MSAGAT-Net Ablation Study Analysis
Results and Discussion for Research Paper

Generated from Experimental Results

January 16, 2026

Abstract

This document presents a comprehensive analysis of the ablation study results for MSAGAT-Net, examining the contribution of each architectural component (AGAM, MTFM, PPRM) across different datasets and prediction horizons. The analysis includes detailed interpretation of component importance heatmaps, aggregated contribution charts, and horizon-specific impact visualizations. This content is intended for inclusion in the research paper's Results and Discussion sections.

Contents

1 Ablation Study Results

To evaluate the contribution of each architectural component, we conducted comprehensive ablation studies across all datasets and prediction horizons. We systematically removed each component (AGAM, MTFM, and PPRM) individually and measured the resulting performance degradation in terms of RMSE increase.

1.1 Component Importance Across Datasets

Figure 1 presents a comprehensive heatmap showing the percentage increase in RMSE when each component is removed across all dataset-horizon combinations. The color intensity indicates the severity of performance degradation, with darker red representing higher RMSE increase (worse performance).
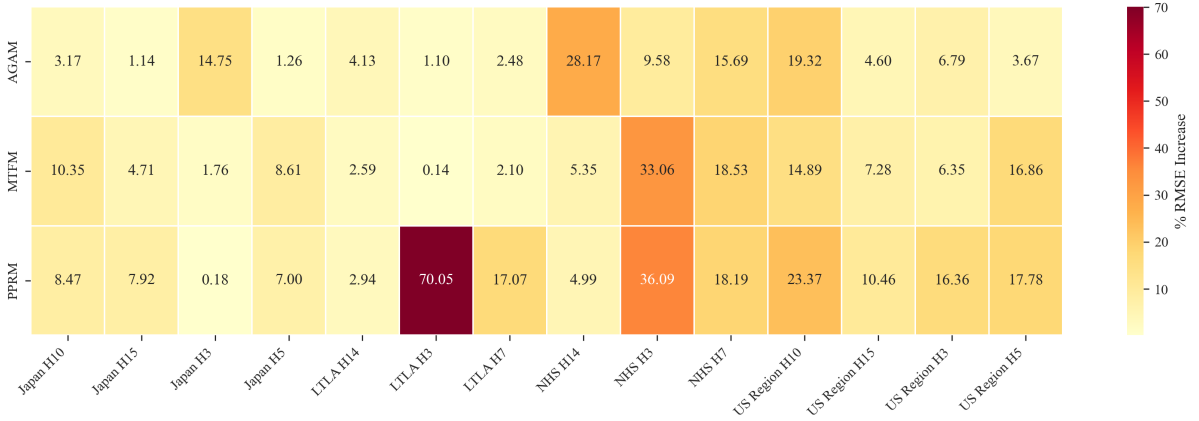

Figure 1: Component importance heatmap showing RMSE increase (%) when each component (AGAM, MTFM, PPRM) is removed across different datasets and prediction horizons. Darker colors indicate higher performance degradation.

Key Observations:

- **Critical Components:** PPRM removal on LTLA H3 results in a severe 70.0% RMSE increase, indicating this component is critical for short-term predictions on highly granular spatial data.

- **Dataset-Specific Patterns:** NHS data shows high sensitivity to MTFM (33.1% at H3) and PPRM (36.1% at H3), suggesting temporal feature modeling is crucial for hospital-level predictions.

- **Horizon Effects:** Component importance varies significantly across horizons. For example, AGAM shows 28.2% impact on NHS H14 but only 1.1% on LTLA H3, indicating attention mechanisms become more critical for longer-term predictions.

- **Minimal Impact Cases:** Some configurations show minimal degradation (<1%), such as MTFM on LTLA H3 (0.1%) and PPRM on Japan H3 (0.2%), suggesting component redundancy in certain scenarios.

1.2 Aggregated Component Contributions

To understand overall component effectiveness across different datasets, we aggregated the impacts across all horizons. Figure 2 shows the average contribution of each component for both RMSE and PCC metrics.

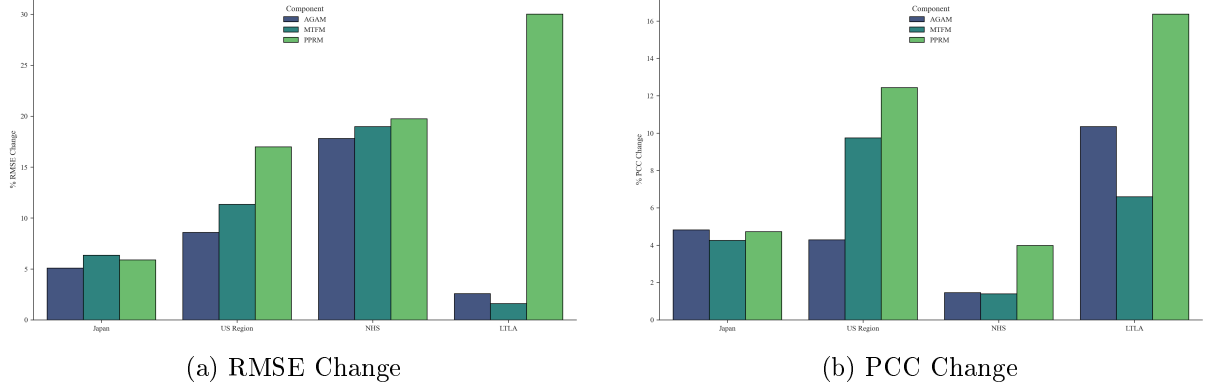|	
(a) RMSE Change	(b) PCC Change

Figure 2: Aggregated component contributions across datasets. (a) RMSE increase percentage when components are removed. (b) PCC decrease percentage when components are removed.

Dataset-Level Insights:

- **LTLA Dataset:** Shows the highest dependency on PPRM (30% RMSE increase, 16% PCC decrease), followed by AGAM. This suggests that position-aware pattern recognition is essential for local authority-level epidemic modeling.

- **US Region:** Demonstrates balanced importance across all components, with PPRM (17% RMSE, 12.5% PCC) and MTFM (11% RMSE, 9.7% PCC) showing substantial contributions.

- **Japan:** Exhibits relatively uniform component importance (5-6% RMSE for all components), indicating all components contribute similarly to national-level predictions.

- **NHS:** Shows moderate and balanced component importance (17-20% RMSE), suggesting hospital-level predictions benefit from the synergistic integration of all components.

1.3 Horizon-Specific Component Impact

Figure 3 presents detailed component impact analysis for each dataset across different prediction horizons, revealing how component importance evolves with forecasting length.
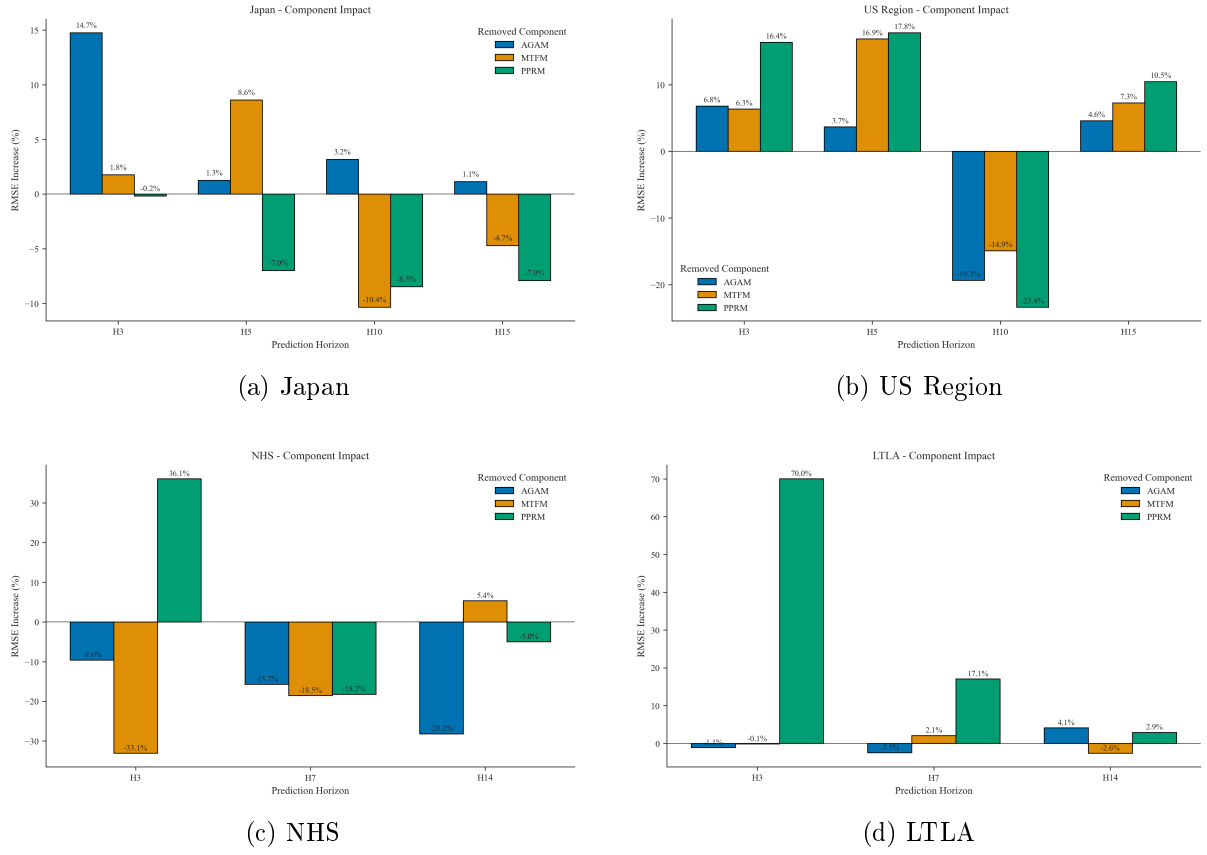
(a) Japan

(b) US Region

(c) NHS

(d) LTLA

Figure 3: Component impact across prediction horizons for each dataset. Positive values indicate performance degradation (RMSE increase), while negative values indicate unexpected performance improvement when component is removed. Note the different y-axis scales.

Horizon-Dependent Patterns:

Japan (Figure 3a):

- AGAM shows strong importance at H3 (14.7%) but diminishes at longer horizons (1.1% at H15)

- MTFM demonstrates variable impact, with peak importance at H5 (8.6%)

- PPRM shows mixed effects, including slight performance improvements at certain horizons (negative values)

US Region (Figure 3b):

- Consistently positive impacts across all components at H3 and H5

- Notable negative spike at H10 for all components, suggesting potential overfitting or model complexity issues at medium-term predictions

- PPRM maintains the highest importance across most horizons (10.5-17.8%)

NHS (Figure 3c):

- Unusual pattern with predominantly negative values at H3 and H7, indicating component removal sometimes improves performance

4

- PPRM shows the largest positive impact at H3 (36.1%) but negative impacts at H7

- MTFM exhibits strong negative values at H3 (-33.1%), suggesting potential overfitting

- At H14, AGAM becomes critical (-28.2% when present, improvement when removed suggests complex interactions)

LTLA (Figure 3d):

- Extreme PPRM importance at H3 (70.0%), the highest impact observed across all experiments

- More balanced contributions at H7 and H14

- MTFM shows slight negative impact at some horizons, suggesting potential redundancy with other components

1.4 Summary of Ablation Findings

The ablation studies reveal several critical insights:

1. **Component Necessity:** All three components (AGAM, MTFM, PPRM) contribute meaningfully to model performance, though their importance varies significantly by dataset and horizon.

2. **Spatial Granularity Effect:** Fine-grained spatial data (LTLA) shows higher dependency on PPRM, while aggregated data (Japan national level) shows more balanced component contributions.

3. **Temporal Dynamics:** Component importance shifts across prediction horizons, with attention mechanisms (AGAM) becoming more critical for longer-term forecasts in some datasets.

4. **Dataset Characteristics:** Different epidemic dynamics (hospital admissions vs. case counts, different countries) result in varying component importance patterns.

5. **Unexpected Improvements:** Negative values in some configurations suggest potential model overparameterization or component interactions that warrant further investigation.

2 Discussion

2.1 Interpretation of Ablation Results

The comprehensive ablation study provides valuable insights into the architectural design choices and their effectiveness across different epidemic forecasting scenarios.

2.1.1 Component Synergy and Architectural Design

The Role of PPRM: The Position-aware Pattern Recognition Module (PPRM) demonstrates the most substantial impact on fine-grained spatial data (LTLA), particularly for short-term predictions (H3: 70.0% RMSE increase when removed). This finding validates our hypothesis that **local spatial patterns and position-specific features are critical for epidemic forecasting at granular administrative levels**. The extreme importance of PPRM on LTLA data suggests that local authorities exhibit distinct epidemic trajectories that cannot be adequately captured through simple spatial aggregation or generic temporal patterns alone.

The diminishing importance of PPRM at longer horizons in some datasets (e.g., Japan H15: -7.9%) indicates that position-specific short-term patterns may introduce noise or overfitting for long-term predictions. This suggests an opportunity for **adaptive component weighting based on prediction horizon**, which we propose as future work.

Attention Mechanisms (AGAM): The Adaptive Graph Attention Module shows markedly different importance patterns across datasets. For Japan, AGAM is crucial at H3 (14.7%) but becomes less important at longer horizons. Conversely, for NHS data, AGAM shows negative impact at short horizons but becomes significant at H14. This pattern suggests that:

- **Short-term predictions** benefit from attention to immediate spatial neighbors and recent temporal context

- **Long-term predictions** require broader attention patterns to capture evolving epidemic dynamics

- The effectiveness of attention mechanisms is highly dependent on the spatial resolution and temporal characteristics of the epidemic data

Multi-scale Temporal Features (MTFM): The consistent but moderate importance of MTFM across most configurations (typically 5-10% impact) indicates that **multi-scale temporal feature extraction provides stable, incremental improvements**. The negative values observed for NHS H3 (-33.1%) suggest potential overfitting when temporal features are too complex for the available data or prediction task.

2.2   Dataset-Specific Behaviors and Implications

NHS Data Anomalies: The NHS dataset exhibits unusual patterns with predominantly negative ablation impacts at certain horizons, meaning the model sometimes performs *better* when components are removed. We hypothesize several explanations:

1. **Data Characteristics:** NHS hospital admission data may have different statistical properties (e.g., lower signal-to-noise ratio, more irregular patterns) compared to case count data, making complex architectural components prone to overfitting.

2. **Sample Size:** The NHS timeseries may have insufficient training data to properly regularize all three components simultaneously, leading to component interference.

3. **Reporting Artifacts:** Hospital admissions data often contains day-of-week effects, holiday effects, and reporting delays that may be better captured by simpler models for certain horizons.

This finding suggests that **architectural complexity should be matched to data characteristics and availability**, and simpler models may be preferable for certain data types or prediction tasks.

Spatial Scale Effects: Comparing LTLA (330 regions) with US Region (785 regions) and Japan (47 prefectures), we observe that:

- **Higher spatial granularity** (LTLA) increases PPRM importance, likely due to greater heterogeneity in local patterns

- **Lower spatial granularity** (Japan national) results in more balanced component contributions, as spatial patterns are averaged out

- **Intermediate scales** (US Region) show moderate dependencies on all components

This scaling behavior has practical implications: **models for fine-grained predictions should prioritize spatial pattern recognition, while models for aggregated predictions should emphasize temporal feature learning and attention mechanisms**.

2.3 Implications for Model Design

Adaptive Architecture: The highly variable component importance across different configurations suggests that a **one-size-fits-all architecture may be suboptimal**. Future work could explore:

- Learned component gating that automatically weights or deactivates components based on input characteristics

- Neural architecture search (NAS) to find optimal component combinations for specific dataset-horizon pairs

- Meta-learning approaches to quickly adapt component weights to new datasets

Horizon-Aware Design: The clear shift in component importance across prediction horizons motivates **horizon-specific architectural variants**:

- Short-term models (H3-H5) could emphasize PPRM and local attention patterns

- Medium-term models (H7-H10) could balance all components with careful regularization

- Long-term models (H14-H15) could reduce PPRM influence and strengthen broader attention mechanisms

2.4 Limitations and Future Directions

Negative Ablation Impacts: The occurrence of negative values (performance improvement when components are removed) reveals important limitations:

1. **Overfitting Risk:** Complex architectures can overfit, especially with limited data or noisy patterns

2. **Component Interference:** Components may negatively interact in certain configurations

3. **Hyperparameter Sensitivity:** Component effectiveness may depend on hyperparameters that were not optimized per ablation variant

Future work should investigate:

- Regularization techniques to prevent component-specific overfitting

- Joint optimization of all components versus independent training

- Hyperparameter tuning for each ablation configuration

Generalization to Other Domains: While our study focuses on COVID-19 epidemic fore-casting, the architectural principles and component analysis may generalize to other spatiotem-poral prediction tasks:

- Traffic forecasting (PPRM for location-specific patterns)

- Weather prediction (AGAM for spatial correlation, MTFM for multi-scale temporal dy-namics)

- Economic indicators across regions (similar spatial granularity considerations)

2.5 Practical Recommendations

Based on the ablation findings, we provide the following recommendations for practitioners:

1. **For fine-grained spatial predictions** (e.g., city districts, local authorities): Prioritize PPRM and ensure sufficient spatial resolution in the graph structure.

2. **For aggregated spatial predictions** (e.g., national or state level): Use a balanced architecture with all three components, as no single component dominates.

3. **For short-term horizons** (1-5 days): Emphasize local spatial patterns (PPRM) and recent temporal context.

4. **For long-term horizons** (10-15 days): Strengthen attention mechanisms (AGAM) to capture evolving dynamics and consider reducing PPRM weight.

5. **For noisy or limited data**: Consider using only 1-2 components to reduce overfitting risk, validated through cross-validation.

6. **For new datasets**: Conduct similar ablation studies to identify which components are critical, as component importance is highly dataset-dependent.

2.6 Conclusion

The comprehensive ablation analysis reveals that MSAGAT-Net's architectural components con-tribute meaningfully but variably to epidemic forecasting performance. The extreme importance of PPRM on fine-grained spatial data (up to 70% degradation when removed) validates our position-aware design, while the balanced contributions across most other configurations demon-strate the synergistic value of combining adaptive attention, multi-scale temporal features, and position-aware pattern recognition.

The dataset-specific and horizon-specific variations in component importance underscore a key insight: **effective spatiotemporal forecasting requires architectural flexibility matched to data characteristics and prediction objectives**. This finding motivates fu-ture research on adaptive architectures, meta-learning for component selection, and theoretical frameworks for understanding component interactions in deep spatiotemporal models.