

Responses to the Editor's and Reviewers' Comments on the Paper
Entitled "*An Approach for Spatio-Temporal Air Quality Prediction
Integrating Swin-LSTM and Kriging Methods*"

(No.: Sustainability - 3510189)

Jiangquan Xie, Fan Liu, Shuai Liu, Xiangtao Jiang

The authors would like to thank the Editor and the anonymous reviewers for their valuable comments and suggestions.

Responses to Reviewer #2:

Comment 1.0:

This paper presents a well-structured and insightful study on spatio-temporal air quality prediction, integrating advanced deep learning techniques with kriging interpolation. The methodology is clearly explained, and the experimental results convincingly demonstrate the effectiveness of the proposed SwinLSTM model. The writing is precise, and the discussion is well-grounded in relevant literature, making the study a valuable contribution to the field. The balance between technical depth and clarity is commendable, making the paper accessible to both specialists and a broader audience. That being said, I do have a few minor suggestions for improvement.

Responses:

Thanks very much for your encouraging comments.

Comment 1.1:

A stronger introduction would benefit from a short paragraph clearly outlining the research gap. Right now, the paper does a great job summarizing existing methods, but it's not entirely clear what's missing from the literature or why a new approach is needed. Adding a brief discussion on what previous studies haven't addressed and how this work fills that gap would make the contribution stand out more clearly.

Responses:

Thank you to the reviewers for their valuable scientific advice. Based on your suggestions, a new section has been added to the introduction of the revised manuscript that clearly describes the shortcomings of existing studies and how this study fills this research gap. The limitations of existing methods in dealing with spatial heterogeneity and temporal continuity simultaneously are pointed out, as well as the fact that most existing models fail to fully utilize the spatial relationships between monitoring sites. By integrating the local-global processing capability of Swin Transformer and the temporal modeling advantage of LSTM, along with the spatial analysis function of Kriging interpolation, this study proposes a novel approach to address these

shortcomings. **Modified introduction content:**

Despite the progress made by deep learning models in regional air quality prediction, challenges remain in effectively capturing the complex spatiotemporal correlations between meteorological conditions and air pollutant concentrations. A critical gap in existing research is the difficulty in simultaneously addressing spatial heterogeneity and temporal continuity in air quality data. Most current models either focus predominantly on temporal patterns using recurrent architectures or prioritize spatial relationships using convolutional approaches, but rarely integrate both dimensions optimally. Additionally, the majority of existing models fail to fully leverage the spatial relationships between monitoring stations, particularly in regions with sparse monitoring networks. These limitations result in prediction inaccuracies, especially for medium to long-term forecasting and in areas with complex terrain or meteorological conditions.

The spatiotemporal air quality prediction task is akin to video prediction, requiring the extraction of temporal and spatial dynamic patterns from continuous time series data. In air quality prediction, the focus is on the diffusion and evolution trends of pollutant concentrations across spatial and temporal dimensions. In the field of computer vision (CV), the Swin Transformer has recently demonstrated outstanding performance in various vision tasks by effectively integrating local spatial information and global context through its innovative shifted window attention mechanism, which computes self-attention within local windows [18][19]. Inspired by this, this study proposes a novel spatiotemporal air quality prediction method that integrates kriging interpolation with a SwinLSTM model. First, kriging interpolates the collected one-dimensional station data onto a two-dimensional spatial plane. Then, a prediction network integrating Swin-Transformer and LSTM modules is constructed to capture more comprehensive spatiotemporal dependencies in the air quality prediction task.

Comment 1.2:

While the paper provides a thorough technical explanation of *kriging interpolation*, it would be helpful to include a brief discussion on why this particular method was chosen over other spatial interpolation techniques. A few sentences comparing its advantages (such as accuracy, adaptability, or computational efficiency-to alternative methods)

Responses:

As suggested by the reviewers, the authors have included in the manuscript a comparison of the specific reasons for choosing kriging interpolation over other spatial interpolation techniques. The modification details the advantages of kriging interpolation over methods such as inverse distance weighting (IDW), spline interpolation, and natural neighbor interpolation, particularly in terms of spatial heterogeneity, prediction error estimation, robustness to outliers, and the ability to account for spatial structure. This additional note helps readers understand the rationale for our method selection and emphasizes the applicability of kriging interpolation in air quality prediction application scenarios. **The updated content is as follows:**

Due to the often spatially unbalanced distribution of monitoring stations, it is necessary to employ reasonable spatial interpolation techniques to estimate the spatial distribution across the entire study area based on data from known monitoring sites [20]. Among these methods, kriging interpolation has been widely applied in this field owing to its rigorous mathematical foundation and the property of providing the best linear unbiased estimate [21][22]. Unlike simpler techniques such as Inverse Distance Weighting (IDW) that assume values are solely determined by distance, or spline interpolation that focuses on smoothness, kriging was selected for this research due to

several key advantages. First, kriging accounts for both the distance and direction of measured points, making it particularly suitable for handling the spatial heterogeneity common in air pollutant distribution. Second, it provides not only predicted values but also error estimates (prediction variance), offering valuable uncertainty quantification not available with deterministic methods. Third, kriging demonstrates superior robustness against outliers compared to polynomial-based methods, an important consideration given the occasional anomalous readings in air quality monitoring data. Finally, kriging's ability to incorporate spatial structure through variogram analysis allows it to adapt to different spatial correlation patterns across various pollutants and geographical regions.

Comment 1.3:

A brief clarification on whether the mean absolute error in Figure 6 accounts for potential biases in the model would be useful.

Responses:

Thanks for your careful review. The mean absolute error (MAE) displayed in Figure 6 accounts for potential biases in the model by capturing both overestimation and underestimation equally through absolute difference calculations. Unlike metrics that may allow positive and negative errors to cancel each other out, MAE considers the magnitude of errors regardless of direction, thereby revealing systematic biases in prediction patterns. The MAE percentage (MAEP) further normalizes these values relative to the magnitude of the true concentrations, enabling fair comparisons across different concentration levels and prediction horizons.

Comment 1.4:

I think the conclusion would benefit from a short section discussing the study's limitations. While the results are promising, it would be helpful to acknowledge any constraints (data availability, potential biases, or the model's applicability to different regions).

Responses: Based on the reviewers' suggestions, the limitations of this study are discussed in detail in the conclusion section. Limitations in data availability, model dependence on specific geographic and meteorological conditions, applicability challenges in areas with different air pollution levels and sparse monitoring networks, and computational resource requirements are clearly identified. **The limitations are updated below:**

Despite the promising results, several limitations of this study should be acknowledged. Data availability constraints restricted the analysis to a three-year period and specific pollutants; longer time series and additional pollutants would enhance model robustness. The model's performance is dependent on the specific geographical and meteorological conditions of the Dongting Lake region, potentially limiting direct transferability to regions with substantially different characteristics without retraining or adaptation. The computational resources required for the SwinLSTM architecture may present implementation challenges for real-time prediction systems with limited processing capabilities.

In the end, we would like to thank the reviewers for their careful reading of our paper and for their valuable suggestions for revision, which make it possible to present our paper better.