Spatio-Temporal Forecasting of Municipal Firearm Events in Colombia: Machine Learning versus Structured Bayesian Modeling

Jorge Isaza[1], Carlos Páez Murillo[1], Erika Villaizón Castro[2], Danna Lesley Cruz-Reyes[3]

[1] Universidad Militar Nueva Granada, Colombia, jorge.isaza@unimilitar.edu.co, 0009-0000-7415-2830.
[2] Universidad Colegio Mayor de Cundinamarca, Colombia, carlos.paez@unimilitar.edu.co, 0000-0002-5221-8437
[3] Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia – Sede Bogotá, Colombia, Correo: dlcruzr@unal.edu.co ORCID: 0000-0002-5977-8162

Abstract
This study evaluates predictive performance in municipal firearm event forecasting in Colombia using a rolling-origin evaluation framework (2011-2024). We compare three approaches: (A) a flexible machine learning model, (B) a Bayesian temporal negative binomial model with AR(1) structure, and (C) a Bayesian spatio-temporal model incorporating BYM2 spatial effects and temporal autoregression. Results show that while machine learning achieves the lowest RMSE overall, structured spatio-temporal modeling substantially improves performance relative to temporal-only models and provides interpretable spatial dependence estimates. The findings clarify when explicit spatial structure contributes to predictive accuracy in high-dimensional municipal forecasting.

1 Introduction
Municipal-level crime forecasting represents a central challenge in applied statistics and public policy. Firearm-related events, in particular, exhibit marked spatial heterogeneity, temporal persistence, and overdispersion, complicating both inference and prediction. Accurate short- and medium-term forecasts at the municipal level are essential for territorial prioritization, resource allocation, and violence prevention strategies. However, methodological approaches to municipal crime forecasting remain fragmented across statistical paradigms.

Recent advances in machine learning (ML) have significantly improved predictive performance in complex high-dimensional settings. Tree-based ensemble methods and gradient boosting algorithms can flexibly capture nonlinear relationships and high-order interactions without imposing distributional assumptions. In crime analytics, ML approaches have demonstrated competitive performance in urban settings, particularly when rich feature sets and lagged structures are available (Gelvez-Ferreira et al., 2022).

Despite their predictive strength, ML models present limitations in interpretability and structural inference. They do not explicitly encode spatial dependence, nor do they naturally provide coherent uncertainty quantification under hierarchical data structures. As a result, while ML models may achieve lower error metrics, they often offer limited insight into the mechanisms driving spatial spillovers and persistence.

Bayesian hierarchical models have become a standard framework for analyzing spatial and spatio-temporal crime data. Their advantages include partial pooling across small areas, explicit uncertainty quantification, and incorporation of structured dependence via random effects (Law et al., 2014; Hu, 2018).

In small-area crime modeling, overdispersion relative to Poisson assumptions is common, motivating the use of Negative Binomial likelihoods. Temporal dependence is typically incorporated through autoregressive processes, while spatial dependence is captured using Conditional Autoregressive (CAR) priors or related Gaussian Markov Random Field formulations.

Spatio-temporal Bayesian models have demonstrated strong inferential performance in urban crime mapping and risk smoothing (Mohammadian and Mateu, 2023). However, much of the literature emphasizes explanatory mapping rather than strict out-of-sample forecasting under rolling-origin evaluation.

Spatial spillovers are particularly relevant in the Colombian context, where municipal boundaries do not isolate socio-economic and criminal dynamics. Evidence suggests that homicide rates and violence indicators exhibit spatial beta-convergence and neighborhood dependence patterns (Santos-Marquez, 2022). Ignoring such dependence may lead to biased risk estimates and suboptimal forecasts.

The BYM2 parameterization offers a principled way to separate structured and unstructured spatial variability while ensuring identifiability and interpretability. When spatial structure is strong, partial pooling can stabilize predictions in low-count municipalities without oversmoothing high-variance urban areas.

Although both ML and Bayesian hierarchical models have been applied to crime data, direct empirical comparisons under unified rolling-origin evaluation frameworks remain limited. In particular:

- Few studies evaluate ML and spatio-temporal Bayesian models under identical forecasting horizons.

- The contribution of explicit spatial structure to predictive accuracy is rarely quantified.
- Operational trade-offs between interpretability and predictive flexibility are not systematically assessed.

Consequently, it remains unclear under what conditions structured spatial modeling meaningfully improves municipal crime forecasting relative to flexible ML baselines.

This study addresses this gap by comparing three forecasting strategies for municipal firearm events in Colombia:

1. A flexible machine learning model (Model A),
2. A Bayesian temporal Negative Binomial model with AR(1) structure (Model B),
3. A Bayesian spatio-temporal model incorporating BYM2 spatial effects and temporal autoregression (Model C).

Using an expanding-window rolling forecasting design from 2011 to 2024, we evaluate out-of-sample predictive performance using RMSE and MAE. Beyond predictive metrics, we quantify spatial dependence via posterior hyperparameters and assess whether incorporating explicit spatial structure reduces aggregate bias and improves forecast stability.

Our contribution is both methodological and substantive. Methodologically, we provide a strict, controlled comparison between ML and structured Bayesian approaches under identical forecasting conditions. Substantively, we quantify the extent to which spatial dependence enhances municipal-level firearm event forecasting in Colombia.

The findings clarify when structural modeling contributes beyond flexible prediction and inform methodological choices in applied crime analytics.

2 Theoretical Framework

Municipal firearm events constitute non-negative count data characterized by three structural features: overdispersion, temporal persistence, and spatial dependence. Modeling such processes requires a framework capable of handling hierarchical structure and structured dependence across both space and time. Hierarchical Bayesian models based on Gaussian Markov Random Fields (GMRFs) provide a principled foundation for this purpose (Rue and Held, 2005).

2.1 Count Processes and Overdispersion

Let Y_{it} denote the number of firearm-related events observed in municipality $i = 1, \dots, N$ at month $t = 1, \dots, T$. Count data are commonly modeled using a Poisson distribution:

$$Y_{it} \sim \text{Poisson}(\mu_{it})$$

with

$$\mathbb{E}(Y_{it}) = \mathbb{V}(Y_{it}) = \mu_{it}$$

However, crime data frequently exhibit overdispersion due to latent heterogeneity, clustering, and omitted covariates (Hilbe, 2011). In such settings, the Negative Binomial distribution provides a flexible alternative:

$$Y_{it} \sim \text{NegBin}(\mu_{it}, \kappa)$$

where κ governs the degree of overdispersion and

$$\mathbb{V}(Y_{it}) = \mu_{it} + \frac{\mu_{it}^2}{\kappa}$$

As $\kappa \to \infty$, the model reduces to the Poisson case.

Negative Binomial specifications are widely used in crime and epidemiological modeling when variance exceeds the mean (Law et al., 2014).

2.2 Population Scaling and Offset Structure

Event counts scale mechanically with population size. To ensure comparability across municipalities of heterogeneous sizes, an offset term is introduced:

$$\log(\mu_{it}) = \log(Pop_{it}) + \eta_{it}.$$

This formulation implies that η_{it} models the log-incidence rate rather than raw counts. The offset approach is standard in disease mapping and small-area risk modeling (Besag et al., 1991; Lawson, 2018).

2.3 Temporal Dependence

Crime processes often exhibit persistence driven by structural socio-economic conditions and local inertia. Temporal autocorrelation is commonly modeled through autoregressive processes. A first-order autoregressive structure is defined as:

$$\delta_t = \rho\delta_{t-1} + \epsilon_t,$$

where $|\rho| < 1$ ensures stationarity and $\epsilon_t \sim \mathcal{N}(0, \tau_t^{-1})$.

AR(1) latent structures are standard in spatio-temporal hierarchical models to capture smooth temporal evolution (Blangiardo and Cameletti, 2015).

2.4 Spatial Dependence

Spatial dependence arises from geographic proximity, social interaction networks, and shared institutional environments. Ignoring spatial correlation may lead to biased variance estimates and inefficient predictions. Conditional Autoregressive (CAR) models provide a classical framework for modeling spatial dependence (Besag, 1974). Let W denote the adjacency matrix defining neighborhood structure. The intrinsic CAR model specifies:

$$v_i \middle| v_{-i} \sim \mathcal{N}\left(\frac{\sum_{j\sim i} v_j}{n_i}, \frac{1}{\tau_v n_i}\right),$$

where $j \sim i$ denotes neighboring municipalities.

The Besag-York-Mollié (BYM) model (Besag et al., 1991) extends this formulation by combining structured and unstructured spatial effects. However, the classical BYM parameterization suffers from identifiability issues. The BYM2 reparameterization proposed by Riebler et al. (2016) resolves this problem by introducing a mixing parameter:

$$\phi_i = \sqrt{1-\phi}u_i + \sqrt{\phi}v_i,$$

where:

- $u_i \sim \mathcal{N}(0, \tau^{-1})$ (unstructured heterogeneity),
- $v_i \sim CAR(W, \tau)$ (structured spatial effect),
- $\phi \in [0,1]$ controls the proportion of spatially structured variance.

The BYM2 formulation ensures interpretability of ϕ as the fraction of spatially structured variability (Riebler et al., 2016).

2.5 Spatio-Temporal Hierarchical Representation

Combining overdispersion, population scaling, spatial dependence, and temporal persistence yields the following hierarchical structure:

$$\log(\mu_{it}) = \log(Pop_{it}) + \alpha + \phi_i + \delta_t.$$

This specification corresponds to a latent Gaussian model under the GMRF framework (Rue and Held, 2005). It enables partial pooling across both municipalities and time periods, stabilizing estimates in low-count areas while preserving structural heterogeneity.

Such models are widely used in disease mapping, environmental risk analysis, and increasingly in crime analytics (Lawson, 2018; Blangiardo and Cameletti, 2015).

3 Data

3.1 Firearm Events Dataset

The empirical analysis is based on a municipal-level panel of firearm-related events in Colombia covering the period January 2010 to December 2024. The data are aggregated at a monthly frequency for each municipality, resulting in a balanced municipality-month panel.

Let Y_{it} denote the total number of firearm-related events observed in municipality $i = 1, \ldots, N$ during month $t = 1, \ldots, T$. The dependent variable therefore represents nonnegative integer counts, with substantial heterogeneity across municipalities and over time.

The original administrative records were cleaned and harmonized to ensure consistent municipal identifiers across the study period. In particular, municipal codes were standardized according to the official DANE classification, and temporal variables were converted into a continuous monthly index. Missing observations were not imputed; municipalities with zero reported events in a given month were recorded explicitly as zero counts.

To capture short- and medium-term temporal dynamics, lagged versions of the outcome variable were constructed. Specifically, we define:

$$Y_{i,t-1}, Y_{i,t-3}, Y_{i,t-6}, Y_{i,t-12}$$

which allow the forecasting models to incorporate persistence at multiple horizons, including seasonal recurrence. These lags are used exclusively as predictors in the machine learning specification (Model A) and serve as diagnostic references for temporal structure in the Bayesian models.

The resulting dataset consists of $N \times T$ observations forming a high-dimensional spatiotemporal panel suitable for rolling-origin forecasting evaluation.

The final working panel includes 1,111 municipalities with complete event and population records for the period 2010-2024. Eleven municipalities present in the official national shapefile were excluded due to the absence of consistent event or population coverage in the source datasets. These municipalities are located primarily in sparsely populated peripheral regions (Amazonas, Guainía, and Vaupés), and represent less than 1

4 Methodology

We compare three forecasting strategies under a unified rolling-origin evaluation framework.

Model A: Machine Learning Baseline

Model A is a flexible supervised learning model trained to predict monthly event counts. The feature set includes:

- Lagged counts $(Y_{i,t-1}, Y_{i,t-3}, Y_{i,t-6}, Y_{i,t-12})$,
- Month-of-year indicators (seasonality),
- Linear time trend,
- Municipal population.

The model is trained using an expanding window and optimized for out-of-sample predictive accuracy. No explicit spatial structure is imposed.

Model B: Temporal Bayesian Negative Binomial

Model B introduces probabilistic structure through a Negative Binomial likelihood with temporal dependence:

$$Y_{it} \sim \text{NegBin}(\mu_{it}, \kappa)$$
$$\log(\mu_{it}) = \log(Pop_{it}) + \alpha + \delta_t$$
$$\delta_t \sim AR(1)$$

Inference is conducted using Integrated Nested Laplace Approximation (INLA).

Model C: Spatio-Temporal Bayesian Model

Model C extends Model B by incorporating spatial effects:

$$Y_{it} \sim \text{NegBin}(\mu_{it}, \kappa)$$
$$\log(\mu_{it}) = \log(\text{Pop}_{it}) + \alpha + \phi_i + \delta_t$$
$$\phi_i \sim BYM2(W), \delta_t \sim AR(1).$$

This model captures both spatial smoothing and temporal persistence.

Figure 1 summarizes the structural differences across the three modeling strategies. The diagram highlights how Model A relies on flexible feature-based learning, Model B introduces probabilistic temporal structure through a Negative Binomial specification with AR(1) dynamics, and Model C further incorporates spatial dependence via a BYM2 component built from the municipal adjacency matrix.

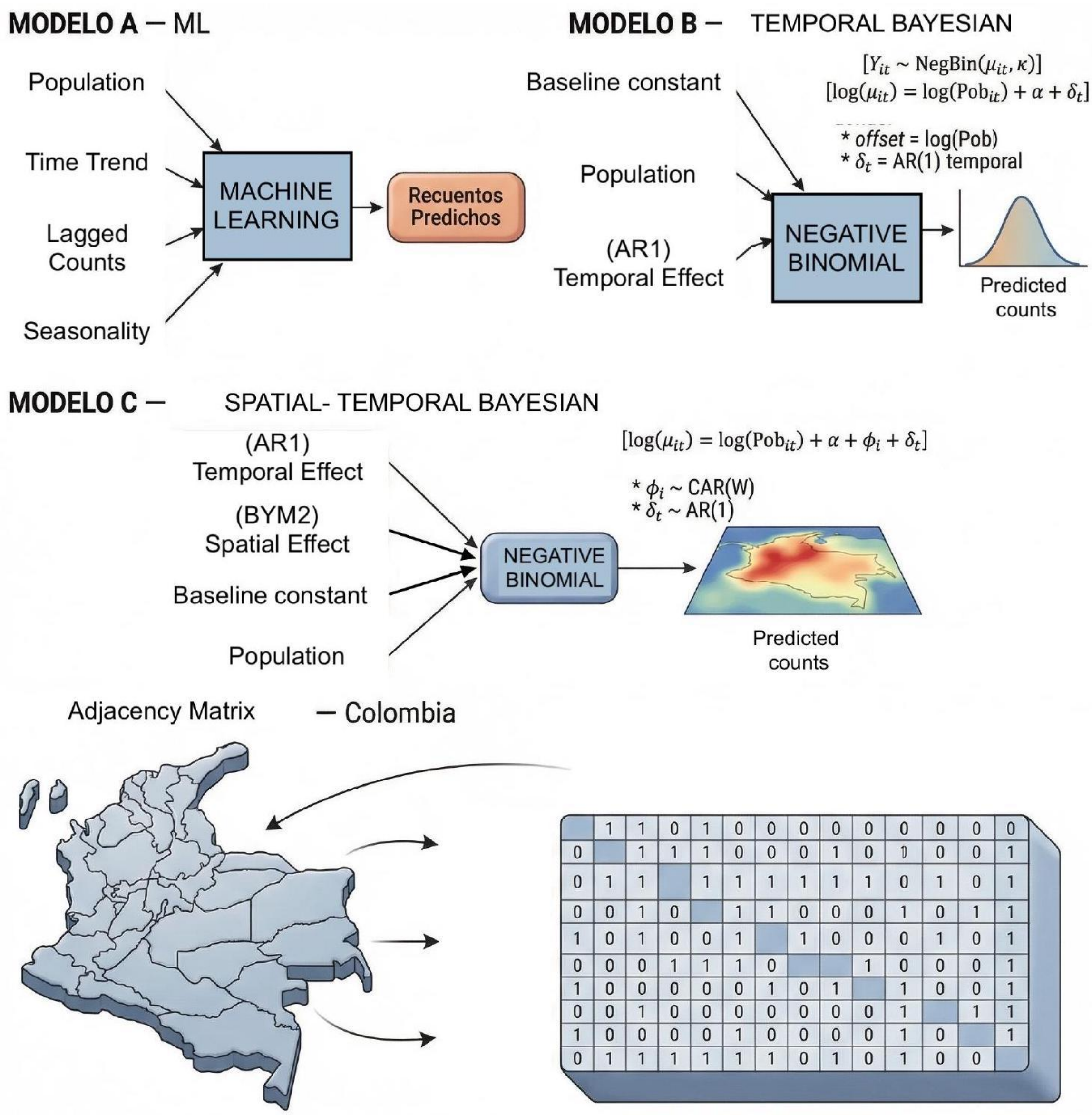

Figure 1: Conceptual architecture of the three forecasting models. Model A (ML) relies on feature-based learning; Model B incorporates temporal structure via AR(1); Model C extends the framework with spatial dependence through a BYM2 (CAR-based) component derived from the municipal adjacency matrix of Colombia.

4.1 Rolling-Origin Forecasting Design

We adopt an expanding-window rolling forecasting scheme:

- Train: 2010-2016 → Test: 2017
- Train: 2010-2017 → Test: 2018
- ...
- Train: 2010-2023 → Test: 2024

Performance metrics:

$$RMSE = \sqrt{\frac{1}{n}\sum \left(Y_{it} - \hat{Y}_{it}\right)^2}$$

$$MAE = \frac{1}{n}\sum \left|Y_{it} - \hat{Y}_{it}\right|.$$

All models are evaluated over identical test periods to ensure strict comparability.

4.2 Hyperparameter Estimation

Inference for Models B and C is conducted within the latent Gaussian modeling framework using Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009). This approach provides accurate deterministic approximations to posterior marginals while avoiding the computational burden of Markov Chain Monte Carlo (MCMC) in high-dimensional spatiotemporal settings.

Let $\boldsymbol{\theta}$ denote the vector of hyperparameters governing the hierarchical structure. In the present models, these include:

- κ : overdispersion parameter of the Negative Binomial likelihood,
- ρ : temporal autocorrelation coefficient in the AR(1) process,
- τ_t : temporal precision parameter,
- τ_s : spatial precision parameter (Model C),
- ϕ : spatial mixing parameter in the BYM2 specification (Model C).

Under the latent Gaussian model representation, the joint posterior distribution factorizes as:

$$\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{Y}) \propto \pi(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \pi(\boldsymbol{x} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}),$$

where $\boldsymbol{x}$ represents the latent field (spatial and temporal random effects) and $\boldsymbol{\theta}$ denotes hyperparameters controlling variance components and dependence structure (Rue and Held, 2005).

Posterior summaries are obtained from the full-sample fit (2010-2024). For each hyperparameter, we report posterior means and standard deviations. In particular:

- The posterior of κ quantifies the degree of overdispersion relative to Poisson.
- The posterior of ρ measures temporal persistence.
- The posterior of ϕ captures the proportion of spatial variance attributable to structured (CAR) dependence under the BYM2 parameterization (Riebler et al., 2016).

5 Results

The final panel contains $\mathbf{200,160}$ observations, consistent with approximately 1,100 municipalities observed over 180 months (2010-2024). This dimensionality provides substantial spatial and temporal variability for structural estimation and out-of-sample validation.

The proportion of zeros is $\mathbf{57\%}$, indicating a substantial presence of null counts, though not an extreme zero-inflation case. However, the combination of: Mean $= 2.01$ and Variance $= 110.88$ yields:

$$\frac{\mathrm{Var}(Y)}{\mathrm{Mean}(Y)} \approx 55$$

This indicates severe overdispersion. Under this scenario:

- The Poisson model is automatically ruled out.
- The Negative Binomial specification is not optional but necessary.
- Any ML approach that does not explicitly model latent heterogeneity is effectively absorbing variance through functional flexibility rather than probabilistic structure.

This point is structural for justifying Models B and C.

5.1 Predictive Performance (Rolling)

Predictive comparison was conducted using a rolling-origin forecasting scheme from 2011 to 2024. At each iteration, the training set expands and a one-year out-of-sample horizon is evaluated. This design prevents information leakage and ensures strict comparability across specifications.

Table 1 reports RMSE and MAE by year and model.

Table 1: Out-of-sample predictive performance (2011-2024)

Year	RMSE_A	RMSE_B	RMSE_C	MAE_A	MAE_B	MAE_C
2011	5.61	12.23	7.94	1.66	2.21	1.74
2012	6.10	12.51	6.69	1.67	2.33	1.71
2013	4.22	12.35	5.36	1.48	2.27	1.64
2014	4.43	10.81	5.53	1.40	2.10	1.57
2015	4.52	12.66	6.28	1.34	2.22	1.63
2016	3.19	11.48	5.57	1.26	2.10	1.59
2017	3.71	11.37	5.76	1.22	2.06	1.55
2018	2.68	11.52	5.25	1.14	2.02	1.50
2019	2.37	10.96	4.57	1.04	1.89	1.38
2020	2.65	9.88	4.24	0.98	1.74	1.29

2021	2.81	9.43	3.66	0.92	1.65	1.18
2022	2.05	8.75	3.46	0.84	1.53	1.08
2023	2.61	7.96	2.95	0.92	1.42	1.02
2024	2.52	6.83	2.62	0.96	1.40	1.03

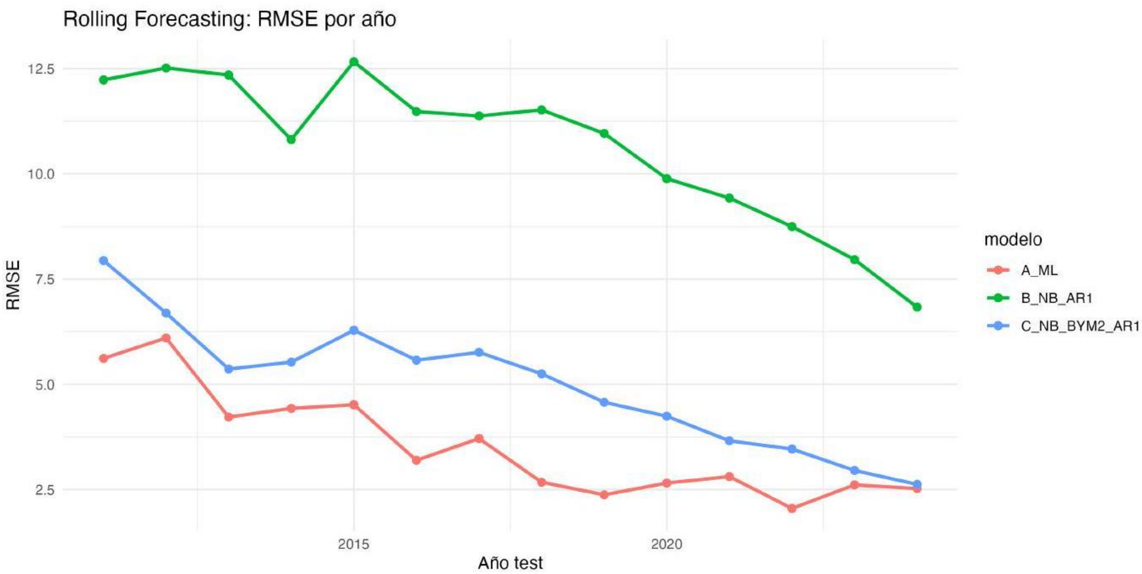

Figure 2: Rolling forecasting RMSE by year (2011-2024).

Model A (ML) achieves the lowest RMSE in every evaluation year. Model C (spatiotemporal) consistently outperforms Model B (temporal-only). The performance gap between A and C narrows substantially after 2021. Model B exhibits systematic underperformance relative to both A and C .

Table 2: Average predictive performance (2011-2024)

Model	RMSE $_{mean}$	MAE $_{mean}$	Mean Abs. Rel. Error
Model A	3.53	1.20	-
Model C	4.99	1.42	0.252
Model B	10.63	1.92	0.398

On average ML reduces RMSE relative to Model C by approximately 29%. Model C reduces RMSE relative to Model B by more than 50%. The purely temporal specification (Model B) is clearly insufficient.

5.2 Contribution of Spatial-Temporal Structure

In Table 3, demonstrate that $\phi = 0.792$ reflects pronounced, structured spatial dependence. The value $\rho \approx 0.85$ further indicates strong temporal persistence. The larger NegBin size parameter implies that, once spatial structure is accounted for, heterogeneity is more effectively absorbed.

Table 3: Posterior summary

Parameter	Mean	SD
Model B (Temporal NegBin + AR1)		
NegBin size (1/overdisp.)	0.843	0.006
Precision (time)	4.079	0.847
ρ (AR1)	0.873	0.027
Model C (Spatio-temporal BYM2 + AR1)		

NegBin size (1/overdisp.)	1.418	0.014
Precision (spatial)	1.834	0.137
ϕ (BYM2 mixing)	0.792	0.042
Precision (time)	5.098	1.448
ρ (AR1)	0.848	0.043

To assess whether the spatio-temporal specification artificially smooths geographic disparities, we compare the territorial concentration of observed and predicted events using the Gini index. The Gini coefficients are: Predicted: 0.7437 and Observed: 0.7137 . The predicted concentration closely mirrors the empirical distribution of events. The slightly higher Gini under the model suggests that spatial smoothing does not homogenize municipalities nor dilute high-intensity clusters. Instead, the structured specification preserves- and in fact slightly sharpens - the underlying geographic inequality.
This result is important because excessive spatial regularization can artificially compress territorial disparities. The BYM2 component appears to borrow strength across neighbors without eliminating genuine spatial concentration.

5.3 Six-Month Forecast Comparison

Table 4: Six-month forecast summary for selected municipalities.

Department	Municipality	Model	$\hat{\mu}_{\text{mean}}$	$\hat{\mu}_{\text{min}}$	$\hat{\mu}_{\text{max}}$	95% CI (width)
Bogotá D.C.	Bogotá D.C.	Model A	119.17	101.62	163.50	-
Bogotá D.C.	Bogotá D.C.	Model B	340.92	299.74	376.25	643.37
Bogotá D.C.	Bogotá D.C.	Model C	145.94	130.04	159.35	269.64
Bolívar	Cartagena	Model A	10.94	0.01	49.34	-
Bolívar	Cartagena	Model B	43.50	38.25	48.01	82.09
Bolívar	Cartagena	Model C	42.92	38.24	46.87	79.31
Norte de Santander	Cúcuta	Model A	22.50	0.01	71.49	-
Norte de Santander	Cúcuta	Model B	34.05	29.94	37.58	64.27
Norte de Santander	Cúcuta	Model C	52.31	46.61	57.12	96.66
Valle del Cauca	Cali	Model A	22.73	0.01	49.34	-
Valle del Cauca	Cali	Model B	98.17	86.31	108.34	185.26
Valle del Cauca	Cali	Model C	125.40	111.74	136.93	231.67

Table 4 reports the six-month ahead predictive summaries for four representative municipalities, while Figures 36 display the corresponding time series together with the forecast trajectories. The selected municipalities differ in scale, volatility, and spatial context, allowing a structured comparison across model specifications.
Several patterns emerge from Table 4. First, in Bogotá (11001), the purely temporal specification (Model B) produces substantially larger projected means relative to both the spatio-temporal model (Model C) and the ML baseline. This suggests that, in the absence of spatial smoothing, temporal persistence alone amplifies recent peaks, leading to upwardbiased forecasts. In contrast, Model C moderates this effect through structured spatial borrowing, producing a lower projected mean and a narrower credible interval (width = 269.64 versus 643.37 under Model B).
Second, in Cartagena (13001), Models B and C generate nearly identical projected means and interval widths. This indicates that spatial structure contributes marginally in this municipality, likely because local dynamics dominate over spatial spillovers. Here, the CAR component adds little beyond temporal persistence.
Third, in Cúcuta (54001) and Cali (76001), the spatio-temporal model produces higher projected means relative to the purely temporal specification. This indicates that neighboring municipalities with elevated

recent activity exert upward pressure through the spatial random effect. In these cases, spatial dependence is not merely smoothing but actively shifting the predictive mean.

Figures 36 visually confirm these mechanisms. In high-variance municipalities (e.g., 54001 and 76001), Model C generates smoother and more stable trajectories than Model B, reflecting the joint contribution of spatial smoothing and temporal persistence. Model B, by contrast, reacts strongly to recent fluctuations without cross-sectional regularization.

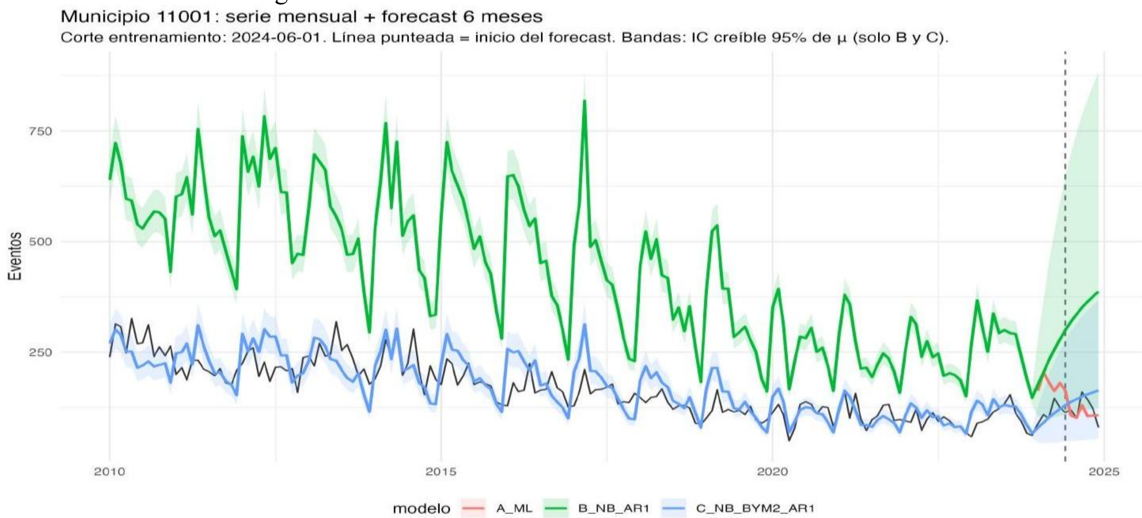

Figure 3: Monthly series and 6 -month forecast - Municipality 11001.

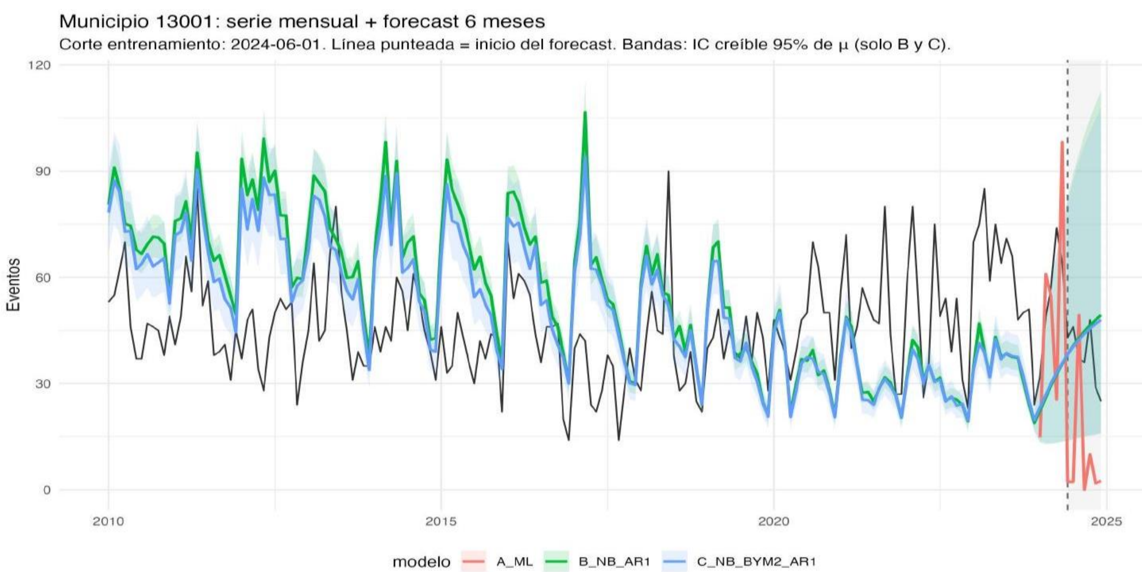

Figure 4: Monthly series and 6-month forecast - Municipality 11001.

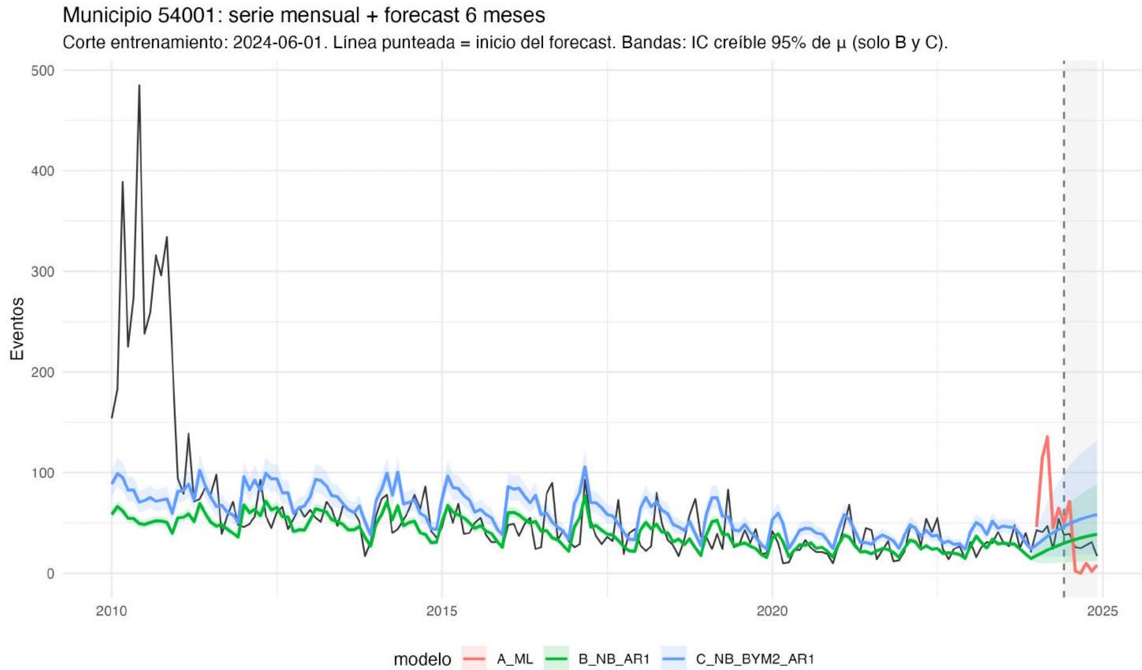

Figure 5: Monthly series and 6-month forecast - Municipality 11001.

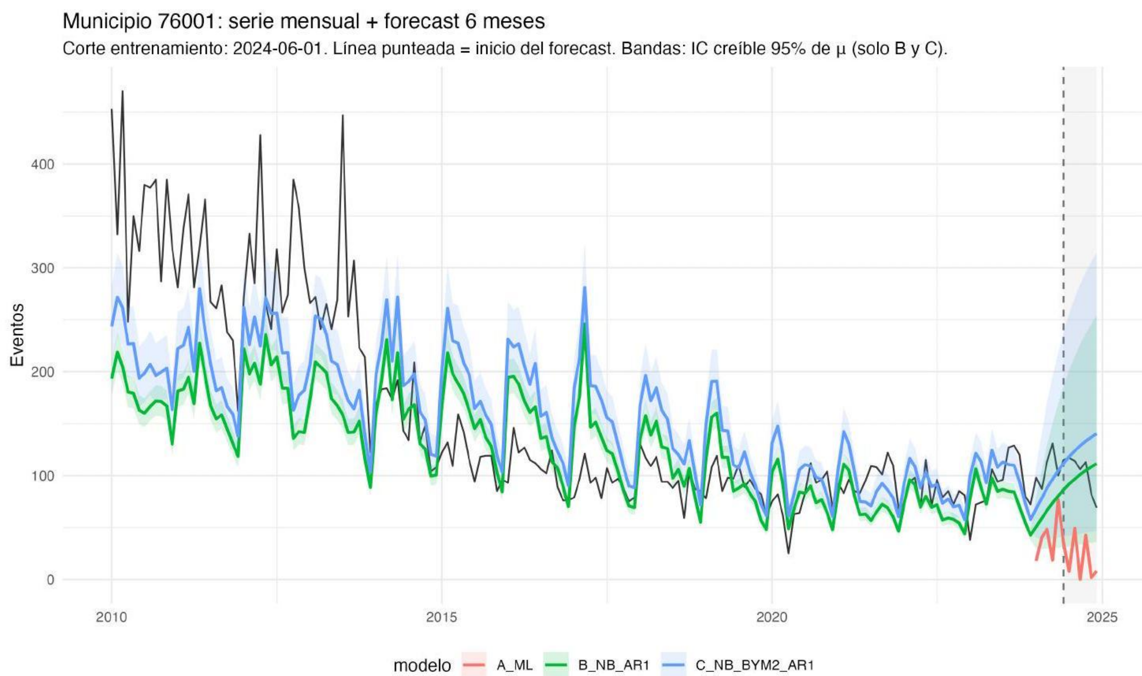

Figure 6: Monthly series and 6-month forecast - Municipality 11001.

The ML baseline (Model A) is included for point prediction comparison; however, it does not report credible intervals in Table 4. This is not a computational omission but a structural limitation: the ML model is trained to minimize prediction error and does not provide a full probabilistic specification of the data-generating process. Consequently, it yields point forecasts but not coherent predictive distributions. Any interval constructed post hoc (e.g., via residual bootstrapping) would not be directly comparable to Bayesian credible intervals derived from the joint posterior distribution. For this reason, uncertainty quantification is reported only for Models B and C.

6 Conclusion

This study provides a structured comparison between flexible machine learning approaches and probabilistic spatio-temporal models for municipal firearm event forecasting in Colombia.

First, machine learning (Model A) consistently achieves the lowest RMSE across the rolling forecasting horizon. In purely point-prediction terms, flexible nonlinear learners dominate both temporal and spatio-temporal Bayesian specifications.

Second, incorporating spatial structure materially improves predictive performance relative to a temporal-only Negative Binomial specification. Model C (BYM2 + AR1) reduces forecasting error by more than 50% compared to Model B, demonstrating that spatial dependence is not merely statistically detectable but predictively consequential.

Third, spatial dependence is both statistically and substantively relevant. The posterior mixing parameter ($\phi \approx 0.8$) indicates that most cross-sectional variation is structured rather than unstructured noise. Moreover, the spatio-temporal model preserves the empirical territorial concentration of events, suggesting that spatial smoothing does not artificially compress geographic inequality.

Fourth, the predictive gap between machine learning and structured spatio-temporal modeling narrows substantially in recent years. This suggests that as the data accumulate and spatial persistence stabilizes, structured probabilistic models become increasingly competitive in forecasting performance.

The results highlight a fundamental trade-off. Machine learning excels at minimizing point prediction error but does not provide coherent uncertainty quantification. In contrast, the spatio-temporal Bayesian specification delivers interpretable spatial effects, structured borrowing of strength, and full predictive distributions, at the cost of slightly higher RMSE.

From a methodological perspective, the findings suggest that explicitly modeling spatial dependence is essential when forecasting territorial event data. Ignoring spatial structure leads to systematic underperformance, while incorporating it yields both interpretative and predictive gains.

Future research should extend this comparison to probabilistic scoring rules (e.g., LogScore or CRPS) and formally test predictive dominance. Additionally, heterogeneity across municipality size and volatility regimes warrants further investigation.

Overall, the evidence indicates that spatial structure is not optional in municipal event forecasting. It is central to both inference and prediction.

Acknowledgments: Optional.

References

1. Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society Series B, 36(2):192-236.
2. Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, 43(1):1-20.
3. Blangiardo, M. and Cameletti, M. (2015). Spatial and Spatio-Temporal Bayesian Models with R-INLA. Wiley.
4. Gelvez-Ferreira, J. D., Nieto-Rodríguez, M.-P., and Rocha-Ruiz, C.-A. (2022). Predicting crime in intermediate cities: A machine learning approach. URVIO: Revista Latinoamericana de Estudios de Seguridad, 34:50-72.
5. Hilbe, J. M. (2011). Negative Binomial Regression. Cambridge University Press. Hu, T. e. a. (2018). Urban crime prediction based on spatio-temporal bayesian models. PLoS ONE, 13(8):e0201234.
6. Law, J., Quick, M., and Chan, P. (2014). Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. Journal of Quantitative Criminology, 30(1):1-27.
7. Lawson, A. (2018). Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. CRC Press.
8. Mohammadian, A. and Mateu, J. (2023). Bayesian approach for modeling spatial-temporal crime data. Journal of Statistical Studies, 16(2):435-448.
9. Riebler, A., Sørbye, S., Simpson, D., and Rue, H. (2016). An intuitive bayesian spatial model for disease mapping that accounts for scaling. Statistical Methods in Medical Research, 25(4):1145-1165.
10. Rue, H. and Held, L. (2005). Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC.
11. Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations. Journal of the Royal Statistical Society Series B, 71(2):319-392.
12. Santos-Marquez, F. (2022). Spatial beta-convergence forecasting models: Evidence from municipal homicide rates in colombia. Journal of Forecasting, 41(2):294-302.