

Response to Journal's 2nd Reviewer Comments
Conditionally Autoregressive Models for House Price Data: Insights From a
Comparative Simulation Study

Indira Puteri Kinasih

1. **The choice of W (whether binary adjacency or row-standardized) is critical. Have the authors explored alternative spatial weight matrices, and how do they impact model performance?**

Thank you for this insightful comment regarding the specification of the spatial weight matrix, W . We fully acknowledge that the choice of W —whether binary adjacency, row-standardized, or distance-based—can substantially influence the performance and interpretation of spatial models. In the present study, our primary objective was to evaluate and compare the performance of CAR, SAR, and GWR models using both simulated data and real-world data from Lombok. To maintain consistency across models and to isolate the structural differences in model behavior, we have, at this stage, limited our analysis to a binary adjacency matrix for W . We agree that exploring alternative specifications of W (e.g., row-standardized or distance-decay matrices) is a valuable direction for further research. We plan to extend our analysis in future work to assess how varying spatial weight structures affect the robustness and predictive accuracy of these spatial models.

2. **While the CAR model shows lower bias and RMSE in the artificial study, does this advantage persist in real-world scenarios, particularly where data sparsity or boundary effects exist?**

Thank you for your thoughtful question. We acknowledge that in small-area data, bias can indeed become a concern due to limited information at the unit level. However, in the case of the multilevel CAR (mlvCAR) model, each area contains multiple observations, allowing us to model area-specific effects—functionally resembling a random effects structure. This hierarchical structure helps mitigate the impact of sparsity within individual areas.

Regarding sample size, prior studies suggest that having approximately 15–30 observations per area is generally sufficient for stable estimation in multilevel frameworks. Given our data structure aligns with this recommendation, we believe the mlvCAR model offers a robust approach, especially in real-world scenarios where heterogeneity and partial pooling across areas are important. We thus recommend the use of mlvCAR to address both the potential bias and boundary effects effectively.

3. **The estimated value of ρ (spatial dependence strength) varies widely (e.g., 0.48 for the multilevel CAR model). Have the authors examined how sensitive the model results are to changes in ρ ?**

In our study, we recognize that the value of ρ plays a crucial role in determining the extent of spatial autocorrelation captured by the model. While we did not conduct a dedicated sensitivity analysis across a range of fixed ρ values, our simulation framework inherently explores variation in ρ through repeated data generation processes, in which ρ is estimated as part of the model fitting. The observed variability in ρ (e.g., 0.48 in the multilevel CAR model) reflects the underlying spatial structure in the data rather than instability in the model.

That said, we agree that a more systematic sensitivity analysis—such as fitting the model under different fixed values of ρ —could offer further insights into model robustness. We consider this a valuable direction for future work, especially in scenarios with varying levels of spatial autocorrelation or data sparsity.

4. **The study discusses the spatial effects ϕ as a key advantage of the CAR model, but how can these effects be meaningfully interpreted for house price determinants? Would spatial effect maps or additional diagnostic tools enhance this understanding?**

Thank you for highlighting the importance of interpreting the spatial effects ϕ in the CAR model. We agree that while ϕ captures unobserved spatial heterogeneity, its interpretability in the context of house price determinants requires careful consideration. Rather than directly reflecting specific covariates, ϕ accounts for spatially structured residual variation—such as omitted neighborhood characteristics, local amenities, or spatial spillovers—that are not captured by observed variables.

To aid interpretation, we acknowledge that visualizing ϕ through spatial effect maps could be a powerful diagnostic tool. Such maps can reveal clusters of under- or over-prediction and provide insights into areas where additional covariates or local effects might be relevant. In the revised

version of this study, we will include a Lombok sub-district-level map displaying the estimated ϕ values to better communicate the spatial structure captured by the model and to guide further interpretation and refinement.

5. **Among the various spatial regression models, Geographically Weighted Regression (GWR) is particularly prominent in property price research (page 2). Before reaching this conclusion, did you compare GWR to other spatial regression models?**

Thank you for this thoughtful question. In our manuscript, we highlight Geographically Weighted Regression (GWR) as a prominent method in the literature, particularly due to its frequent application in house price studies. This observation is based on a review of previous works rather than a performance-based comparison. While GWR is indeed widely used for its ability to model spatial heterogeneity at the local level, our study goes further by systematically comparing GWR with CAR and SAR models through simulation and real data applications. This comparative analysis aims to evaluate their relative strengths and limitations more objectively.

6. **The scope of properties analysed includes residential houses and commercial real estate (page 16). How do the authors reconcile this with the abstract, which mentions only house price data?**

Thank you for your observation. We acknowledge that the description in the manuscript may not have been sufficiently clear. The original dataset obtained through web-scraping did include a broad range of property types—such as residential houses, villas, boarding houses, and vacant land. However, for the purposes of this study, we restricted our analysis to residential houses only. We will revise the manuscript accordingly to ensure that this scope is clearly stated in both the abstract and the data collection section.

7. **After data cleaning, the final dataset included 598 entries representing multiple observations per sub-district (page 16). Why does the distribution of these data not appear on the map (Figure 6)?**

Thank you for the observation. For clarity, we note that the Lombok house price dataset used in this study does not contain precise geographic coordinates for each observation. As a result, it is not possible to accurately display the distribution of the data points on a map. While each observation is linked to a sub-district, the absence of point-level location data limits our ability to visualize them spatially within those areas. Therefore, for a more informative visualization, we are preparing a heatmap of house prices at the sub-district level, which will reflect the spatial variation in prices across Lombok using aggregated data.

8. **The authors have applied point-level data using three different models (GLMM, GWR, and multilevel CAR) (page 17). How was this approach implemented, given that there is no explanation of how each observation was georeferenced?**

Thank you for this important observation. We clarify that GLMM and multilevel CAR models do not require explicit geographic coordinates for each individual observation. Instead, they rely on an area-level identifier (ID) that links each observation to a corresponding spatial unit (e.g., sub-district).

In contrast, GWR requires spatial coordinates for each observation. To meet this requirement, we generated random points around the centroid of each sub-district, using `st_sample()`, where the number of points corresponds to the number of observations in that area. Each point was placed within a 0.5 km radius from the centroid, ensuring a reasonable spatial spread while maintaining the area-specific context. This approach enabled us to simulate realistic spatial positions for property transactions in the absence of precise geolocation data.

9. **Why is Table 5 not referred to in the text?**

Yes. We already include it in the text, in the revise version.

10. **The authors have argued that point-level models like GWR are limited when precise location data are unavailable (page 18). However, for the Lombok House Price Data, the authors have used point-level models. How do you reconcile this contradiction?**

Thank you for highlighting this important point. As we noted in our response to Comment 8, the Lombok house price data do not come with precise geographic coordinates. In order to apply the `gwr()` function, which requires point-level location data, we generated synthetic coordinates by simulating random points around the centroid of each sub-district. While this workaround allowed us to run the GWR model, we acknowledge that it is not an ideal or practical solution for real-world applications. This limitation reinforces our argument that point-level models such as GWR can be challenging to apply meaningfully when precise location data are not available.

11. **In the Lombok house price analysis, no sensitivity analysis was conducted to test how different definitions of adjacency (W matrix) influence the results. Would the findings remain consistent if a different spatial weights structure was used?**

Thank you for your thoughtful observation. As mentioned in our response to Comment 1, in this study we used a binary adjacency matrix to define spatial relationships, primarily to maintain consistency across the CAR and SAR models and to focus on comparing their general performance. We acknowledge that the choice of spatial weights matrix (W) can significantly influence model outcomes. While we have not conducted a sensitivity analysis in this current version, we agree that testing alternative specifications—such as row-standardized or distance-based matrices—would be valuable. We consider this an important direction for future work to assess the robustness of our findings under different definitions of spatial connectivity.

12. **The study primarily evaluates CAR models within the Lombok context, but how well does it generalize to larger, high-density urban areas (e.g., Jakarta)? Would additional testing in a more complex metropolitan setting strengthen the validity of the conclusions?**

Thank you for this important point. Indeed, numerous studies have already explored property price modeling in large, high-density urban areas such as Jakarta, where spatial data availability and heterogeneity are typically greater. In contrast, there is relatively limited research focused on low-density or semi-rural contexts such as Lombok. Our study aims to address this gap by evaluating the applicability and performance of spatial models, particularly the CAR model, in a setting that has different data characteristics and spatial structures compared to major metropolitan areas.

We also acknowledge that in richer datasets—common in urban contexts where more granular, high-frequency, and multidimensional data are available—machine learning approaches may offer competitive or even superior predictive performance. However, our focus here is to assess interpretable spatial models in contexts where data limitations remain a challenge.