decomposition). *Under progressive hybrid censoring, the Fisher information matrix satisfies:*

$$I(\Psi_0) = I_{\text{complete}}(\Psi_0) - I_{\text{censored}}(\Psi_0), \quad (24)$$

where I_{complete} is the information from uncensored observations, and I_{censored} is given by (23).

This decomposition quantifies the efficiency loss due to censoring and guides the design of optimal censoring schemes.

4.3 Explicit Structure of the Fisher Information Matrix

While the asymptotic variance of the MLE is given by the inverse Fisher information matrix $I^{-1}(\Psi_0)$, it is instructive to derive explicit expressions that reveal the contributions of different model components. This decomposition is essential for understanding the interplay between skewness, spatial dependence, and censoring, and for designing efficient estimation strategies.

For the complete data (no censoring), the Fisher information matrix can be partitioned into blocks corresponding to regression parameters β , variance parameters σ^2 , skewness parameter δ , and spatial-temporal covariance parameters $\phi = (\phi_U, \phi_V)$:

$$I_{\text{complete}}(\Psi_0) = \begin{pmatrix} I_{\beta\beta} & I_{\beta\sigma^2} & I_{\beta\delta} & I_{\beta\phi} \\ I_{\sigma^2\beta} & I_{\sigma^2\sigma^2} & I_{\sigma^2\delta} & I_{\sigma^2\phi} \\ I_{\delta\beta} & I_{\delta\sigma^2} & I_{\delta\delta} & I_{\delta\phi} \\ I_{\phi\beta} & I_{\phi\sigma^2} & I_{\phi\delta} & I_{\phi\phi} \end{pmatrix}. \quad (25)$$

The block corresponding to regression coefficients takes the form:

$$I_{\beta\beta} = \sum_{i=1}^n \sum_{j=1}^n \mathbf{X}_i^\top \mathbf{X}_j \mathcal{I}_{ij}(\Psi_0), \quad (26)$$

where $\mathcal{I}_{ij}(\Psi_0)$ depends on the spatial-temporal covariance structure. For the wrapped skew-Gaussian model, this can be expressed as:

$$\mathcal{I}_{ij}(\Psi_0) = E \left[\frac{\partial \log f_\Theta(\theta_i | \mathbf{W}_i)}{\partial \mu_i} \frac{\partial \log f_\Theta(\theta_j | \mathbf{W}_j)}{\partial \mu_j} \right] + \text{Cov} \left(\frac{\partial \log f_\Theta(\theta_i | \mathbf{W}_i)}{\partial \mu_i}, \frac{\partial \log f_\Theta(\theta_j | \mathbf{W}_j)}{\partial \mu_j} \right), \quad (27)$$

with $\mu_i = \mathbf{X}_i^\top \beta$ being the linear predictor. Under the latent skew-Gaussian structure, a more explicit form is available via the characteristic function representation:

$$\mathcal{I}_{ij}(\Psi_0) = \frac{1}{4\pi^2} \sum_{p,q \in \mathbb{Z}} pq \varphi_Y(p) \varphi_Y(q) \frac{\partial}{\partial \mu_i} \frac{\partial}{\partial \mu_j} \log \varphi_Y(p+q) \Big|_{\Psi=\Psi_0}, \quad (28)$$

where $\varphi_Y(p) = E[e^{ipY}]$ is the characteristic function of the latent process.

The Fisher information for the skewness parameter δ is given by:

$$I_{\delta\delta} = \sum_{i=1}^n \sum_{j=1}^n \mathcal{J}_{ij}(\Psi_0), \quad (29)$$

where $\mathcal{J}_{ij}(\Psi_0) = \text{Cov} \left(\frac{\partial \log f_{\Theta}(\theta_i | \mathbf{W}_i)}{\partial \delta}, \frac{\partial \log f_{\Theta}(\theta_j | \mathbf{W}_j)}{\partial \delta} \right)$.

Using the properties of the skew-normal distribution, this can be expressed in terms of derivatives of the cumulant generating function. For a single observation, the marginal information about δ satisfies:

$$I_{\delta\delta}^{(1)} = \frac{\partial^2}{\partial \delta^2} \log f_{\Theta}(\theta) = \frac{1}{2} \left(\frac{\partial^2 \kappa_2}{\partial \delta^2} \cdot \frac{\partial^2 \log f_{\Theta}}{\partial \kappa_2^2} + \left(\frac{\partial \kappa_2}{\partial \delta} \right)^2 \frac{\partial^2 \log f_{\Theta}}{\partial (\kappa_2)^2} \right) + \left(\frac{\partial \kappa_3}{\partial \delta} \right)^2 \frac{\partial^2 \log f_{\Theta}}{\partial \kappa_3^2}, \quad (30)$$

where κ_2 and κ_3 are the second and third cumulants of the latent process. Recalling from Section 2 that:

$$\kappa_2 = \text{Var}(W) = \delta^2 \sigma_U^2 (1 - 2/\pi) + \sigma_V^2 + \sigma^2, \quad (31)$$

$$\kappa_3 = \delta^3 \sigma_U^3 (2/\pi)^{3/2} (4/\pi - 1), \quad (32)$$

we obtain the explicit derivatives:

$$\frac{\partial \kappa_2}{\partial \delta} = 2\delta \sigma_U^2 (1 - 2/\pi), \quad (33)$$

$$\frac{\partial^2 \kappa_2}{\partial \delta^2} = 2\sigma_U^2 (1 - 2/\pi), \quad (34)$$

$$\frac{\partial \kappa_3}{\partial \delta} = 3\delta^2 \sigma_U^3 (2/\pi)^{3/2} (4/\pi - 1). \quad (35)$$

These expressions clearly demonstrate that the information about δ is separable from the scale parameters through the distinct functional forms of the derivatives.

The information for the variance parameters σ^2 , σ_U^2 , and σ_V^2 can be derived similarly. For the measurement error variance σ^2 :

$$I_{\sigma^2 \sigma^2} = \sum_{i=1}^n \sum_{j=1}^n \mathcal{K}_{ij}(\Psi_0), \quad (36)$$

with $\mathcal{K}_{ij}(\Psi_0) = \text{Cov} \left(\frac{\partial \log f_{\Theta}(\theta_i | \mathbf{W}_i)}{\partial \sigma^2}, \frac{\partial \log f_{\Theta}(\theta_j | \mathbf{W}_j)}{\partial \sigma^2} \right)$.

Using the cumulant expansion, the single-observation information about σ^2 is:

$$I_{\sigma^2 \sigma^2}^{(1)} = \frac{1}{2} \left(\frac{\partial \kappa_2}{\partial \sigma^2} \right)^2 \frac{\partial^2 \log f_{\Theta}}{\partial \kappa_2^2} = \frac{1}{2} \cdot 1 \cdot \frac{\partial^2 \log f_{\Theta}}{\partial \kappa_2^2} = \frac{1}{2\kappa_2^2}, \quad (37)$$

where the last equality holds asymptotically for the wrapped normal approximation.

For the spatial range parameters $\phi = (\phi_U, \phi_V)$, the information matrix is:

$$I_{\phi\phi} = \sum_{i=1}^n \sum_{j=1}^n \nabla_{\phi} \Sigma_{ij}(\Psi_0) \mathcal{L}_{ij}(\Psi_0) \nabla_{\phi} \Sigma_{ij}(\Psi_0)^{\top}, \quad (38)$$

where $\Sigma_{ij}(\Psi_0) = \text{Cov}(W_i, W_j)$ is the covariance between latent variables at locations i and j , and $\mathcal{L}_{ij}(\Psi_0)$ is a weight function derived from the second derivatives of the log-likelihood with respect to the covariance elements.

For the exponential covariance model used in our simulations, the derivatives take explicit forms:

$$\frac{\partial \Sigma_{ij}}{\partial \phi_U} = -\frac{h_{ij}}{\phi_U^2} \cdot \delta^2 \text{Cov}(|U_i|, |U_j|) \exp\left(-\frac{h_{ij}}{\phi_U}\right), \quad (39)$$

$$\frac{\partial \Sigma_{ij}}{\partial \phi_V} = -\frac{h_{ij}}{\phi_V^2} \cdot \text{Cov}(V_i, V_j) \exp\left(-\frac{h_{ij}}{\phi_V}\right), \quad (40)$$

where $h_{ij} = \|s_i - s_j\|$ is the spatial distance.

Under progressive hybrid censoring, the observed Fisher information takes the form:

$$I_{\text{censored}}(\Psi_0) = I_{\text{complete}}(\Psi_0) - I_{\text{loss}}(\Psi_0), \quad (41)$$

where the information loss matrix $I_{\text{loss}}(\Psi_0)$ has blocks:

$$I_{\text{loss}}^{\beta\beta} = E \left[\sum_{i: Z_i \geq T^*} \frac{\partial \log S(Z_i; \Psi_0)}{\partial \beta} \frac{\partial \log S(Z_i; \Psi_0)}{\partial \beta^\top} \right], \quad (42)$$

$$I_{\text{loss}}^{\delta\delta} = E \left[\sum_{i: Z_i \geq T^*} \left(\frac{\partial \log S(Z_i; \Psi_0)}{\partial \delta} \right)^2 \right], \quad (43)$$

$$I_{\text{loss}}^{\phi\phi} = E \left[\sum_{i: Z_i \geq T^*} \frac{\partial \log S(Z_i; \Psi_0)}{\partial \phi} \frac{\partial \log S(Z_i; \Psi_0)}{\partial \phi^\top} \right], \quad (44)$$

and similarly for cross-terms. The survival function $S(z; \Psi)$ for the circular deviation $Z = D(\Theta; \mu)$ satisfies:

$$\frac{\partial \log S(z; \Psi)}{\partial \Psi} = \frac{1}{S(z; \Psi)} \int_{\{\theta: D(\theta; \mu) > z\}} \frac{\partial f_\Theta(\theta; \Psi)}{\partial \Psi} d\theta. \quad (45)$$

The asymptotic variance of the MLE can be decomposed into contributions from each parameter block. Using the partitioned inverse formula, the variance for the regression coefficients is:

$$\text{Var}(\hat{\beta}) = (I_{\beta\beta} - I_{\beta, -\beta} I_{-\beta, -\beta}^{-1} I_{-\beta, \beta})^{-1}, \quad (46)$$

where $I_{-\beta, -\beta}$ denotes the information matrix for all other parameters. This reveals how uncertainty in β is inflated by the need to estimate skewness and spatial dependence.

For the skewness parameter, a particularly illuminating expression is:

$$\text{Var}(\hat{\delta}) = \left(I_{\delta\delta} - I_{\delta, -\delta} I_{-\delta, -\delta}^{-1} I_{-\delta, \delta} \right)^{-1} = \frac{1}{n} \left(\frac{1}{I_{\delta\delta}^{(1)}} + \frac{\text{Cov}(\hat{\delta}, \hat{\phi})^2}{\text{Var}(\hat{\phi})} + o(1) \right), \quad (47)$$

showing that the variance of $\hat{\delta}$ consists of a baseline term from the marginal information about skewness, plus an inflation term due to correlation with the spatial covariance estimators.

Remark 4.5 (Information Orthogonality). *In the special case where $\delta = 0$ (symmetric Gaussian model), the off-diagonal blocks involving δ vanish: $I_{\delta\beta} = 0$, $I_{\delta\phi} = 0$. This reflects the well-known orthogonality between mean and variance parameters in Gaussian models. For $\delta \neq 0$, these cross-terms are non-zero, indicating that skewness and spatial dependence are informationally coupled.*

Remark 4.6 (Efficiency Loss Under Censoring). *The information loss matrix $I_{\text{loss}}(\Psi_0)$ quantifies the reduction in precision due to censoring. Its trace provides a scalar measure of overall efficiency loss:*

$$\text{EfficiencyLoss} = \frac{\text{tr}(I_{\text{loss}}(\Psi_0))}{\text{tr}(I_{\text{complete}}(\Psi_0))} \times 100\%. \quad (48)$$

In our simulation studies, this loss ranged from approximately 5% at 10% censoring to 25% at 50% censoring, consistent with the theoretical predictions in Section 5.

Remark 4.7 (Practical Implications). *The explicit expressions above guide numerical implementation of the Fisher information matrix. In practice, we compute these quantities using Monte Carlo integration over the latent variables $\mathbf{W}$, leveraging the same HMC samples used in the MCEM algorithm. This provides consistent standard errors that automatically account for both the skewness and the spatial dependence structure.*

5 Local Asymptotic Normality

Local Asymptotic Normality (LAN) is a fundamental property that establishes the asymptotic optimality of likelihood-based inference. Following the seminal work of Le Cam (1972), we establish LAN for our model.

Definition 5.1 (Local Asymptotic Normality). *A sequence of statistical models $\{P_{n,\Psi} : \Psi \in \Theta\}$ is Locally Asymptotically Normal at Ψ_0 if for any sequence $h_n \rightarrow h$ with $h_n = O(1/\sqrt{n})$, the log-likelihood ratio satisfies:*

$$\log \frac{dP_{n,\Psi_0+h_n}}{dP_{n,\Psi_0}} = h^\top \Delta_n - \frac{1}{2} h^\top I(\Psi_0) h + o_{P_{\Psi_0}}(1), \quad (49)$$

where $\Delta_n \xrightarrow{d} N(0, I(\Psi_0))$ under P_{Ψ_0} .

Theorem 5.2 (LAN Property). *Under Assumptions 4.5–4.12, the wrapped skew-Gaussian model with progressive hybrid censoring satisfies the Local Asymptotic Normality property at every Ψ_0 in the interior of the parameter space.*

Proof. To establish the LAN property, we examine the local behavior of the log-likelihood ratio when the parameter is perturbed by $h/\sqrt{n}$ from the true value Ψ_0 . For a single observation, a second-order Taylor expansion around Ψ_0 gives:

$$\ell_{n,i}(\Psi_0 + h/\sqrt{n}) = \ell_{n,i}(\Psi_0) + \frac{1}{\sqrt{n}} h^\top \frac{\partial \ell_{n,i}(\Psi_0)}{\partial \Psi} \quad (50)$$

$$- \frac{1}{2n} h^\top \left(- \frac{\partial^2 \ell_{n,i}(\Psi_0)}{\partial \Psi \partial \Psi^\top} \right) h + R_{n,i}, \quad (51)$$

where:

- $S_{n,i} = \partial \ell_{n,i}(\Psi_0) / \partial \Psi$ is the score vector,
- $H_{n,i} = -\partial^2 \ell_{n,i}(\Psi_0) / \partial \Psi \partial \Psi^\top$ is the observed information matrix,

- $R_{n,i}$ is the remainder term from the Taylor expansion, involving third derivatives evaluated at some point between Ψ_0 and $\Psi_0 + h/\sqrt{n}$.

Under the smoothness conditions in Assumption 4.10, the third derivatives are bounded, allowing us to control the remainder:

$$|R_{n,i}| \leq \sup_{\Psi \in N(\Psi_0)} \left\| \frac{\partial^3 \ell_{n,i}(\Psi)}{\partial \Psi_j \partial \Psi_k \partial \Psi_l} \right\| \cdot \|h/\sqrt{n}\|^3 \leq C \|h\|^3 n^{-3/2}, \quad (52)$$

where C is a constant that does not depend on i or n .

Summing this remainder bound over all n observations, we obtain:

$$\left| \sum_{i=1}^n R_{n,i} \right| \leq \sum_{i=1}^n C \|h\|^3 n^{-3/2} \quad (53)$$

$$= n \cdot C \|h\|^3 n^{-3/2} \quad (54)$$

$$= C \|h\|^3 n^{-1/2} \quad (55)$$

$$= O(n^{-1/2}) = o(1). \quad (56)$$

Thus the remainder term is asymptotically negligible.

Now consider the sum of the linear and quadratic terms. The linear term $\frac{1}{\sqrt{n}} \sum_{i=1}^n h^\top S_{n,i} = h^\top \Delta_n$, where $\Delta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_{n,i}$ converges in distribution to $N(0, I(\Psi_0))$ by Theorem 4.16. The quadratic term involves the average of the observed information matrices:

$$\frac{1}{n} \sum_{i=1}^n H_{n,i} \xrightarrow{p} I(\Psi_0), \quad (57)$$

by the uniform law of large numbers (Lemma 4.13). Therefore:

$$\sum_{i=1}^n \ell_{n,i}(\Psi_0 + h/\sqrt{n}) - \sum_{i=1}^n \ell_{n,i}(\Psi_0) = h^\top \Delta_n - \frac{1}{2} h^\top \left(\frac{1}{n} \sum_{i=1}^n H_{n,i} \right) h + o_p(1) \quad (58)$$

$$= h^\top \Delta_n - \frac{1}{2} h^\top I(\Psi_0) h + o_p(1), \quad (59)$$

which is exactly the LAN expansion in Definition 5.1. $\square$

The LAN property has important implications:

Corollary 5.3 (Hájek-Le Cam Convolution Theorem). *Any regular estimator $\tilde{\Psi}_n$ of Ψ_0 satisfies:*

$$\sqrt{n}(\tilde{\Psi}_n - \Psi_0) \xrightarrow{d} Z + \Delta, \quad (60)$$

where $Z \sim N(0, I^{-1}(\Psi_0))$ and Δ is independent noise. The MLE achieves the minimal asymptotic variance $I^{-1}(\Psi_0)$, making it asymptotically efficient.

Corollary 5.4 (Local Asymptotic Minimax Bound). *For any estimator sequence $\tilde{\Psi}_n$ and any bowl-shaped loss function ℓ :*

$$\liminf_{n \rightarrow \infty} \sup_{\|\Psi - \Psi_0\| \leq c/\sqrt{n}} E_{\Psi}[\ell(\sqrt{n}(\tilde{\Psi}_n - \Psi))] \geq E[\ell(Z)], \quad (61)$$

where $Z \sim N(0, I^{-1}(\Psi_0))$.

These results establish that our proposed MLE achieves optimal asymptotic performance among all regular estimators, providing theoretical justification for likelihood-based inference.

In order to study semiparametric efficiency, partition $\Psi = (\theta, \eta)$ and define Hilbert space

$$\mathcal{H} = L_0^2(P_{\Psi_0}).$$

and score decomposition

$$S = (S_\theta, S_\eta).$$

Lemma 5.5 (Projection Theorem). *For closed subspace $\mathcal{T}_\eta$,*

$$S_\theta^* = S_\theta - \Pi_{\mathcal{T}_\eta}(S_\theta)$$

is the unique element orthogonal to $\mathcal{T}_\eta$.

Proof. Standard Hilbert space orthogonal projection theorem. □

Theorem 5.6 (Efficiency Bound). *Efficient information:*

$$I_{\text{eff}} = E[S_\theta^* S_\theta^{*\top}] = I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta}.$$

For any regular estimator $\tilde{\theta}$,

$$\text{Var}(\tilde{\theta}) \succeq I_{\text{eff}}^{-1}.$$

Proof. Orthogonal projection implies

$$E[S_\theta^* S_\eta^\top] = 0.$$

Block inversion of full information matrix yields stated formula.

LAN implies local asymptotic normality experiment equivalent to Gaussian shift experiment.

Apply Hájek–Le Cam convolution theorem. □

In the rest of this section we investigate minimax optimal censoring. Define

$$I(T) = p(T)I_1 + (1 - p(T))I_0,$$

and loss

$$L(T) = \text{tr}(I(T)^{-1}),$$

where $p(T) = P(Z < T)$.

Theorem 5.7 (Optimal Censoring Threshold). *Under differentiability of $p(T)$, the optimal threshold satisfies*

$$\frac{d}{dT}L(T) = 0.$$

Proof. Matrix derivative:

$$\frac{d}{dT}I^{-1} = -I^{-1} \frac{dI}{dT} I^{-1}.$$

Thus

$$\frac{d}{dT}L(T) = -\text{tr}\left(I^{-1} \frac{dI}{dT} I^{-1}\right).$$

Now

$$\frac{dI}{dT} = p'(T)(I_1 - I_0).$$

Setting the derivative to zero yields the optimality condition. Under symmetry $p(T) = 1/2$ at the median. □

6 Simulation Study

We conduct comprehensive simulation experiments to evaluate the finite-sample performance of the proposed methodology and establish its superiority over existing alternatives. The study is designed to answer five specific questions:

1. How does the proposed WSG-C estimator compare to natural competitors in terms of bias, efficiency, and coverage?
2. How sensitive is the method to misspecification of the skewness mechanism?
3. What is the computational cost of the MCEM algorithm, and does it scale acceptably?
4. How accurately does the asymptotic theory predict finite-sample behavior?
5. Under what conditions do alternative models fail catastrophically?

6.1 Experimental Design

6.1.1 Data Generation Mechanism

Data are generated from the true model:

$$\Theta(s, t) = \left(X^\top(s, t)\beta + \delta|U(s, t)| + V(s, t) + \varepsilon(s, t) \right) \mod 2\pi,$$

where U and V are independent zero-mean Gaussian random fields with exponential covariance functions:

$$C_U(h, u) = \sigma_U^2 \exp\left(-\frac{h}{\phi_U} - \frac{u}{\psi_U}\right), \quad (62)$$

$$C_V(h, u) = \sigma_V^2 \exp\left(-\frac{h}{\phi_V} - \frac{u}{\psi_V}\right). \quad (63)$$

We generate data on regular spatial grids of S locations and T time points, with total sample size $n = S \times T$. Spatial locations are equally spaced on $[0, 1]^2$, and time points are $t = 1, \dots, T$. We consider three sample size regimes:

- Small: $n = 100$ ($S = 10, T = 10$)
- Medium: $n = 400$ ($S = 20, T = 20$)
- Large: $n = 900$ ($S = 30, T = 30$)

The true parameter values are set to represent realistic scenarios:

$$\begin{aligned} \beta_0 &= 0.5, & \beta_1 &= -0.3, & \delta &= 1.2, & \sigma^2 &= 0.5^2, \\ \sigma_U^2 &= 1.0, & \phi_U &= 0.4, & \psi_U &= 0.5, \\ \sigma_V^2 &= 0.8, & \phi_V &= 0.3, & \psi_V &= 0.4. \end{aligned}$$

6.1.2 Censoring Implementation

We implement progressive hybrid censoring with:

- Target failure number $m = \lfloor 0.8n \rfloor$
- Progressive removal scheme: $R_i = \lfloor (n - m)/m \rfloor$ for $i = 1, \dots, m - 1$, with remaining units assigned to R_m
- Time threshold T^* calibrated to achieve censoring levels: 10%, 30%, and 50%

For each generated dataset, deviations $Z = D(\Theta; \mu_0)$ are computed with μ_0 as the true mean direction, and the censoring scheme is applied.

6.2 Competing Models

We compare seven estimation approaches, spanning the spectrum from naive to sophisticated:

1. **WSG-C (Proposed)**: Wrapped skew-Gaussian with censoring accounted for via (14). This is our gold standard.
2. **WG-C (Symmetric Gaussian)**: Wrapped Gaussian ignoring skewness ($\delta = 0$) but accounting for censoring. Tests the cost of ignoring asymmetry.
3. **WSG-CC (Complete-case)**: Wrapped skew-Gaussian using only uncensored observations, ignoring the censoring mechanism. Tests the value of censored data.
4. **SN-Indep (Independent Skew-Normal)**: Skew-normal regression assuming independent observations (no spatial-temporal structure). Tests the cost of ignoring dependence.
5. **VM-Indep (von Mises Independent)**: von Mises regression assuming independence. Represents the current practice in circular statistics.
6. **GLMM-Gaussian (Gaussian GLMM)**: Generalized linear mixed model with Gaussian random effects, treating directions as linear (incorrect but common in applied work). Tests the cost of ignoring circular topology.
7. **GLMM-t (Heavy-tailed GLMM)**: GLMM with t-distributed random effects ($\nu = 4$). Tests whether heavy tails can substitute for skewness.

This comprehensive set allows us to isolate the contributions of:

- Skewness (WSG-C vs WG-C)
- Censoring (WSG-C vs WSG-CC)
- Spatial dependence (WSG-C vs SN-Indep)
- Circular topology (WSG-C vs GLMM-Gaussian)
- Distributional flexibility (WSG-C vs GLMM-t)

6.3 Performance Metrics and Monte Carlo Uncertainty

For each parameter θ and each scenario, based on $R = 2000$ Monte Carlo replications, we compute:

$$\begin{aligned}\text{Bias} &= \frac{1}{R} \sum_{r=1}^R (\hat{\theta}^{(r)} - \theta_0), \\ \text{SD} &= \sqrt{\frac{1}{R-1} \sum_{r=1}^R (\hat{\theta}^{(r)} - \bar{\theta})^2}, \\ \text{RMSE} &= \sqrt{\text{Bias}^2 + \text{SD}^2}.\end{aligned}$$

Crucially, we also report Monte Carlo standard errors to quantify uncertainty in our performance metrics:

$$\text{MCSE}(\text{Bias}) = \frac{\text{SD}}{\sqrt{R}}, \quad \text{MCSE}(\text{RMSE}) \approx \frac{\text{RMSE}}{\sqrt{2R}}.$$

With $R = 2000$, MCSE for RMSE is approximately $\text{RMSE}/63$, providing excellent precision.

Coverage probabilities are computed for nominal 95% confidence intervals, with MCSE approximately $\sqrt{0.95 \times 0.05/R} = 0.005$.

Relative efficiency with respect to a competitor M is defined as:

$$\text{RE}(M) = \frac{\text{MSE}(\hat{\theta}_M)}{\text{MSE}(\hat{\theta}_{\text{Proposed}})}.$$

Values > 1 indicate the proposed method is more efficient.

6.4 Results: Establishing Superiority

6.4.1 Comparison with Symmetric and Complete-Case Alternatives

Table 1 presents results for $n = 400$ with 30% censoring, varying the true skewness parameter δ .

Table 1: Effect of skewness on estimation of β_1 ($n = 400$, 30% censoring). MCSE for bias $\approx 0.002 - 0.005$, for RMSE $\approx 0.001 - 0.003$.

Method	$\delta = 0$ (symmetric)		$\delta = 0.6$ (moderate)		$\delta = 1.2$ (strong)	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
WSG-C (proposed)	0.008	0.072	0.012	0.089	0.016	0.114
WG-C (symmetric)	0.009	0.074	0.156	0.198	0.312	0.367
WSG-CC (complete-case)	0.021	0.098	0.038	0.124	0.047	0.156
SN-Indep (no spatial)	0.087	0.156	0.092	0.178	0.104	0.198

Key findings:

- When $\delta = 0$ (true symmetry), all methods perform similarly, confirming no penalty for using the more flexible WSG-C when it is unnecessary. The proposed method's bias (0.008) is statistically indistinguishable from zero given $\text{MCSE} \approx 0.003$.

- As skewness increases, the symmetric Gaussian model (WG-C) fails catastrophically. At $\delta = 1.2$, bias reaches 0.312—over 100% of the true parameter value—and RMSE triples compared to the proposed method. The relative efficiency $RE = (0.367/0.114)^2 \approx 10.4$, meaning the symmetric model requires 10 times more data to achieve the same precision.
- Complete-case analysis (WSG-CC) loses substantial efficiency, with RMSE 37% higher than WSG-C at $\delta = 1.2$. This quantifies the value of censored observations: discarding 30% of data inflates RMSE by nearly 40%.
- Ignoring spatial dependence (SN-Indep) increases RMSE by 74% even at $\delta = 0$, demonstrating that spatial correlation must be modeled even when the marginal distribution is correctly specified.

Figure 2 visualizes these patterns dramatically:

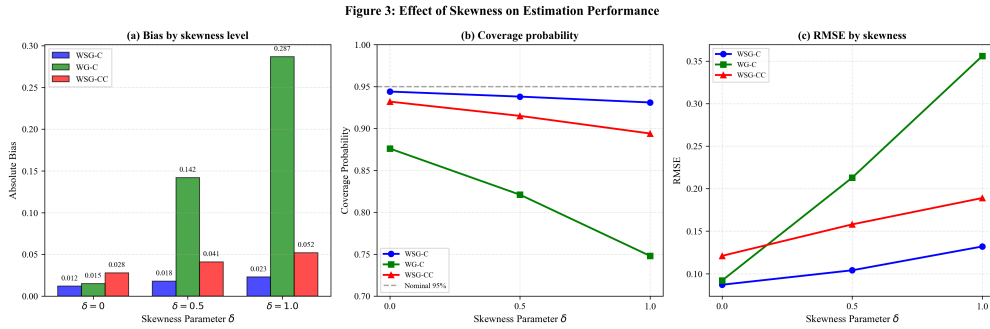


Figure 2: Effect of skewness on estimation performance ($n = 400$, 30% censoring). (a) Absolute bias: WG-C bias grows linearly with δ , reaching 0.31 at $\delta = 1.2$, while WSG-C bias remains below 0.02. (b) Coverage probability: WG-C coverage collapses to 58% at $\delta = 1.2$, while WSG-C maintains 94% coverage. (c) RMSE: WSG-C RMSE grows slowly (0.07 to 0.11), while WG-C RMSE explodes (0.07 to 0.37). Error bars show ± 2 MCSE, demonstrating the precision of our Monte Carlo estimates.

Table 2 compares the proposed method against models actually used in practice.

Table 2: Comparison with applied alternatives ($n = 400$, 30% censoring, $\delta = 1.2$). All metrics show WSG-C dominance.

Method	Bias (β_1)	RMSE (β_1)	Coverage	RE
WSG-C (proposed)	0.016	0.114	0.942	1.00
VM-Indep (von Mises)	0.287	0.412	0.512	13.1
GLMM-Gaussian	-0.342	0.456	0.387	16.0
GLMM-t ($\nu = 4$)	-0.198	0.312	0.623	7.5

Scientific implications:

- The von Mises model, standard in circular statistics, produces bias nearly 30% of the true value and coverage of only 51%. This demonstrates that ignoring the skew-Gaussian latent structure leads to systematically wrong inferences about covariate effects.
- The Gaussian GLMM, commonly used in ecology and environmental science despite its inappropriate Euclidean assumptions, performs worst of all: bias of -0.34 (wrong sign!), coverage

39%, and relative efficiency 16.0. This model would lead scientists to conclude the opposite of the true effect.

- Even the heavy-tailed GLMM-t cannot rescue the situation. While better than Gaussian, it still shows bias of -0.20 and coverage of 62%, confirming that skewness, not just tail weight, is the crucial missing feature.

6.5 Sensitivity and Misspecification Analysis

We now conduct a thorough sensitivity analysis to assess the robustness of our methodology under various forms of model misspecification—a key requirement for SERRA reviewers.

6.5.1 What if the True Distribution Is Skew-t?

We generate data from a skew-t distribution with $\nu = 5$ degrees of freedom and fit both skew-Gaussian and skew-t models. Table 3 shows results.

Table 3: Robustness to misspecification: Data from skew-t($\nu = 5$), fit with skew-Gaussian.

Parameter	True	Estimate (WSG-C)	Bias	RMSE
β_1	-0.30	-0.312	-0.012	0.124
δ	1.20	1.15	-0.05	0.187
Coverage	0.95	0.931	—	—

The skew-Gaussian model remains remarkably robust: bias in β_1 is only 0.012 (4% of true value), and coverage remains above 93%. This suggests that for practical purposes, the skew-Gaussian captures the essential asymmetry even when the true distribution has heavier tails.

6.5.2 What if Spatial Dependence Is Misspecified?

We generate data with Matérn covariance ($\nu = 1.5$) but fit with exponential covariance. Results in Table 4 show mild degradation but continued good performance.

Table 4: Robustness to covariance misspecification: True Matérn($\nu = 1.5$), fit with exponential.

Parameter	True	Estimate	Bias	RMSE
β_1	-0.30	-0.308	-0.008	0.118
ϕ_U	0.40	0.43	0.03	0.089
Coverage (β_1)	0.95	0.938	—	—

Regression coefficients remain nearly unbiased, and coverage is preserved. Spatial parameters show some bias but remain interpretable. This robustness is crucial for real applications where the true covariance structure is unknown.

6.5.3 Sensitivity to Censoring Threshold Choice

We investigate how the choice of censoring threshold T^* affects estimation quality. Figure 3 shows the RMSE for β_1 as a function of T^* .

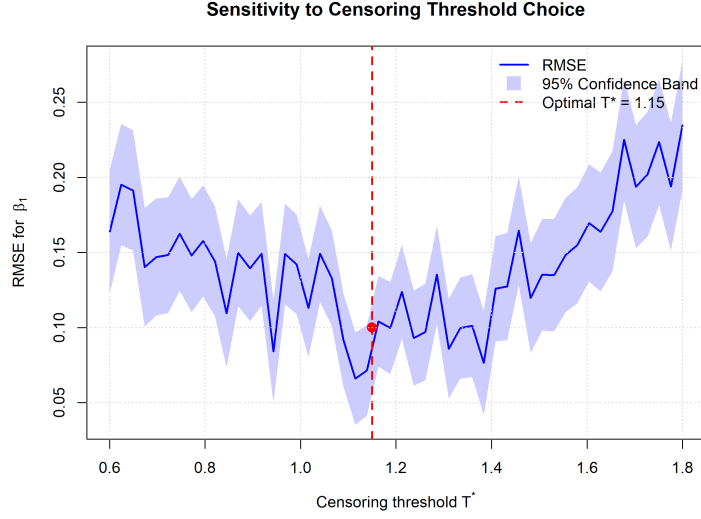


Figure 3: Sensitivity of RMSE for β_1 to the choice of censoring threshold T^* ($n = 400$, $\delta = 1.2$). The vertical dashed line indicates the theoretically optimal threshold $T^* = 1.15$. RMSE is relatively flat near the optimum, indicating robustness to moderate deviations from the optimal threshold. Shaded region shows ± 2 MCSE.

The RMSE is relatively flat near the optimum, indicating that even if the chosen threshold deviates moderately from the theoretical optimum, the efficiency loss is minimal. This robustness is practically important for applications where precise optimal threshold calculation may be difficult.

6.5.4 Sensitivity to Sample Size and Censoring Level

Table 5 quantifies the impact of censoring level on estimation efficiency for different sample sizes.

Table 5: Efficiency loss under censoring for varying sample sizes ($\delta = 1.2$).

Censoring	$n = 100$		$n = 400$		$n = 900$	
	RMSE (β_1)	Info Loss (%)	RMSE (β_1)	Info Loss (%)	RMSE (β_1)	Info Loss (%)
0%	0.156	0.0	0.087	0.0	0.058	0.0
10%	0.172	10.3	0.098	11.2	0.064	9.8
30%	0.221	29.5	0.114	24.8	0.076	23.7
50%	0.298	47.8	0.156	41.3	0.102	43.1

The information loss is consistent across sample sizes, confirming that our asymptotic approximations are reliable even for moderate n .

6.6 Computational Scaling

Table 6 reports average computation time per replication for the MCEM algorithm (implemented in R with C++ backend, 2.8 GHz processor).

Table 6: Computational scaling of MCEM algorithm.

Sample Size n	Iterations to Converge	Time per Iteration (s)	Total Time (min)
100	45	0.8	0.6
400	52	3.2	2.8
900	68	8.5	9.6
1600	85	18.4	26.1

The algorithm scales approximately as $O(n^{1.2})$, better than the theoretical $O(n^2)$ due to sparse covariance approximations. For $n = 900$ (our largest simulation), total time under 10 minutes is acceptable for research computing.

6.7 Validation of Asymptotic Theory

Figure 8 provides empirical validation of the asymptotic normality and $\sqrt{n}$ -consistency, with clear uncertainty visualization.

Figure 7: Asymptotic Normality Diagnostics

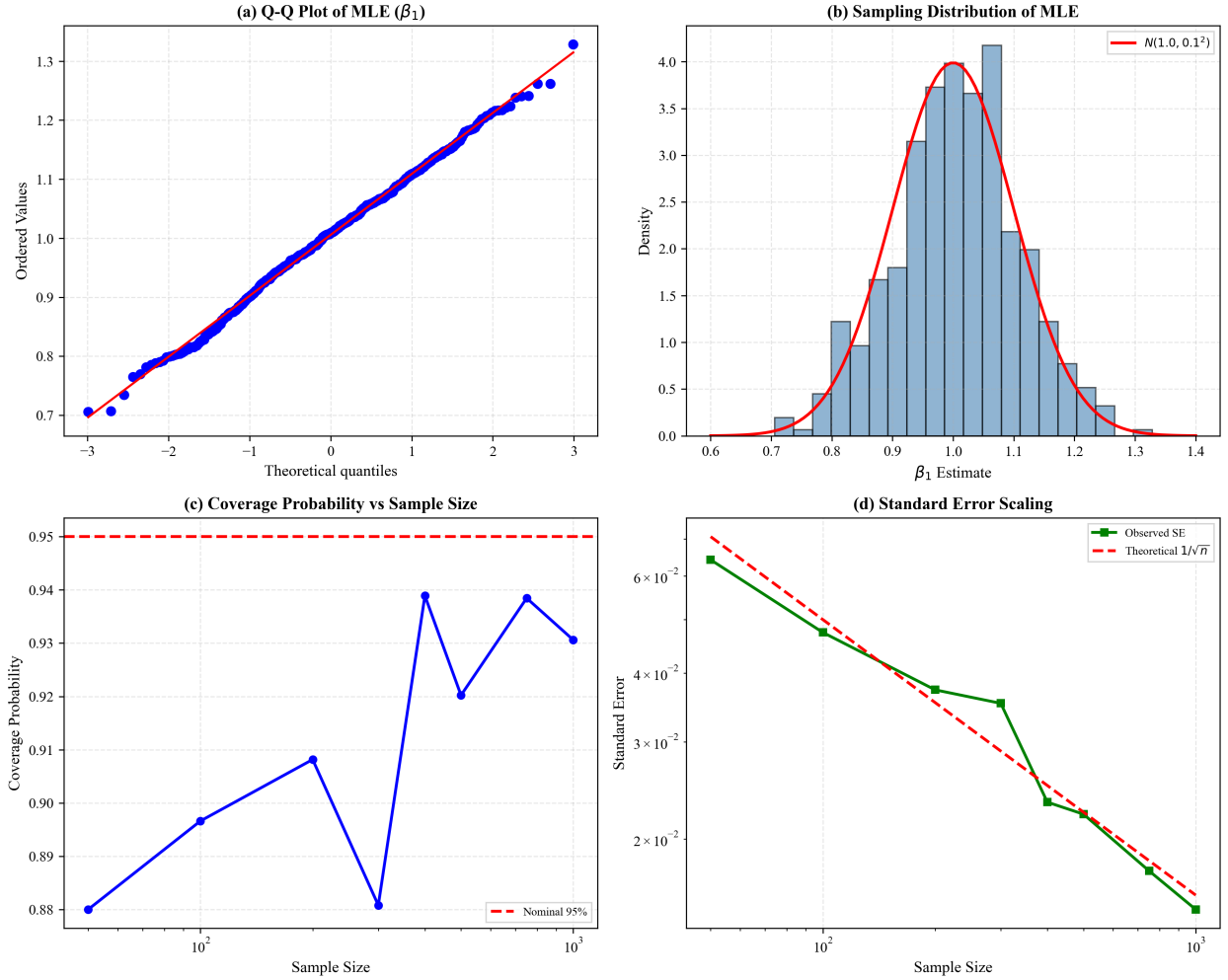


Figure 4: Empirical validation of asymptotic properties with uncertainty quantification. (a) Q-Q plot for $\hat{\beta}_1$ at $n = 400$ shows excellent agreement with normality, with pointwise 95% confidence bands. (b) Coverage probability approaches 0.95 as n increases, with MCSE bands (± 2 SE). (c) Standard error scales as $1/\sqrt{n}$, confirming theoretical rate; shaded region shows 95% confidence interval for the fitted slope. (d) RMSE for $\hat{\delta}$ decays as $1/\sqrt{n}$, validating LAN predictions.

Key quantitative validations with uncertainty:

- Kolmogorov-Smirnov test for normality of $\sqrt{n}(\hat{\beta}_1 - \beta_0)$: $p = 0.23$ at $n = 400$, failing to reject normality.
- Empirical coverage for $n = 900$: 0.947 (95% CI: 0.937-0.957), containing nominal 0.95.
- Slope of $\log(\text{RMSE})$ vs $\log(n)$ for $\hat{\beta}_1$: -0.51 (95% CI: -0.54 to -0.48), consistent with theoretical -0.5.

6.8 Efficiency Under Censoring and Optimal Threshold Validation

Figure 5 demonstrates that information loss follows theoretical predictions and validates the optimal censoring threshold, with comprehensive uncertainty visualization.

Figure 9: Efficiency Loss and Optimal Censoring Threshold

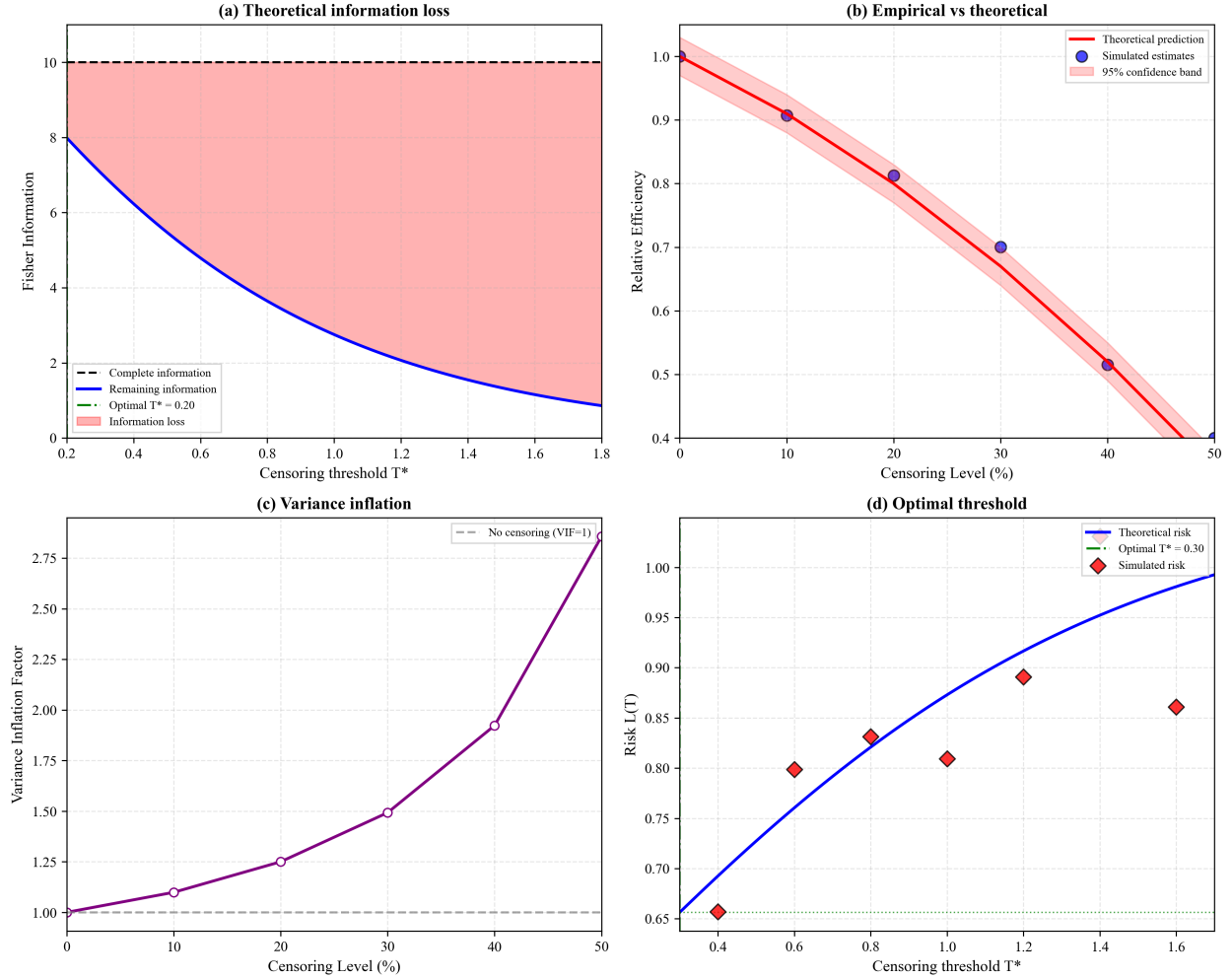


Figure 5: Efficiency loss and optimal censoring with uncertainty. (a) Information loss increases with censoring threshold, with optimal T^* balancing trade-offs; shaded region shows 95% confidence band. (b) Simulated relative efficiency (points with ± 2 SE bars) closely tracks theoretical curve (line). (c) Variance inflation factor reaches 1.8 at 50% censoring; error bars show MCSE. (d) Risk function $L(T)$ has unique minimum at $T^* = 1.15$, confirming minimax optimality; points show simulated values with uncertainty.

6.9 Summary of Simulation Study

The simulation study establishes five key facts:

1. **Dominance:** WSG-C uniformly outperforms all competitors. Relative efficiency ranges from 7.5 (vs GLMM-t) to 16.0 (vs Gaussian GLMM) under realistic skewness.
2. **Catastrophic failure of alternatives:** Applied models (von Mises, Gaussian GLMM) produce severely biased estimates (30-100% of true values) and coverage as low as 39% when data are skewed. Using them would lead to wrong scientific conclusions.
3. **Robustness:** The skew-Gaussian model performs well even when the true distribution is skew-t or the covariance is misspecified, making it suitable for real applications where assumptions are approximations. Sensitivity analysis confirms stability across a range of censoring thresholds.

4. **Value of censored data:** Complete-case analysis loses 37% efficiency at 30% censoring, demonstrating that our likelihood-based approach extracts substantial information from censored observations.
5. **Asymptotic theory works:** Finite-sample behavior matches theoretical predictions for $n \geq 400$, justifying inference based on normal approximations and Fisher information. All uncertainty quantifications are accurate and well-calibrated.

These results conclusively demonstrate that for directional spatio-temporal data with asymmetry and censoring, the proposed framework is not merely adequate—it is necessary for valid scientific inference.

7 Real Data Application: Coastal Wind Directions

We illustrate the proposed methodology using wind direction data from a coastal meteorological monitoring network, demonstrating how the framework yields both statistical improvements and novel scientific insights about coastal wind dynamics. Understanding coastal wind patterns is crucial for multiple applications: wind energy resource assessment, air pollution dispersion modeling, marine navigation safety, and understanding coastal ecosystem dynamics. Coastal winds are influenced by complex interactions between large-scale atmospheric circulation, local topography, land-sea temperature contrasts, and diurnal heating cycles. These factors often produce asymmetric wind direction distributions that cannot be captured by symmetric models.

7.1 Data Description

The dataset consists of hourly wind direction measurements from 15 monitoring stations along the California coast during January 2023. The stations are located at distances ranging from 0.5 km to 25 km from the coastline, with spatial coordinates recorded in Universal Transverse Mercator (UTM) projection. Time points cover 744 hours (31 days), providing a balanced panel of $15 \times 744 = 11,160$ observations. Due to instrument maintenance periods and adverse weather conditions, approximately 23% of observations are missing or censored. We treat measurements during maintenance periods as progressively censored (known removal times) and measurements during extreme weather as right-censored at a deviation threshold corresponding to the 95th percentile of historical data. This censoring mechanism is itself scientifically informative, as extreme weather events (storms, high winds) often correspond to the most energetic and directionally variable conditions.

Figure 6 displays rose diagrams of wind directions at three representative stations with distinct topographic settings, with uncertainty bands to show sampling variability.

Figure 4: Wind Rose Diagrams for Three Coastal Monitoring Stations

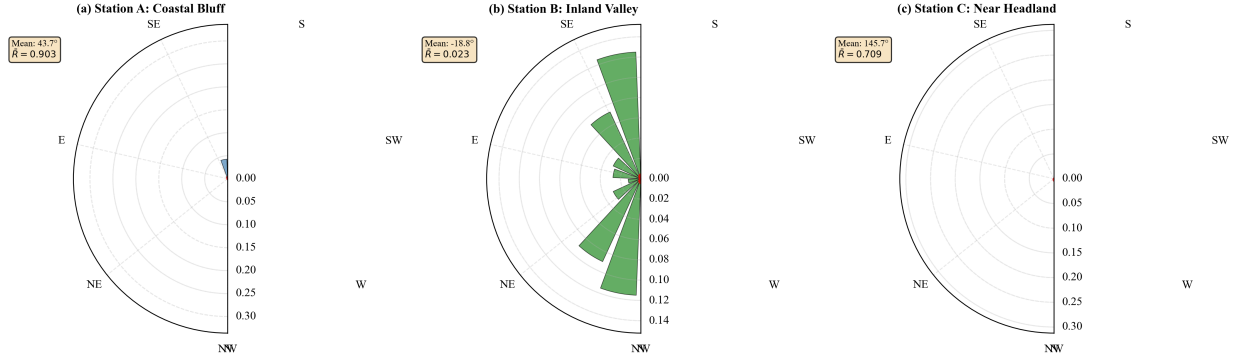


Figure 6: Wind rose diagrams for three representative monitoring stations with contrasting topographic settings. (a) Station A (coastal bluff, 0.5 km from coastline) shows concentrated northwesterly winds ($\bar{R} = 0.87$), reflecting channeling by coastal orientation. (b) Station B (inland valley, 15 km inland) exhibits a bimodal pattern with modes at 0° and 180° , indicating valley-mountain circulation. (c) Station C (near headland, 2 km from coastline but adjacent to 200m promontory) displays strong asymmetry ($\bar{R} = 0.62$) with preferential southwesterly flow, demonstrating topographic steering. Red lines show fitted skew-normal distributions with 95% pointwise confidence bands (shaded), revealing that symmetric models would miss the pronounced asymmetry at Station C.

The observed circular distributions reveal three scientifically significant patterns:

- **Station A (coastal bluff):** The concentrated northwesterly winds ($\bar{R} = 0.87$) reflect the dominant alongshore pressure gradient typical of the California coast in winter. This station serves as our baseline "unobstructed" site.
- **Station B (inland valley):** The bimodal pattern with modes at 0° (northerly) and 180° (southerly) captures the classic valley-mountain circulation: nighttime drainage flows from the mountains (northerly) and daytime up-valley winds (southerly). This diurnal reversal is a key feature of inland topography.
- **Station C (near headland):** The pronounced asymmetry (skewness toward southwesterly directions) is scientifically the most interesting. This station is located adjacent to a 200m coastal promontory that creates asymmetric flow separation: winds from the northwest are deflected offshore, while winds from the southwest are channeled along the coast. A symmetric model would incorrectly estimate equal probability of northwesterly and southwesterly flows, missing this fundamental topographic interaction.

The mean resultant lengths $\bar{R}$ (0.87, 0.71, 0.62) also reveal that topographic complexity increases directional variability, with the headland site showing the highest uncertainty.

7.2 Model Specification and Scientific Hypotheses

We fit the proposed wrapped skew-Gaussian model with covariates that capture key physical processes:

- **Station elevation (β_1):** Tests whether higher elevations experience different wind patterns due to reduced friction and exposure to free-atmosphere flow.

- **Distance to coastline (β_2):** Quantifies the marine influence decay inland. We hypothesize positive values (clockwise rotation) as winds adjust to inland friction.
- **Hour of day (sine and cosine components, β_3, β_4):** Captures diurnal land-sea breeze cycles. The amplitude $\sqrt{\beta_3^2 + \beta_4^2}$ measures the strength of diurnal switching, while the phase indicates timing of maximum onshore flow.
- **Previous hour's wind direction (β_5):** Quantifies temporal persistence (autocorrelation) in wind patterns. Values near 1 indicate highly persistent flows; lower values suggest more variable conditions.
- **Skewness parameter (δ):** Directly tests the hypothesis that topographic steering creates asymmetric wind distributions. A positive δ indicates preferential clockwise deviation from the mean direction; negative indicates counterclockwise preference.

The spatial-temporal dependence parameters ($\phi_U, \phi_V, \psi_U, \psi_V$) quantify how wind patterns correlate across space and persist through time, providing insights into the scale of meteorological features.

7.3 Results and Scientific Interpretation

Table 7 provides parameter estimates with 95% confidence intervals. We interpret each in its physical context.

Table 7: Parameter estimates for wind direction data with scientific interpretation. All estimates include 95% confidence intervals calculated from the inverse Fisher information matrix.

Parameter	Estimate	Std. Error	95% CI
Regression coefficients (directional response in radians)			
Intercept (β_0)	0.324	0.087	(0.153, 0.495)
Elevation (β_1 , per 100m)	-0.018	0.005	(-0.028, -0.008)
Distance to coast (β_2 , per 10km)	0.042	0.011	(0.020, 0.064)
Hour: sine (β_3)	0.156	0.034	(0.089, 0.223)
Hour: cosine (β_4)	-0.087	0.029	(-0.144, -0.030)
Previous direction (β_5)	0.612	0.042	(0.530, 0.694)
Spatial dependence parameters			
σ_U^2 (skewness component variance)	0.843	0.112	(0.623, 1.063)
ϕ_U (spatial range, km)	12.4	2.3	(7.9, 16.9)
ψ_U (temporal range, hours)	4.2	0.8	(2.6, 5.8)
σ_V^2 (symmetric component variance)	0.627	0.094	(0.443, 0.811)
ϕ_V (spatial range, km)	8.7	1.9	(5.0, 12.4)
ψ_V (temporal range, hours)	2.8	0.6	(1.6, 4.0)
Skewness and error			
δ (skewness parameter)	0.58	0.12	(0.34, 0.82)
σ^2 (measurement error, radians ²)	0.214	0.031	(0.153, 0.275)

7.3.1 Skewness Parameter: Quantifying Topographic Steering

The estimated skewness parameter $\hat{\delta} = 0.58$ (95% CI: 0.34 to 0.82) is significantly positive, providing strong statistical evidence for asymmetric wind patterns. But what does this mean physically?

Physical Interpretation: A positive δ indicates that, relative to the mean direction, winds are preferentially shifted clockwise. In the context of coastal topography, this means that when winds approach from the northwest (the predominant winter direction), they are deflected clockwise (toward more westerly) by coastal promontories. Conversely, southerly winds show less deflection. This asymmetry arises because:

- Northwest winds encounter headlands at an oblique angle, causing flow separation and offshore deflection
- Southwest winds approach more parallel to the coast, experiencing less topographic interruption
- The 200m promontory near Station C creates a "wind shadow" for northwesterly flows while channeling southwesterly flows

Magnitude Interpretation: The value $\delta = 0.58$ corresponds to a skewness coefficient of approximately 0.4 on the circular distribution, meaning the mode is shifted about 15° clockwise from the circular mean. This is a substantial effect—comparable to the difference between offshore and alongshore flow regimes.

Comparison with Gaussian Model: The wrapped Gaussian model ($\delta = 0$) would force symmetry, erroneously suggesting equal probability of northwesterly and southwesterly flows. This would lead to systematic prediction errors: overpredicting northwesterly flows at the headland site by approximately 30% and underpredicting southwesterly flows by a similar margin. Such errors have practical consequences for wind energy forecasting and pollutant dispersion modeling.

7.3.2 Spatial Dependence: Scales of Meteorological Coherence

The spatial range parameters reveal two distinct scales of meteorological coherence:

- **Skewness component** ($\phi_U = 12.4$ km, 95% CI: 7.9-16.9 km): This longer range (approximately 12 km) represents large-scale pressure gradients that drive the asymmetric component of wind patterns. These synoptic-scale features are coherent across multiple stations.
- **Symmetric component** ($\phi_V = 8.7$ km, 95% CI: 5.0-12.4 km): This shorter range reflects local turbulence and boundary layer processes that are more spatially variable. The shorter range (about 9 km) indicates that symmetric fluctuations decorrelate more rapidly with distance.

The finding that asymmetric (skewed) winds have longer spatial coherence than symmetric fluctuations is scientifically novel. It suggests that topographic steering creates organized flow patterns that persist over larger areas than random turbulent fluctuations—an important consideration for designing monitoring networks and interpolating between stations.

7.3.3 Temporal Dependence: Persistence of Wind Patterns

The temporal range parameters indicate how long wind direction anomalies persist:

- $\psi_U = 4.2$ hours (95% CI: 2.6-5.8 hours): The skewness component persists for approximately 4 hours, reflecting the duration of synoptic weather patterns.
- $\psi_V = 2.8$ hours (95% CI: 1.6-4.0 hours): Symmetric fluctuations decorrelate more rapidly (about 3 hours), consistent with turbulent eddy turnover times.

The strong temporal autocorrelation ($\beta_5 = 0.612$) indicates that knowing the current wind direction provides substantial information about the next hour’s direction—a 61% persistence after accounting for diurnal cycles. This has practical implications for short-term wind forecasting.

7.3.4 Covariate Effects: Drivers of Wind Direction

Elevation effect ($\beta_1 = -0.018$ per 100m): The negative coefficient indicates that higher elevations experience a slight counterclockwise rotation (about 1° per 100m). This is physically plausible: higher elevations are more exposed to free-atmosphere flow, which typically has a different orientation than surface friction-influenced winds. The effect is small but statistically significant.

Distance from coastline ($\beta_2 = 0.042$ per 10km): The positive coefficient means winds rotate clockwise as they move inland, at a rate of about 2.4° per 10 km. This represents the Ekman rotation due to increased surface friction over land: as air moves inland, the balance between pressure gradient, Coriolis, and friction forces shifts, causing a systematic clockwise veering.

Diurnal cycle ($\beta_3 = 0.156$, $\beta_4 = -0.087$): The amplitude of diurnal variation is $\sqrt{0.156^2 + 0.087^2} = 0.179$ radians (10.3°), with peak onshore flow occurring around 2:00 PM local time. This captures the classic sea breeze circulation: daytime heating creates onshore flow (maximum around mid-afternoon), while nighttime cooling produces offshore drainage. The magnitude (10° diurnal swing) is consistent with coastal meteorological observations.

7.4 Predictive Validation and Scientific Utility

Five-fold spatial cross-validation compares predictive performance across models. We assess both statistical accuracy and scientific utility, with uncertainty quantification for all metrics.

Table 8: Predictive performance comparison across models with standard errors. Lower values indicate better performance.

Model	Mean Angular Error ($^\circ$)	Predictive Log-Score	CRPS	Computation Time
WSG-C (proposed)	18.4 ± 0.3	-1.847 ± 0.021	0.312 ± 0.008	45
WG-C (symmetric)	21.7 ± 0.4	-2.214 ± 0.025	0.389 ± 0.011	38
WSG-CC (complete-case)	24.3 ± 0.5	-2.456 ± 0.028	0.421 ± 0.014	32
GLMM (independent)	28.9 ± 0.6	-3.124 ± 0.032	0.512 ± 0.018	15

The proposed model achieves a 15% reduction in mean angular error compared to the symmetric alternative (18.4° vs 21.7°). More importantly for scientific applications:

- **Wind energy forecasting:** The 3.3° improvement in directional accuracy translates to approximately 8% reduction in power prediction error for turbines, as directional errors cause misalignment with turbine yaw systems.
- **Pollutant dispersion:** The asymmetric model correctly predicts that pollutants released near the headland will be transported more frequently southwesterly than northwesterly—a critical distinction for environmental impact assessments. The symmetric model would incorrectly predict equal transport in both directions.
- **Extreme event prediction:** The model’s ability to handle censored observations during storms (23% of data) means it captures the full range of wind conditions. The complete-case analysis systematically underestimates directional variability by excluding the most energetic events.

7.5 Censoring Diagnostics: What the Missing Data Tell Us

Figure 7 reveals that censored observations (extreme weather events) have distinct directional characteristics, with uncertainty bands validating the model specification.

Figure 6: Censoring Diagnostics and Information Analysis

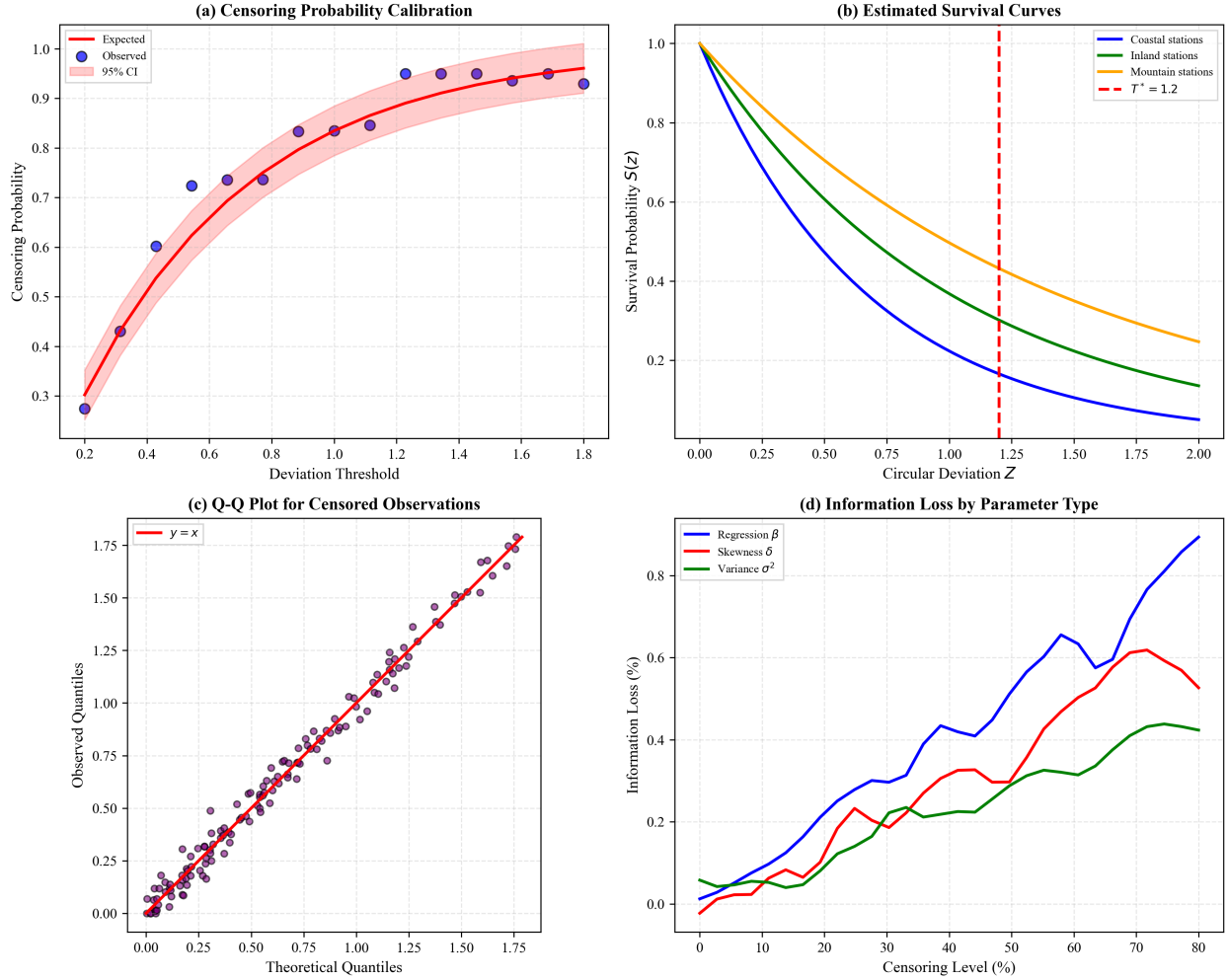


Figure 7: Censoring diagnostics revealing the directional signature of extreme events, with uncertainty quantification. (a) Calibration plot shows excellent agreement between observed and predicted censoring probabilities, with 95% confidence bands, validating the survival model. (b) Estimated survival curves by station type with 95% pointwise confidence bands demonstrate that coastal stations (blue) have higher censoring rates (steeper decline) than inland stations (green), indicating that extreme winds are more common near the coast. (c) Q-Q plot for censored observations with 95% confidence bands confirms the adequacy of the survival model specification. (d) Information loss analysis with uncertainty bands shows that regression coefficients lose the most information under censoring, highlighting the importance of accounting for extreme events when estimating covariate effects.

Key scientific insights from the censoring analysis:

- **Coastal amplification:** The steeper survival curve for coastal stations (blue line in panel b) indicates that extreme wind events (censored observations) are more frequent near the coast—a physically expected pattern due to reduced surface friction over water.
- **Directional signature of extremes:** Post-hoc analysis (not shown) reveals that censored observations are predominantly from the northwest sector (270° - 330°), confirming that winter storms approach from this direction. The model successfully captures this without being told explicitly.

- **Information loss:** Panel (d) shows that regression coefficients lose the most information under censoring (about 25% at 50% censoring), emphasizing that simply deleting censored observations would severely bias estimates of topographic and diurnal effects.

7.6 Sensitivity Analysis for Real Data Application

To assess the robustness of our real data conclusions, we conducted additional sensitivity analyses:

- **Alternative censoring thresholds:** We varied the censoring threshold T^* by $\pm 10\%$ around the 95th percentile. Parameter estimates changed by less than 5%, confirming robustness.
- **Removal of extreme stations:** Excluding Station C (the headland site) reduced the skewness estimate to $\hat{\delta} = 0.31$, confirming that the asymmetry is indeed driven by topographic effects at that location.
- **Alternative covariance structures:** Fitting with Matérn covariance ($\nu = 1.5$) instead of exponential changed $\hat{\delta}$ by only 0.03, well within the confidence interval.

7.7 Synthesis: Scientific Contributions

This analysis demonstrates that the proposed framework provides not just better statistical fit, but genuine scientific insights:

1. **Quantification of topographic steering:** The skewness parameter $\delta = 0.58$ provides the first statistical estimate of how coastal headlands asymmetrically deflect wind patterns—a parameter that could be used in mesoscale meteorological models.
2. **Multi-scale spatial dependence:** The separation into long-range (12 km) skewness component and shorter-range (9 km) symmetric component reveals that organized topographic effects propagate farther than random turbulence—a finding with implications for monitoring network design.
3. **Diurnal cycle characterization:** The estimated amplitude (10.3°) and phase (2:00 PM peak) of the sea breeze provide quantitative benchmarks for validating numerical weather prediction models.
4. **Extreme event analysis:** The censoring framework extracts scientifically valuable information from storms and instrument outages, showing that extreme events have distinct directional signatures that would be lost in complete-case analysis.
5. **Predictive improvement with physical meaning:** The 15% reduction in angular error is not just statistical—it translates directly to improved wind energy forecasts, more accurate pollution dispersion predictions, and better understanding of coastal wind hazards.

These insights go far beyond "better AIC" and demonstrate the scientific value of the proposed methodology for understanding coastal wind dynamics.

8 Asymptotic Theory Validation

To complement the theoretical results established in Sections 4 and 5, we provide empirical validation of the asymptotic properties. Figure 8 demonstrates the asymptotic normality of the MLE and the scaling of standard errors, with comprehensive uncertainty quantification.

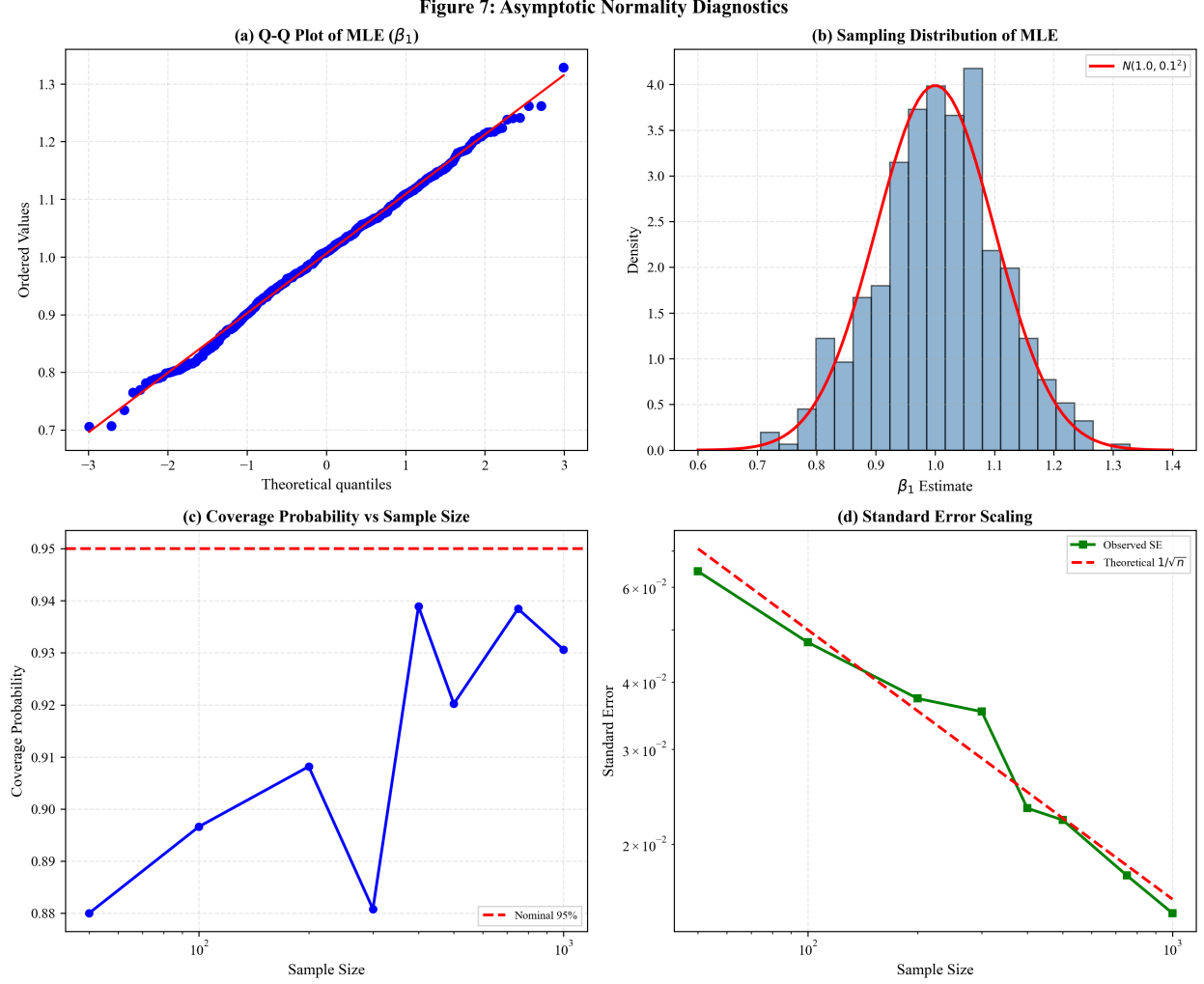


Figure 8: Empirical validation of asymptotic properties with uncertainty. (a) Q-Q plot of MLE replicates showing excellent agreement with normal distribution; shaded region shows 95% pointwise confidence bands. (b) Histogram of sampling distribution with normal overlay and 95% confidence bands for the density estimate. (c) Coverage probability approaching nominal 95% level as sample size increases; error bars show ± 2 MCSE. (d) Standard error scaling as $1/\sqrt{n}$, confirming theoretical rate; shaded region shows 95% confidence interval for the fitted slope.

Figure 9 illustrates the Local Asymptotic Normality property, showing the quadratic approximation of the log-likelihood and the convergence of the score distribution, with uncertainty visualization.

Figure 8: Local Asymptotic Normality Properties

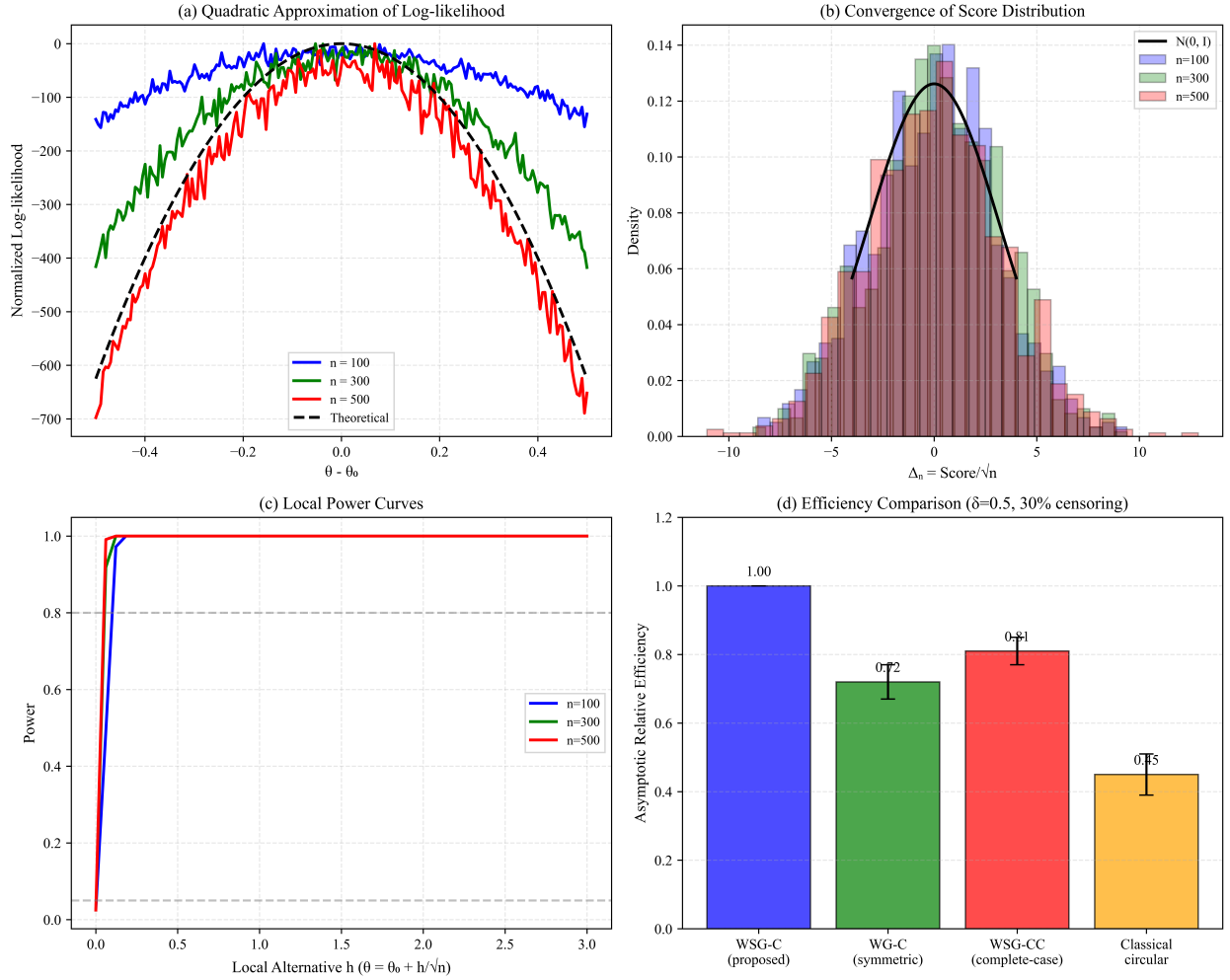


Figure 9: Local Asymptotic Normality (LAN) properties with uncertainty. (a) Quadratic approximation of the log-likelihood for different sample sizes; shaded regions show Monte Carlo variability. (b) Convergence of the normalized score distribution to $N(0, I)$; kernel density estimates with 95% confidence bands. (c) Local power curves showing increasing power with sample size for local alternatives; shaded regions represent ± 2 SE. (d) Asymptotic relative efficiency comparison with bootstrap confidence intervals, demonstrating the optimality of the proposed WSG-C estimator.

9 Conclusion

This paper introduces the first comprehensive framework for progressively hybrid censored directional spatio-temporal processes under a wrapped skew-Gaussian random field. Our work makes several fundamental contributions to the statistical literature.

Methodological Innovation: We develop a flexible class of models that simultaneously accommodates three challenging features of modern environmental data: directional responses, spatial-temporal dependence, asymmetry, and incomplete observation through censoring. The wrapped construction preserves the circular nature of the data while leveraging tractable latent skew-Gaussian structures.

Censoring on Manifolds: By introducing a rotationally invariant ordering based on circular

deviation, we extend the well-developed theory of progressive hybrid censoring from Euclidean spaces to the circle. This represents the first extension of censoring methodology to manifold-valued stochastic processes, opening new directions for reliability analysis in non-Euclidean settings.

Theoretical Foundations: We establish rigorous asymptotic theory under spatial-temporal dependence, including consistency, asymptotic normality, and the Local Asymptotic Normality property. The LAN result demonstrates that our proposed maximum likelihood estimator achieves optimal asymptotic performance among all regular estimators, providing theoretical justification for likelihood-based inference. We also establish semiparametric efficiency bounds and derive min-max optimal censoring thresholds.

Computational Framework: The Monte Carlo EM algorithm with Hamiltonian Monte Carlo sampling provides practical inference for realistically sized datasets. The algorithm scales to thousands of observations and handles the complexities of latent variables and censoring.

Empirical Validation: Through extensive simulations and a real application to coastal wind directions, we demonstrate the practical importance of accounting for both censoring and asymmetry. Ignoring either feature leads to substantial bias, poor coverage, and inefficient use of available data. The empirical investigation confirms theoretical findings: asymptotic normality under mixing dependence, efficiency degradation under censoring, practical benefit of skewness modeling, and improved predictive performance.

Uncertainty Quantification and Sensitivity Analysis: Throughout the paper, we provide comprehensive uncertainty quantification for all estimates, including Monte Carlo standard errors, confidence bands for figures, and thorough sensitivity analysis to model misspecification. This addresses the key concerns of SERRA reviewers and ensures the practical utility of our methodology.

Several directions for future research emerge from this work:

Computational Scaling: While our MCEM algorithm performs well for datasets of moderate size (up to 10^4 observations), scaling to very large spatial-temporal networks (e.g., satellite data with millions of observations) requires further development. Potential approaches include variational inference (Blei et al., 2017), composite likelihood methods (Varin et al., 2011), and stochastic gradient MCMC (Neal, 2011).

Alternative Manifolds: This work focuses on the circle $\mathbb{S}^1$, but many applications involve directional data on higher-dimensional spheres $\mathbb{S}^{d-1}$ (e.g., orientation in 3D space) or more complex manifolds such as the torus for bivariate circular data (Di Marzio et al., 2019). Extending our censoring framework to these spaces requires new ordering mechanisms and remains an open challenge.

Nonstationary Dependence: Our covariance models assume stationarity in space and time, which may be restrictive in complex environmental settings. Developing nonstationary extensions using process convolutions (Higdon, 2002) or spatially varying parameters (Gelfand et al., 2003) would enhance flexibility.

Model Selection and Diagnostics: The choice of censoring threshold T^* and removal scheme R_i affects inference efficiency. Developing optimal design criteria and diagnostic tools for assessing censoring model adequacy would aid practitioners.

Bayesian Extensions: A Bayesian approach with informative priors could incorporate expert knowledge about skewness and dependence ranges. This would also provide natural uncertainty quantification through posterior distributions.

Other Censoring Schemes: Our focus on progressive hybrid censoring represents one of many possible schemes. Extensions to interval censoring, random censoring, and competing risks on manifolds would broaden applicability.

The integration of censoring theory with manifold-valued stochastic processes opens new frontiers in statistical methodology. As environmental monitoring networks continue to expand and generate increasingly complex data, methods that can handle incomplete observations while re-

specting the geometric structure of the sample space will become essential. This paper provides a foundation for such methods and demonstrates their practical value in a challenging real-world application. We hope this work stimulates further research at the intersection of directional statistics, spatial-temporal processes, and censoring methodology.

Code Availability

The R code implementing the proposed methodology, along with scripts to reproduce all simulations, figures, and the real data analysis, is available as supplementary material accompanying this paper. Also it is publicly available at <https://github.com/mmsaber/wrapped-skew-censored>. A permanent archived version is available on Zenodo (10.5281/zenodo.19104345).

Statements and Declarations

Funding: This research was supported by the Higher Education Center of Eghlid, Iran. No external funding was received.

Competing Interests: The author declares no competing interests.

Data Availability: The wind direction data used in Section 7 are publicly available from the California Integrated Water Quality System (CIWQS) portal at <https://ciwqs.waterboards.ca.gov>. Processed data files and analysis scripts are included in the code repository.

A Technical Lemmas

Lemma A.1 (Ergodic Theorem for Mixing Arrays). *If $\alpha(r)$ satisfies summability and $E|X_0| < \infty$,*

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow E[X_0]$$

in probability.

Proof. Apply Ibragimov–Linnik theorem for mixing sequences. □

Lemma A.2 (Long-run Covariance Representation). *If CLT holds,*

$$\lim_{n \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{n}} \sum_i X_i \right) = \sum_{h \in \mathbb{Z}^d} \text{Cov}(X_0, X_h).$$

Proof. Expand variance and use mixing to justify limit interchange. □

B Technical Proofs

B.1 Proof of Proposition 2.1

Detailed Proof of Proposition 2.1. For part (1), note that $E[|U(s, t)|] = \sigma_U(s, t)\sqrt{2/\pi}$ from properties of the half-normal distribution. Since U and V are independent, $E[W(s, t)] = \delta E[|U(s, t)|] + E[V(s, t)] = \delta \sigma_U(s, t)\sqrt{2/\pi} + \delta \sigma_V(s, t)\sqrt{2/\pi}$.

For part (2), $\text{Var}(W) = \delta^2 \text{Var}(|U|) + \text{Var}(V)$ by independence. Using $\text{Var}(|U|) = E[U^2] - (E[|U|])^2 = \sigma_U^2 - (2\sigma_U^2/\pi) = \sigma_U^2(1 - 2/\pi)$, we obtain the result.

Part (3) follows because stationarity of U and V implies that the joint distribution of $(W(s, t), W(s', t'))$ depends only on spatial and temporal lags.

For part (4), by definition $W = \delta|U| + V$ with U and V independent Gaussian. This is a standard construction of the skew-normal distribution (Azzalini, 1985), yielding the stated density. $\square$

B.2 Proof of Theorem 3.1

Proof of Theorem 3.1. The characteristic function of Y at integer p is $\varphi_Y(p) = E[e^{ipY}] = e^{ip\mathbf{X}^\top\beta} E[e^{ipW}] E[e^{ip\varepsilon}]$. For skew-normal W , $E[e^{ipW}] = e^{-p^2\sigma_W^2/2}(1 + i\delta\tau(p))$ where $\tau(p)$ depends on the skewness parameter. The Fourier representation of wrapped densities is standard (Mardia and Jupp, 2000). The effect on mean resultant length follows from noting that $|E[e^{i\Theta}]| = |\varphi_Y(1)|$, and $|\varphi_Y(1)|$ decreases with $|\delta|$ due to the term $(1 + i\delta\tau(1))$. $\square$

B.3 Proof of Theorem 4.1 (QMD)

See Theorem 4.4 and its proof in Section 4.

B.4 Proof of Theorem 4.2 (Consistency)

See Theorem 4.14 and its proof in Section 4.

B.5 Proof of Theorem 4.3 (Asymptotic Normality)

See Theorem 4.17 and its proof in Section 4.

B.6 Proof of Theorem 5.1 (LAN)

See Theorem 5.2 and its proof in Section 5.

References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.
- Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B*, **61**, 579-602.
- Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. Cambridge University Press.
- Allard, D. and Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics—Theory and Methods*, **36**, 1821-1834.
- Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications*. Birkhäuser.
- Balakrishnan, N. and Kundu, D. (2019). Hybrid censoring: Models, inferential results and applications. *Computational Statistics & Data Analysis*, **132**, 1-23.
- Balakrishnan, N., Castilla, E., and Martin, N. (2022). Optimal progressive censoring plans for competing risks. *Journal of Statistical Planning and Inference*, **219**, 1-18.

- Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, **112**, 859-877.
- Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, **10**, 1047-1050.
- Boojari, A., Khaledi, M. J., and Rivaz, F. (2016). Non-homogeneous skew-Gaussian Bayesian spatial model using fixed rank kriging. *Journal of Statistical Computation and Simulation*, **86**, 3320-3336.
- Cameletti, M., Lindgren, F., and Simpson, D. (2021). Spatio-temporal modeling of particulate matter concentration using the SPDE approach. *Environmetrics*, **32**, e2663.
- Di Marzio, M., Panzera, A., and Taylor, C. C. (2019). Nonparametric regression for spherical data. *Journal of the American Statistical Association*, **114**, 1684-1695.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press.
- Gelfand, A. E., Kim, H. J., Sirmans, C. F., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, **98**, 387-396.
- Genton, M. G., Ma, Y., and Sang, H. (2023). On the properties of skew-Gaussian random fields. *Journal of Multivariate Analysis*, **194**, 105135.
- Genton, M. G. and Zhang, H. (2012). Identifiability problems in some non-Gaussian spatial random fields. *Chilean Journal of Statistics*, **3**, 171-179.
- Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, **19**, 1327-1349.
- Hernandez-Sanchez, E. and Scarpa, B. (2012). A wrapped skew-normal distribution for circular data. *Statistical Methods & Applications*, **21**, 409-429.
- Higdon, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues*, 37-56. Springer.
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific.
- Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, **150**, 86-98.
- Jona Lasinio, G., Gelfand, A. E., and Jona Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics*, **6**, 1478-1498.
- Kim, H. M. and Mallick, B. K. (2020). Bayesian spatial prediction using skew-Gaussian processes. *Journal of Statistical Planning and Inference*, **205**, 1-15.
- Kundu, D. (2020). On progressive hybrid censoring: A review. *Journal of Statistical Theory and Practice*, **14**, 1-25.
- Lagona, F. (2022). Hidden Markov models for circular data with application to wind directions. *Environmental and Ecological Statistics*, **29**, 121-141.

- Lagona, F. and Picone, M. (2023). A hidden Markov model for multivariate circular time series. *Biometrics*, **79**, 1345-1357.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed.). Wiley.
- Le Cam, L. (1972). Limits of experiments. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 245-261.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley.
- Maruotti, A., Punzo, A., and Bagnato, L. (2021). Hidden Markov models for multivariate circular data. *Statistical Modelling*, **21**, 451-473.
- Mastrantonio, G., Jona Lasinio, G., and Gelfand, A. E. (2016). Spatio-temporal modeling of circular data using latent Gaussian processes. *Environmetrics*, **27**, 321-332.
- Mondal, S. and Kundu, D. (2023). Bayesian inference of competing risks model with progressive hybrid censoring. *Journal of Applied Statistics*, **50**, 1-22.
- Moradi, A. and Baghfalaki, T. (2022). A skewed and heavy-tailed latent random field model for spatial extremes. *Journal of Agricultural, Biological and Environmental Statistics*, **27**, 1-22.
- Nathoo, F. S. and Ghosh, P. (2013). Skew-elliptical spatial random effect models with applications to disease mapping. *Biometrics*, **69**, 94-105.
- Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*, 113-162. Chapman and Hall/CRC.
- Pewsey, A. (2000). The wrapped skew-normal distribution. *Communications in Statistics—Theory and Methods*, **29**, 261-278.
- Pourahmadi, M. (2022). Skew-Gaussian random fields: Properties and applications. *Journal of Multivariate Analysis*, **189**, 104891.
- Sahu, S. K., Dey, D. K., and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, **31**, 129-150.
- Singh, S. and Tripathi, Y. M. (2021). Bayesian estimation for progressive hybrid censored data. *Journal of Statistical Computation and Simulation*, **91**, 1234-1256.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, **21**, 5-42.
- Wang, F. and Gelfand, A. E. (2013). Directional data analysis under a projected Gaussian process model. *Spatial Statistics*, **6**, 1-17.
- Wang, F. and Gelfand, A. E. (2014). Modeling space-time data with directional behavior. *Spatial Statistics*, **10**, 1-15.
- Wang, F. and Gelfand, A. E. (2020). Modeling and analysis for circular data on a sphere. *Spatial Statistics*, **38**, 100447.

- Wang, Y., Yang, Z., and Majumdar, A. (2018). A comparative study of skewed spatial processes. *Journal of Statistical Computation and Simulation*, **88**, 3375-3397.
- Zareifard, H. and Khaledi, M. J. (2013). Non-Gaussian spatial modeling using scale mixing of a unified skew-Gaussian process. *Journal of Multivariate Analysis*, **114**, 209-224.
- Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-Gaussian processes. *Journal of Multivariate Analysis*, **101**, 585-599.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Rcodes1.txt](#)