FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 15, August, 2026, pp 1 – 4
DOI: https://doi.org/10.33003/fjs-2026-1015-5489

Comparative Analysis of the Modified Spatial Variance Shift Outlier Model

[1]Ehiosasu Idusuyi, [2]Adesupo Adeoye Akinrefon, [2]Emmanuel Torsen & [4]Abraham Okolo

[1]Department of Statistics, Abubakar Tafawa Balewa University, Bauchi, Nigeria.
[2]Department of Statistics, Modibbo Adama University, Adamawa, Nigeria

*Corresponding authors' email: Osasu20@gmail.com

ABSTRACT

Spatial outliers pose substantial challenges in spatial data analysis, often distorting parameter estimates and leading to misleading conclusions. Various accommodation strategies have evolved through variance-shift frameworks and hierarchical robust approaches, but comprehensive evaluations across multiple contamination levels remain limited. While standard frameworks such as Mixed Models (MM), Spatial Autoregressive (SAR) models and Spatial Generalized Linear Mixed Models (SGLMM) address specific data complexities like random effects or spatial dependencies, they generally fail to account for spatial outliers concurrently. This study evaluates the performance of the Modified Spatial Variance Shift Outlier Model (m-SVSOM) in comparison with these earlier methods and the existing Spatial Variance Shift Outlier Model (SVSOM). A simulation study was conducted across various sample sizes (n = 16, 100, 256, 1000) and varying outlier proportions (5%, 10%, 15%, and 20%). The estimation accuracy of the proposed model was assessed using the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The results consistently demonstrate that the m-SVSOM outperforms the Mixed Model, SAR, SGLMM, and the SVSOM across all scenarios. Specifically, the Modified SVSOM maintained the lowest error metrics even as contamination levels increased, indicating superior robustness in accommodating outliers and managing spatial dependencies. These findings suggest that the m-SVSOM provides a more robust and efficient framework for accommodating spatial outliers while retaining the information contained in influential observations.

Received: 10 July 2026
Accepted: 12 Aug. 2026
Published: 27 Aug. 2026

Keywords: Outliers, m-SVSOM, MSE, SGLMM, SAR, Spatial dependency

INTRODUCTION

The presence of outliers can substantially affect model performance, resulting in biased parameter estimates and misleading conclusions (Kowalskie, 2025). Model performance is typically assessed using Mean Squared Error (MSE), which measures the relative difference between predicted and true values within a dataset (Hodson *et al.*, 2021). Statistical models describe the presumed impact of explanatory variables and the probability distributions associated with aspects of the process which are assumed to be characterized by random variation (Stroup, 2016).

While Mixed Models incorporate both fixed effects, representing population-level parameters, and random effects, which capture variation among groups or subjects (Breslow, 1984; Breslow & Clayton, 1993; Jiang, 2021; Williams, 1982), spatial models extend these frameworks to account for geographic dependencies. The Spatial generalized linear mixed model (SGLMM) which was first introduced by Diggle *et al.*, (1998) captures either random effect or within group error which may or may not be spatially correlated (Boubeta *et al.*, 2023; Vernebles & Ripley, 2002) and the Spatial autoregressive model (SAR) captures the dependency in the outcome variable of one region influenced by the outcome of surrounding or neighboring regions (Sirait, 2023). Chavez-Chong *et al.* (2025) demonstrated that SAR models suffer from multicollinearity issues and proposed ridge regularization with spatial cross-validation to improve estimation and prediction.

One critical limitation of the MM, SAR, and SGLMM frameworks is that they do not explicitly accommodate spatial outliers. While MM and SGLMM include fixed and random effects, MM fails to account for spatial dependence, and neither framework systematically addresses outliers. SAR and SGLMM models incorporate spatial dependencies but remain highly sensitive to contamination. Recent literature has begun to address this gap by developing specialized outlier-accident strategies. Cheng and Song (2023) proposed simultaneous outlier detection and variable selection for spatial Durbin models, integrating these steps to improve interpretability. In the mixed-model domain, Sugasawa *et al*. (2025) introduced hierarchical gamma-divergence for linear mixed models, down-weighting outliers at both the error and random-effects levels, with automated tuning and scalable algorithms.

Traditional approaches often discard outlying data points as noise. However, a more efficient strategy is outlier accommodation, which adjusts the model to incorporate these values without losing the unique information they may provide (Akeyede *et al.*, 2015; Baba *et al.*, 2022; Dai *et al.*, 2016; Moore & McCabe, 1999). Baba *et al.* (2022) introduced the Spatial Variance Shift Outlier Model (SVSOM), which adapts the variance-inflation idea to spatial regression by revising the model for flagged outlier observations and estimating via REML, Idusuyi *et al.*, (2026) introduced a modified spatial variance shift outlier model (m-SVSOM).

Although several studies have employed outlier detection and accommodation in the linear regression model or the spatial model (Baba *et al*., 2022; Cook & Weisberg, 1982; Dai *et al.*, 2016, Gumedze, *et al*., 2010; Song *et al.*, 2024), an extensive evaluation of the performance of these existing models with the robust outlier detection models have not been dealt with where outliers are introduced at various levels. This forms the basis of this study as we seek to assess the performance across models (MM, SGLMM, SAR, SVSOM) and the m-SVSOM at various sample sizes while varying outlier proportion

MATERIALS AND METHODS

The Statistical Models

Mixed Model (MM)

The Mixed model can be expressed as

$y = X\beta + Zb, \quad b \sim N(0, G)$ (1)

Where y= (n x1) column vector of response variable, X= $nx\,p$ matrix of fixed effects, β = Coefficient of the fixed effect, Z = $n\,xq$ matrix of random effects and b = coefficient of the random effect and G is the variance covariance matrix of the random effect.

Spatial Generalized Linear Mixed Model (SGLMM)

$g(\mu_i) = \eta = X\beta + Zb, \quad ; b = \lambda Wb + u$ (2)

Where $g(.)$ = Link function, μ_i= Expected mean of the observations, $X = n\,xp$ matrix of fixed effect, β = Coefficient of the fixed effect, $Z = nxq$ matrix of random effect, b = coefficient of the random effect , λ = the spatial autocorrelation coefficient in the random effect and $W = n \times n$ Spatial weights matrix.

Spatial Autoregressive Model (SAR)

$y = \rho Wy + X\beta + e$ (3)

Where y is an (nx1) vector of response variable, X is an (n x k) matrix of independent variables, β is a (kx1) vector of regression parameters, $W = n \times n$ Spatial weights matrix, Wy= spatial lagged dependent variable and e is the (n x 1) vector of the error term.

Spatial Variance Shift Outlier Model (SVSOM)

Baba et al., (2022) proposed the SVSOM and is given below as

$A\eta = X\beta + \delta_i\varphi_i + B^{-1}e$ (4)

Where A is the spatial lag dependent matrix operator and B spatial autocorrelation matrix operator and $A = (I - \rho W_1)$, B = $(I - \lambda W_2)$, ρ= the spatial autoregressive coefficient,$W_1 and\ W_2 = n \times n$ Spatial weights matrix, λ = coefficient on the spatial autocorrelations in the random error term, X is an (n x k) matrix of independent variables, β is a (kx1) vector of regression parameters. $\varphi_i \sim N(0, \sigma^2\delta_i\omega_{si}\delta_i^T)$ is defined as the unknown random effect and ω_{si} is the weight measure that determines the inflation in the variance parameter, δ_i is the dummy variable that assigns 1 if it is an outlier and zero otherwise.

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

Following Idusuyi *et al*., (2026), the modified model is given by:

$A\eta = X\beta + Zb + \delta_i\varphi_i + B^{-1}e$ (5)

Where A is the spatial lag dependent matrix operator and B spatial autocorrelation matrix operator and $A = (I - \rho W_1)$, B = $(I - \lambda W_2)$, ρ= the spatial autoregressive coefficient, $W_1 and\ W_2 = n \times n$ Spatial weights matrix, λ = coefficient on the spatial autocorrelations in the random error term, X is an (n x k) matrix of independent variables, β is a (kx1) vector of regression parameters, Z is an (n x q) matrix of random effects, and b is a (qx1) vector of random effects. $\varphi_i \sim N(0, \sigma^2\delta_i\omega_{si}\delta_i^T)$ is defined as the unknown random effect and ω_{si} is the weight measure that determines the inflation in the variance parameter, δ_i is the dummy variable that assigns 1 if it is an outlier and zero otherwise.

Numerical Analysis Study

A comprehensive simulation study was undertaken to evaluate the performance of the modified SVSOM model relative to the SVSOM, MM, SGLMM, and SAR models. The analysis considered a range of sample sizes (n = 16, 100, 256, 1000) and varying outlier proportions (5%, 10%, 15%, and 20%). Spatial weight matrices ($W_1 = W_2$) were specified using row-standardized Queen's contiguity. Random effects were generated from a normal distribution, while outliers were introduced via variance inflation at randomly selected locations. The initial parameter values were set to $\beta_0 = 2.0$, $\beta_1 = 0.5, \beta_2 = 0.5, \sigma^2 = 1.0, \rho = 0.5,$ and $\lambda = 0.3$. For each simulation replicate, two independent random effects, $b_1$and b_2, were drawn from $N(0,1)$. The experiment was repeated 10,000 times, with model performance assessed using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE)

RESULTS AND DISCUSSION

Table 1: Performance of the Models under Outlier Levels at n=16 (Mean Results over 10,000 Simulation Replications)

Error	Outlier	Mixed Model	m-SVSOM	SVSOM	SAR	SGLMM
MSE	5%	0.3105	**0.1519**	0.2005	0.3621	0.2585
	10%	0.3345	**0.1632**	0.2153	0.3882	0.2828
	15%	0.3579	**0.1744**	0.2338	0.4251	0.3032
	20%	0.3915	**0.1886**	0.2489	0.4494	0.3253
RMSE	5%	0.5572	**0.3897**	0.4477	0.6016	0.5083
	10%	0.5783	**0.4039**	0.4639	0.6230	0.5317
	15%	0.5982	**0.4176**	0.4834	0.6519	0.5505
	20%	0.6256	**0.4343**	0.4988	0.6703	0.5703

Table 1 shows that the best-performing model at the various outlier levels was the m-SVSOM with the lowest MSE values at 5% (MSE=0.1519), (RMSE=0.3897), at 10% (MSE=0.1632), (RMSE=0.4039), at 15% (MSE=0.1744), (RMSE=0.4176), at 20% (MSE=0.1886), (RMSE=0.4343) as compared with SVSOM, Mixed Model, SAR and Spatial GLMM.

Table 2: Performance of the Models Under Outlier Levels at n=100 (Mean Results over 10,000 Simulation Replications)

Error	Outlier	Mixed Model	m-SVSOM	SVSOM	SAR	SGLMM
MSE	5%	0.1151	**0.0482**	0.0696	0.1395	0.0930
	10%	0.1257	**0.0525**	0.0756	0.1493	0.0995
	15%	0.1336	**0.0564**	0.0803	0.1617	0.1096
	20%	0.1438	**0.0602**	0.0862	0.1734	0.1177
RMSE	5%	0.3393	**0.2195**	0.2638	0.3735	0.3050
	10%	0.3545	**0.2290**	0.2749	0.3864	0.3155

Error	Outlier	Mixed Model	m-SVSOM	SVSOM	SAR	SGLMM
	15%	0.3655	**0.2375**	0.2834	0.4021	0.3311
	20%	0.3792	**0.2454**	0.2936	0.4163	0.3430

Table 2 shows that the best-performing model at the various outlier levels was the m-SVSOM with the lowest MSE values at 5% (MSE=0.0482), (RMSE=0.2195), at 10% (MSE=0.0525), (RMSE=0.2290), at 15% (MSE=0.0564), (RMSE=0.2375), at 20% (MSE=0.0602), (RMSE=0.2454), The other models considered were SVSOM, Mixed Model, SAR, and SGLMM.

Table 3: Performance of the Models under Outlier Levels at n=256 (Mean Results over 10,000 Simulation Replications)

Error	Outlier	Mixed Model	m-SVSOM	SVSOM	SAR	SGLMM
MSE	5%	0.0695	**0.0268**	0.0400	0.0844	0.0545
	10%	0.0748	**0.0290**	0.0434	0.0923	0.0600
	15%	0.0805	**0.0317**	0.0467	0.0994	0.0648
	20%	0.0876	**0.0339**	0.0500	0.1076	0.0690
RMSE	5%	0.2637	**0.1638**	0.1999	0.2905	0.2334
	10%	0.2735	**0.1703**	0.2084	0.3038	0.2449
	15%	0.2838	**0.1779**	0.2160	0.3153	0.2545
	20%	0.2960	**0.1842**	0.2235	0.3280	0.2626

Table 3 shows that the best-performing model at the various outlier levels was the m-SVSOM with the lowest MSE values at 5% (MSE=0.0268), (RMSE=0.1638), at 10% (MSE=0.0290), (RMSE=0.1703), at 15% (MSE=0.0317), (RMSE=0.1779), at 20% (MSE=0.0339), (RMSE=0.1842), performing groups are the SVSOM, Mixed Model, SAR and SGLMM.

Table 4: Performance of the Models under Outlier Levels at n=1000(Mean Results over 10,000 Simulation Replications)

Error	Outlier	Mixed Model	m-SVSOM	SVSOM	SAR	SGLMM
MSE	5%	0.0334	0.0116	0.0180	0.0418	0.0254
	10%	0.0354	0.0126	0.0198	0.0458	0.0279
	15%	0.0389	0.0134	0.0210	0.0488	0.0298
	20%	0.0419	0.0146	0.0228	0.0523	0.0321
RMSE	5%	0.1828	0.1078	0.1341	0.2044	0.1593
	10%	0.1881	0.1121	0.1408	0.2140	0.1671
	15%	0.1972	0.1158	0.1449	0.2208	0.1727
	20%	0.2048	0.1206	0.1510	0.2287	0.1790

Table 4 shows that the best-performing model at the various outlier levels was the m-SVSOM with the lowest MSE values at 5% (MSE=0.0116), (RMSE=0.1078), at 10% (MSE=0.0126), (RMSE=0.1121), at 15% (MSE=0.0134), (RMSE=0.1158), at 20% (MSE=0.0146), (RMSE=0.1206) when compared to the SVSOM, Mixed Model, SAR and Spatial GLMM.

CONCLUSION

This study evaluated the m-SVSOM against traditional Mixed Models (MM), SVSOM, SAR models and SGLMM under varying sample sizes and levels of outlier contamination. Across all tested sample sizes from as low as 16 to a sample size of 1000, and all outlier levels (5%, 10%, 15%, and 20%), the m-SVSOM consistently yielded the lowest MSE and RMSE. This indicates that the model effectively accommodates outliers rather than simply discarding them as noise. While all models showed improved accuracy (lower MSE) as the sample size increased, the m-SVSOM maintained a significant gap relative to the other models, even at a high contamination level of 20%. For instance, at n=1000, the m-SVSOM achieved an MSE of 0.0146 under 20% contamination, whereas the SAR and Mixed Models recorded significantly higher MSEs of 0.0523 and 0.0419, respectively. From a computational perspective, the study used 10,000 simulation replications, indicating that the reported performance comparisons are based on repeated estimation rather than on a single simulated sample.

Future research should extend the comparison to non-normal error distributions and a real-data application to evaluate predictive performance and substantive interpretability in addition to MSE and RMSE. These extensions would provide stronger evidence concerning the generalizability, robustness, and practical utility of the m-SVSOM.

REFERENCES

Akeyede, I., Saleh, I.M. & Babalola, O.A. (2015). Proportional Effect of outliers on over dispersion. *FULafia Journal of Science and Technology*, 1(1), 95-100. https://lafiascijournals.org.ng/index.php/fjst/article/view/16.

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. https://doi.org/10.3390/math10173182.

Boubeta, M., Lombardía, M.J. & Morales, D. (2023). Small Area Prediction of Proportions and Counts Under a Spatial Poisson Mixed Model. *Statistical Methods & Applications*, 33, 1193-1215. DOI:10.1007/s10260-023-00729-7.

Breslow, N. E. (1984). Extra-Poisson Variation in Log-linear models. *Applied Statistics,* 33,38–44. https://doi.org/10.2307/2347661

Breslow, N. E. & D. G. Clayton (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the*

American Statistical Association, 88, 9–25. https://doi.org/10.2307/2290687

Chavez-Chong, C.O., Hardouin, C. & Fermin, A. (2025). Ridge Regularization for Spatial Autoregressive Models with Multicollinearity Issues. *AStA Advances in Statistical Analysis*, 109(1), 25–52. DOI:10.1007/s10182-024-00496-0.

Cheng, Y. & Song, Y. (2023). Simultaneous Outlier Detection and Variable Selection for Spatial Durbin Model. *Brazilian Journal of Probability and Statistics*, *37*(3), 596–618. https://doi.org/10.1214/2-BJPS583

Cook, R.D. &Weisberg, S.(1982). Residuals and Influence in Regression; Chapman and Hall: New York, NY, USA. https://doi.org/10.1002/bimj.4710270110

Dai, X., Jin, L., Shi, A. & Shi, L.(**2016).** Outlier Detection and Accommodation in General Spatial Models. *Statistical Methods Application*, 25, 453–475. DOI: 10.1007/s10260-015-0348-1

Diggle, P. J., Tawn, J. A. & Moyeed, R. A. (1998). Model-Based Geostatistics. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 47(3), 299–350. https://www.jstor.org/stable/2986101

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. *Computational Statistics & Data Analysis*, 54, 2128–2144. https://doi.org/10.1016/j.csda.2010.03.019

Hodson, T.O., Over, T.M. & Foks, S.S. (2021). Mean Squared Error , Deconstructed. *Journal of Advances in Modeling Earth Systems*, 13, 1-10. https:// doi.org/10.1029/2021MS002681

Idusuyi, E., Akinrefon, A.A., Torsen, E. & Okolo, A. (2026). A Modified Spatial Variance Shift Outlier Model. *International Journal of Development Mathematics,* 3(2),351 – 361. DOI: 10.62054/ijdm/0302.22.

Jiang, J. (2021) Linear and Generalized Linear Mixed Models and Their Applications, 2nd ed. New York: Springer. DOI: 10.1007/978-1-0716-1282-8.

Kowalskie, J. (2025). The Impact of Outliers on Linear Regression Models: Detection and Correction Strategies. *OTS Canadian Journal*, 4(6), 108-118. DOI: 10.58840/fzbcv732

Moore, D.S. & McCabe, G.P. (1999). Introduction to the practice of Statistics, 3rd edition. W.H.Freeman, New York. DOI:https://doi.org/10.2307/3619120.

Sirait, T. (2023). Spatial Autoregressive Modelling on Linear Mixed Models for Dependency Between Regions, *Aceh International Journal of Science and Technology*, 12(1), 40-52. DOI: 10.13170/aijst.12.1.30403.

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(2), 1-14. DOI: 10.1016/j.spasta.2024.100834.

Stroup, W.W. (2016). Generalized Linear Mixed Models: Modern Concepts, Methods and Applications, Boca Raton, FL: CRC Press. https://doi.org/10.1177/0962280214563202

Sugasawa, S., Hui, F.K.C. & Welsh. (2025). Robust Linear Mixed Models using Hierarchical Gamma-Divergence. *Journal of Computational and Graphical Statistics*, *35*(2), 720–733. https://doi.org/10.1080/10618600.2025.2565235.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics* with S. Springer, New York, Fourth Edition. DOI: 10.1007/b97626.

Williams, D. A. (1982). Extra-Binomial Variation in Logistic Linear Models. *Applied Statistics*, 31, 144–148. https://doi.org/10.2307/2347977.