International Journal of Development Mathematics Vol 3 Issue 2 Page No. 351 - 361

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

# A Modified Spatial Variance Shift Outlier Model

**Ehiosasu Idusuyi[a,*], Adesupo, A. Akinrefon[b], Emmanuel Torsen[b] and Abraham Okolo[b]**

*[a]Department of Statistics, Abubakar Tafawa Balewa University, Bauchi, Nigeria.*
*[b]Department of Statistics, Modibbo Adama University, Yola, Adamawa State, Nigeria*

**A R T I C L E I N F O**

***Article history:***
Received 02 March 2026
Received in revised form 20 May 2026
Accepted 29 May 2026

**Keywords:**
SVSOM, REML, m-SVSOM, Spatial Autocorrelation

**MSC 2020 Subject classification:**
62F35, 62H11, 62J20, 62M30.

**A B S T R A C T**

The Variance Shift Outlier Model (VSOM) and its spatial extension (SVSOM) for outlier accommodation have been developed in linear and spatial frameworks; however, they often fail to account simultaneously for spatial disparities and correlated measures within groups. This study proposes a Modified Spatial Variance Shift Outlier Model (m-SVSOM) within a Generalized Linear Mixed Model (GLMM) framework to address these limitations. The proposed model incorporates spatially lagged dependent variables and spatial autocorrelation in the residuals, while accounting for random effects to capture repeated measures. Variance parameter estimation is performed using the Restricted Maximum Likelihood (REML) method. To evaluate the model's performance, a simulation study was conducted across varying sample sizes (n = 16, 36, 144, 400, 900, 1000). The results show that the m-SVSOM consistently outperforms the SVSOM, achieving lower Mean Squared Error (MSE = 0.0522) and Root Mean Squared Error (RMSE = 0.2108). The findings suggest that the proposed model yields more robust parameter estimates, making it a valuable tool for analyzing complex spatial datasets with hierarchical structures.

## 1. Introduction

An outlier is an observation that deviates from the distributional behaviour of the remaining dataset (Barnett & Lewis, 1994; Moore & McCabe, 1999; Monhor & Takemoto, 2005; Petronilla & Chinaka, 2023; Vijayalakshmi *et al*., 2025). The concept of outliers originated from the mathematical analysis of geodetic and astronomical measurement data according to Monhor & Takemoto (2005). Identifying the specific cause requires a thorough investigation of the process, from data collection to processing, to minimize variability in the data, avoid overfitting of regression models, reduce bias in parameter estimation, and prevent misleading conclusions (Gumedze, 2018). It is also important to note that outliers in a dataset are not necessarily bad points; they can reveal the most informative phenomena in a study.

Galton, a British statistician in the 19th century, developed several statistical tools and concepts that are fundamental for identifying outliers. Over time, various outlier detection methods have evolved and are generally categorized into three classes: statistical (model-based), proximity-based, and clustering-based (Smiti, 2020).
Detecting and accommodating outliers can enhance the accuracy of regression coefficient estimates. Evidence suggests that spatial outliers hold valuable information, leading to models that more accurately reflect the underlying data and improve predictive performance (Baba *et al*., 2022).
Two principal outlier formulations have emerged: the Mean-Shift Outlier Model (MSOM) (Dai *et al*., 2016; Lehmann *et al*., 2020; Song *et al*., 2024) and the Variance-Shift Outlier Model (VSOM) (Babadi *et al*., 2016; Gharbaghi *et*

* **Corresponding author. Tel.**: +2347034912387
**E-mail address:** osasu20@gmail.com (Idusuyi, E.)
https://doi.org/10.62054/ijdm/0302.22

*al.*, 2026;Gumedze & Chatora, 2014; Gumedze *et al*., 2010; Ismadyaliana *et al*., 2024; Thompson, 1985). The MSOM assumes that outliers arise from additive shifts in the mean of specific observations, effectively treating them as case-deletion scenarios (Cook & Weisberg, 1982). In contrast, the VSOM posits that outliers are observations with inflated variances relative to the bulk of the data, a formulation particularly suited to situations where extreme observations reflect increased measurement uncertainty or heterogeneous error structures (Thompson, 1985; Gumedze *et al*., 2010). These outlier models were developed within the framework of either linear models or general spatial models. The choice between mean-shift and variance-shift formulations depends on the nature and relevance of the outlier mechanism in the substantive context, with the VSOM having a clear advantage over the MSOM, where observations detected as outliers are down-weighted and their weights are automatically incorporated into the parameter estimation (Beckman and Cook, 1983).

Thompson (1985) adopted the VSOM within a linear regression model and employed Restricted Maximum Likelihood (REML) for parameter estimation. Gumedze *et al*. (2010) adopted the VSOM within a linear mixed model to detect and accommodate outliers, and employed REML to estimate the variance parameters. Dai *et al*., (2016) proposed the MSOM and VIOM in a spatial regression model to detect and accommodate outliers; Maximum Likelihood (ML) was employed for parameter estimation.

The variance shift outlier model (VSOM) was later extended by Baba *et al*., (2022), who formulated the VSOM in the general spatial regression model (GSM) and called it the Spatial Variance Shift Outlier Model (SVSOM), which was employed to further accommodate spatial outliers based on spatial weights that are said to contain neighborhood information both in the dependent and residual variables, REML was employed for parameter estimation.

In addition some literatures on outlier detection and accommodation that have applied various estimation methods include:(Cai *et al*., 2024; Gumedze & Dunne, 2011) employed the Restricted Maximum Likelihood to estimate the parameters in their models, (Dai *et al*., 2016; Torabi, 2014) employed the maximum likelihood, Melo *et al*., (2013) employed the Markov Chain Monte Carlo maximum likelihood, and Shukla *et al*., (2023) employed the Robust Modified Maximum Likelihood (MML) estimation.

Despite these advances, existing spatial outlier models exhibit important limitations. Firstly, extensions of VSOM to linear models or linear mixed models have failed to account for spatial disparities, treating spatial structure as absent or secondary (Babadi *et al*., 2016; Gumedze *et al*., 2010; Kowalskie, 2025). Secondly, spatial models such as SVSOM do not accommodate correlated or repeated measures within groups at specific locations, thereby ignoring hierarchical data structures common in longitudinal studies, clustered designs and multi-site investigations. This forms the basis of this study, as it addresses the limitations of these existing outlier and accommodation models (Baba *et al.,* 2022; Babadi *et al*., 2016; Gumedze et al., 2010) by proposing a Modified Spatial Variance Shift Outlier Model (m-SVSOM) embedded within a Generalized Linear Mixed Model (GLMM) framework. The m-SVSOM integrates three key components: (1) spatial lagged dependent variables and spatial autocorrelation in residual terms to capture spatial dependence, (2) random effects to model intra-group correlation arising from repeated measures or hierarchical grouping, and (3) variance-shift outlier parameters estimated via REML to accommodate extreme observations without deletion. This unified model extends the SVSOM of Baba et al. (2022) by incorporating the random-effects structure of GLMMs, thereby enabling robust outlier accommodation in spatially dependent, hierarchically structured data.

## 2. Methodology

### 2.1 Existing Spatial variance shift outlier model

Following Baba *et al.*, (2022), the general spatial model is given as

$$\boldsymbol{y} = \rho \boldsymbol{W}_1 \boldsymbol{y} + \boldsymbol{X}\beta + \boldsymbol{u}, \qquad \boldsymbol{u} = \lambda \boldsymbol{W}_2 \boldsymbol{u} + \boldsymbol{e} \tag{1}$$

Where **y** is an (nx1) vector of response variable, **X** is a (n x k) matrix of explanatory variables, $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are (n x n) spatial weights square matrices which are known. $\lambda$ a coefficient on the spatial autocorrelation in the residual term. $\boldsymbol{e}$ is the (n x 1) vector of the error term, $\boldsymbol{e} \sim MVN(0, \sigma^2 I_n)$.

The Spatial Variance Shift Outlier Model (SVSOM) is given below as

$$\boldsymbol{y}^* = \boldsymbol{X}^*\beta + \boldsymbol{\delta_i}^* \varphi_i + e \quad (2)$$

$i = 1,2,\dots,n,$ $\boldsymbol{\delta_i}^*$ a dummy variable that assigns

$$\boldsymbol{\delta_i}^* = \begin{cases} 1 & \text{if the ith observation is an outlier} \\ 0, & \text{otherwise} \end{cases}$$

$\varphi_i$ is an unknown random coefficient $\varphi_i \sim N\left(0, \sigma^2 \boldsymbol{\delta_i}^* \omega_{si} \boldsymbol{\delta_i}^{*\intercal}\right)$ ,$\omega_i$ is the measure that determines inflation in the variance.

Using the Restricted maximum likelihood estimation method the outlier parameter estimate was obtained below as

$$\hat{\varphi}_i = \hat{\omega}_{si} \boldsymbol{\delta_i}^{*\intercal} \boldsymbol{P}\hat{\boldsymbol{B}}\hat{\boldsymbol{A}}\boldsymbol{y} \quad (3)$$

**2.2 The proposed Modified Spatial Variance Shift Outlier Model (m-SVSOM)**

To capture the random effect in Equation (2), the equation is re-expressed in the GLMM framework for normal responses, identity link g(μ)=μ is used.

The GLMM can be expressed as

$$g(\mu_i) = \boldsymbol{\eta}_i = \boldsymbol{x}_i^\intercal \beta + \boldsymbol{z}_i^\intercal b, \quad b \sim N(0, G) \quad (4)$$

Where $g(.)$ = Link function; $\mu_i$= Expected mean of the observations; $\boldsymbol{x_i}$ =Vectors of $p$ fixed effect; $\beta$ = Coefficient of the fixed effect; $\boldsymbol{z_i}$ = Vectors of $q$ random effect and $b$ = coefficient of the random effect and $G$ is the variance-covariance matrix of the random effect.

Following Maestrini *et al.,* (2024), let **X** denote $n \times p$ and **Z** denote $n \times q$ matrices, respectively, which are obtained by stacking the $\boldsymbol{x}_i$'s and $\boldsymbol{z_i}$'s respectively as row vectors and $\boldsymbol{\eta} = (\eta_{1,\dots,}\eta_n)^\intercal$.

Considering the special case of the spatial GLMM with normally distributed responses and an identity link function $\eta \equiv y$, incorporating both spatial dependence between regions (i, j) and an error term with spatial effects, yields Equation (4) below

$$\boldsymbol{\eta_i} = \rho \boldsymbol{W_1}\boldsymbol{\eta_j} + \boldsymbol{X}\beta + \boldsymbol{Z}b + \boldsymbol{u}; \quad \boldsymbol{u} = \lambda \boldsymbol{W_2}\boldsymbol{u} + \boldsymbol{e} \quad ; \boldsymbol{e} \sim N(0, \sigma^2) \quad (5)$$

Where:

$\rho$= the spatial autoregressive coefficient

$W_1 and\ W_2 = n\ x\ n$ Spatial weights matrix

$\lambda$ = coefficient on the spatial autocorrelations in the random error term

$\boldsymbol{W\eta_j}$= spatial lagged dependent variable

From Equation (5), let $\boldsymbol{A} = (I - \rho W_1)$ , $\mathbf{B} = (I - \lambda W_2)$

Equation (5) can be re-parameterized as

$$\boldsymbol{A\eta} = \boldsymbol{X}\beta + \boldsymbol{Z}b + \boldsymbol{B^{-1}e} \tag{6}$$

$$\boldsymbol{A\eta} \sim MVN(\boldsymbol{X}\beta, \sigma^2(\boldsymbol{ZGZ^T} + (\boldsymbol{B^TB})^{-1})$$

Introducing spatial outliers via variance shift and adopting the Spatial Variance Shift Outlier Model (SVSOM) gives the Modified Spatial Variance Shift Outlier Model (m-SVSOM) as

$$\boldsymbol{A\eta} = \boldsymbol{X}\beta + \boldsymbol{Z}b + \boldsymbol{\delta_i}\varphi_i + B^{-1}e \tag{7}$$

Pre-multiplying Equation (7) by B gives

$$\boldsymbol{BA\eta} = \boldsymbol{BX}\beta + \boldsymbol{BZ}b + \boldsymbol{B\delta_i}\varphi_i + e, \tag{8}$$

$$let \quad (\boldsymbol{y}^* = \boldsymbol{BA\eta},\ \boldsymbol{X}^* = \boldsymbol{BX},\ \boldsymbol{Z}^* = \boldsymbol{BZ},\ \boldsymbol{\delta_i}^* = \boldsymbol{B\delta_i})$$

Where $\boldsymbol{\delta_i}^* = \boldsymbol{B\delta_i}$is the n×1 transformed outlier indicator vector

Thus, $\boldsymbol{y}^* = \boldsymbol{X}^*\beta + \boldsymbol{Z}^*b + \boldsymbol{\delta_i}^*\varphi_i + e$ (9)

$$\boldsymbol{y}^* \sim MVN\left(\boldsymbol{X}^*\beta, \sigma^2\left(\boldsymbol{Z}^*G\,\boldsymbol{Z}^{*\top} + \boldsymbol{\delta_i}^*\omega_{si}\boldsymbol{\delta_i}^{*\top} + R\right)\right)$$

$$\boldsymbol{y}^* \sim MVN\left(\boldsymbol{X}^*\beta, \sigma^2\left(H + \boldsymbol{\delta_i}^*\omega_{si}\boldsymbol{\delta_i}^{*\top}\right)\right)$$

where G, $\omega_{si}$and R represent the Variance- covariance matrix of the random effect, outlier and error term in the model.

## 2.3 Parameter Estimation

Setting the dummy variable $\boldsymbol{\delta_i}^* = 0$ and $\varphi_i = 0$ in Equation (9), implying no outlier term gives Equation (10) below

$$\boldsymbol{y}^* = \boldsymbol{X}^*\beta + \boldsymbol{Z}^*b + e \tag{10}$$

$$\boldsymbol{y}^* \sim MVN\ (\boldsymbol{X}^*\beta, \sigma^2H)$$

The marginal log likelihood function of the random variable is obtained as

$$\mathcal{L}(b) = -\frac{p}{2}log(2\pi) - \frac{1}{2}log|G| - \frac{1}{2}b^\top G^{-1}b$$

And the conditional distribution of $\boldsymbol{y}^*|b$ is

$$\mathcal{L}(\boldsymbol{y}^*|b) = -\frac{n}{2}log(2\pi) - \frac{1}{2}log|R| - \frac{1}{2}\boldsymbol{e}^\top R^{-1}\boldsymbol{e}$$

$$\mathcal{L}(\boldsymbol{y}^*, b) = \mathcal{L}(\boldsymbol{y}^*|b) + \mathcal{L}(b)$$

$$= \left[-\frac{n}{2}log(2\pi) - \frac{1}{2}log|R| - \frac{1}{2}\boldsymbol{e}^{\top}R^{-1}\boldsymbol{e}\right]$$
$$-\left[\frac{p}{2}log(2\pi) + \frac{1}{2}log|G| + \frac{1}{2}b^{\top}G^{-1}b\right]$$
$$= -\frac{1}{2}[\boldsymbol{n}log(2\pi) + log|R| + \boldsymbol{e}^{\top}R^{-1}\boldsymbol{e}]$$
$$-\frac{1}{2}[\boldsymbol{p}log(2\pi) + log|G| + b^{\top}G^{-1}b] \quad (11)$$

Assuming that the variances of the random and error terms are known, following Gumedze & Dunne (2011), Equation (11) yields.

let $e = \boldsymbol{y}^* - \boldsymbol{X}^*\beta - \boldsymbol{Z}^*b$

$$\mathcal{L}(\boldsymbol{y}, b) = -\frac{1}{2}[\boldsymbol{n}log(2\pi) + log|R| + (\boldsymbol{y}^* - \boldsymbol{X}^*\beta - \boldsymbol{Z}^*b)^{\top}R^{-1}(\boldsymbol{y}^* - \boldsymbol{X}^*\beta - \boldsymbol{Z}^*b)]$$
$$-\frac{1}{2}[\boldsymbol{p}log(2\pi) + log|G| + b^{\top}G^{-1}b] \quad (12)$$

Ignoring terms that do not depend on β and b, Equation (12) can be re-expressed as

$$\mathcal{L}(\boldsymbol{y}, b) = -\frac{1}{2}(\boldsymbol{y}^* - \boldsymbol{X}^*\beta - \boldsymbol{Z}^*b)^{\top}R^{-1}(\boldsymbol{y}^* - \boldsymbol{X}^*\beta - \boldsymbol{Z}^*b) - \frac{1}{2}b^{\top}G^{-1}b$$
$$= -\frac{1}{2}[(\boldsymbol{y}^* - \boldsymbol{X}^*\beta - \boldsymbol{Z}^*b)^{\top}R^{-1}(\boldsymbol{y}^* - \boldsymbol{X}^*\beta - \boldsymbol{Z}^*b) + b^{\top}G^{-1}b] \quad (13)$$

Taking the partial derivatives of the joint distribution in Equation (13) with respect to $\beta$ and equating to zero results in Equation (14) below

$$\boldsymbol{X}^{*\top}R^{-1}\boldsymbol{X}^*\beta + \boldsymbol{X}^{*\top}R^{-1}\boldsymbol{Z}^*\mathrm{b} = \boldsymbol{X}^{*\top}R^{-1}\boldsymbol{y}^* \quad (14)$$

Similarly, taking the partial derivative of Equation (13) with respect to b and equating it to zero results in Equation (15) below

$$\boldsymbol{Z}^{*\top}R^{-1}\boldsymbol{X}^*\boldsymbol{\beta} + (\boldsymbol{Z}^{*\top}R^{-1}\boldsymbol{Z}^* + G^{-1})b = \boldsymbol{Z}^{*\top}R^{-1}\boldsymbol{y}^* \quad (15)$$

The estimate of $\beta$ and $b$ are obtained using the Henderson Mixed model equation, which is obtained by arranging Equation (14) and Equation (15) in a matrix form below as

$$\begin{pmatrix} \boldsymbol{X}^{*\top}R^{-1}\boldsymbol{X}^* & \boldsymbol{X}^{*\top}R^{-1}\boldsymbol{Z}^* \\ \boldsymbol{Z}^{*\top}R^{-1}\boldsymbol{X}^* & (\boldsymbol{Z}^{*\top}R^{-1}\boldsymbol{Z}^* + G^{-1}) \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^{*\top}R^{-1}\boldsymbol{y}^* \\ \boldsymbol{Z}^{*\top}R^{-1}\boldsymbol{y}^* \end{pmatrix} \quad (16)$$

The estimates are obtained below as

$$\widehat{\boldsymbol{\beta}} = ((BX)^{\top}\boldsymbol{H}^{-1}(BX))^{-1}(BX)^{\top}\boldsymbol{H}^{-1}BA\boldsymbol{y} \quad (17)$$

Similarly, $$\hat{b}^{(i)} = G(BZ)^{\top}\boldsymbol{H}^{-1}(A\boldsymbol{y} - BX\widehat{\boldsymbol{\beta}}) \quad (18)$$

### 2.4 Variance Estimation

The REML for the m-SVSOM is given below as

$$-2R\mathcal{L}(\omega_{si}, \sigma^2; \boldsymbol{y}^*) = (n-p)log\sigma^2 + \log|H| + \log|X^{*\top}\boldsymbol{H}^{-1}X^*| + \frac{(\boldsymbol{Ay})^{\top}P(\boldsymbol{Ay})}{\sigma^2} \quad (19)$$

For which the estimated variance can be shown to be

$$\hat{\sigma}^2 = \frac{(\boldsymbol{Ay})^{\top}P(\boldsymbol{Ay})}{(n-p)} \quad (20)$$

### 2.5 Parameter Estimation with the Outlier Parameter

The estimates of the fixed and random effects are analogous to Equations (17) and (18) above. Following (Gilmour, 1997; Meyer, 1989), the mixed model equation from Equation (9) is analogous to the Mixed Model equation of Linear Mixed Model in Equation (16) and can be expressed below as;

$$F_i\binom{\varphi_i}{\theta} = \begin{pmatrix} \boldsymbol{\delta_i}^{*\top} R^{-1} y^* \\ W^\top R^{-1} y^* \end{pmatrix} \tag{21}$$

where $W = [\boldsymbol{X}^* \ \boldsymbol{Z}^*]$, $\hat{\theta} = (\hat{\beta}^\top, \hat{b}^\top)$ with the solutions for θ corresponding to Equations (17) and (18), respectively and $\varphi_i$ is similar to Equation (3).

Where $F_i = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}$

The REML log-likelihood function is given below as

$$-2R\mathcal{L}(\omega_{si}, \sigma^2; \boldsymbol{y}^*) = (n-p)log\sigma^2 + \log|H_i| + \log|\boldsymbol{X}^{*\top} H_i^{-1} \boldsymbol{X}^*| + \frac{(\boldsymbol{Ay})^\top P_i(\boldsymbol{Ay})}{\sigma^2} \tag{22}$$

Equation (22) can be re-expressed following Gumedze *et al*., (2010) as

$$+\frac{(\boldsymbol{Ay})^\top P(\boldsymbol{Ay})}{\sigma^2} - \frac{(\boldsymbol{Ay})^\top P\boldsymbol{\delta_i}^*(\boldsymbol{\delta_i}^{*\top} P\boldsymbol{\delta_i}^* + \omega_{si}^{-1})^{-1}\boldsymbol{\delta_i}^{*\top} P(\boldsymbol{Ay})}{\sigma^2} + \log(\boldsymbol{\delta_i}^{*\top} P\boldsymbol{\delta_i}^* \omega_{si} + 1) \tag{23}$$

Substituting $\boldsymbol{e_i} = (\boldsymbol{Ay})^\top P\boldsymbol{\delta_i}^*$ which is the conditional residual for unit i observation and $\boldsymbol{a_{si}} = \boldsymbol{\delta_i}^{*\top} P\boldsymbol{\delta_i}^*$ into Equation (23) results in

$$-2R\mathcal{L}(\omega_{si}, \sigma^2; \boldsymbol{y}^*) = (n-p)log\sigma^2 + \log|R| + \log|G| + \log|F| + \frac{(\boldsymbol{Ay})^\top P(\boldsymbol{Ay})}{\sigma^2} + \log(\boldsymbol{a_{si}}\omega_{si} + 1) - \frac{\boldsymbol{1}}{\sigma^2}\left(s_i^2 - \frac{s_i^2}{(\boldsymbol{a_{si}}\omega_{si}+1)}\right) \tag{24}$$

And can be further rewritten as

$$-2R\mathcal{L}(\omega_{si}, \sigma^2; \boldsymbol{y}^*) = (n-p-1)log\sigma^2 + \log|H| + \log|\boldsymbol{X}^* H^{-1} \boldsymbol{X}^*| + \frac{(\boldsymbol{Ay})^\top P(\boldsymbol{Ay}) - s_i^2}{\sigma^2} + \log\sigma^2(\boldsymbol{a_{si}}\omega_{si} + 1) + \frac{s_i^2}{\sigma^2(\boldsymbol{a_{si}}\omega_{si}+1)} \tag{25}$$

Simplifying further let $\sigma^2 = \Phi_1$ and $\sigma^2(\boldsymbol{a_{si}}\omega_{si} + 1) = \Phi_2$

Replacing back in Equation (25) gives

$$-2R\mathcal{L}(\Phi_1, \Phi_2; \boldsymbol{y}^*) = (n-p-1)log\Phi_1 + \log|H| + \log|\boldsymbol{X}^* H^{-1} \boldsymbol{X}^*| + \frac{(\boldsymbol{Ay})^\top P(\boldsymbol{Ay}) - s_i^2}{\Phi_1} + \log\Phi_2 + \frac{s_i^2}{\Phi_2} \tag{26}$$

Take partial derivatives of with respect to $\Phi_1$ gives

$$\hat{\Phi}_1 = \frac{(\boldsymbol{Ay})^\top P(\boldsymbol{Ay}) - s_i^2}{(n-p-1)}$$

substituting $s_i^2 = t_{si}^2\,\hat{\sigma}^2$ ; $(n-p)\hat{\sigma}^2 = (\boldsymbol{Ay})^\top P(\boldsymbol{Ay})$ gives

$$\hat{\Phi}_1 = \widehat{\sigma^2}_i = \frac{(n-p-t_{si}^2)\hat{\sigma}^2}{(n-p-1)} \tag{27}$$

Similarly, taking the partial derivative with respect to $\Phi_2$ gives

$\hat{\Phi}_2 = s_i^2$ (28)

Re-expressed as

$\sigma^2(\boldsymbol{a_{si}}\omega_{si} + 1) = s_i^2$

And can be rewritten below as

$\hat{\omega}_{si} = \frac{(n-p)(t_{si}^2-1)}{(n-p-t_{si}^2)(\mathbf{1}-\boldsymbol{v_{si}})}$ (29)

**Simulation Study**

A simulation study is conducted to evaluate the performance of the m-SVSOM model relative to the SVSOM. The models are simulated for several sample sizes, considering n=16, 36, 144, 400, 900,1000, spatial weights $W_1 = W_2$ were constructed using a row-standardized Queen's contiguity. Random effects were generated from a normal distribution, and outliers were introduced via variance inflation at randomly selected indices. Initial values of $\beta_0 = 2.0$, $\beta_1 = 0.5$,$\beta_2 = 0.5$, $\sigma^2 = 1.0$ ,$\rho = 0.5$ and $\lambda = 0.3$ . Two random effects, $b_1$ and $b_2$, were independently drawn from N (0, 1) for each simulation replicate. The simulation was run 10,000 times, and the performance metric used is the Mean Squared Error.

## 3. Results

Table 1: Model Performance Under Different Sample Sizes

| | **MSE** | |
|---|---|---|
| N | **m-SVSOM** | SVSOM |
| 16 | **0.1687** | 0.2254 |
| 36 | **0.1021** | 0.1399 |
| 144 | **0.0431** | 0.0627 |
| 400 | **0.0230** | 0.0347 |
| 900 | **0.0138** | 0.0217 |
| 1000 | **0.0131** | 0.0205 |

(*Based on 10,000 simulations; MSE averaged across replicates)*[1]

Table 1 above presents the MSE for the two models across sample sizes (n=16, 36, 144, 400, 900, 1000) in a simulation study focused on the models' outlier-detection capabilities. The results showed a consistent decline in MSE, with the m-SVSOM achieving the lowest MSE across varying sample sizes compared to the SVSOM model.

Table 2: Overall Average Performance of the Models

| Model | MSE | RMSE |
|---|---|---|
| m-SVSOM | 0.0522 | 0.2108 |
| SVSOM | 0.0735 | 0.2525 |

(*Lower MSE and RMSE indicate better predictive accuracy)*[2]

Table 2 shows that the m-SVSOM dominates on the two-error metrics, i.e., (MSE = 0.0522) and (RMSE = 0.2108), followed by the SVSOM with (MSE = 0.0735) and (RMSE = 0.2525)

Table 3: Parameter Estimation

| | | Parameters | | | | | MSE |
|---|---|---|---|---|---|---|---|
| **Model** | **(λ, ρ)** | $\hat{\beta}_0$ $(s.e)$ | $\hat{\beta}_1$ $(s.e)$ | $\hat{\beta}_2$ $(s.e)$ | $\hat{b}_1$ $(s.e)$ | $\hat{b}_2$ $(s.e)$ | |
| **m- SVSOM** | **(0.3, 0.5)** | **2.0009 (0.2048)** | **0.4984 (0.0977)** | **0.5069 (0.1044)** | **-0.0026 (0.3031)** | **0.0011 (0.3028)** | **0.0522** |
| SVSOM | (0.3,0.5) | 2.0106 (0.2024) | 0.4963 (0.1004) | 0.5000 (0.0997) | | | 0.0735 |

From Table 3, it can be deduced that the SVSOM does not incorporate a random effect, but the m-SVSOM does. It can be deduced that the m-SVSOM achieves the lowest prediction error (MSE=0.0522), implying that its parameter estimates converge more closely to the true parameters than those of other models with λ=0.3 and ρ=0.5.

## 4. Discussion

The proposed m-SVSOM consistently outperformed the SVSOM across all sample sizes, with lower MSE and RMSE (Table 2). The inclusion of random effects and spatial lag terms allowed the model to account for hierarchical data structures that SVSOM ignores. These results align with Gumedze et al. (2010) and Baba et al. (2022), but extend their work by simultaneously handling spatial dependence and within-group correlation. The decreasing MSE with larger n (Table 1) supports the asymptotic efficiency of REML estimation. However, the model assumes normality and a single spatial weights matrix, which may not hold in all applied settings. Future work should explore non-Gaussian responses (e.g., Poisson, binomial) and time-varying spatial weights.

## 5. Conclusion

This research successfully extended the Spatial Variance Shift Outlier Model (SVSOM) by integrating spatial weight matrices with fixed- and random-effects components, the m-SVSOM provides a comprehensive approach to handling data that exhibit both spatial dependence and intra-group correlation.

The simulation results show that the m-SVSOM outperforms SVSOM across all tested sample sizes. Specifically, the m-SVSOM achieved a significantly lower MSE (0.0522 vs 0.0735 for the SVSOM) and had parameter estimates (fixed and random effects) that more accurately approached their true values. The steady decrease in error metrics with increasing sample size further validates the asymptotic efficiency of REML estimation in this m-SVSOM.

In conclusion, the m-SVSOM accounts for spatial disparities and correlated measures simultaneously, thus enhancing the precision of regression coefficient estimates and improving the overall predictive performance of spatial statistical analyses.

**Limitation and Future Research**

While m-SVSOM captures both spatial disparities and random effects within a group, it assumes a normally distributed response with an identity link function, a single spatial weight structure, and a constant ρ; this can be extended to other distributions in the exponential family.

**References**

Baba, A.M., Midi, H., Adam, M.B. & Rahman, N.H.A.(2022). Spatial Outlier Accommodation Using a Spatial Variance Shift Outlier Model. *Mathematics*, 10,1-19.

Babadi, B., Rasekh, A., Zare, K. & Rasekhi, A.A. (2016). A Variance Shift Model for Detection of Outliers in the Linear Mixed Measurement Error Models. *Communications in Statistics - Theory and Methods*, 45(24), 7350–7366.

Barnett V, Lewis T (1994) Outliers in statistical data. Wiley, New York

Beckman, R.J. & Cook, R.D. (1983). Outliers. *Technometrics*, 25, 119–149.

Cai, J., Hu, W., Yang, Y., Yan, H. & Chen, F. (2024). Outlier detection in spatial error models using modified thresholding-based iterative procedure for outlier detection approach. *BMC Medical Research Methodology*,24(1), 1-14.

Cook, R.D. &Weisberg, S.(1982). Residuals and Influence in Regression; Chapman and Hall: New York, NY, USA.

Dai, X., Jin, L., Shi, A., & Shi, L.(**2016).** Outlier Detection and Accommodation in General Spatial Models. *Statistical Methods Application*, 25, 453–475.

Gharbaghi, K., Babadi, B. & Rasekh, A. (2026). Identifying Outliers Using the Variance Shift Method in the Linear Mixed Model Under Linear Stochastic Constraint. *Journal of the Indian Society for Probability and Statistics*

26(2) 1–21.

Gilmour, A.R., Cullis, B.R., & Verbyla, A.P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 51, 269–273.

Gumedze, F.N. (2018). Use Of Likelihood Ratio Tests to Detect Outliers Under the Variance Shift Outlier Model. *Applied Statistics*,44,1-24.

Gumedze F.N & Chatora,T.D. (2014). Detection of outliers in longitudinal count data via Overdispersion. *Computational Statistics and Data Analysis*, 79, 192-202.

Gumedze, F.N. & Dunne, T.T. (2011). Parameter Estimation and Inference in the Linear Mixed Model. *Linear Algebra and its Applications*,435, 1920–1944

Gumedze, F.N., Welham, S.J., Gogel, B.J. & Thompson. R. (2010). A Variance Shift Model for Detection of Outliers in the Linear Mixed Model. *Computing Statistical Data Analysis*, 54, 2128–2144.

Ismadyaliana, S., Setiawan & Purnomo, J.D.T. (2024). Panel Data Modelling: Identifying and Handling Outliers with the VSOM Approach. *MethodsX,* 13, 1-15.

Kowalskie, A. (2025). The Impact of Outliers on Linear Regression Models: Detection and Correction Strategies. *OTS Canadian Journal*, *4*(6), 108-118.

Lehmann, R., Lösler, M. & Neitzel, F.(2020). Mean Shift versus Variance Inflation Approach for Outlier Detection—A Comparative Study. *Mathematics*, 8, 1-21.

Maestrini, L., Hui1, F. K.C. & Welsh, A. H.(2024). Restricted Maximum Likelihood Estimation in Generalized Linear Mixed Models. *Biometrika.* 1-59.

Melo, O. O., Mateu, J. & Melo, C. E.(2013). Spatial Generalized Linear Mixed Models Based on Distances. *Statistical Methods in Medical Research*, 1-23.

Moore, D.S. & McCabe, G.P. (1999). Introduction to the Practice of Statistics, 3rd edition. W.H.Freeman, New York.

Monhor, D & Takemoto, S (2005). Understanding the Concept of Outlier and Its Relevance to The Assessment of Data Quality: Probabilistic Background Theory. *Earth Planets Space*, **57**, 1009–1018.

Meyer K (1989) Restricted Maximum Likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet Select Evol* 21, 317–340.

Petronilla, O.U. & Chinaka, N.E. (2023). A review on the advances of outliers. *Biometrics & Biostatistics International Journal*, 12(1), 8–13.

Shukla, S., Lalitha, S. & Srivastava, P. (2023). Accommodation Of Outliers by Robust MML Estimation for Spatial Autoregressive Model. *International Journal of System Assurance Engineering and Management*, 14, 293–306.

Smiti, A. (2020). A Critical Overview of Outlier Detection Methods. *Computer Science Review*, 38,1-11.

Song, Y., Fang, M., Wang, Y. & Hou, Y. (2024). Rapid Outlier Detection, Model Selection, and Variable Selection Using Penalized Likelihood Estimation for General Spatial Models. *Journal of Spatial Statistics, 61,* 1-14.

Thompson, R. (1985). "A Note On Restricted Maximum Likelihood Estimation with An Alternative Outlier Model". *Journal of the Royal Statistical Society: Series B (Methodological)* 47(1), 53–55.

Torabi, M. (2014). Spatial Generalized Linear Mixed Models with Multivariate CAR Models for Areal Data. *Spatial Statistics,* 1-33.

Vijayalakshmi S.,Nicy, S. & Sajesh T.A (2025). Unveiling Outliers with Robust Covariance Matrix Estimation: A Shrinkage Approach. *Stat Papers,* 66, 1059–1085.