Improving crime count forecasts in the city of Rio de Janeiro via reconciliation

Leonardo Moog *

Abstract

Crime forecasts based on reliable data and sound statistical methodologies can provide valuable input for tactical deployment of police resources in so-called crime *hot spots* and effective planning of police operations. In Rio de Janeiro, the Public Safety Secretariat uses criminal activity forecasts to detect crime patterns and evaluate police performance. This paper evaluates the impact of reconciliation on the forecasts of 271 series of registered criminal occurrences in the city of Rio de Janeiro on a monthly basis from January 2003 to December 2019. We verify that reconciliation improves crime count forecasts and may provide valuable input for the optimal deployment of police resources.

1 Introduction

Among the chief concerns of policymakers is providing efficient law enforcement and promoting public safety. To achieve this, tactical deployment of police resources in so-called crime hotspots and effective planning of police operations are crucial.

Rio de Janeiro's history with violence began in the 1980s when drug dealers occupied the favelas for drug trafficking purposes. This led to a sharp increase in urban violence that persisted through the 1990s despite significant public outcry and intense coverage from media outlets.

*I would like to thank Marcus Nascimento, who wrote this article in collaboration with me

Driven by the necessity of violence reduction due to the international events that would take place in Rio, such as the 2014 World Cup and the 2016 Olympic Games, an initiative to recuperate urban areas dominated by organized crime was implemented in the 2000s: the Unidades de Polícia Pacificadora or Pacifying Police Units (UPPs).

Beginning in 2009, the UPP policy consisted of three stages. First, elite police units would execute large operations in chosen favelas to retake the territory. Then, the favela would be temporarily occupied by police forces in a “stabilization” phase. Lastly, occupation would become permanent through the construction of a UPP unity, that is, a physical infrastructure in the favela to accommodate permanently assigned police officers.

Despite some initial success and public support, the UPP policy ultimately failed for several reasons, from the lack of investments that garnered social inclusion in the pacified favelas, to the abusive and overly violent behavior of police authorities, and also a severe fiscal crisis that restricted the government’s budget (Ferraz, Monteiro, and Ottoni, 2021). As of 2016, in the midst of a national economic crisis, crime indicators started to deteriorate again in the city, which eventually led to a federal intervention by army forces in 2018 in an attempt to contain the escalation of violence. This period also coincided with the emergence of another type of criminal group, particularly on the outskirts of the city: the militias, paramilitary groups formed mostly by police officers that extorted the population for protection and for the provision of services.

In this scenario, according to Seixas et al. (2010), the Secretaria de Segurança Pública (Public Safety Secretariat) of Rio de Janeiro implemented a ”Plan-Do-Study-Act” (Tague, 2023, PDCA) model in its management process, and data is essential to this process. In particular, data enables a better comprehension of public safety issues and helps in identifying their primary causes. Forecasts of criminal activity based on reliable data and sound statistical methodologies can provide valuable input to this managerial cycle (Berk, 2008; Gorr and Harries, 2003; Harries, 2003), since the Public Safety Secretariat uses criminal activity forecasts to detect crime patterns and evaluate police performance. Seixas et al. (2010) establishes a performance measurement procedure based on a public safety index. Every quarter, the index is analyzed in order to guide the optimal allocation of police in different city areas. The index is comprised of four dimensions, one of which is the crime forecast for the next quarter. Hence, accurate forecasts of crime series can provide a valuable contribution to public policy decisions regarding police activity.

Our paper is connected to two strands of the criminology literature: police performance measurement and crime forecasting. Regarding the former, Sonnichsen (1997) analyzed identified the main difficulties in measuring police effectiveness, as well as government efforts to implement improved approaches in six countries: Canada, Germany, the United Kingdom, Sweden, The Netherlands, and the United States. Filstad and Gottschalk (2009) suggested five potential indicators to build a framework that takes into account the complaint-handling process in police oversight bodies: quality and quantity of complaints received; complaints completion process and time; conviction rate from complaints charges; learning and advice for police agencies and confidence in the police oversight agency.

Davis (2012) reviewed some of the international best practices for measuring police performance and described some fundamental concerns in designing benchmarks to evaluate law enforcement agencies. The author also underlines some of the benefits of performance measurement, such as guiding officials in monitoring department operations, promoting adherence to policies and strategic plans, and detecting patterns of bias or misconduct. Fielding and Innes (2006) focused on policing strategies in the United Kingdom that incorporate ideas and insights from social research methodology, arguing that qualitative approaches offer more meaningful evaluation forms. For a more comprehensive review of police performance measurements, we refer to Tiwana, Bass, and Farrell (2015).

Despite the connection with police performance measurement, we do not intend to propose new methods or to supply a critical view of the ones already adopted by the Secretaria de Segurança Pública of Rio de Janeiro in this article. Instead, we wish to introduce innovative statistical tools to improve the calculation of crime-focused indicators. Specifically, this paper aims to provide a practical, easy-to-implement, and computationally efficient tool in line with recent statistical developments to support decision-makers and subsidize performance evaluation mechanisms in the public administration of Rio de Janeiro.

In our application, we have multiple time series that are hierarchically arranged and can be aggregated at different geographical levels that are based on an administrative organization formulated by the Secretaria de Segurança Pública of Rio de Janeiro. Multiple time series structured this way are often called "hierarchical time series" (Fliedner, 2001). To incorporate this specificity in a practical, easy-to-implement, and computationally efficient way, we apply reconciliation techniques (Hyndman et al., 2011; Wickramasuriya, Athanasopoulos, and Hyndman, 2019). Hence, our approach yields forecast estimates in two stages. In the first stage, crime count forecast series from all hierarchical levels are obtained by applying state space methods (Durbin and Koopman, 2012). Since crime count series usually suffer from issues related to the infrequent occurrences of criminal phenomena, we depart from the standard Gaussian estimation techniques and assume a Poisson distribution for our process instead (Durbin and Koopman, 2000). This approach enables us to deal more appropriately with data sets filled with observations that are either equal to zero or very close to it since the Poisson distribution is better suited for processes of rare events.

The reconciliation techniques are employed in the second stage. Reconciliation methods were initially proposed by Hyndman et al. (2011). The authors introduced a method based on independently forecasting series from all hierarchical levels and used a regression model to reconcile the forecasts optimally. The revised forecasts are computed via Ordinary Least Squares and the method is shown to improve forecast accuracy compared to traditional (often statistically unfounded) reconciliation methods such as the so-called "top-down" and "bottom-up" approaches.[1]Wickramasuriya, Athanasopoulos, and Hyndman (2019) further formalized this reconciliation method by minimizing the mean squared error of coherent forecasts under the property of unbiasedness. Their approach provides a closed-form solution to this problem, which they refer to as the "trace minimization" (MinT) solution. As it turns out, however, the MinT approach does not guarantee that the revised forecasts, although coherent, will be non-negative, even if all of the

[1]The "bottom-up" approach consists in forecasting series at the lowest hierarchical level and then simply summing them according to the hierarchical structure in order to obtain aggregated forecasts. The "top-down" approach, on the other hand, consists of forecasting only the most aggregated series and then using ad-hoc weights to produce disaggregated forecasts.

original base forecasts satisfy this property. Since in a variety of applications, the one in our paper included, forecasts should be inherently non-negative, this posed as a potential problem in applications. Wickramasuriya, Turlach, and Hyndman (2020) then proposed an extension to the MinT approach that allowed for non-negativity constraints in the revised forecasts.

To the best of our knowledge, forecast reconciliation techniques have not yet been employed to forecast crime, even though crime data is a perfect suitor for such techniques. Indeed, in crime forecasting exercises it is often difficult to fully explore the informational content of a large number of series since the data is both serially and spatially correlated. This means that in order to make full use of the information available, one would need to take into account the complex covariance structure that permeates the data. In a setting where the number of crime series is large (say, when dealing with a variety of geographical locations, such as precincts or neighborhoods within a city) and the number of temporal observations is substantial, specifying multivariate models that can handle all these particularities presented by the data might become unfeasible. As a result, one is often forced to aggregate the data, thus losing relevant information. Moreover, since the hierarchical structure of crime data is often unaccounted for in most empirical exercises, this results in even further loss of information.

The main advantage of this methodology is that reconciliation methods enable us to deal with serial correlation and implicitly handle spatial autocorrelations, incorporating the geographical hierarchical structure in the data through a practical, easy-to-implement, and computationally efficient approach. Therefore, the contribution this paper aims to make is twofold. First, it hopes to inform public safety policy-making by providing data based on novel statistical techniques. This will be done by forecasting crime occurrences in different municipal areas. Second, on the methodological side, it intends to introduce reconciliation methods to crime forecasting and to compare some of the different approaches proposed so far.

The remainder of the paper is organized as follows. Section 2 provides a brief review of crime forecasting. Our data set, including the geographical hierarchy, is described in Section 3. A basic setup of hierarchical time series, notation, and relevant departures from the employed methods, such as non-Gaussian state space models for count time series, forecast reconciliation, and its alternative estimators for the covariance matrix are presented in Section 4. Section 5 presents the results obtained from our forecasts and the subsequent reconciliation, while Section 6 concludes.

2 A brief review of crime forecasting

Crime forecasting is strongly supported by Crime Pattern Theory (Brantingham and Brantingham, 1984) and Environmental Criminology (Bruinsma and Johnson, 2018). These theories aim to explain the variation in the distribution of criminal activity across time and space by providing a theoretical foundation for how an offender decides whether to perpetrate an offense. They establish that criminal activity results from a decision-making process where the offender responds to environmental cues and opportunities (Brantingham, 2016). Such cues may be physical features of the environment or local social conditions, and their interaction determines why an offender commits specific crimes at specific places and times.

These theories imply that criminal activity is not a simple white noise process (although it does have a significant random component) but rather a complex phenomenon that follows certain patterns of concentration in both the spatial and the temporal dimensions. Hence, one may employ statistical methods to identify such patterns and forecast criminal activity. Therefore, in the rest of this section, we briefly review the different methodological approaches used in the literature to that end.

First, we have works that are based on time series methods. For example, Deadman and Pyle (1997) and Deadman (2003) used economic data to forecast criminal activity in England and Wales a year ahead. Gorr, Olligschlaeger, and Thompson (2003) used univariate time-series models to forecast crime a month ahead in police precincts in Pittsburgh, PA. Jiang and Barricarte (2011), in their turn, applied time-series methods to forecast age-specific crime rates, thus enabling them to forecast crime by using those rates on the projected population.

Spatial models are also widely used, particularly to determine crime hotspots. A crime hotspot is a “geographic location of high crime concentration, relative to the distribution of crime across the whole region of interest” (Chainey and Ratcliffe, 2005). For instance, Cohen, Gorr, and Olligschlaeger (2007) worked with leading indicators to obtain 1-month-ahead forecasts of specific crimes over a grid system covering the city of Pittsburgh, PA. Chainey, Tompson, and Uhlig (2008) compared different mapping techniques (grid thematic mapping, kernel density estimation, spatial ellipses, and others) in their performance to determine hotspots for criminal activity and showed that kernel density estimation (KDE) consistently outperforms other mapping methods. Hart and Zandbergen (2014) then analyzed how user-defined parameters affect the predictive accuracy of models within a KDE context.

There are also spatio-temporal methods that arise when data are observed over time as well as space. Aldor-Noiman et al. (2016) modeled violent crime counts in Washington, D.C., between 2001-2008, using a first-order integer-valued autoregressive process and discovered clusters of region-specific crime series. Liesenfeld, Richard, and Vogler (2017) introduced a latent spatio-temporal heterogeneous state process to predict monthly crimes in Pittsburgh, PA. Hu et al. (2018) applied a spatio-temporal kernel density estimation (STKDE) method to predict hotspot mapping and proposed a new metric to evaluate such hotspots.

Lastly, several recent contributions have applied machine learning techniques to forecast crime (Corcoran, Wilson, and Ware, 2003; Jha et al., 2021; Shamsuddin, Ali, and Alwee, 2017). Recent research also includes social media (Gerber, 2014; Vomfell, Härdle, and Lessmann, 2018; Wang, Brown, and Gerber, 2012) and taxi pick-up and drop-off (Kadar and Pletikosa, 2018; Wang et al., 2016; Zhao and Tang, 2017) information to boost their predictions. The examples and the list of papers cited here are not exhaustive, so we refer to Kounadi et al. (2020) for a more comprehensive review.

3 Data set

Our data set contains the number of monthly registered criminal occurrences in the city of Rio de Janeiro from January 2003 to December 2019. This information is public, and it is available on the Instituto de Segurança Pública (Public Safety Institute) website (`http://www.ispdados.rj.gov.br/`). For the period under analysis, the criminal occurrences had to be registered exclusively in the police precincts. More recently, for crimes like threats, assault, violence against women, and street theft and robbery, online registration has been made available as well.

The data set comprises which and how many crimes occurred in any given month, as well as where they occurred. For registering and planning purposes, the Secretaria de Segurança Pública (Public Safety Secretariat) divides the city of Rio de Janeiro in minor administrative zones called "public safety integrated circumscriptions", or CISPs. These CISPs can be aggregated into larger areas which are denominated "public safety integrated areas", or AISPs. These AISPs can also be further aggregated into major regions called "public safety integrated regions", or RISPs. Finally, the city of Rio de Janeiro encompasses 41 CISPs, 17 AISPs, and 2 RISPs. Figure 1 illustrates these administrative zones.

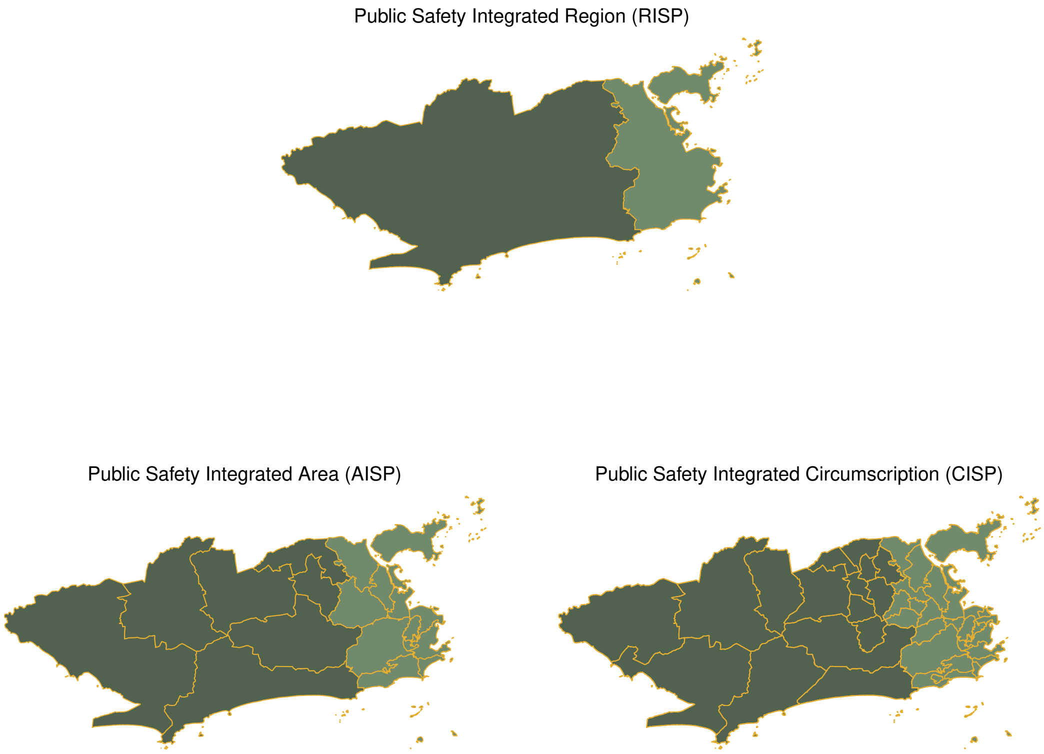

Figure 1: Administrative zones according to the Public Safety Secretariat.

We restrict our attention to five crimes: intentional lethal and violent crimes (CVLI), attempted homicide, rape, extortion, and residential burglary. Intentional lethal and violent crimes include murder, robbery, and bodily injury resulting in death. Although theft and robbery data are also commonly considered in crime forecasting exercises, we choose not to include these crimes in our analysis due to the poor quality of registers in Brazil. For example, in 2021, the Brazilian Institute of Geography and Statistics (IBGE), through the Continuous National Household Sample Survey (Continuous PNAD), reported that only 44.8% street theft victims and 42.1% street robbery victims registered the incidence. Hence, since these crimes tend to be significantly under reported, we chose not to include them.

Figure 2 represents the spatial distribution of the total number of reported crimes (in the square root scale) in the year 2019 across the public safety integrated circumscriptions (CISPs). For crimes like rape, extortion, and residential burglary, we observe a concentration of registers in four CISPs (35, 32, 34, 36), all located on the West side of Rio de Janeiro. These CISPs encompass neighborhoods like Campo Grande, Santa Cruz, Cidade de Deus, and Bangu and jointly concentrate more than 34%, 27%, and 29% of the rapes, extortions, and residential burglaries, respectively, registered in the city.

As for intentional lethal and violent crimes (CVLI) and attempted homicides, the aforementioned CISPS also concentrate a significant percentage of registers (more than 24% and 21%, respectively). However, there are also CISPs on the city's North side that present a high incidence, such as CISPs (21, 22, 39), which encompass neighborhoods like Bonsucesso, Penha, and Pavuna. Table 3 in Appendix B describes the relations among CISPs, AISPs, and RISPs and how they are related to the neighborhoods in Rio de Janeiro.

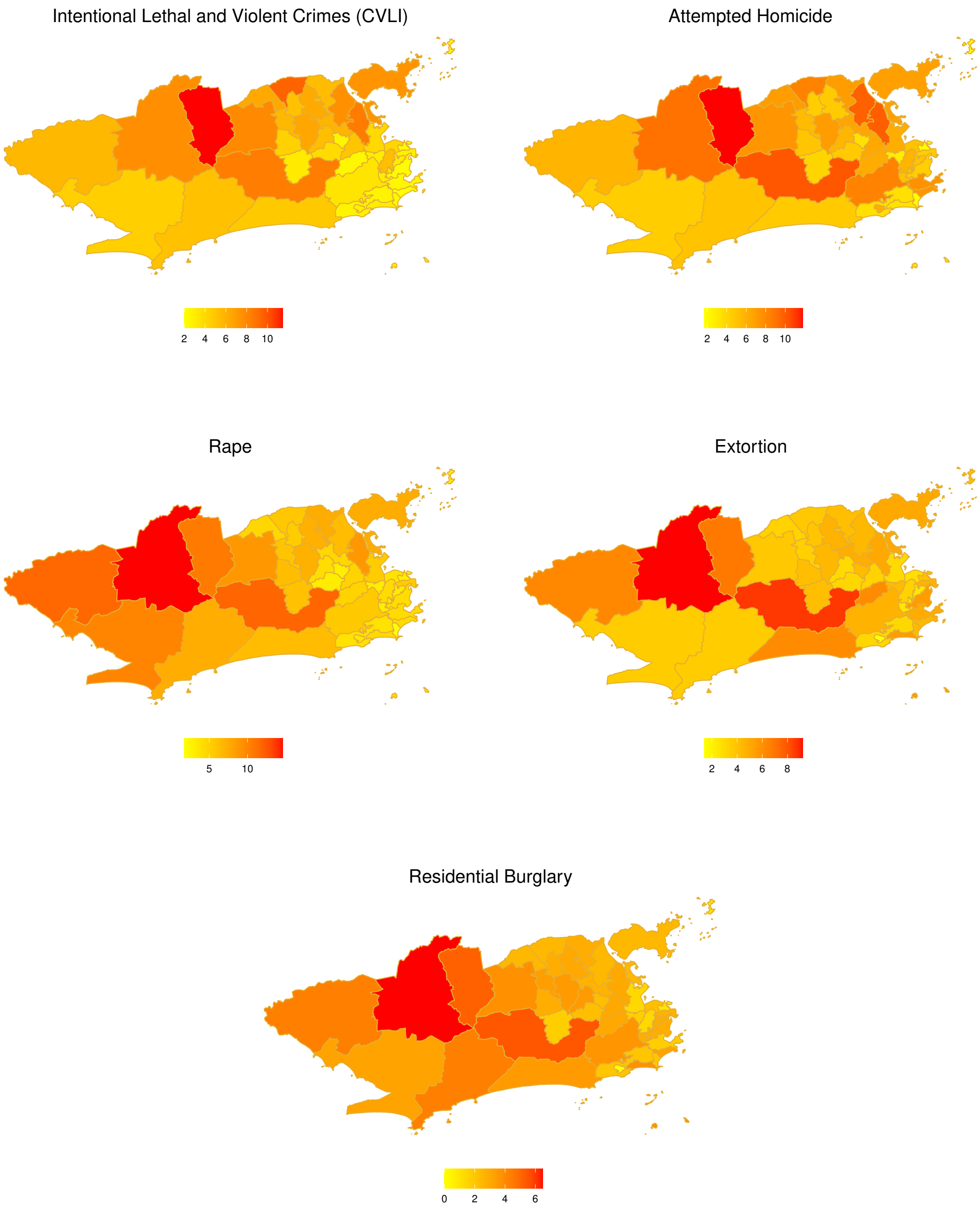

Figure 2: Spatial distribution of the total number of reported crimes (in the square root scale) in the year 2019 across the public safety integrated circumscriptions of the city of Rio de Janeiro.

Due to discontinuities in some administrative units, our forecasting exercise in this paper focuses on 37 CISPs, 14 AISPs, and 2 RISPs. Since for each one of those geographical units we consider the five crime series mentioned above, we have 5 $\times$ (37+14+2) = 265 time series. We also have the (five) crime series for the entire city of Rio, which takes us to 270 series. Finally, we have a series for the total number of crimes in the city. Therefore, we work with 271 series that span over a four-level hierarchy. Table 1 summarizes the hierarchical structure of our data set.

Table 1: Hierarchy for our data set.

Level	Number of series per crime	Total series per level
Total: Rio de Janeiro	1	1
Level 1: Crimes	5	5
Level 2: RISPs	2	10
Level 3: AISPs	14	60
Level 4 (bottom): CISPs	37	185

Figure 3 summarizes our data set's most disaggregated series. Each graph compiles a crime series. The ribbons in gray represent an interval from the 10% quantile to the 90% quantile, and the solid black lines represent the median, considering the crime counts for all 37 CISPs in a specific month. It is possible to observe that for all crimes, in most of the time series, we have low count-valued observations. This particularity implies that applying Gaussian models to fit our time series is not appropriate, as the Normal approximation to the Poisson distribution is unsuitable. Thus, we use Poisson models as suggested by Aldor-Noiman et al. (2016).

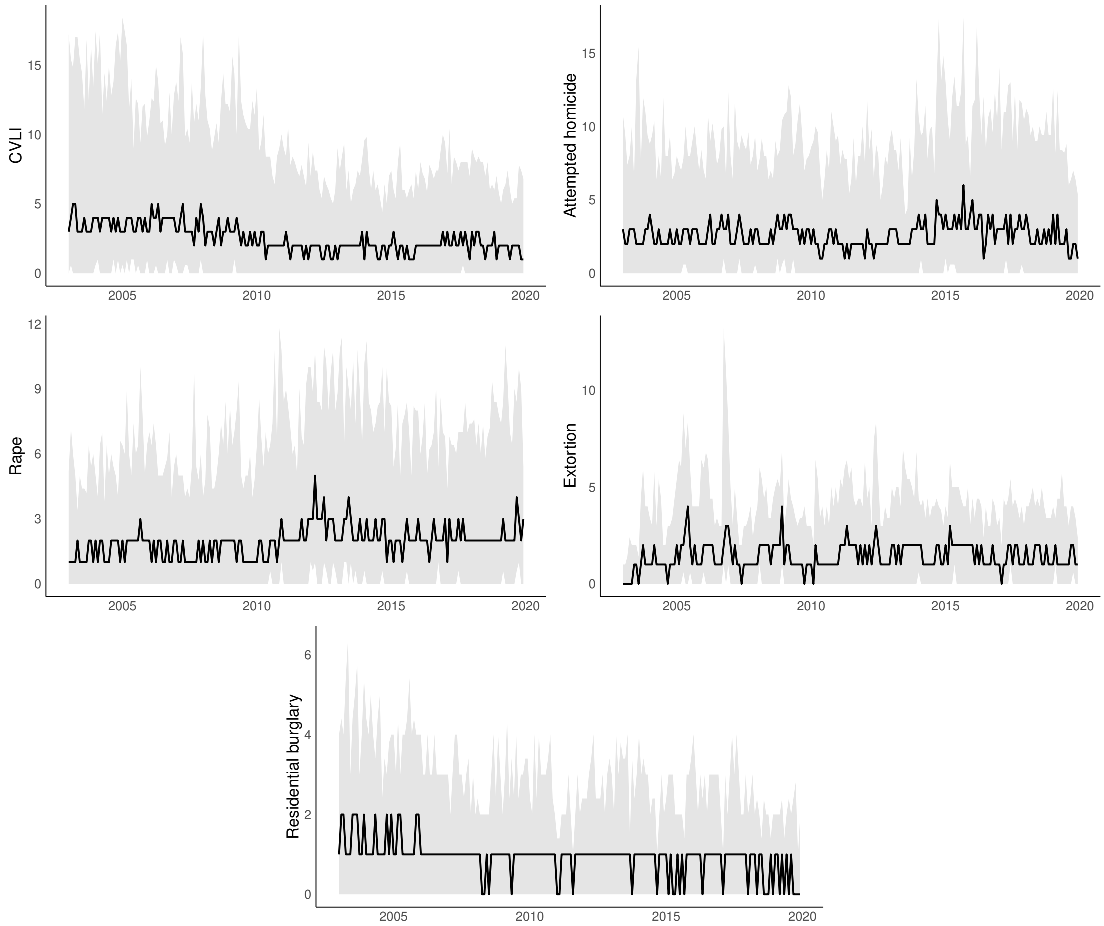

Figure 3: Graphical representation of the most disaggregated time series. Each graph compiles a crime series; the ribbons in gray represent an interval from the 10% quantile to the 90% quantile, and the solid black lines represent the median, considering the crime counts for all 37 CISPs in a specific month.

Since we have a geographical hierarchy and each CISP represents an area of Rio de Janeiro, one question that arises in our framework is that of spatial correlation. However, as stated by Wickramasuriya, Athanasopoulos, and Hyndman (2019), “One advantage of using a reconciliation approach to forecasting is that it implicitly models spatial autocorrelations in the data. With such a large collection of time series, it would be challenging to model the spatial autocorrelations directly, but through reconciliation, we can implicitly account them”. Hence, in our case, the reconciliation approach is very attractive as it not only incorporates the hierarchical structure of our data - which by itself should lead to forecasting gains - but it also allows us to implicitly incorporate spatial correlations in a way that is feasible. Since fitting 271 univariate models is less computationally costly than fitting a single multivariate model of the same dimension, the estimation process provided by reconciliation techniques is much more efficient.

Considering these characteristics of the data, the following section introduces some preliminary ideas and notations, describes the non-gaussian state-space model used for obtaining the h-step-ahead base forecasts, and summarizes the reconciliation methods we will employ.

4 Methodology

4.1 The setup

Let $\boldsymbol{y}_t \in \mathbb{R}^m$ denote an m-dimensional vector of the observations at time t of *every* time series in our dataset - including all levels of the hierarchy presented before. Let $\boldsymbol{b}_t \in \mathbb{R}^n$ denote an n-dimensional vector of only the most disaggregated series at time t. The "summing" matrix $\boldsymbol{S}$ of order $m \times n$ connects $\boldsymbol{y}_t$ and $\boldsymbol{b}_t$ via

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t.$$

The entries of $\boldsymbol{S}$ consist of zeros and ones and determine how the time series in $\boldsymbol{y}_t$ are obtained from the disaggregated series in $\boldsymbol{b}_t$. So $S_{ij} = 1$ means that the ith aggregated series contains the jth most disaggregated series and $S_{ij} = 0$ means that it does not, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

Taking an example similar to that of Wickramasuriya, Athanasopoulos, and Hyndman (2019), we let x_t be the data observed at time t at its most aggregated level; $x_{A,t}$ and $x_{B,t}$ be the observations at an intermediary aggregation level; and $x_{AA,t}, x_{AB,t}, x_{AC,t}, x_{BA,t}, x_{BB,t}$ be the disaggregated observations at the lowest hierarchical level. Figure 4 is a tree diagram of this structure. Thus we have $n = 5$, $m = 8$, $\boldsymbol{x}_t = (x_t, x_{A,t}, x_{B,t}, x_{AA,t}, x_{AB,t}, x_{AC,t}, x_{BA,t}, x_{BB,t})'$ and $\boldsymbol{b}_t = (x_{AA,t}, x_{AB,t}, x_{AC,t}, x_{BA,t}, x_{BB,t})'$. This hierarchical structure implies the following summing matrix

$$
\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ & & \boldsymbol{I}_n & & \end{bmatrix},
$$

where $\boldsymbol{I}_n$ is an identity matrix of dimension $n = 5$.

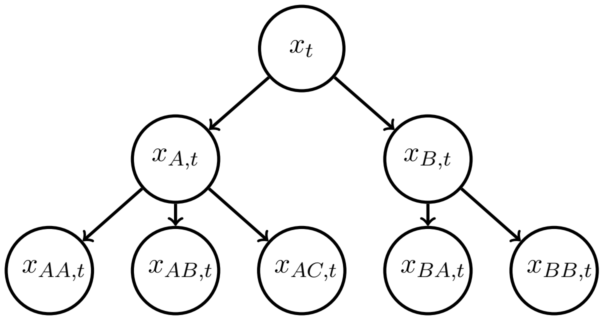

Figure 4: Example of a tree diagram for hierarchical time series.

Now, let $\hat{\boldsymbol{y}}_T(h)$ be an m-dimensional vector of h-step-ahead base forecasts for every series in the our dataset using data up to time T, and stacked in the same order as $\boldsymbol{y}_t$. Note that since any method may have been applied to obtain these predictions, they will generally not be *coherent*, which means that the forecasts of aggregate variables are not equal to the sum of the corresponding disaggregated forecasts. In this way, the forecasts do not incorporate the properties of the real data.

4.2 Base forecasts

In our data set we have count-valued time series, meaning that the vector $\boldsymbol{y}_t$ is constrained to the non-negative integers, i.e., $\boldsymbol{y}_t \in \mathbb{Z}^m_{\geq 0}$. The h-step-ahead base forecasts $\hat{\boldsymbol{y}}_T(h)$ are obtained by fitting independent state space models for each time series $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$ through the *KFAS* package (Helske, 2017) for R (R Core Team, 2013).

Our choice to model our base forecasts via state-space methods is motivated by a few reasons.

First, it is computationally inexpensive. This is important since we are working with hundreds of series in our paper. Moreover, the Kalman filter has desirable statistical properties (Alsadik, 2019) and is broadly available through packages for R, which further facilitates its implementation. It is also worth noting that although Kalman filter estimates are derived from a maximum likelihood problem, the optimality of base forecasts is not indispensable within the context of forecast reconciliation, since regardless of how base forecasts are obtained, reconciliation procedures *already* impose optimality during aggregation. This is a result that applies more broadly to forecast combinations in general: as shown in Kolassa (2011), forecast combinations limit the need for model selection as they reduce the uncertainty of model specification and estimation. Therefore, although our base forecasts and the forecasts used by the Secretaria de Segurança Publica might not be estimated by similar methods, reconciliation can still generate gains - even if base forecasts are not derived from an optimization problem or formal model at all!

Our state-space model takes the form:

$$p(y_{it}|\theta_{it}) = p(y_{it}|\boldsymbol{Z}_{it}\boldsymbol{\alpha}_{it}) \tag{1}$$

$$\boldsymbol{\alpha}_{it+1} = \boldsymbol{V}_{it}\boldsymbol{\alpha}_{it} + \boldsymbol{U}_{it}\boldsymbol{\eta}_{it}, \tag{2}$$

where 1 is the Observation Equation and 2 is the State Equation. In this model, $\boldsymbol{\alpha}$ is the unobserved state vector, $\theta_{it} = \boldsymbol{Z}_{it}\boldsymbol{\alpha}_{it}$ is a signal of this state and $p(y_{it}|\theta)$ is a density function, which implies a relationship between y_{it} and $\boldsymbol{\alpha}_{it}$. We assume that $\boldsymbol{\eta}_{it} \sim N(0, \boldsymbol{Q}_{it})$ and that y_{it} has a Poisson distribution with parameter λ_{it}. This implies that the signal θ_{it} is connected to λ_{it} via a logarithmic link function $\theta_{it} = \log(\lambda_{it})$.

We specify system matrices $\boldsymbol{Z}_{it}$, $\boldsymbol{U}_{it}$, $\boldsymbol{V}_{it}$, and the covariance matrix $\boldsymbol{Q}_{it}$, as time-invariant structures, i.e., do not depend on time, and also identical structures for all i. This implies that the state process is a random walk with a drift, as follows

$$
\begin{aligned}
\boldsymbol{Z} &= \begin{pmatrix} 1 & 0 \end{pmatrix} \qquad \boldsymbol{\alpha}_{it} = \begin{pmatrix} \mu_{it} \\ \nu_{it} \end{pmatrix} \\
\boldsymbol{V} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \boldsymbol{U} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\boldsymbol{Q}_i &= \tau^2.
\end{aligned}
$$

The algorithm we use to estimate our Poisson model comes from (Helske, 2017). It actually finds a Gaussian model that has the same conditional posterior mode $p(\theta|\boldsymbol{y})$ through an iterative process where $p(\theta|\boldsymbol{y})$ is approximated via the Laplace method. Then, θ_t estimates are updated by applying a Kalman filter to this approximating Gaussian model. The final estimates $\hat{\theta}_t$ correspond to the mode of $p(\theta|\boldsymbol{y})$.[2] Once we have the estimates for $\hat{\theta}_t$, the algorithm recovers the values of $\hat{\lambda}_{it}$ in a way that minimizes the bias introduced by the direct transformation $\hat{\lambda}_{it} = e^{\hat{\theta}_{it}}$. We refer the interested reader to (Helske, 2017) for a more detailed exposition of this procedure.

4.3 Forecast Reconciliation

The class of linear reconciliation methods can be written as

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h),$$

where $\boldsymbol{P}$ is a projection matrix of order $n \times m$ and $\tilde{\boldsymbol{y}}_T(h)$ is a vector of reconciled forecasts that are coherent by construction. The main idea of linear reconciliation methods is to map a vector of base forecasts into a set of reconciled ones. The simplest approach, in this case, is known as "bottom-up" and consists of $\boldsymbol{P} = \boldsymbol{I}_n$, which means that the forecasts of the various aggregated series are simply summations of the most disaggregated series (Dunn, Williams, and DeChaine, 1976; Orcutt, Watts, and Edwards, 1968; Shlifer and Wolff, 1979).

[2] In models such as ours, the difference between the mode and the mean is negligible.

In this article, we focus on the least squares reconciled forecasts. Consider first the following regression model, as in Hyndman et al. (2011)

$$\hat{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{\beta}_T(h) + \boldsymbol{\epsilon}_h,$$

where $\boldsymbol{\beta}_T(h) = \mathrm{E}\left[\boldsymbol{b}_{T+h}|\boldsymbol{\mathcal{I}}_T\right]$ is the unknown conditional mean of the future values of the most disaggregated series at the bottom level of the hierarchy, $\boldsymbol{\mathcal{I}}_t = (\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots)$ is the set of all available information up to time t, and $\boldsymbol{\epsilon}_h$ is the ”coherency error”, that is, the difference between the base forecasts $\hat{\boldsymbol{y}}_h$ and their reconciled values $\tilde{\boldsymbol{y}}_h$. We assume $\mathrm{E}\left[\boldsymbol{\epsilon}_h|\boldsymbol{\mathcal{I}}_T\right] = 0$ and $\mathrm{var}\left(\boldsymbol{\epsilon}_h|\boldsymbol{\mathcal{I}}_T\right) = \boldsymbol{W}_h^R$, a positive definite matrix. In this case, least squares reconciled forecasts are characterized by the following projection matrix

$$\boldsymbol{P} \;=\; \left(\boldsymbol{S}'\boldsymbol{W}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W} \tag{3}$$

where the weighting matrix $\boldsymbol{W}$ is equal to $(\boldsymbol{W}_h^R)^{-1}$

If $\boldsymbol{W}_h^R$ were known, the generalized least squares (GLS) estimator of $\boldsymbol{\beta}_T(h)$ would be the minimum variance unbiased estimator. But in practice $\boldsymbol{W}_h^R$ is unknown and needs to be estimated. As shown by Wickramasuriya, Athanasopoulos, and Hyndman (2019), however, it is impossible to estimate $\boldsymbol{W}_h^R$ due to identifiability conditions. Thus, the solution found by Hyndman et al. (2011) and Athanasopoulos et al. (2017) was to use alternative weighting matrices that led to OLS and WLS estimators that we detail further on.

Another solution was later proposed by Wickramasuriya, Athanasopoulos, and Hyndman (2019). They develop a more general framework that encompasses the OLS and WLS estimators we just mentioned. Moreover, instead of using $\boldsymbol{W}_h^R$, the covariance matrix of *coherency* errors, they use the covariance matrix of the base *forecast* errors $\boldsymbol{W}_h^F = var[\boldsymbol{y}_T(h) - \hat{\boldsymbol{y}}_T(h)]$ as the weighting matrix. They show that the correspondent projection matrix $\boldsymbol{P}$ minimizes the trace of var $(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_T(h)|\boldsymbol{\mathcal{I}}_t)$ among all possible unbiased reconciliations. In other words, their solution gives the minimum variance linear unbiased reconciled forecasts. On a subsequent paper, Wickramasuriya, Turlach, and Hyndman (2020) further extend their framework to also include non-negativity constraints in their forecasts.

4.4 Alternative estimators for $\boldsymbol{W}_h$

As shown by Athanasopoulos et al. (2017), the identification of $\boldsymbol{W}_h^R$ is impossible. Therefore, this section presents four alternatives for the weighting matrix that have been put forth in the literature:

1. $\boldsymbol{W} = \boldsymbol{I}_n$. This was introduced by Hyndman et al. (2011), who avoid the estimation of $\boldsymbol{W}_h$ altogether by using a simple ordinary least squares (OLS) estimator, thus replacing $\boldsymbol{W}_h$ by $\boldsymbol{W} = \boldsymbol{I}_n$.

2. $\boldsymbol{W} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{S}\mathbf{1})$, with $\mathbf{1}$ being a unit n-dimensional column vector. This weighted least squares (WLS) estimator was proposed by Athanasopoulos et al. (2017) in the context of temporal hierarchies and it possesses some desirable properties, such as the facts that it is computationally simple to implement and it depends only on the hierarchical structure of the data, which is given by the summation matrix $\boldsymbol{S}$. They call it a "*structural scaling*" estimator. Hereafter, this is denoted by WLS$_s$;

3. $\boldsymbol{W} = \text{diag}(\widehat{\boldsymbol{W}}_1^F)$, where

$$\widehat{\boldsymbol{W}}_1^F = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{e}}_t(1)\hat{\boldsymbol{e}}_t'(1),$$

   is the one-step-ahead base forecast error sample covariance matrix. This WLS estimator was initially used by Athanasopoulos et al. (2017), who called it a *series variance scaling* estimator. It was also used by Hyndman, Lee, and Wang (2016). The idea is to take only the diagonal of $\widehat{\boldsymbol{W}}_1^F$ due to the fact that using the entire matrix can generate poor estimates in applications such as ours, where the number of series is large compared to the number of time periods. Hereafter, we denote this as WLS$_v$;

4. $\boldsymbol{W} = \lambda \operatorname{diag}(\widehat{\boldsymbol{W}}_1^F) + (1-\lambda)\widehat{\boldsymbol{W}}_1^F$, where λ_D is a shrinkage intensity parameter. This was suggested by Wickramasuriya, Athanasopoulos, and Hyndman (2019) and it is a more general case of the previous estimator. Like before, the idea is to use the diagonal matrix $\widehat{\boldsymbol{W}_1^F}$, which leaves variances unchanged, and to shrink its off-diagonal elements towards zero. If λ = 1, we get complete shrinkage, thus obtaining the WLS$_v$ estimator from before. In our exercise, we follow Schafer and Strimmer (2005) and use

$$\hat{\lambda} = \frac{\sum_{i \neq j} \widehat{\operatorname{var}(\hat{r_{ij}})}}{\sum_{i \neq j} \hat{r}_{ij}^2},$$

where $\hat{r}_{ij}$ is the ijth entry of 1-step-ahead sample correlation matrix. This is shown to be the analytic solution for the optimal λ parameter when minimizing a mean squared error risk function. We denote this as MinT.

5 Results

In this section, we evaluate the impact of reconciliation on base forecasts considering the time series of crimes in the city of Rio de Janeiro. To that end, we use a rolling window with initial training data equal to 108 observations and generate 1- to 3-steps-ahead base forecasts for each 271 series. This procedure results in 96 1-step-ahead reconciled forecasts, 95 2-step-ahead reconciled forecasts, and 94 3-step-ahead reconciled forecasts.

The base forecasts were obtained through the state space model for count-valued time series presented in Subsection 4.2. Furthermore, for comparison purposes, we consider the “bottom-up” approach, the least squares reconciliation methods presented on the previous section, and the same least squares methods with non-negativity constraints, as proposed by Wickramasuriya, Turlach, and Hyndman (2020). The least squares methods were implemented via *hts* package for R (R Core Team, 2013). We apply the four alternative estimators for the weighting matrix $\boldsymbol{W}_h$ described in

Subsection 4.4 for the least squares and the non-negative least squares reconciliations. The results are compared using the root mean squared error (RMSE).

Figure 5 and Table 2 in Appendix A summarize the results for all hierarchical levels, methods, and forecast horizons under analysis. The highlighted (larger) circles indicate the best methods. RMSE differences in scales among the hierarchies are due to differences in scales in data per se, which means that the RMSE should not be compared among levels but only within the same level. Moreover, the RMSE should only be compared for a fixed number of h steps ahead. Thus, our analysis sticks to comparisons across methods within levels concerning specific steps ahead.

For the higher hierarchy levels of RISP, Crime, and the city of Rio de Janeiro, there seems to be a pattern of dominance by the OLS approach over other methods that employ more sophisticated covariance structures. Observing Figure 5 and Table 2, the OLS approach minimizes the root mean squared error (RMSE) of the RISP, Crime, and Rio de Janeiro forecasts for all three forecasting horizons - the only exception being the three-steps-ahead forecast for Crime. Moreover, at these hierarchical levels the non-negativity constraint is irrelevant: regardless of whether or not it is implemented, the OLS root mean squared errors are essentially the same. This is because as base forecasts are aggregated, reconciled forecasts will rarely be equal to zero at higher levels of aggregation, so the constraint becomes non-binding.

As pointed out by Wickramasuriya, Athanasopoulos, and Hyndman (2019), the OLS approach used by Hyndman et al. (2011) is theoretically optimal only under certain conditions which are rarely verified in practice. Nevertheless, it is one of the two dominant estimators in our exercise. This is in line with a similar result often found in the forecast combination literature: simple approaches, such as the unweighted average of forecasts, tend to perform at least as well as more sophisticated combination schemes Timmermann (2006).

On the lower hierarchy levels, the OLS dominance disappears, and the most efficient method seems to be the WLS_v estimator proposed by Hyndman, Lee, and Wang (2016) and Athanasopoulos et al. (2017). Indeed, as can be seen in the results, this method is the most efficient one for forecasting crimes in the AISP and CISP hierarchy levels for all forecasting horizons - the one exception being the one-step-ahead forecast for the AISP level, where the OLS method is slightly superior. As mentioned before, using the base forecast errors sample covariance matrix as the weighting matrix provides the minimum variance linear unbiased reconciled forecasts. Since the WLS_v estimator uses the diagonal of that matrix (due to dimensionality reasons), strong forecasting performance is not surprising. Interestingly, the MinT approach is not the most efficient method for any of our forecasts.

As expected, the further in time one tries to forecast crime, the higher the RMSE of the estimates. This reflects the greater uncertainty associated with forecasts for many periods ahead. Also, according to intuition, the popular "bottom-up" reconciliation approach performs significantly worse than every other method we test - especially as the hierarchy levels increase. For example, take some of the forecasts for AISP and Rio de Janeiro. For $h = 3$, the optimal forecast for Rio de Janeiro has an RMSE of 52.9 (OLS), whereas the Bottom-up RMSE is almost three times larger, at 150.4. On the other hand, for $h = 1$, the AISP Bottom-up forecast has a real mean squared error of 5.3, while the errors for the optimal forecast (OLS) are only slightly lower at 4.9. This illustrates how pervasive aggregation errors associated with statistically unfounded reconciliation techniques can be when dealing with hierarchical time series. Comparing the results from the least squares reconciliation methods and the base forecasts, we observe that the OLS and the WLS_s consistently overcome the base forecasts regardless of the case under analysis.

The effects of non-negativity constraints are also noteworthy. For the lower levels of hierarchy in our data, the non-negativity constraint seems to provide not only coherent forecasts but also improvements in efficiency, since among the methods we tested, for the CISP and AISP hierarchies, the non-negative WLS_v is the one with the smallest RMSE. However, for higher hierarchical levels, including a non-negativity constraint seems not to improve accuracy in any significant way, as we have already seen in the RISP, Crime, and Rio de Janeiro results. This is intuitive: many crime series have several observations that are either equal to zero or have very low counts that are close to zero. In lower hierarchical levels, this fact increases the probability that the non-negativity constraint will bind, thus providing gains in forecasting accuracy. As we move up the hierarchical order and aggregate our series, the observations close to zero become more infrequent, so adding a non-negativity constraint becomes innocuous.

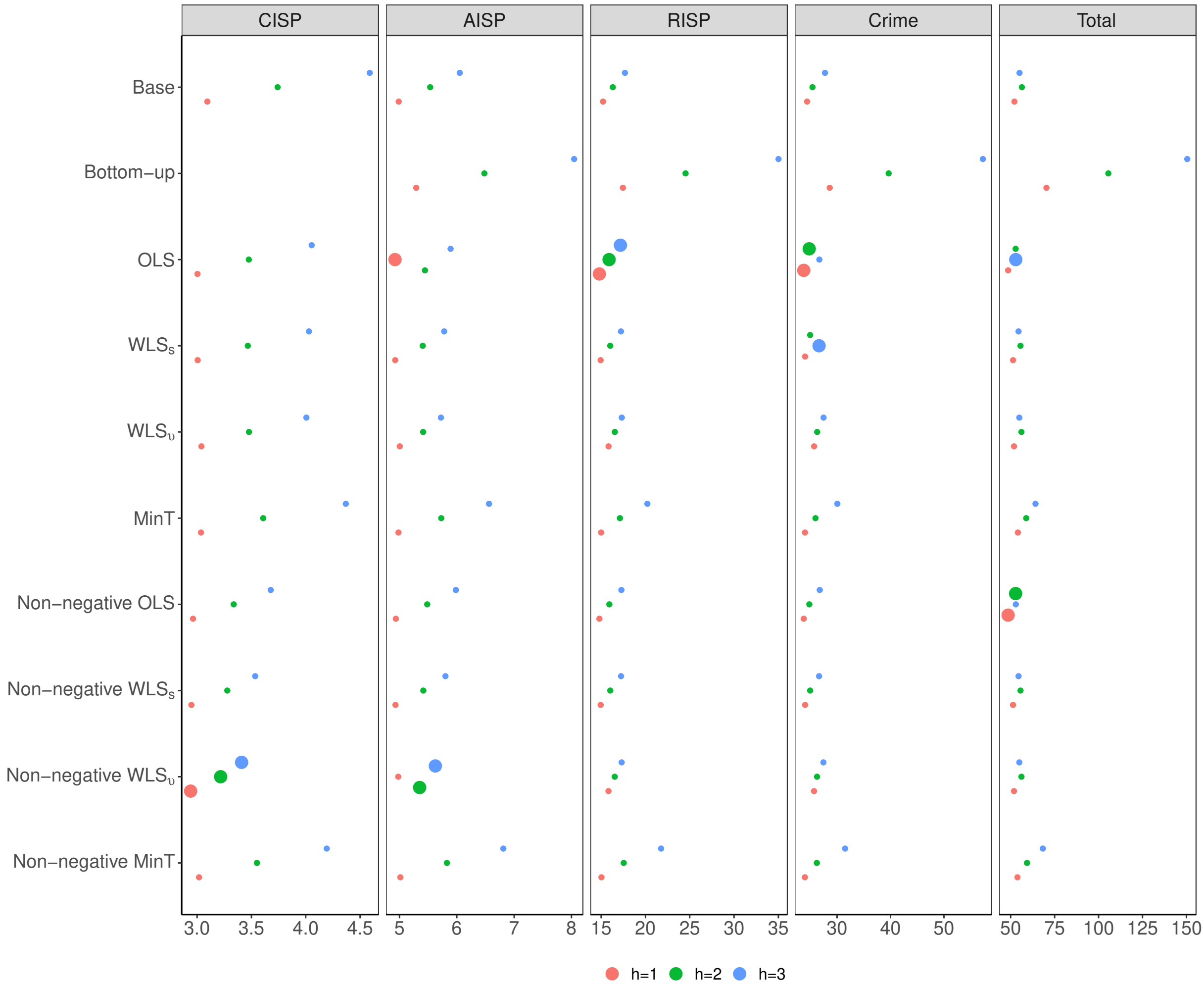

Figure 5: Out-of-sample forecast performances based on the root mean square error (RMSE). Axis x and y represent the RMSE and the reconciliation methods, respectively. The highlighted (larger) circles indicate the best methods, and the circles to the right indicate worse performances.

6 Final Remarks

In this paper, we introduce reconciliation techniques to crime forecasting. To the best of our knowledge, such techniques have not yet been utilized in the criminology literature. Reconciliation allows us to work with hundreds of crime series and to implicitly explore information related to both time and space correlation, as well as to the hierarchical structure of the data, which leads to accuracy gains in prediction exercises. It does so in a simple and computationally efficient manner. As underlined by Wickramasuriya, Turlach, and Hyndman (2020), this is because reconciliation techniques allow the reduction of a complex multivariate forecasting problem to a collection of simple univariate forecasting problems, which are then reconciled in a manner that captures the relationships between the series through a weighting matrix.

Crime forecasts are used by authorities in order to evaluate and optimally design public safety policies. In Rio de Janeiro, the Institute of Public Safety (ISP) uses crime forecasts to subsidize police deployment decisions across the city, as shown in Seixas et al. (2010). We show that reconciliation techniques greatly improve the accuracy of such forecasts in comparison to standard aggregation techniques, such as the bottom-up approach.

We compare the performance of different reconciliation approaches and make use of 271 different time series that span over a four-level hierarchy and that are possibly highly correlated among themselves due to the spatial nature of the data. This is a typical situation where reconciliation techniques may prove useful since they allow us to individually forecast each of the series and then optimally combine them instead of modeling every series in a single multivariate model that would need to take into account a complex covariance structure, with prohibitive computational costs.

Our base forecasts are obtained through standard state-space models. Since crime count series typically have several observations equal to zero (or very close to zero), especially in lower hierarchical levels, we assume a Poisson distribution for our data-generating process and then use a Kalman filter to obtain our estimates.

We evaluate five reconciliation methods: an intuitive (but statistically unfounded) method called the “bottom-up” approach, where we sum our base forecasts and four least-squares reconciliation methods. For each of these models, we run regressions both with and without the inclusion of a non-negativity constraint.

The results show that for higher hierarchical levels, the simple OLS approach is usually preferred over others with more sophisticated covariance structures. The non-negativity constraint seems to be of little importance in such cases. As one turns his attention to crime series on lower hierarchical levels, the OLS estimator ceases to be optimal, and it is replaced by WLS_v estimator instead. Moreover, since these are the series with more observations equal or close to zero, the non-negativity constraint becomes active more often, thus providing efficiency gains in the estimation process. The bottom-up approach is dominated by every other method.

Therefore, our paper introduces a simple yet effective manner of improving the accuracy of crime forecasts in a high-dimensional setting. Doing so provides higher-quality information to be utilized in the decision-making process of police authorities. As Seixas et al. (2010) explained, the city's Public Safety Secretariat implemented a "Plan-Do-Study-Act" management model, where police action plans are designed based on data regarding criminal activity and then subsequently analyzed to detect patterns or causes for criminality. Within this setup, among the data analyzed by authorities, criminal activity forecasts, in particular, are one of the four measures used in the construction of an index that orients police deployment throughout the city every quarter. The results in our paper could be used to improve police deployment and thus help alleviate, at least to some extent, the issue of violent crime.

Lastly, all data sets and codes are available at `https://github.com/marcuslavagnole` for reproducibility purposes.

References

Aldor-Noiman, Sivan, Lawrence D. Brown, Emily B. Fox, and Robert A. Stine. 2016. "Spatio-temporal low count processes with application to violent crime events." *Statistica Sinica* 26 (4): 1587–1610.

Alsadik, Bashar. 2019. "Kalman Filter." In *Adjustment Models in 3D Geomatics and Computational Geophysics*, edited by Bashar Alsadik, Vol. 4 of *Computational Geophysics*, 299–326. Elsevier.

Athanasopoulos, George, Rob J.Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. 2017. "Forecasting with Temporal Hierarchies." *European Journal of Operational Research* 262 (1): 60–74.

Berk, Richard. 2008. "Forecasting Methods in Crime and Justice." *Annual Review of Law and Social Science* 4: 219–238.

Brantingham, Paul J., and Patricia L. Brantingham. 1984. *Patterns in Crime*. 1st ed. New York: Macmillan.

Brantingham, Paul Jeffrey. 2016. "Crime diversity." *Criminology* 54 (4): 553–586.

Bruinsma, Gerben J. N., and Shane D. Johnson. 2018. *The Oxford Handbook of Environmental Criminology*. 1st ed. New York: Oxford University Press.

Chainey, Spencer, and Jerry Ratcliffe. 2005. *GIS and Crime Mapping*. 1st ed. New Jersey: John Wiley & Sons.

Chainey, Spencer, Lisa Tompson, and Sebastian Uhlig. 2008. "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime." *Security Journal* 21: 4–28.

Cohen, Jacqueline, Wilpen L. Gorr, and Andreas M. Olligschlaeger. 2007. "Leading Indicators and Spatial Interactions: A Crime-Forecasting Model for Proactive Police Deployment." *Geographical Analysis* 39 (1): 105–127.

Corcoran, Jonathan J., Ian D. Wilson, and J.Andrew Ware. 2003. "Predicting the geo-temporal variations of crime and disorder." *International Journal of Forecasting* 19 (4): 623–634.

Davis, Robert C. 2012. *Selected International Best Practices in Police Performance Measurement*. Technical Report. RAND Corporation.

Deadman, Derek. 2003. "Forecasting residential burglary." *International Journal of Forecasting* 19 (4): 567–578.

Deadman, Derek, and David Pyle. 1997. "Forecasting recorded property crime using a time-series econometric model." *The British Journal of Criminology* 37 (3): 437–445.

Dunn, D. M., William H. Williams, and T. L. DeChaine. 1976. "Aggregate Versus Subaggregate Models in Local Area Forecasting." *Journal of American Statistical Association* 71 (353): 68–71.

Durbin, James, and Siem Jan Koopman. 2000. "Time Series Analysis of Non-Gaussian Observations Based on State Space Models from Both Classical and Bayesian Perspectives." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 62 (1): 3–56.

Durbin, James, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*. 2nd ed. New York: Oxford University Press.

Ferraz, Claudio, Joana Monteiro, and Bruno Ottoni. 2021. "Regaining the Monopoly of Violence: Evidence from the Pacification of Rio de Janeiro's Favelas." Preprint on webpage at `https://joanacmmonteiro.com/pesquisa/`.

Fielding, Nigel, and Martin Innes. 2006. “Reassurance Policing, Community Policing and Measuring Police Performance.” *Policing and Society* 16 (2): 127–145.

Filstad, Cathrine, and Petter Gottschalk. 2009. “Performance evaluation of police oversight agencies.” *Policing and Society* 21 (1): 96–109.

Fliedner, Gene. 2001. “Hierarchical forecasting: issues and use guidelines.” *Industrial Management & Data Systems* 101 (1): 5–12.

Gerber, Matthew S. 2014. “Predicting crime using Twitter and kernel density estimation.” *Decision Support Systems* 61: 115–125.

Gorr, Wilpen, and Richard Harries. 2003. “Introduction to crime forecasting.” *International Journal of Forecasting* 19 (4): 551–555.

Gorr, Wilpen, Andreas Olligschlaeger, and Yvonne Thompson. 2003. “Short-term forecasting of crime.” *International Journal of Forecasting* 19 (4): 579–594.

Harries, Richard. 2003. “Modelling and predicting recorded property crime trends in England and Wales—a retrospective.” *International Journal of Forecasting* 19 (4): 557–566.

Hart, Timothy, and Paul Zandbergen. 2014. “Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting.” *Policing: An International Journal* 37 (2): 305–323.

Helske, Jouni. 2017. “KFAS: Exponential Family State Space Models in R.” *Journal of Statistical Software* 78 (10): 1–39.

Hu, Yujie, Fahui Wang, Cecile Guin, and Haojie Zhu. 2018. “A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation.” *Applied Geography* 99: 89–97.

Hyndman, Rob J., Roman A. Ahmed, George Athanasopoulos, and Han Lin Shang. 2011. “Optimal combination forecasts for hierarchical time series.” *Computational Statistics and Data Analysis* 55: 2579–2589.

Hyndman, Rob J., Alan J. Lee, and Earo Wang. 2016. “Fast computation of reconciled forecasts for hierarchical and grouped time series.” *Computational Statistics and Data Analysis* 97: 16–32.

Jha, Sudan, Eunmok Yang, Alaa Omran Almagrabi, Ali Kashif Bashir, and Gyanendra Prasad Joshi. 2021. “Comparative analysis of time series model and machine testing systems for crime forecasting.” *Neural Computing and Applications* 33: 10621–10636.

Jiang, Quanbao, and Jesús Javier Sánchez Barricarte. 2011. "A Crime Rate Forecast and Decomposition Method." *International Journal of Criminology and Sociological Theory* 4 (2): 648–656.

Kadar, Cristina, and Irena Pletikosa. 2018. "Mining large-scale human mobility data for long-term crime prediction." *EPJ Data Science* 7: 1–27.

Kolassa, Stephan. 2011. "Combining exponential smoothing forecasts using Akaike weights." *International Journal of Forecasting* 27 (2): 238–251.

Kounadi, Ourania, Alina Ristea, Adelson Araujo Jr., and Michael Leitner. 2020. "A systematic review on spatial crime forecasting." *Crime Science* 9: 1–22.

Liesenfeld, Roman, Jean-François Richard, and Jan Vogler. 2017. "Likelihood-Based Inference and Prediction in Spatio-Temporal Panel Count Models for Urban Crimes." *Journal of Applied Econometrics* 32 (3): 600–620.

Orcutt, Guy H., Harold W. Watts, and John B. Edwards. 1968. "Data Aggregation and Information Loss." *The American Economic Review* 58 (4): 773–787.

R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. `http://www.R-project.org/`.

Schafer, Juliane, and Korbinian Strimmer. 2005. "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics." *Statistical Applications in Genetics and Molecular Biology* 4 (1).

Seixas, Luis Antônio Fonseca, Ricardo Ribas da Costa, Vinícius Augusto R. Chelotti Brum de Almeida, Maíra Alarcon Salomão, and Greice Maria Silva da Conceição. 2010. *Manual de procedimentos para o sistema de metas e acompanhamento de resultados*. Secretaria de Segurança Pública do Estado do Rio de Janeiro.

Shamsuddin, Nurul Hazwani Mohd, Nor Azizah Ali, and Razana Alwee. 2017. "An overview on crime prediction methods." In *2017 6th ICT International Student Project Conference (ICT-ISPC)*, 1–5.

Shlifer, E., and R. W. Wolff. 1979. "Aggregation and Proration in Forecasting." *Management Science* 25 (6): 594–603.

Sonnichsen, Richard C. 1997. "Measuring Police Performance." In *Monitoring Performance in the Public Sector*, Routledge.

Tague, Nancy R. 2023. *The Quality Toolbox*. 3rd ed. Milwaukee: ASQ Quality Press.

Timmermann, Alan. 2006. "Forecast Combinations." *Handbook of Economic Forecasting* 1: 135–196.

Tiwana, Neena, Gary Bass, and Graham Farrell. 2015. “Police performance measurement: an annotated bibliography.” *Crime Science* 4: 1–28.

Vomfell, Lara, Wolfgang Karl Härdle, and Stefan Lessmann. 2018. “Improving crime count forecasts using Twitter and taxi data.” *Decision Support Systems* 113: 73–85.

Wang, Hongjian, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. “Crime Rate Inference with Big Data.” In *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 635–644.

Wang, Xiaofeng, Donald E. Brown, and Matthew S. Gerber. 2012. “Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information.” In *2012 IEEE International Conference on Intelligence and Security Informatics*, 36–41.

Wickramasuriya, Shanika L., George Athanasopoulos, and Rob J. Hyndman. 2019. “Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization.” *Journal of the American Statistical Association* 114 (526): 804–819.

Wickramasuriya, Shanika L., Berwin A. Turlach, and Rob J. Hyndman. 2020. “Optimal non-negative forecast reconciliation.” *Statistics and Computing* 30: 1167–1182.

Zhao, Xiangyu, and Jiliang Tang. 2017. “Modeling Temporal-Spatial Correlations for Crime Prediction.” In *2017 ACM on Conference on Information and Knowledge Management*, 497–506.

A Out-of-sample forecast performances

Table 2: Out-of-sample forecast performances based on the root mean square error (RMSE).

		Base	Bottom-up	Real				Non negative real			
				OLS	WLS_s	WLS_v	MinT	OLS	WLS_s	WLS_v	MinT
Total	$h=1$	52.134	70.329	48.545	51.343	51.893	54.099	**48.539**	51.343	51.893	53.859
	$h=2$	56.374	105.553	52.808	55.580	56.127	58.834	**52.800**	55.580	56.127	59.315
	$h=3$	55.039	150.442	**52.904**	54.475	54.924	64.106	52.906	54.475	54.923	68.288
Crime	$h=1$	24.395	28.627	**23.767**	24.026	25.709	24.001	23.769	24.026	25.688	23.985
	$h=2$	25.406	39.650	**24.789**	24.963	26.279	25.963	24.808	24.964	26.254	26.215
	$h=3$	27.740	57.280	26.689	**26.633**	27.474	30.054	26.762	26.634	27.448	31.512
RISP	$h=1$	15.217	17.456	**14.797**	14.939	15.833	14.999	14.799	14.939	15.824	15.029
	$h=2$	16.312	24.525	**15.893**	16.023	16.542	17.116	15.921	16.025	16.531	17.544
	$h=3$	17.684	35.017	**17.175**	17.238	17.326	20.216	17.281	17.242	17.311	21.764
AISP	$h=1$	4.986	5.292	**4.921**	4.925	5.005	4.981	4.936	4.930	4.978	5.015
	$h=2$	5.535	6.480	5.444	5.405	5.412	5.728	5.483	5.415	**5.351**	5.827
	$h=3$	6.051	8.046	5.891	5.779	5.723	6.561	5.983	5.802	**5.625**	6.811
CISP	$h=1$	3.094	-	3.003	3.005	3.039	3.036	2.962	2.947	**2.940**	3.018
	$h=2$	3.740	-	3.476	3.466	3.477	3.608	3.337	3.277	**3.216**	3.551
	$h=3$	4.589	-	4.054	4.029	4.005	4.368	3.677	3.534	**3.409**	4.193

B RISP x AISP x CISP x Neighborhood

Table 3: Relations among CISPs, AISPs, RISPs, and neighborhoods in the city of Rio de Janeiro.

RISP	AISP	CISP	Unidade Territorial
1	5	1	Centro (parte)
1	5	4	Centro (parte), Gamboa, Santo Cristo, Saúde
1	5	5	Centro (parte), Lapa, Paquetá
1	4	6	Catumbi, Cidade Nova, Estácio, Rio Comprido, Centro (parte)
1	5	7	Santa Teresa
1	2	9	Catete, Cosme Velho, Flamengo, Glória, Laranjeiras
1	2	10	Botafogo, Humaitá, Urca
1	23	11	Rocinha
1	19	12	Copacabana (parte), Leme
1	19	13	Copacabana (parte)
1	23	14	Ipanema, Leblon
1	23	15	Gávea, Jardim Botânico, Lagoa, São Conrado, Vidigal
2	31	16	Barra da Tijuca, Itanhangá, Joá
1	4	17	Caju, Mangueira, São Cristóvão, Vasco da Gama
1	6	18	Maracanã, Praça da Bandeira, Tijuca (parte)
1	6	19	Alto da Boa Vista, Tijuca (parte)
1	6	20	Andaraí, Grajaú, Vila Isabel
1	22	21	Benfica, Bonsucesso, Higienópolis, Manguinhos, Maré, Ramos
1	16	22	Brás de Pina (parte), Olaria, Penha, Penha Circular (parte)
1	3	23	Cachambi, Méier (parte), Todos os Santos (parte)
1	3	24	Abolição, Água Santa (parte), Encantado, Engenho de Dentro (parte), Pilares, Piedade
1	3	25	Engenho Novo, Jacaré, Jacarezinho, Riachuelo, Rocha, Sampaio, São Francisco Xavier
1	3	26	Água Santa (parte), Engenho de Dentro (parte), Lins de Vasconcelos, Todos os Santos
2	41	27	Colégio (parte), Irajá, Vicente de Carvalho, Vila Kosmos, Vila da Penha, Vista Alegre
2	18	28	Vila Valqueire, Praça Seca, Tanque (parte)
2	9	29	Cavalcanti, Engenheiro Leal, Madureira, Turiaçu, Vaz Lobo, Cascadura, Oswaldo Cruz (parte), Quintino Bocaiúva
2	9	30	Bento Ribeiro, Campinho, Marechal Hermes, Oswaldo Cruz (parte)
2	41	31	Anchieta, Guadalupe, Parque Anchieta, Ricardo de Albuquerque
2	18	32	Anil, Cidade de Deus, Curicica, Gardênia Azul, Jacarepaguá, Taquara
2	14	33	Campo dos Afonsos, Deodoro, Jardim Sulacap, Magalhães Bastos, Realengo, Vila Militar
2	14	34	Bangu, Gericinó, Padre Miguel, Senador Camará
2	40	35	Campo Grande, Cosmos, Inhoaíba, Santíssimo, Senador Vasconcelos
2	27	36	Paciência e Santa Cruz
1	17	37	Bancários, Cacuia, Cidade Universitária, Cocotá, Freguesia, Galeão, Jardim Carioca, Jardim Guanabara, Moneró, Pitangueiras, Portuguesa, Praia da Bandeira, Ribeira, Tauá, Zumbi
1	16	38	Brás de Pina (parte), Cordovil, Jardim América, Parada de Lucas, Penha Circular (parte), Vigário Geral
2	41	39	Acari, Barros Filho, Costa Barros, Parque Colúmbia, Pavuna
2	9	40	Coelho Neto, Colégio (parte), Honório Gurgel, Rocha Miranda
2	18	41	Freguesia (Jacarepaguá), Pechincha e Tanque (parte)
2	31	42	Recreio dos Bandeirantes, Barra de Guaratiba, Camorim, Grumari, Vargem Grande, Vargem Pequena
2	27	43	Guaratiba, Pedra de Guaratiba, Sepetiba
1	3	44	Del Castilho, Engenho da Rainha, Inhaúma, Maria da Graça, Tomás Coelho