



RADIOSCIENCES AU SERVICE DE L'HUMANITÉ

Anticiper l'évolution de la délinquance par les mathématiques *Forecasting criminal patterns for decision making*

Nicolas Valescant (*), Daniel Camara (*), Patrick Perrot (*)

** Service Central de Renseignement Criminel, Gendarmerie Nationale
Criminal Intelligence Service, Gendarmerie Nationale*

Forecasting, pattern analysis, model selection, criminality; Prédiction, profil de criminalité, sélection de modèle

Résumé/Abstract

Reacting to criminality is not enough in our increasingly digitalized societies. In a context of budget limitations, law enforcement agencies have the obligation to optimize the use their forces and in order to achieve this optimum, a new pro-active approach must be used. Criminality evolves fast, and offenders now easily have technical means to extend both their sphere of operation and the speed. Developing a proactive approach means anticipating criminality and acting according to the new information prediction gives us. Given these facts, it appears vital to offer means of action based on data science and artificial intelligence to decision makers, directly in contact to the field. The purpose of this article is to introduce such a tool, based on the time-series analysis of criminality and developed for professionals responsible for public safety.

Introduction

Fighting crime means acting on two fronts: repression and prevention. If crime is represented on a time scale, repression is a downstream action. Law-enforcement agencies make investigations related to a criminal activity, the goal being to arrest the perpetrator. Prevention takes criminality from the other end of this time scale and the objective is different: analyzing past data of criminality in order to understand its deterministic component well enough to extract patterns and relevant information used to put countermeasures into action. Origins of crime lie in numerous factors, sometimes there is an inner logic of seasonality or repeatability that can be understood and forecasted. There is however a part of randomness in criminality, which is driven by endogenous and exogenous characteristics within what we will call a pseudo-random process.

An analysis method based on risk evaluation appears particularly adapted to prevention. Crime has to be understood not only alone but also more broadly inside the context of society, it cannot be dealt with if not considered as a social mechanism, influenced or influencing socioeconomic factors. The point is therefore not to give specific instructions to patrols like PredPol (Predictive Policing) or other similar software solutions do, but to provide both an operational help and an accurate analysis of criminality to decision makers. We will not give a specific level for the use of our application, but based on the nature of the study, we believe that it would be the most effective used by professionals with levers of action and a true decision power.

The goal is to bring retrieve the relevant information lying in the relation between past and future facts by computing a probability. This analysis will help implementing relevant countermeasures based on a more objective decision-making thanks to the heritage of past events. Decision is made on a triple vision: prediction, prospection and situation. Prospection is a scenario-oriented analysis, it tries to think ahead low-probability

events and how to deal with them given the means of action available. An expert or a group of expert work on the scenarios and submit their practical solutions, based on a consensus within this group. Prediction on the other hand is not adapted to low probability events, because its strength is to catch patterns within recurrent events. Both predictive and situational analyses can be modeled and the quality of generated models can be assessed through mathematical methods.

In the first section, we'll discuss the rise and contribution of mathematics in the policing field, then predictive and regressive methods used for the situational and social prevention will be detailed in a second section and in the last section we'll offer a specific metric to evaluate our results.

1. Mathematics in the service of security

The contribution of mathematics in terms of policing is no recent event; pioneers such as Edmond Locard and Alphonse Bertillon applied scientific methods to forensic science during the late 19th century. However over the last few years the development of data science opened a new branch in the application of mathematical methods. The United States have been at the origin of new analytical approaches at the beginning of the 21st century. The core of these approaches is predictive analysis and the use of computerized cartography. [1] Early tools such as CrimeStat opened the way for some more advanced programs like PredPol, PreCobs or KeyCrime in Italy. These software solutions aim at providing specific instructions to patrols on the field by indicating high-risk streets and neighborhoods. Most compute these risks using Jerry Ratcliffe's spatial repeatability applied to criminality: "If your house has been broken into, there is an increased risk that your house will be broken into again within a very short time" [3]

For this application, we chose to use monthly data at the department level. The result of this choice is a macroscopic view, useful for simple and quick decision-making and whose results can be easily evaluated. When it comes to law-enforcement, prediction cannot be compared to meteorological prediction, the reason lies in the fact that once a prediction is made, and specific areas are targeted by law-enforcement agencies, perpetrators are disrupted [5] This means that the only possible evaluation can be made on the actions and countermeasures taken, not on the result of the prediction, since a validation period was used during the model selection process. We use numerous methods such as exponential smoothing; several autoregressive integrated moving average (ARIMA) models and neural networks. These methods are then used in a process inspired by boosting. Each one being a weak learner that is applied to the specific time series, the error algorithm then gives a weight to each weak learner and the best one is chosen to be the prediction.

2. Methodology : Theoretical principles and experimentation

2.1 Scope of the study

The study focuses on a wide range of infractions (107), different by nature, these are standardized by the Etat 4001 indexes, a database made of all facts handled by French law-enforcement agencies (Gendarmerie and Police Nationale, Préfecture de Police) [6] A fact is listed within the database if a judicial procedure has been transmitted to the Prosecutor (generally because of a complaint or a police investigation). In this work these infractions are clustered on 46 more general clusters called aggregates. The method is applied to France and its departments, the department in France is an administrative division of the territory and the characteristics of each department can greatly vary. The sizes goes from 86.00 Km² to 105 Km² and the population distribution varies greatly, from 2.5 Millions to 77 Thousand of habitants. France is made of 101 departments.

2.2 Theoretical principles

This work uses the statistical analysis of past crimes to determine tendencies and try to forecast future crime patterns. Criminal acts are not deterministic facts; it is hardly possible to determine that a

specific criminal act will happen at a specific place and date. However, they are not completely random facts either. A kind of logic exists, linked to opportunity, social and economic situations. We may not be able to predict the behavior of one specific person, but statistically speaking, we can determine the tendencies of a social behavior. Moreover criminal activity, as any other human activity, is influenced by external facts such as climatic conditions, seasonal activities and holidays, among others. Many of these influences are seasonal, and linked to the population's movements. For example, during the summer, places near the sea are more prone to have pickpockets, in the wintertime we find that there is a rise of this criminal activity in regions near ski resorts. When we order this activity by time we can use time series analysis to forecast the behavior of criminals and put counter measures to constraint its spread.

The main objective of this work is to have the most accurate criminality forecast one year ahead. This is important for the management point of view to balance forces and schedule reinforcements. However, in terms of criminality the concept of error in the prediction is different from other fields. We cannot just use a standard error measure such as Root Mean Square Error (RMSE) or Mean Absolute Error (MAE). These focus on the prediction error of each term and prime the ones that are, on average, closest to real data. In terms of lawbreaking analysis, the trend of the curve is as important, if not even more, as the simple distance between forecasted data and reality. For public safety forces it is important to know the tendency of a given criminal activity, i.e. if it is an increasing or a decreasing phenomenon. The perception of the error is also different, for example, having a forecasting that is 1% wrong, but that inverts all the points of the curve, in general, is perceived as worse than the one that got a relatively higher average error but got all the tendencies right.

Criminal activities vary greatly among the different types, for example, pickpocketing activities and bomb attacks have not the same distribution, neither in terms of time nor in terms of most exposed places. The geographical distribution also have a great influence on the crime, the same kind of infraction, in two different regions, may behave completely differently.

Hardly one single time forecasting method would be capable to handle the large variability in the parameters this project intends to deal with. Considering this, the method developed here was to try to find the method that fits the better the characteristics of the tuple crime aggregate/French department.

2.3 The developed method

Taking into account the complexity of the work to be done, instead of trying to fit all the possible forecasts, we decide to use different methods. For each tuple aggregate and department we collect the pertinent historical data of previous years ($Y-N$ to $Y-1$) to predict the behavior of the last year we have data (Y), where N is the number of years we have in our historical database. Then we use different time series methods to forecast the behavior of the year Y . The real data of Y is then used to verify which method represented Y the better. After deciding which method is probably the most adapted to the characteristics of the tuple aggregate and department, we use the data of $Y-N$ to Y to predict $Y+1$. We cannot be sure at 100% that the chosen method is the one that will express the behavior of $Y+1$ the better, but is the best guess we have right now.

Apart from the choice of which methods to use, the most important part of the work is to choose the error function that will find the appropriate method. The error function is the one that will chose one method in detriment of other and, in the end, ensure the quality of the whole system. Different error functions are described in the literature and different works compare error functions trying to define the advantages and disadvantages of each different methods [7][8][9] However, in the best of our knowledge, none of existent methods is able to express the specificities of the forecasting for criminality data.

3. Evaluation: defining a specific metric is mandatory

In order to measure the performance level of the analysis, we must use an objective metric, adapted to our problem. A simple error level between real and predicted value is useful for a first approach, but when it comes to the needs of law enforcement, the user interest lies somewhere else. This error metrics does not take the future risk into account, this is why we chose to combine error level and the curve slope to define a new error metric. This slope shows if the risk for the following period of time (month in our data) is higher, lower or stable. This information is more important for our users than having the lowest distance between predicted and real values.

The application of this error to the method aims at anticipating the changes of infraction levels. This type of tool can, if used at a hierarchical level where decision making and means of actions are real, help having a broader and more objective view of criminality based on past data and more importantly help optimize the action of law-enforcement agencies in various fields such as the force deployment. Mathematical methods used to develop this application open relatively unexplored perspectives in French law-enforcement agencies.

Given the fact that no error method was adapted to our field, we formulate a new error function that takes into account not only the distance between the points of the curves, but also the tendencies. The method is immune to zero values, and is capable to test errors in different scales, what are not possible for some error methods.

The Mean Absolute Percentage and Tendency Error (MAPTE) is defined as:

$$MAPTE = \frac{\sum_{t=1}^T \left(\frac{|Y_t - F_t|}{\frac{1}{(T-1)} \sum_{i=2}^T |Y_i - Y_{i-1}|} \right)}{T} \times w_1 + \frac{N}{T} \times w_2 + \frac{\sum_{j=1}^T (|Y_j - F_j|)}{\sum_{j=1}^T (|Y_j - Y_{j-1}|)} \times w_3 + C \times w_4$$

The size of the forecast (F) and the real series (Y) we are testing it against is the same, T. In the first part of the series, we evaluate the average error between the two series making, it independent of the scale and putting it between 0 and 1. In the second term, the N represents the number of times the slopes of two consecutive points of the two series. The third term in the top we have the average error between real and forecasted data, divided by the envelop of the real data. C, is the number of times tree points were collinear in the forecasted curve, divided by the size of the series. This is used to minimize the impact of methods that forecast just a straight line, even if they are, on average, correct they are not useful for criminality forecasting, from the point of view of law enforcement forces. The purpose of MAPTE is to be independent of the scales of Y and F and have an error, that in the basic form, should be between 0 and 4. However, as MAPTE has four distinct and independent components, we have also weights for each one of these components, (w_1 to w_4). One can use these set a weight to each component to increase/decrease its importance in the error counting. For example, if for some reason one wants to decrease the importance of the collinearity, it can easily add a weight factor to it, i.e. multiply for a value smaller than 1, or greater than one to increase even more its weight in the equation.

4. Implementation in real life

For each department we apply 29 different time series analysis, 15 different methods and plus 14 variants of these methods which in some cases changes greatly the format of the forecast curve, the result is presented in table 1.

Method (number of different configurations)	MAPE [9]	MASE [9]	MAPTE	Mean non matched slops	Parentage of usage
ARIMA[11] (6)	26.65	0.72	1.909	3.50	19.00%
Average naive method (2)	29.89	0.75	1.842	3.54	13.66%
Exponential smoothing state space model [17] (4)	30.88	0.76	1.993	3.89	11.99%
Neural Network Time Series Forecasts [18] (3)	31.54	0.71	1.929	3.40	11.30%
Singular Spectrum Analysis [19] (1)	35.65	0.72	2.022	3.75	8.15%
Seasonal Decomposition of Time Series object [14] (2)	30.02	0.78	1.927	3.56	7.98%
Seasonal naive[14](1)	28.4	0.7	1.788	3.38	7.27%
Holt Winters [10](3)	30.1	0.93	2.17	3.69	6.79%
Long term forecasts[21] (1)	18.95	0.5	1.896	2.28	4.32%
Exponential smoothing state space [13] (2)	24.44	0.75	1.912	3.67	3.30%
Naive and Random Walk Forecasts [14](1)	17.96	0.5	1.676	2.90	2.54%
Trigonometric terms for seasonality[14] (1)	26.98	0.77	1.959	3.81	2.47%
Support Vector Machines [20] (1)	31.23	0.69	2.333	3.61	0.69%
Cubic Spline Forecast [16](1)	28.33	0.73	2.291	4.07	0.46%
Theta method forecast [15] (1)	27.01	0.59	1.949	3.75	0.09%

Table 1 – distribution of chosen methods, taking into account the MAPTE error and compared with other error methods

We can see in *Table 1* that the most used forecasting method is ARIMA, its six variants count for 19% of the forecasts considered the best possible forecasts, using MAPTE error metric for the whole dataset. The dataset consists of all aggregates and departments put together (46 aggregates * 100 departments = 46000 executions of the meta model, thus each individual forecasting method). The second method is the average naive method with 13.66%. The average naive method is an adaptation of a traditional naïve, but instead of just repeating the value of the last year and say it is the value of the next one, we get the monthly average of the last N years and use that as the forecast of the next year. Even if simple, this heuristic presents good results, and it is quite well adapted to our requirements. It considers seasonality and the specific characteristics of the focused tuple, department/aggregate). We can observe it in action in the graph of *Figure 1*, where the average naive is chosen by MAPTE as the one that best describes the real occurrences for the last year.

Figure 1 shows the chosen forecasting curves for different minimum error, in the light of different error methods. The represented aggregate is the deliberate attack to physical integrity for one given department. MAPTE favors methods that try to follow the tendencies in detriment of being, on average, closer to the real result. We can see that random walk, would be chosen, for this specific configuration, by

Root Mean Squared Error (RMSE) [8], Mean Square Error (MSE) [9], Mean Symmetric Percentage Error (sMAPE) [22] and Geometric Mean Absolute Error (GMAE) [9] It is true that Random Walk give a result that is, on average, close to the real one, however, for the law enforcement planning for the next year, this flat behavior is not quite interesting. The same can be said for Mean Error (ME) [23] and Mean Percentage Error (MPE) [23] that choose Cubic Spline method. Mean Absolute Percentage Error (MAPE) [9], Mean Absolute Error (MAE) [8] and Mean Absolute Scaled Error (MASE) [9] would have, on their turn chosen a Holt-Winters forecasting. This method, even if show the tendency of the real curve, what is better than the other two other forecasts, it does not try to forecast the variance between months.

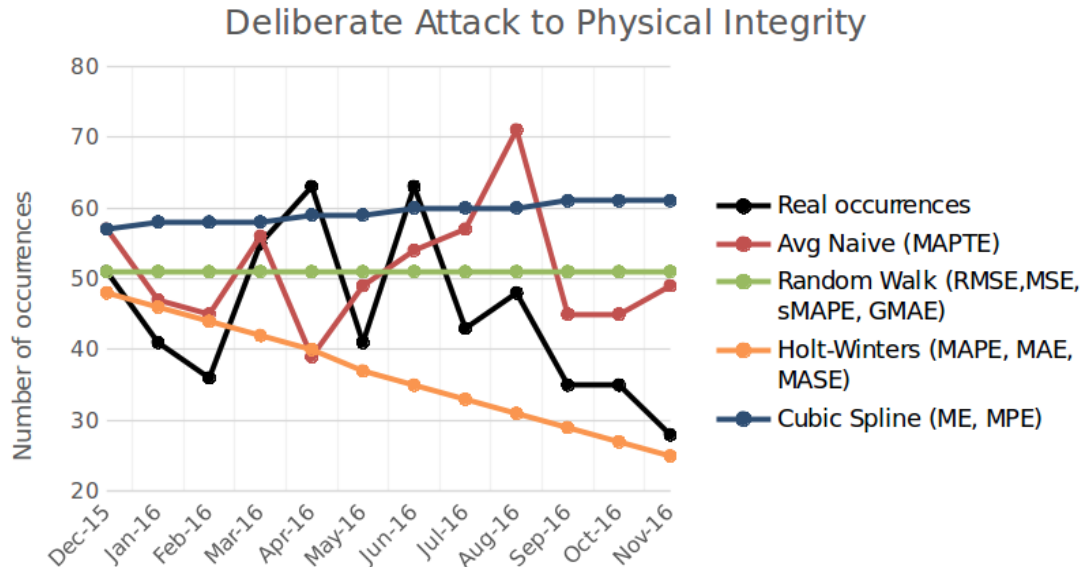


Figure 1 – Example of forecasting for deliberate attack to physical integrity set of infractions for a given department. The graph shows that different error calculations would choose different forecasting methods.

Figure 2 shows how the values of Figure 1 are presented to the user. The forecasting is a help to give more data to improve the decision making process in the law enforcement field. The graph shows the values of the last slide window year and the forecasting for that period, and the expected forecast for the next year, now taking into account the last year data and using the same method that was pointed as the closest one to the last year.

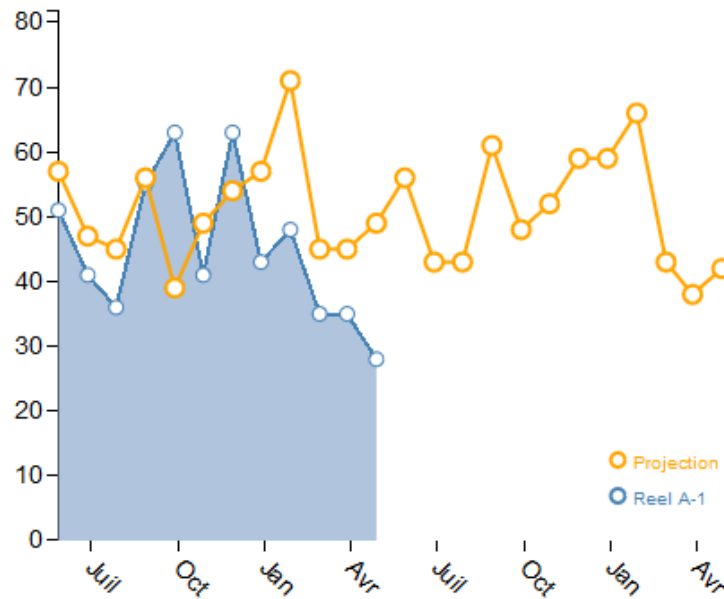


Figure 2 – Real graph of the developed application to present to the user the forecasted curve, in yellow, and the real occurrences for the last year, in blue. In this way the user may have a hint of the probable quality of the forecast for this index/department. .

5. Conclusions

This work presented a meta-method for forecasting criminal data, it combines different time-series analysis methods and uses an original error metric to determine which forecasting method adapts better to the tuple crime type/focused territory at hand. However, the method itself is agnostic to the type of data being analyzed, as well as the error function. The error function can be used in any case where the tendency is important and we need to evaluate data that different characteristics and bring it to the same scale. Even if the method itself is not restricted to this granularity level, the data handled here intends to provide a macroscopic tool for fighting against crime, it aims at anticipating the changes of infraction intensities. If used at a hierarchy level, where decision-making and means of actions are real, it provides a broader and more objective view of criminality, based on past data and more importantly it may help to optimize the action of law-enforcement agencies in various fields such as the force deployment. Mathematical methods used to develop this application open relatively unexplored perspectives in French law-enforcement agencies and are an effort of the Criminal Intelligence Service of the French Gendarmerie in the sense of understanding crime and preventing it from happen.

Bibliography

- [1] John Eck, Spencer Chainey, James Cameron, R Wilson, « Mapping crime : understanding hotspots » In National Institute of Justice, 2005
- [2] C. P. Haberman, J. H. Ratcliffe, The predictive policing challenges of near repeat armed street robberies, *Policing: A Journal of Policy and Practice*, 6(2): 151-166, 2012
- [3] J.H. Ratcliffe, Crime mapping and the training needs of law enforcement, *European Journal on Criminal Policy and Research*, 10 (1): 65-83, 2004
- [4] P. Perrot, K. T. Achi, « Forecasting analysis in a criminal intelligence context », Proc. International Crime and Intelligence Analysis Conference, 2015
- [5] P. Perrot, « L'analyse du risque criminel : l'émergence d'une nouvelle approche », *Revue de l'électricité et de l'électronique*, 2014
- [6] Interstats Méthode N° 2 : l'enregistrement des crimes et délits non routiers par la police et la gendarmerie, Interstats, novembre 2015
- [7] C. Willmott, K. Matsuura, Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance, *Clim. Res.*, 30, 79–82, 2005
- [8] T. Chai, R. R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE) - Arguments against avoiding RMSE in the literature - *Geosci. Model Dev.*, 7, 1247–1250, 2014
- [9] Rob J. Hyndman, Another Look at Forecast-Accuracy Metrics for Intermittent Demand, *FORESIGHT Issue 4*, June 2006
- [10] P.R. Winters, Forecasting sales by exponentially weighted moving averages, *Management Science* 6 324-342, 1960
- [11] J. Haslett and A. E. Raftery, Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource (with discussion); *Applied Statistics* 38, 1-50, 1989
- [12] R.J. Hyndman, A.B. Koehler, J.K. Ord, R.D. Snyder, *Forecasting with exponential smoothing: the state space approach*, Springer-Verlag: New York, 2008
- [13] A. M. De Livera, R. J. Hyndman, R. D. Snyder, Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association*, 106 (496), 1513-1527, 2011
- [14] Rob Hyndman, *Forecasting Functions for Time Series and Linear Models*, available online at <http://github.com/robjhyndman/forecast>, version 7.3, October 13, 2016
- [15] V. Assimakopoulos and K. Nikolopoulos, The theta model: a decomposition approach to forecasting, *International Journal of Forecasting*, 16, 521-530, 2000
- [16] Rob J Hyndman, Maxwell L. King, Ivet Pitrun, Baki Billah, Local linear forecasts using cubic smoothing splines, *Australian and New Zealand Journal of Statistics*, 47 (1), 87-99, 2005
- [17] R.J. Hyndman, A.B. Koehler, R.D. Snyder, and S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, *International J. Forecasting*, 18 (3), 439–454, 2002

- [18] Rob Hyndman, Forecasting Functions for Time Series and Linear Models, available online at <http://github.com/robjhyndman/forecast>, version 7.3, October 13, 2016
- [19] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, 35, 395–424, 1989
- [20] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 3, Article 27, May 2011
- [21] P. Cortez. Sensitivity Analysis for Time Lag Selection to Forecast Seasonal Time Series using Neural Networks and Support Vector Machines, In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN 2010)*, pp. 3694-3701, Barcelona, Spain, July, 2010
- [22] S. Makridakis, M. Hibon, The M3-competition: Results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476, 2000
- [23] Aman U. Khan, W. Bartley Hildreth, *Case studies in public budgeting and financial management*. New York, N.Y: Marcel Dekker. ISBN 0-8247-0888-1, 2003