Empirical Calibration of an Agent-Based Model of Car Theft

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Social and Economic Data Analysis
at the Department of Economics of the University of Konstanz

By: Angelica Maria Becerra Garnica
Rietholzstrasse 30
8125 Zollikerberg
Switzerland

Period of completion: 26th Feb 2019 – 23rd Jul 2019

1st Referee: Assistant Prof. Dr. Karsten Donnay
2nd Referee: Prof Dr. Susumu Shikano

Konstanz, 23 Jul 2019

Contents

Introduction

In the first semester of 2017, the police captured twelve criminal organizations concentrated on car thefts in Bogota de Bogotá (2017). Despite police efforts, car theft has been increasing between January and April of 2018 (Police, 2018), but only few works have tried to identify relevant factors that might help police to reduce this kind of crime.

Ospina Baena, Peña, and Mauricio (2015) presented a qualitative work evaluating different actors involved in car theft, and proposed some strategies based on information provided by experts, laws, and reports from diverse sources. Norza Céspedes, Velásquez, Andrés, Castillo Romero, and Torres Guzmán (2013) identified some variables that influence this kind of crime using the theory of rational choice and opportunity, and confront them with data between 2008 and 2012 and interviews to convicted people. They inferred some patterns and strategies to fight the occurrence of the event, and also they point out how the geospatial analysis is necessary to find archetypes with a specific behaviour. Their work highlights the importance of modelling the process of crime identifying different sources and patterns that are involved in crime, and why it is essential to understand the profiles of those events to implement actions that can reduce the number of incidents.

Under this scenario of understanding pattern crime, Groff (2014) describes the use of agent-based models (ABM) under many contexts of crime. One of them is the utilization of ABM to test theories and mechanisms, and they gave the specific example with the approach of rational choice mentioning: *"Agents can be created that use the logic of rational choice perspective to evaluate a situation. The modeller can examine what combination of situational* *characteristics translates into a decision to commit a crime.".* Given the advantages of ABM to study this kind of phenomena, the aim of this work is to implement an ABM that reproduces car thefts in Bogotá having into account some geospatial variables at the level of neighborhoods, and also to find which are the relevant parameters that describe better the spatial distribution of the crime.

Felson and Clarke (1998) specify that some theories concentrate on the description of people and how likely they commit crime, but theories do not put effort on the environment that might have bigger influence on the probabilities to commit crime. Under this scenario theories like routine activity have been take at bigger role on the description of crime.

Chapter 1

Related work

1.1 Crime theories

Clarke and Felson (2017) discusses the similarities and the differences of routine activity and rational choice theories. Both theories complement each other, so they are sometimes implemented together to evaluate factors that might influence crime. For instance, Norza Céspedes et al. (2013) made a qualitative evaluation using these theories and they identified some relevant factors that have an impact on car thefts in Colombia. First, they mentioned a variable related with the motivation: the profit. Second, they describe three factors that are related with mechanisms to commit crime: security of the car, informal surveillance from people on the surroundings, and security measures from the environment (police, cameras, secure parking). Third, in the analyses they mentioned that thieves usually studied two environmental variables before committing crimes: common crowded places like commercial spots, and places with low police presence and where people leave unattended vehicles. Finally, some variables describe the process after theft occurs like: penalties, how to make profit, or the organizational level of the criminals.

There exist multiple crime theories as framework to analyze why individuals commit crime. Groff, Johnson, and Thornton (2018) mentioned nineteen different theories that were used to implement an agent based model depending on the type of urban crime. The most relevant theories are: 1) routine activity (Cohen & Felson, 2016), 2) crime pattern (Brantingham & Brantingham, 1984), 3) rational choice perspective (Clarke & Cornish, 1985), and 4) social disorganization (Clarke & Cornish, 1985).

Routine activity theory is based on human ecology and rational choice theory, and do not focus on the characteristics of the offender, instead it focuses on the crime rate trends and cycles (Cohen & Felson, 2016) having into account three main factors: "likely offenders, suitable targets and the absence of capable guardians". Also under micro level perspective, this theory states that the routine of daily activities may influence or increase the crimes.

Crime pattern theory (Brantingham & Brantingham, 1984) uses elements from the routine activity theory because clearly, they might create suitable environments to commit crimes at the same space or at the same time, creating patterns of the offender behavior. Additionally, thieves also create similar mechanism to operate creating spatio-temporal patterns.

Rational choice theory (Clarke & Cornish, 1985) is mainly used in microeconomic models due to the convergence of social and economic behaviors, and it is based on the preferences of people. Related to crime, this theory explores how criminals may prefer rationally and intentionally one target over another, and also the decision making process to perpetrate a crime.

In social disorganization theory (Bursik Jr, 1988), the focus is on the offender. The theory studies how characteristics of people and their surroundings might influence their behavior to be involved in criminal activities, like location, age, or race.

Repeat victimization theory works over the hypothesis that when there is a crime in one place, it will be highly likely the repetition of the crime in the same spot, and usually under this theory the offender is the same in all the events(Farrell & Pease, 2014).

1.2 Agent Based Models

Some of the most interesting topics related to the subject are compiled in the *Encyclopedia of Criminology and Criminal Justice* Bruinsma and Weisburd (2014). For instance, Malleson and Evans (2014) show an overview of ABM applied to prediction of crime, and collect the definitions of agents under this context. Birks and Elffers (2014) give five frameworks at micro-level: 1) *"Routine activity approach"*, 2) *"The rational choice perspective"*, 3) *"Social learning theory"*, 4) *"Network theory of peer association"* and 5) *"Social disorganization theory"*. Groff (2014) describe that ABM has been widely applied to crime in general and different types of crime like residential burglary and street robbery.

Groff (2007) present the use of agent-based models applied to street robbery focusing on routine activity theory. Its aim is to present a simulation of the micro-level data that is not commonly available, and link every operationalized variable with the crimes theories. Based on that they present an ABM with 12 parameters grouped by three levels: society, agent, and situational using the street network Seattle. Hayslett-McCall et al. (2008) used cellular automata and multi-agent systems to residential burglary based on social disorganization and routine activity theories using real crime data to initialize and parameterized the model. Malleson, See, Evans, and Heppenstall (2012) also used an agent-based model to residential burglary, however, they used Physical conditions, Emotional state, Cognitive Capabilities and Social Status (PECS) as a reference model for human behaviour.

One of the most recent and complete articles of ABM in crime is *"State of the Art in Agent-Based Modeling of Urban Crime: An Overview"* from Groff et al. (2018) where they study 45 papers and summarize the main authors, a development of interest on the topic, thy type of crimes, the theories and software used, the implementation details of the simulation process and the methods to evaluate the models.

A relevant characteristic of the articles analyzed by Groff et al. (2018) is that 40% of them implemented models to evaluate policies related to the crime, and 60% used ABM to assess crime theories. Also, approximately one-third of the articles focused on burglary, and only one of the 45 empathized on *"auto theft, robbery, shoplifting"*.

Like Groff et al. (2018) show, there is not much literature available that focus in the application of ABM and specifically in car theft. Berger and Borenstein (2013) applied an ABM based on the rational choice theory, and its purpose was to evaluate policies and criminal laws defining that someone commit a crime if the gain is bigger that a function in terms of the probability of detection, a fine, and the imprisonment term.

1.3 Genetic Algorithms

Given an optimization problem, i.e., a problem where we want to find the best solution from the set of all feasible solutions. The best solution given a objective function is the instance over the domain of the function that maximizes(minimizes) the function. Often, the solution of an optimization problem involves a random guided search over the domain of possible solutions because analytically the solution is hard or even impossible to compute. Consequently, we can apply a metaheuristics optimization techniques to find a solution. These techniques are inspired and try to imitate principles of biology or physics, to guide the search of the solution (Kruse, Borgelt, Braune, Mostaghim, & Steinbrecher, 2016). The genetic algorithms belongs to this family an imitates the evolution of the genetic material in a population following principles of differential reproduction and better adaptation of individuals with higher fitness with respect to the environment. The basic principle is that individuals with high fitness (high values in the objective function when is a maximization problem) have better change to survive and get offspring in the next generation.

Algorithm 1: Evolutionary algorithm scheme (adapted from (Kruse et al., 2016, Chapter 11))

```
input  : sol_per_pop, size_mutate, size_crossover, num_generations
output: pop[num_generations] after num_generations
/* Create the initial population                                     */
initialize pop[0] ;
/* Evaluate initial population (compute fitness)                     */
fitness[0] ⟵ cal_pop_fitness(pop[0]) ;
for t ⟵ 1 to num_generations do
    /* Select individuals based on fitness                           */
    pop[t] ⟵ tournament(pop[t − 1]) ;
    /* Apply genetic operators and evaluate the new
       individuals                                                   */
    crossoverPop ⟵ crossover(pop[t], size_crossover) ;
    fitnessCrossover ⟵ cal_pop_fitness(crossoverPop) ;
    mutatePop ⟵ mutate(pop[t], size_mutate) ;
    fitnessMutate ⟵ cal_pop_fitness(mutatePop) ;
    /* Update population with new individuals from genetic
       operators                                                     */
    pop[t] ⟵ pop[t] ∪ crossoverPop ∪ mutatePop ;
    fitness[t] ⟵ fitness[t] ∪ fitnessCrossover ∪ fitnessMutate ;
end
```

Algorithm 1 presents the general scheme of a genetic algorithm where we can highlight the building blocks that are necessary to develop a genetic algorithm. An extensive description with a helpful discussion of the common pitfalls, a rich description of each component and helpful guidelines of the common definitions can be found in Kruse et al. (2016).

- Encoding: The initial step in the design of the genetic algorithm consist in define how the solution candidates are gonna be encoding. The encoding balance a trade off between simplicity and some desirable properties in terms of the algorithm. In general, it is desirable that similar solutions have a similar fitness, this ensures that the objective function represents a smooth surface where guided random search is useful.

- Creation of initial population: In this block we want to create the initial population where the genetic algorithm starts the search. This step needs to ensure that is probable to generate candidates solutions from all the domain of possible solutions, and it potentially avoids to consider solutions that does not fulfill the constrains of the problem.

- Fitness function: The fitness values of the candidate solutions guide the selection of the individuals. In most of the cases the fitness function corresponds with the objective function. When the optimization problem have constraints is possible to represent those constraints in the fitness function to penalize candidate solution that does not meet those.

- Selection method: ensure the selection of promising candidate solution to be keep in following generations. This selection method requires to keep a balance between exploitation and exploration. The former refers to the effect of duplicate in a short time the most prominent candidates with the risk to get a local maximum. The later ensures that not promising candidates have a probability bigger than zero to be selected, this with the hope that the genetic operators helps to find unexplored promising regions of candidate solutions.

- Genetic operators: Once the population for the next generation is selected. New solution candidates are created from the selected pool through modifications of their genetic material.

  - Crossover: The usual case for this operator consider the selection of two candidates and the creation of one or two new candidates by mixing the chromosomes (term to refer to the actual encoding of a solution).

  - Mutate: It is possible that some specific value in one of the components of the solution candidate are missing in the initial population. The mutation operator takes randomly one or multiples alle (term to refer the value of an specific component of a candidate solution) and change the values with some random values in the domain range of the alle. This mutation creates could enrich the population with some new areas in the domain search that where visited before.

Schutte (2010) presents an calibration of an ABM applied to aviation, and describe how important is to compare real data with the ABM models. Given that the dynamics of aviation in the model implemented is very known, Schutte (2010) concentrate the analysis over each component in the model,and then evaluates every dynamic critically. Additionally, a genetic algorithm is used to estimate the parameters of the ABM. The genetic optimization is evaluated a different levels and confronted with the real phenomenon.

Chapter 2

Data description

2.1 Data Sources

Car data was chosen among different types of theft, because it has better quality on the reports. The first reason for the high quality is that people report the car theft to the authorities to avoid being incriminated for crimes that the offender might effectuate with the car. Another reason is that the victim should report to the police in order to have the compensation from the insurance companies.

2.1.1 Theft reports

The focus of the analysis is on the city of Bogota restricted to the urban area given its particular characteristics: Bogota is the capital, it is the densest city, the crime rates are the highest, and the spatial information of neighborhoods and their delimitation are more precise.

The information of crimes is public available, and is taken from the National Police of Colombia Police (2018) for the whole country with the following variables between January 2010 and April 2018:

- Date and time of the crime
- Method of theft
- Administrative areas (There are 19 in Bogota)

- Neighborhood
- Place characteristic (street, malls, private parking places, condominium, etc.)
- Owners age
- Owners sex
- Owner marital status
- Owner level of education
- Owner profession
- Type of car
- Car brand
- Car model
- Car registration year
- Car colour

A good characteristic of this database is that information is updated continually, and the data previously registered do not have notable changes. Additionally, the number of cases are highly accurate to represent the phenomenon because the majority of people report thefts of vehicles to police given the high value of the good, to avoid being involved in criminal activities that thieves may do with the car, and also insurance companies require it (Norza Céspedes et al., 2013).

However, one main problem of the database is that main variables related to the description of the crime do not have standardized values, so it was essential to define unique values that help the analysis of the data. For all the variables the unification was made manually, except for neighborhood which was a semi-automatized process.

The majority of neighborhood names registered do not match exactly with the official database of neighborhoods, and only 24.5% of the reports can be linked directly. To solve this, it was useful to calculate string distances using the package van der Loo (2014) and the approach presented in Doctor (2015). This procedure gives different candidates based on the minimum measure from eight string distances. As a result 782 neighborhoods (45.7%) were easily linked with an official name.

After the implementation of this method, 510 neighborhood remained without an official name, but those were checked manually using Google maps and its coordinates to know the official name. Consequently, 432 were founded, 65 were not founded, and 13 were in the rural area. Finally, those who did not have information of the neighborhood and were in those rural areas, were eliminated and represented 145 theft reports, i.e. only 0.7% of the total of theft reports.

2.1.2 Geographic information

The geospatial information of the city of Bogotá is public on the webpage of the Special Administrative Unit for the District's Cadaster (Cadaster-Bogota, 2018). The four shape files used were:

- Buildings
- Lot area registers with associated tables of uses, and socioeconomic stratum per lot.
- Blocks
- Neighborhoods

The information of socioeconomic stratum was processed per neighborhood having into account the number of lots per stratum. And with the table of uses was possible to determine the area designated to private parking spaces.

2.2 Descriptive statistics

In 2017 in Colombia, 7.700 cars were stolen, and 34.6% of the cases were in the capital. Consequently, on average seven vehicles have been reported stolen daily last year in Bogota Police (2018). Based on the information for the whole period between January 2010 and April 2018, the report of thefts has the following common characteristics:

- 83.9% were on the street and 74.4% were non-violent. Consequently, 62.1% were on the street and non-violent.

- The preferred class to commit a crime were cars with 61.1%. SUV, Vans and Pick up represent 29.9% of the crimes, and other classes of vehicles represents 9%.

- The sex of the owner was mostly male (83.6%).

- The three most desired colors for thieves were gray (25.5%), white (24.3%) and red(17.7%). However, this is an expected result because the preferred colors in Colombia are gray, black, white and red.

- The three most desired colors for thieves were gray (25.5%), white (24.3%) and red(17.7%). However, this is an expected result because the preferred colors of cars in Colombia are gray, black, white and red (Finanzas-personales, 2018).

- The preferred brand selected by thieves was Chevrolet (Figure 2.1), but again is the favorite brand for cars in Colombia.

- In figure 2.2 is possible to observe that criminals prefer cars from recent years, like Norza Céspedes et al. (2013) described.

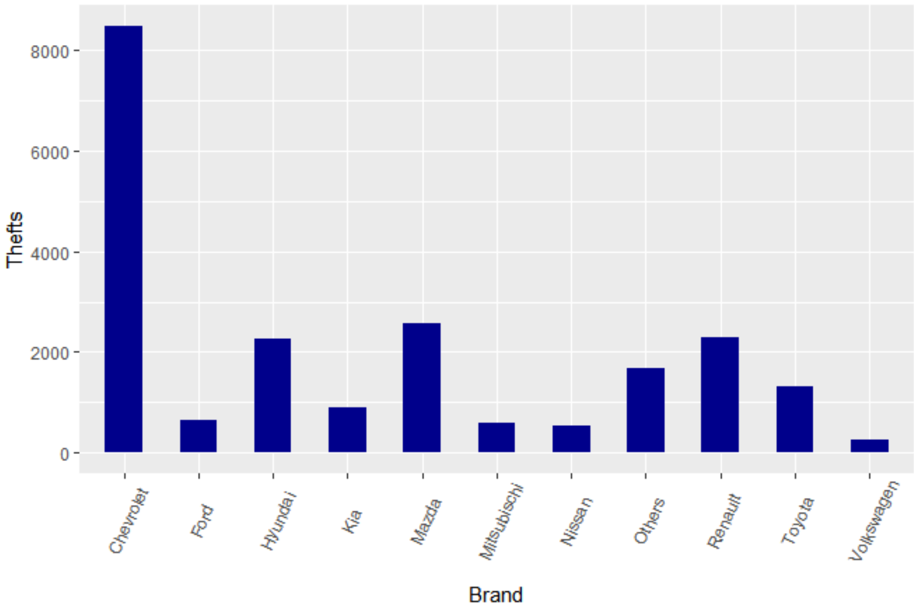


Figure 2.1: Thefts by brand

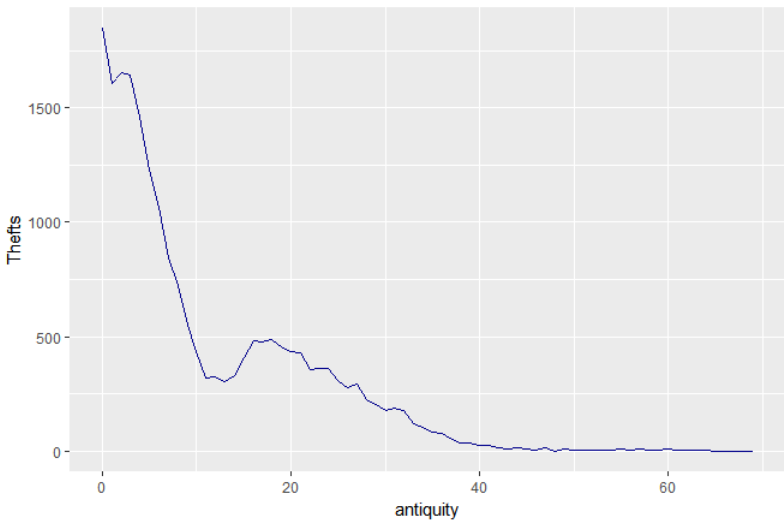


Figure 2.2: Thefts by antiquity

- Figure 2.3 shows that the phenomenon has more prevalence in neighborhoods of stratum medium.

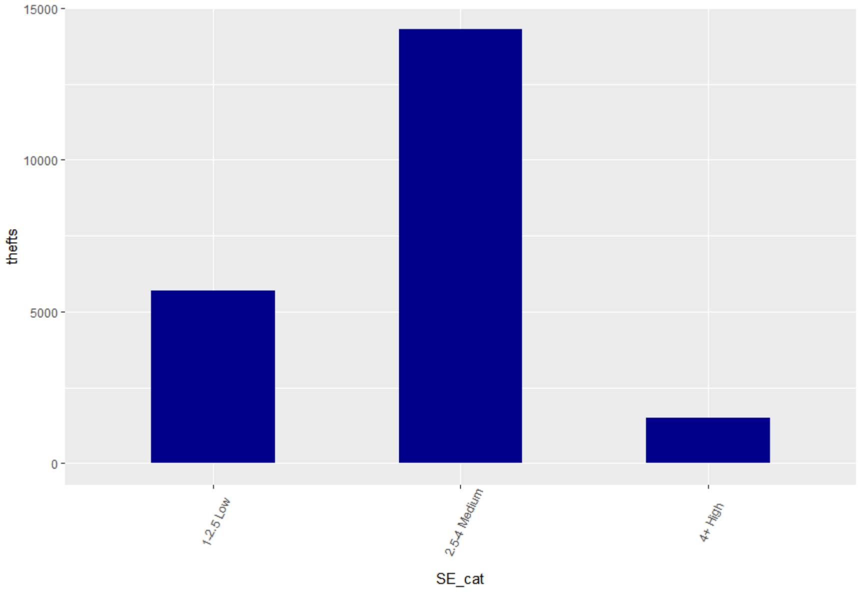


Figure 2.3: Thefts by socioeconomic stratum

2.2.1 Over time

Over time the main findings are:

- When Gustavo Petro was mayor between 2012 and 2015, car thefts were reduced (Figure 2.4).

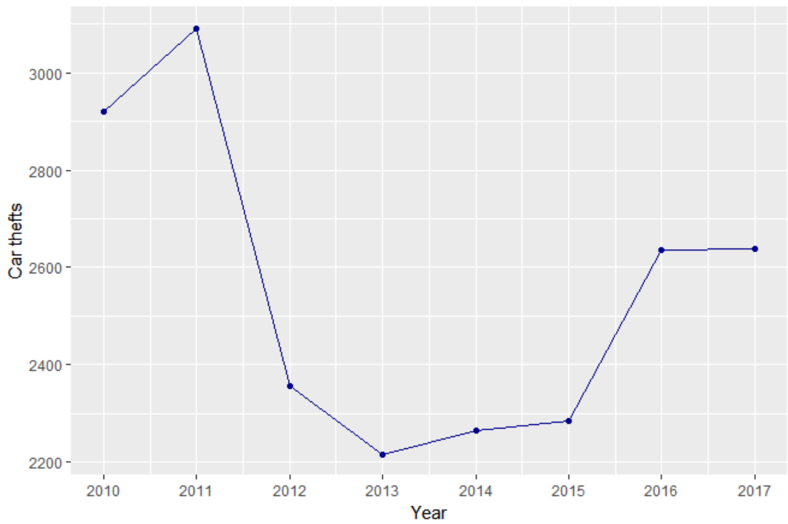


Figure 2.4: Thefts per year

- Figure 2.5 shows that this type of crime does not have a particular behaviour by month. However, it is possible to observe that after one or two months when the crime was high, there is a reduction and then increases again.

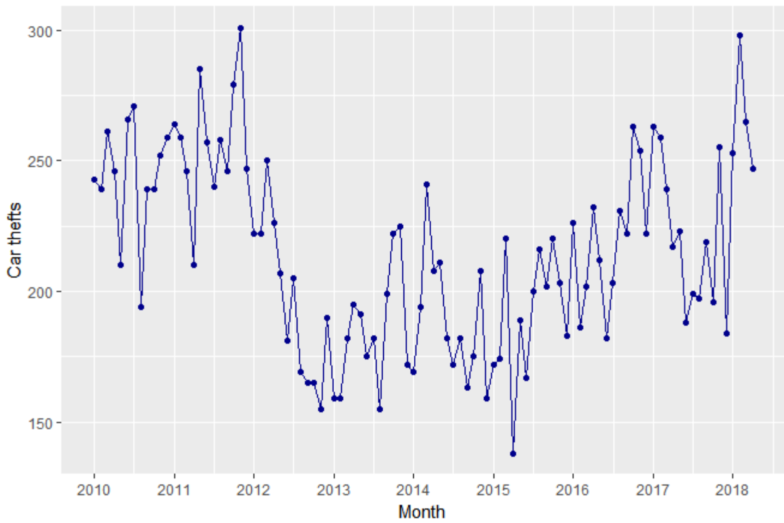


Figure 2.5: Thefts per month

- Daily behaviour does not have a particular characteristics. However, it is possible to observe one particular outlier; this was a massive theft in one car dealership (Figure 2.6).

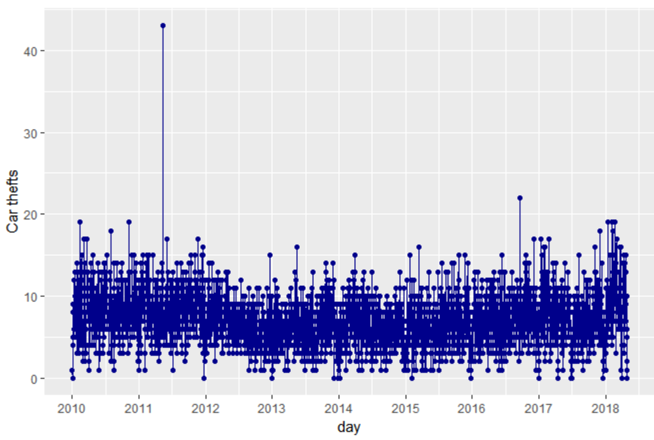


Figure 2.6: Thefts per day

- Figures 2.7 and 2.8 show that per year and month crimes were mostly committed on Saturdays and Sundays.

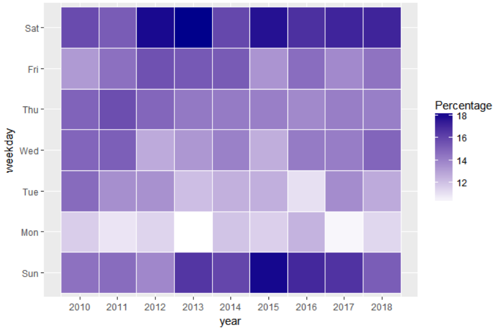


Figure 2.7: Weekday per year

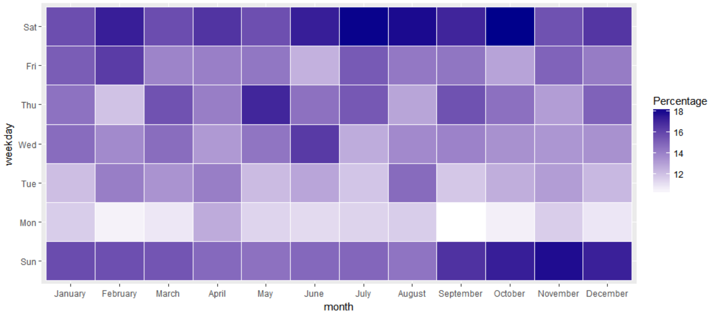


Figure 2.8: Weekday per month

2.2.2 Target population

The implementation of the model focused on specific reports for unattended vehicles:

- Car thefts considering three type of brand cars: 1) Chevrolet, 2) Renault and 3) Mazda. Those are the most common brands in Bogotá.
- Car thefts considering two classes:1) cars and 2) SUV,Vans, Pick up.
- Car thefts that were non-violent.
- Crimes committed on the street.
- Cars that have antiquity of 15 years or less.

Figure 2.9 show the distribution by neighborhood of the target thefts.

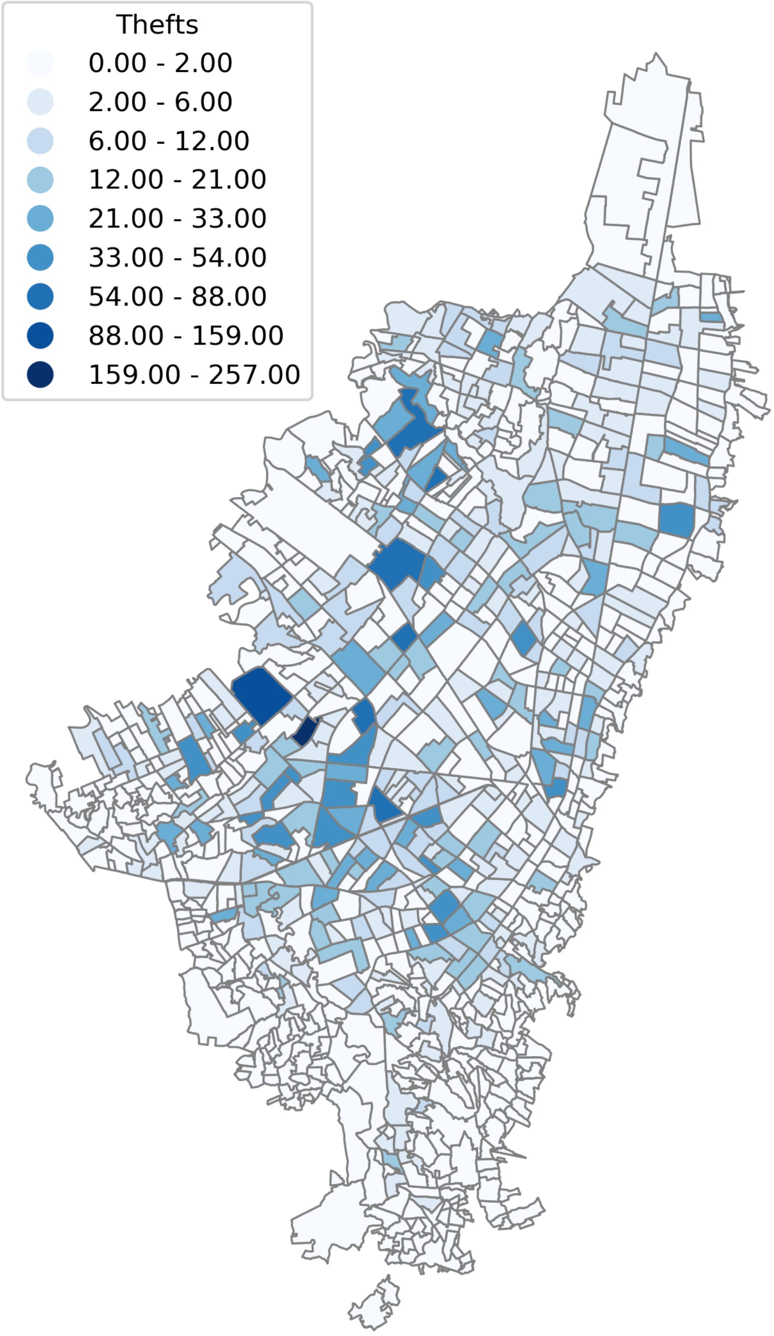


Figure 2.9: robos python

2.3 Multivariate description

Before the implementation of the ABM model, it is important to review the relationship of the variables proposed to evaluate the dynamics of car theft -at the neighborhood level-. One of the most frequent techniques to evaluate association of the features is through a simple linear regression. However, crime counts usually does not have a normal distribution, and they usually are discrete and positive variables with many zeros (Lindsey, 2000). So, dealing with this kind of data usually breaks the assumptions over the distribution of the errors in the model (Coxe, West, & Aiken, 2009).

Counts of crime in Bogota are not the exception. We can see on Figure 2.10 that car thefts at the level of neighborhood does not have a normal distribution. The histogram shows a considerable amount of neighborhoods having zero or only one theft during the observed time period. Specifically for the target population, 40.14% of neighborhoods did not report car thefts and only two neighborhoods accounted more than one hundred cases. Therefore based on a basic description of the data, a simple linear regression would not be a suitable model.

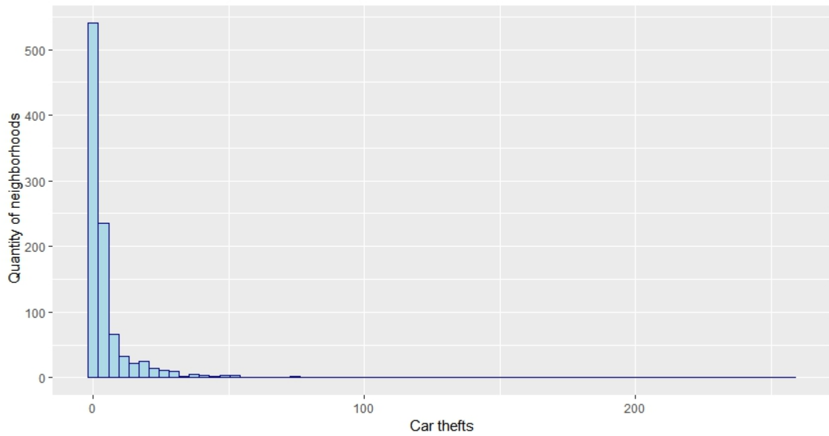


Figure 2.10: Histogram car thefts per neighborhood - Target population

A more elaborated way to model count crimes would be with a Poisson regression (MacDonald & Lattimore, 2010). It is widely used in criminology studies given that the Poisson distribution is skewed to lower values, like count of crimes. However, this distribution uses one single parameter for the expected value and the variance of counts, and when we are analyzing crime data, this assumption is usually not fulfilled. Notably, the concentration of crime on a few spots might cause overdispersion, and under this scenario, a better option would be a variation of the Poisson model: a negative binomial model.

Table 2.1 shows the estimations using R Core Team (2017) with the information at the level of neighborhoods. The Poisson model are based on the function `glm`, and negative binomial models with the function `glm.nb` from the package `MASS` (Venables & Ripley, 2002). The dependent variable was the count of car thefts using the target population and the explanatory variables used were: if the neighborhood has a police station, the parking and commercial area designated into the neighborhood, socio-economic stratum of the neighborhood weighted by quantity of dwellings per level, and a proxy of population given by the quantity of blocks or the quantity of dwellings. All variables, except police station and socio-economic stratum, were rescalated in the range $0 - 1$ to facilitate the comparisons.

As MacDonald and Lattimore (2010) mentioned, an additional challenge for count models is when zero rates are prevalent (underdispersion). On this case zero inflated models are helpful to deal with it. Table 2.2 shows that poisson and negative binomial zero inflated models have a better fit comparing them with its counterparts; this is based on the AIC criteria and the Log-Likelihood. Despite the zero inflated negative binomial has estimation problems on the zero model, we can observe that in general the influence of the explanatory variables do not change across all different models. All zero-inflated models were estimated with the function `zeroinfl` from the package `pscl` (Zeileis, Kleiber, and Jackman (2008)).

In both tables (2.1 and 2.2), the presence of police stations increases the expected rate of observed car thefts presumably because the location of them are in the most risky neighborhoods. As expected, parking area is a protective factor; more parking spaces reduces the risk of car theft for unattended vehicles. Concerning the socio-economic status, the results show that higher

Table 2.1: Poisson and Negative Binomial models

	Poisson		Negative Binomial	
	Blocks	Dwellings	Blocks	Dwellings
(Intercept)	-0.30^{***}	0.05	-1.38^{***}	-1.09^{***}
	(0.06)	(0.04)	(0.18)	(0.15)
Police station	1.07^{***}	1.00^{***}	1.13^{***}	0.90^{***}
	(0.03)	(0.03)	(0.15)	(0.14)
Parking area	-0.79^{***}	-4.14^{***}	-1.64^{**}	-6.24^{***}
	(0.14)	(0.16)	(0.61)	(0.85)
Commercial area	-0.41^{**}	0.65^{***}	-0.72	1.83^{**}
	(0.15)	(0.14)	(0.64)	(0.63)
SE Stratum	0.40^{***}	0.39^{***}	0.80^{***}	0.73^{***}
	(0.01)	(0.01)	(0.06)	(0.06)
Total of blocks	1.28^{***}		1.30^{***}	
	(0.06)		(0.21)	
Total of dwellings		3.71^{***}		5.74^{***}
		(0.08)		(0.71)
AIC	12315.95	11430.95	4560.08	4519.95
BIC	12345.30	11460.30	4594.32	4554.19
Log Likelihood	-6151.98	-5709.48	-2273.04	-2252.97
Deviance	10354.55	9469.55	978.74	982.67
Num. obs.	984	984	984	984

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

values of this variable increases the expected rates; probably, because they might have bigger acquisition capacity which leads to more cars going between those neighborhoods. About the influence of the density variables of blocks and dwellings, we can see that both increases the expected rate of the dependent variable but the impact of dwellings is bigger than the impact of blocks. The only variable that has a contradictory behavior is commercial area; nevertheless, the difference is given the association of the density variables of blocks or dwellings, and not with the zero-inflated models.

Although the results are pretty consistent, and we had an overview of the factors that might influence the car theft phenomenon in Bogota, count models have the disadvantage of assuming independence of the events (Lindsey, 2000). In this case, it is highly probable that the information at the neighborhood level is not independent, so it will be relevant to compare these results with the ABM model that have into account the spatial distribution.

Last but not least, it is important to clarify that the histogram and the models presented in this chapter used the data based on the target population of 5.135 car thefts. However, there are not meaningful differences when the response variable has into account all 21.468 car thefts (see appendix A).

Table 2.2: Zero inflated models

	Poisson		Negative Binomial	
	Blocks	Dwellings	Blocks	Dwellings
Count model:				
(Intercept)	0.90^{***}	0.98^{***}	-0.03	0.18
	(0.06)	(0.05)	(0.25)	(0.23)
Police station	0.92^{***}	0.83^{***}	1.10^{***}	0.95^{***}
	(0.03)	(0.03)	(0.14)	(0.14)
Parking area	-0.19	-3.05^{***}	-0.68	-4.45^{***}
	(0.14)	(0.17)	(0.55)	(0.71)
Commercial area	-1.12^{***}	-0.17	-1.05	0.80
	(0.16)	(0.15)	(0.60)	(0.68)
SE Stratum	0.22^{***}	0.25^{***}	0.39^{***}	0.36^{***}
	(0.02)	(0.02)	(0.08)	(0.08)
Total of blocks	0.83^{***}		1.13^{***}	
	(0.06)		(0.21)	
Total of dwellings		3.00^{***}		4.59^{***}
		(0.08)		(0.69)
Log(theta)			-0.62^{***}	-0.57^{***}
			(0.07)	(0.07)
Zero model:				
(Intercept)	2.09^{***}	1.70^{***}	2.21^{***}	2.43^{***}
	(0.25)	(0.20)	(0.53)	(0.48)
Police station	-0.62^{**}	-0.59^{*}	0.02	0.12
	(0.24)	(0.24)	(0.51)	(0.55)
Parking area	1.75	9.11^{***}	-301.71	-315.10
	(0.97)	(1.60)	(183.03)	(215.92)
Commercial area	-4.21^{*}	-9.04^{**}	-95.77^{**}	-85.39^{*}
	(2.00)	(2.89)	(36.08)	(36.69)
SE Stratum	-0.70^{***}	-0.56^{***}	-1.15^{***}	-0.96^{***}
	(0.09)	(0.09)	(0.28)	(0.29)
Total of blocks	-1.45^{***}		0.65	
	(0.28)		(0.74)	
Total of dwellings		-8.58^{***}		-8.85
		(1.41)		(4.87)
AIC	9859.92	9183.70	4483.77	4447.61
Log Likelihood	-4917.96	-4579.85	-2228.88	-2210.80
Num. obs.	984	984	984	984

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Chapter 3

Agent Based Model

Groff (2014) recommend: *"The guiding principle of ABM is* ***simple is better****"*, and partly Miller and Page (2009) agree with this statement. Last authors argue that if we can understand the phenomenon employing a simplistic version of the model, then it would be possible to implement more complex dynamics based on the information we got from a simple version. But in the same article they explicitly mention: *"Even if we know the fundamentals of a particular system, we may not be able to use that knowledge to reconstruct higher-level systems"* (section 3.3., page 41, Miller and Page (2009)). Nevertheless, it could be worthless to build a completely different setting every time we want to make a more complex model. For this reason in the article, they reconcile both points of views giving a meaningful role to the theories that explain the transformation from simple dynamics to the complicated ones.

Based on the characteristics of the crime, implementations of ABMs can go from basic models to very sophisticated scenarios. Following simplicity, the majority of the ABMs for thefts analyze the dynamics focusing on burglary because it might be easier to define compared with other types of robbery (Devia & Weber, 2013). As a result, Groff et al. (2018) show that there are not many studies evaluating the dynamics associated with different kinds of crime, for instance, car theft. Generally, it is hard to associate car theft with the guiding principle. Therefore, one challenge is to represent the combination of different crime theories, and the complexity of the system with some simple dynamics.

3.1 Model Dynamics

Different versions of models might be possible only changing the configuration of one of the main components for an ABM: the types and number of agents, the environment, and the set of behavior rules (Gilbert, 2008). Generally, the theories define the variables and the behavior rules that govern the model. In chapter 1, we identify the significant characteristics of the theories that have been mostly implemented in ABM for crime. The model uses four out of the six most common theories: 1) Routine activity, 2) Crime pattern theory, 3) Rational choice perspective, and 4) repeat and near-repeat victimization; specific features try to represent each approach in the implementation to decode the underlying dynamics easily. The main reason for not considering the social disorganization theory and the BDI framework is because this approaches needs another type of agent: `the offender`, and there is not enough quality information for Bogota to formulate proper dynamics from this perspective.

It is relevant to have in mind that if the model takes into account different mechanisms, then the interpretation of the results becomes difficult because one rule might have a link to two or more different theories (Birks, Townsley, and Stewart (2012)). Consequently, it is essential to differentiate each feature implemented in the model and their connection with the assumptions.

The first theory, routine activity, is based on the integration of three attributes: offender motivation, suitable targets, and absence of capable guardian (Cohen & Felson, 2016). In the model proposed, the offender motivation is associated with a gain related to commit the crime. The second attribute, the target, is represented by the `cars` located in the environment. And finally, the guardian is associated with four variables at the level of the neighborhood. Two of them are exogenous variables: the presence of police stations, and the area of secured parking. However, the model presented in 2.3 shows that police stations do not prevent crime. Thus, the model will include an endogenous variable represented by increments of security when there is one car theft. Nevertheless, this variable is initialized with the socio-economic (SE) stratum based on the assumption that neighborhoods with high SE has better security than neighborhoods with low SE.

For the second theory, it is possible to consider many criteria related to crime pattern in terms of time, places, targets, and methods. We can see on Norza Céspedes et al. (2013) and in chapter 2 some clear patterns for all qualities. At the level of time, thefts commonly occurred over the weekends and in specific times of the day. However, the variable time is very complex to model according to the reality in the simulations; Groff et al. (2018) confirm this: "*if the agents made decisions as frequently as they do in real life, the model would run too slowly or not at all*". Consequently, the model considers the entire period observed, and it will not include the evaluation of time patterns. At the level of places, the majority of thefts are on the streets. Regarding targets and methods, the most common targets chosen are new cars that belong to three brands, and car thefts are generally non-violent. This patterns will be considered in the ABM, but additionally the model will evaluate the influence of the commercial area at the neighborhood level.

The basis of the third theory, rational choice, is in economics. One natural way to represent this theory is to construct the probabilities based on the cost-benefit of the actions, and indeed Berger and Borenstein (2013) uses this approach. Consequently, the model will set these probabilities using the difference between the gain of the agent `car` and the difficulty (or risk) to commit the crime using the information from the cell (`neighborhood`).

Finally, the near-repeat victimization theory is one of the arguments which takes more advantage from the ABM model. Endogeneity is the main characteristic of this construct, and for that reason, it increases the difficulty to test it with classical models. What we checked in 2.2 was the association by year of the counts per neighborhood getting high correlations. Therefore, the assumption included in the model is that the probability of "successful" crime is higher after each theft in the same location (repeat-victimization), or on the surrounding areas (near-repeat victimization). It is possible to see this approach as a spreading phenomenon, like gossip model, but using an extension of a basic cellular automata where each cell can have many agents (Gilbert & Troitzsch, 2005). So in the model, many cars are associated with one neighborhood, and thefts inside of it will influence the cells itself and the surroundings. To sum up, the goal is to evaluate the impact of repeat victimization theory at the neighborhood level. So the model will test if thefts spread across neighborhoods and if this effect reflects the real geospatial pattern.

Cells and Agents

Batty (2005) describes cellular models as particular cases of agent-based models. Typically, depending on the type of agent different rules define the communication, the actions, and how agents move. In burglary scenarios, it is common to observe ABM implementations combining cellular automata for the target, and multi-agent strategies where the offenders move throughout the environment. However, the movement might become a demanding feature in the model. For instance, the architecture presented in Hayslett-McCall et al. (2008) requires more knowledge, and the adoption of additional assumptions to recreate the rules that govern the mobility of agents.

Starting from the cellular automata approach, the model considers as cells each `"neighborhood"`; Batty (2005) indicates that *"in cellular models, cells are agents"*. In addition to this type of agent, the model includes the agents `"car"` where c_i is the `car` c located in the `neighborhood` i. Naturally, with this type of agent, motion becomes an essential characteristic to describe the dynamics of the phenomenon(Batty, 2005). However, in order to keep the simplicity of the model, there will not be assumptions over the trajectory and the movement of `"cars"`.

The state of each `neighborhood` or cell i is a function denoting the counts of thefts t_i based on the simulation step s, and each step can be seen as attempts to commit crime. For the initial step $s = 0$ we will have that $t_i(0) = 0 \ \forall \ i$, and the last step is reached when the sum of all thefts over the total of neighborhoods (n_{nb}) is equal to the total cars (n_{cars}) in the entire city, i.e.

when $\sum_{i=1}^{n_{nb}} t_i = n_{cars}$. However, a global parameter $target$ is defined as a stop criteria, i.e. the simulation never reaches its maximum simulation step, and then the final simulation step is achieved when $\sum_{i=1}^{n_{nb}} t_i = target$.

Furthermore, $t_i(s+1)$, with $i = 1, \ldots, n_{nb}$, is a function based on the quantity of past thefts $t_i(s)$ plus a function $f(x, p_{c_i}(s+1))$ that increases the count of thefts with probability p_{c_i}. The function $t_i(s+1)$ can be described as follow:

$$t_i(s+1) = \begin{cases} t_i(s) + f(x, p_{c_i}(s+1)) & t_i(s) < n_{cars_i} \\ n_{cars_i} & t_i(s) = n_{cars_i} \end{cases} \tag{3.1.1}$$

In 3.1.1, the function $f(x, p_{c_i}(s+1))$ determines if there is a change inside the cell i in the state $s+1$, i.e. if the count of thefts is incremented inside i. First, it selects randomly a car c_i from the entire set of vehicles without having into account the cell for the selection, and then a random number $x \in [0, 1]$ is compared with the probability $p_{c_i}(s+1)$ of c_i being stolen; c located in the `neighborhood` i. As a result the function can be describe as:

$$f(x, p_{c_i}(s+1)) = \begin{cases} 0 & x \geq p_{c_i}(s+1) \\ 1 & x < p_{c_i}(s+1) \end{cases} \tag{3.1.2}$$

The decision of committing a crime is a function of the gain g_{c_i}, and a difficulty (δ_i) associated with characteristics of the environment (cell i). Berger and Borenstein (2013) use the same factors gain and risk, to execute the car theft based on a deterministic decision: $gain > risk$. However, it does not consider the stochastic behavior, which is an important characteristic of this phenomenon. For instance, there are cases where the offender might underestimate the real risk, and although $risk > gain$, the thieve commits the crime.The other problem is that the approach in Berger and Borenstein (2013) does not consider the magnitude of the cost-benefit. For instance, two cars in the same cell i but $gain_1$ ¿ $gain_2$; under this scenario, the probability of being stolen for the vehicle with $gain_1$ should be higher than the likelihood for the car with

$gain_2$. To avoid this problem, Devia and Weber (2013) propose a probability that includes variables the agent observes from the environment and a factor of noise to have offenders variety into the likelihood. However, Bhavnani, Donnay, Miodownik, Mor, and Helbing (2014) present a better approach to introduce noise and also it consider the cost-benefit difference. Consequently, I use in the model what Bhavnani et al. (2014) and Hayslett-McCall et al. (2008) proposes. They compute the probability through a combination of weighted parameters, and then it is compared with a random number to decide whether to commit a crime or not. So, the probability in the model I proposed is:

$$p_{c_i}(s+1) = \left(1 + exp\left[\frac{-(g_{c_i} - \delta_i(s))}{\lambda}\right]\right)^{-1} \tag{3.1.3}$$

with g_{c_i} gain from crime, $\delta_i(s)$ the difficulty associated, and λ regulating factor of the probability. Specifically, the magnitude of the difference between benefit (gain g), and cost (difficulty δ) influences the likelihood of car c being robbed or not.

The distribution of the real cost of the cars registered in the city by brand would be a good approximation of the gain g_{c_i}, but there is no public information about the type of cars registered for the entire city of Bogota, just aggregated information by year. Consequently, g_{c_i} are generated randomly from a truncated normal distribution $N \sim (0.5, 0.09)$ between 0.2 and 0.8. I chose these values under the assumption that outside the range $[0.2, 0.8]$, g_{c_i} is not associated with crime because low values of gain do not generate benefit, and huge values are associated with expensive cars that have technology anti-theft.

$$\delta_i(s) = \frac{[\alpha * Pol] + [\beta * (1 - Co)] + [\gamma * P] + [\eta * Sec(s)] + [\kappa * (1 - RV(s))]}{\alpha + \beta + \gamma + \eta + \kappa} \tag{3.1.4}$$

Then difficulty (equation 3.1.4) is a weighted combination of five variables:

- Pol: Increases the difficulty δ_i, and it takes the value 1 if the neighborhood i has a police station, otherwise is 0.

- Co: Decreases the difficulty δ_i, and is a linear transformation of the commercial area in the neighborhood i with values between 0 and 1.

- P: Increases the difficulty δ_i, and is a linear transformation of the secured parking area in the neighborhood i with values between 0 and 1.

- $Sec(s)$: Increases the difficulty δ_i. It is initialized with a linear transformation of socio-economic stratum in the neighborhood i with values between 0 and 0.6; traditionally it exists a classification with seven strata in Bogota. This values is updated every time there is a car theft, i.e. when $f(x, p_{c_i}(s+1)) = 1$, then $Sec(s)$ increases by a factor of r but never goes beyond one.

- $RV(s) = [\tau * pt_i(s)] + [(1-\tau) * pt_{i,k}(s)]$ with $\tau \in [0,1]$: Decreases the difficulty δ_i and it simulates the dynamic for repeat victimization. The parameter τ defines if the impact of previous thefts is given by the cell i itself or by the surroundings of i. Based on the function 3.1.1, it is possible to define the proportion of past thefts for each neighborhood i as:

  $$pt_i(s) = \frac{t_i(s)}{\max_{j=1,\ldots,n_{nb}}(t_j(s))}$$

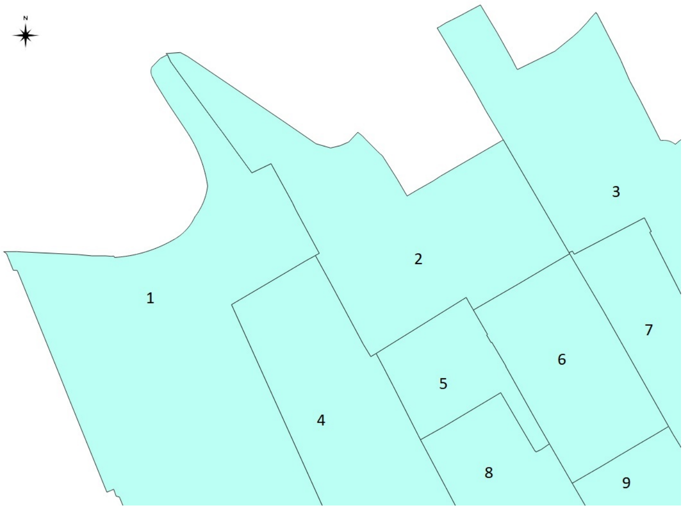
  And the proportion of past thefts for the surroundings k of the neighborhood i as:

  $$pt_{i,k}(s) = \frac{\sum_{k=1}^{n_{nb_i}} t_k(s)}{n_{nb_i} * \max_{j=1,\ldots,n_{nb}}(t_j(s))}$$

  During the initial steps, when $max(t_j(s)) < 1$ then $pt_i(s) = 0$ and $pt_{i,k}(s) = 0$. Finally, as a result both values are between 0 and 1.

  In addition, the surroundings are define by the closest neighbors of the cell i. For instance, below in the figure 3.1 we can see that the cell $i = 5$ has $n_{nb_5} = 4$ neighbors with $nb_5 = \{2, 4, 6, 8\}$, and cell $i = 2$ has $n_{nb_2} = 6$ with $nb_2 = \{1, 3, 4, 5, 6, 7\}$.

Figure 3.1: Example of cells and neighborhoods



To sum up, all variables and the theories associated are:

	Variables	**Values**	**Associated theory**
Explanatory variables	g_c: Gain obtained from the crime	$0 \leq v_p \leq 1$	Rational choice
	POL: Presence of police in the neighborhood	0 or 1	Routine activity
	Co: Commercial area in neighborhood	$0 \leq v_p \leq 1$	Crime pattern
	P: Secured parking area in neighborhood	$0 \leq v_p \leq 1$	Routine activity
Endogenous variables	Sec: Secured intialized by SE of the neighborhoods. It is incremented by r when there is a car theft	$0 \leq v_p \leq 1$	Routine activity
	RV: Weighted combination of past thefts in the neighborhood itself and its surroundings	$0 \leq v_p \leq 1$	Repeat victimization

Additional to the previous variables, there are other attributes and characteristics considered in the model.

- $\lambda \in [0.3, 1]$: regulating factor of the probability 3.1.3 this factor was restricted to this range to avoid solutions .

- τ: weight included in $RV(s)$ defining the influence of the surroundings. If $\tau = 1$ then $RV(s)$ only has into account past thefts in the cell i but not in the surroundings.

- $n_{cars} = 200.000$: the total of cars generated during the simulation for the entire city. Despite Movilidad-Bogota (2015) informed that in 2015 there were more than two million of cars in Bogota, I generated a representative number of 200.000 cars for the simulation.

- r: Increasing rate in $Sec(s)$ every time there is a crime in the cell i.

- n_{cars_i}: the total of cars in neighborhood i depending on the chosen density. For simplicity, the model does not consider the movement of the vehicles; therefore, the distribution to allocate the cars inside each cell becomes a relevant factor. As a result, 200.000 cars generated were allocated randomly based on the density of blocks inside the neighborhoods, and also I evaluated a variation where the allocation is based on the dwellings inside neighborhoods.

- $gain - SE\ allocation$: Additional to the random allocation of cars, I will consider a variation where the cars are allocated according to the gain and the socio-economic (SE) stratum. Consequently, cars with low values of gain will be in neighborhoods with low SE, and vehicles with high values of gain will be in cells with high SE.

- $target = 5.135$: The model stops when it reaches this quantity. This is what in previous chapter I called target population, an it is a subset of car thefts events based on the pattern of non-violent episodes committed on the streets, and including only the most common brands and classes with vehicles that have antiquity of 15 years or less.

Small scale example

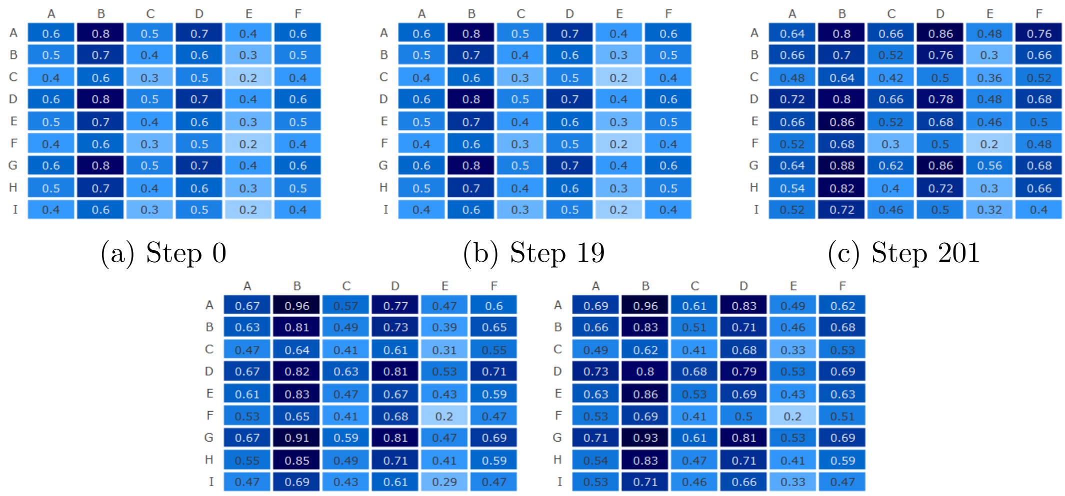
To illustrate the general behavior of the model, I recreated a "city" with random allocation of cars and the following characteristics:

- $n_{nb} = 54$ cells or neighborhoods.

- $n_{nb_i} = 20$ for all cells.

- $n_{cars} = 1080.$
- $target = 300$

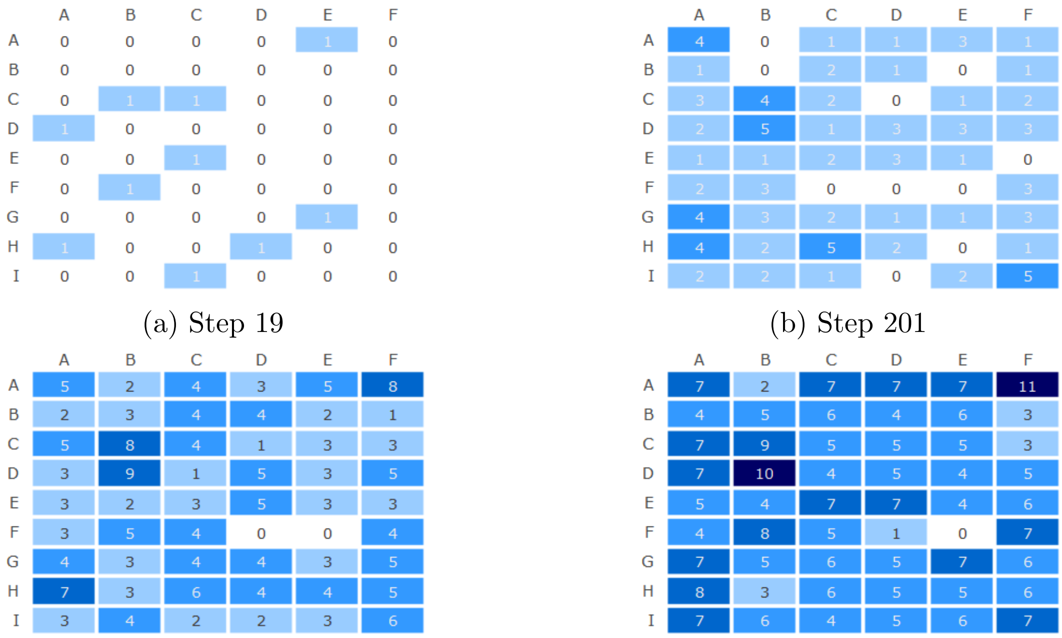
Using two different sets of parameters, we can see changes through the simulation steps or attempts to commit car theft. Figures 3.2 and 3.4 show the evolution of delta, and figures 3.3 and 3.5 present the number of thefts.



(a) Step 0 (b) Step 19 (c) Step 201

(d) Step 417 (e) Step 610

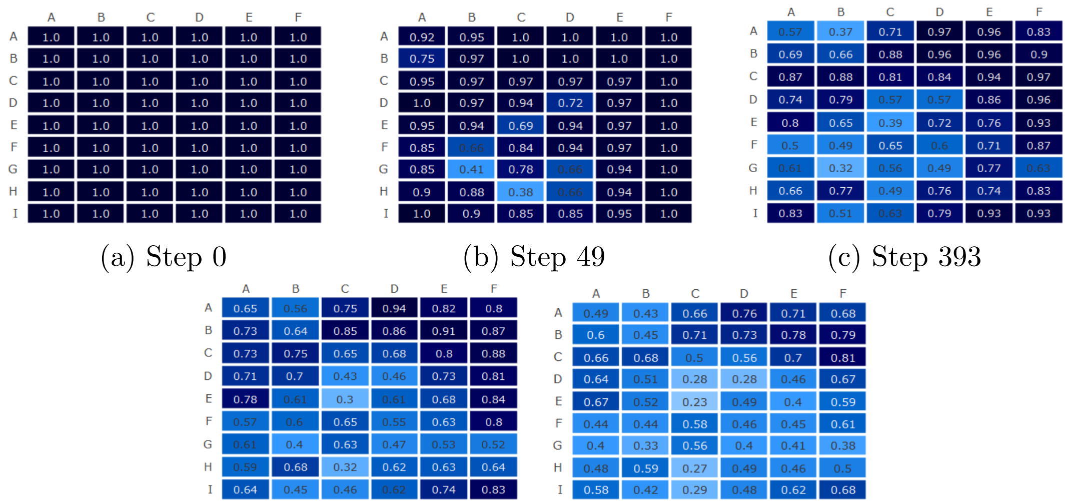
Figure 3.2: δ_i with $\alpha, \beta, \gamma, \eta, \kappa, \tau, \lambda, r = 1$



(a) Step 19 (b) Step 201

(c) Step 417 (d) Step 610

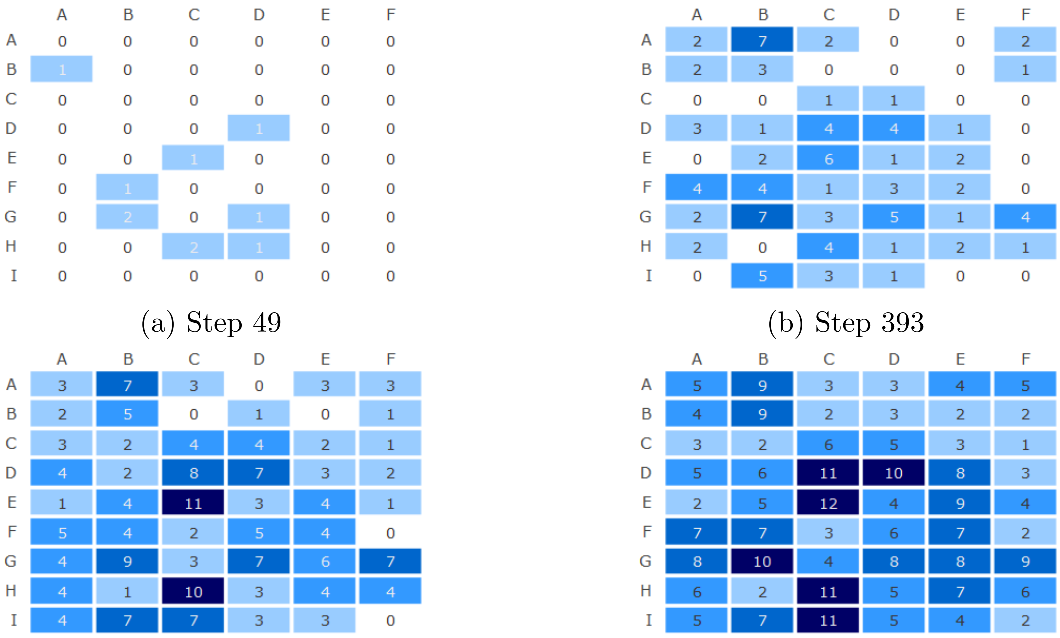
Figure 3.3: Thefts with δ_i with $\alpha, \beta, \gamma, \eta, \kappa, \tau, \lambda, r = 1$



(a) Step 0 (b) Step 49 (c) Step 393

(d) Step 665 (e) Step 943

Figure 3.4: δ_i with $\alpha, \beta, \gamma, \eta, r = 0$ and $\kappa = 1, \tau = 0.5, \lambda = 0.3$



(a) Step 49 (b) Step 393

(c) Step 665 (d) Step 943

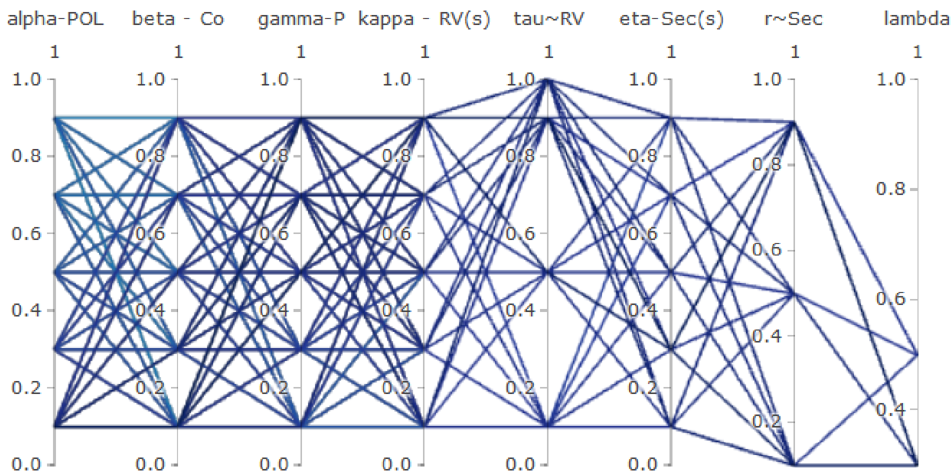
Figure 3.5: Thefts with δ_i with $\alpha, \beta, \gamma, \eta, r = 0$ and $\kappa = 1, \tau = 0.5, \lambda = 0.3$

3.2 Model Estimation

The model proposed in section 3.1 includes nine parameters that reflect the weight of each mechanism. A simple approach to get a suitable set of parameters that reproduce closely the real counts of thefts is through a combination of a single values. For instance in the figure 3.6 we can see the complete space and the coverage with the following sets:

- $\alpha, \beta, \gamma, \eta, \kappa \in 0.1, 0.3, 0.5, 0.7, 0.9$
- $\tau \in 0.1, 0.5, 0.9, 1$
- $\lambda \in 0.3, 0.5$

Figure 3.6: Space of the parameters



Based on the variation we can have a simple grid to cover partly the space of possible combinations. However, this basic example requires the execution of 75.000 models and it might be possible that this approach do not take into account the best solution. On this situations genetic algorithms is a very useful methodology. With genetic algorithms it is possible to cover a wider space, and it exploits sets that have more potential as suitable solutions based on the fitness function. For the implementation, the fitness function is the correlation of thefts by neighborhood with the empirical data.

To execute the genetic algorithm the number of generations was between 30 and 60. For the initial population, I considered three values: 640, 1.000, and 2.000. The size for tournament was 3 but the algorithm always selected the best solution over all solutions. The values of the population were equal to execute crossover and also mutation, but when the initial population was 640 and 2.000 these values were 128, and when the initial population was 1.000 the number for crossover and mutation was 250.

The ABM model and the genetic algorithm were implemented in python from scratch. The codes are included in the Appendix.

Chapter 4

Results

4.1 Optimization based on GA

This section present the results under four scenarios:

1. Allocation of cars randomly, and the distribution of blocks in the neighborhood to generate the number of cars.

2. Allocation of cars based on the association of gain and socio-economic stratum, and the distribution of blocks in the neighborhood to generate the number of cars..

3. Allocation of cars randomly, and the distribution of dwellings in the neighborhood to generate the number of cars.

4. Allocation of cars based on the association of gain and socio-economic stratum, and the density of dwellings in the neighborhood to generate the number of cars.

On figures 4.1, 4.2, 4.3 and 4.4 presented below, it is possible to observe that in general, all scenarios converge to specific solutions. Scenarios 1 and 2 have particularly very low fitness with values between -0.01 and 0.24, but the allocation of cars according to SE increases the range of the correlation in 0.02. This is not a substantial change, but might indicate that this configuration, when the density of cars is related to the number of blocks, reflects better the reality.

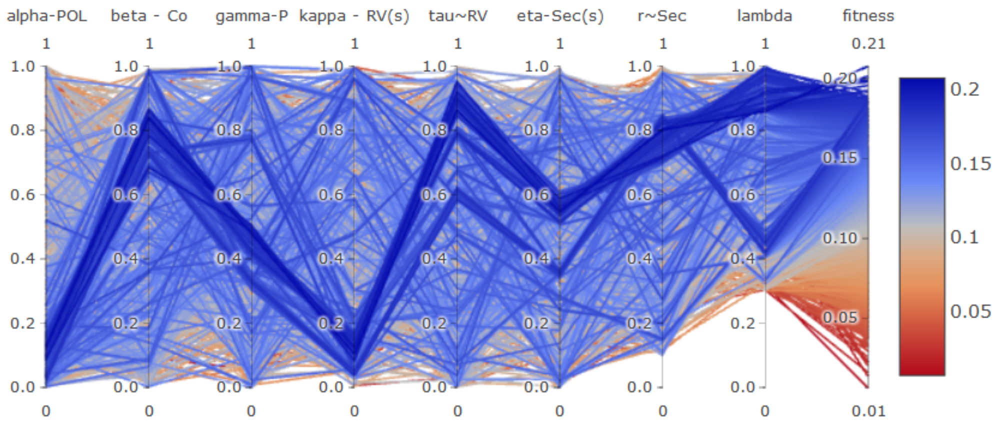


Figure 4.1: Scenario 1. Initial population = 640

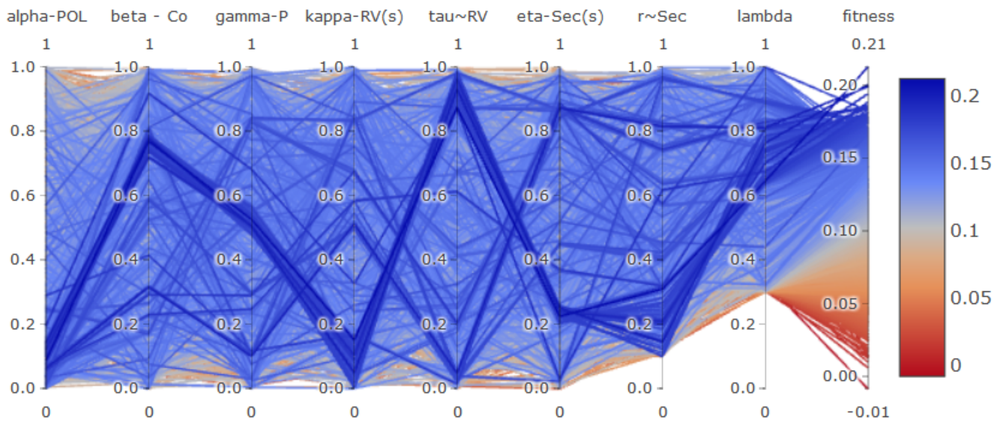


Figure 4.2: Scenario 1. Initial population = 2000

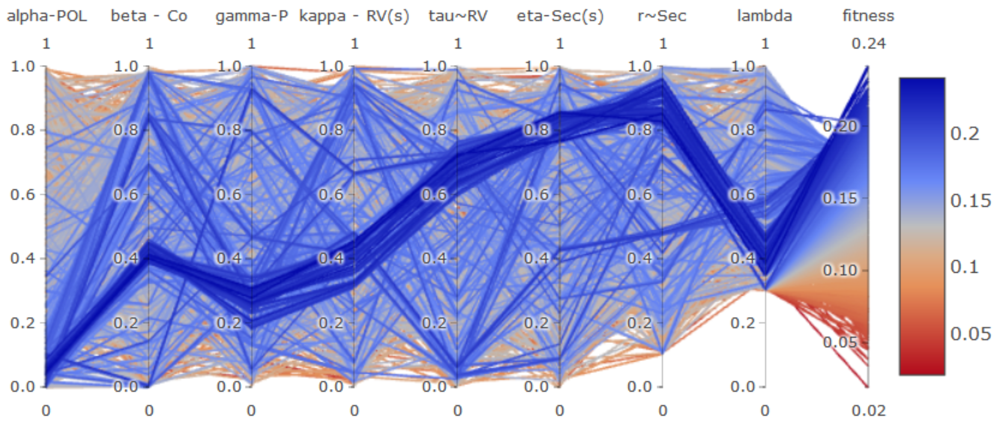


Figure 4.3: Scenario 2. Initial population = 640

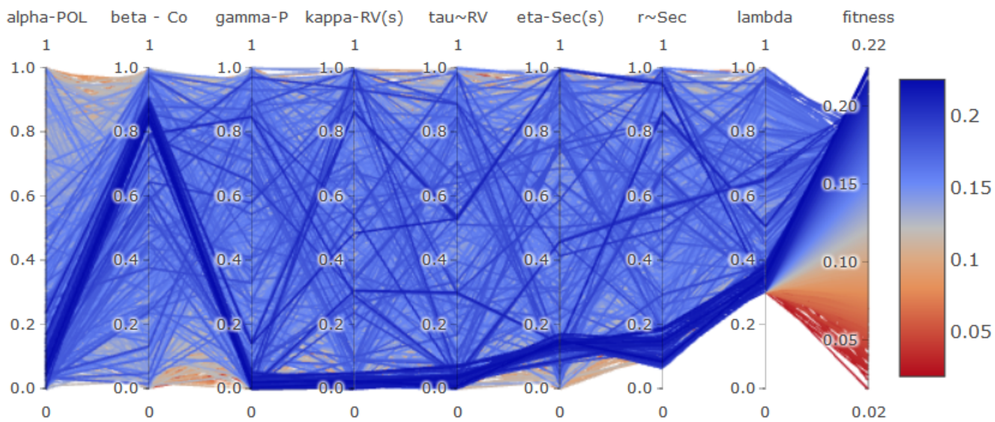


Figure 4.4: Scenario 2. Initial population = 2000

However, when we see 4.5, 4.6, 4.7, and 4.8 where allocation of cars according to the number of blocks was used, then, scenarios 1 and 2 does not look accurate.

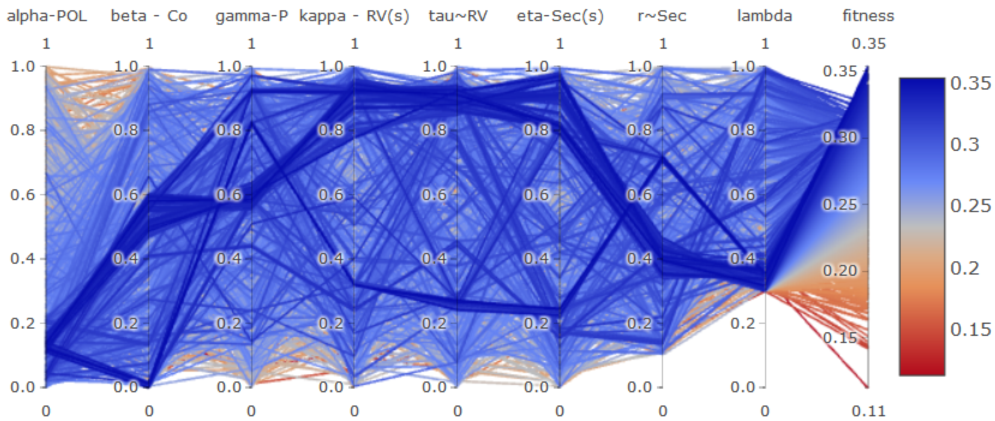


Figure 4.5: Scenario 3. Initial population = 640

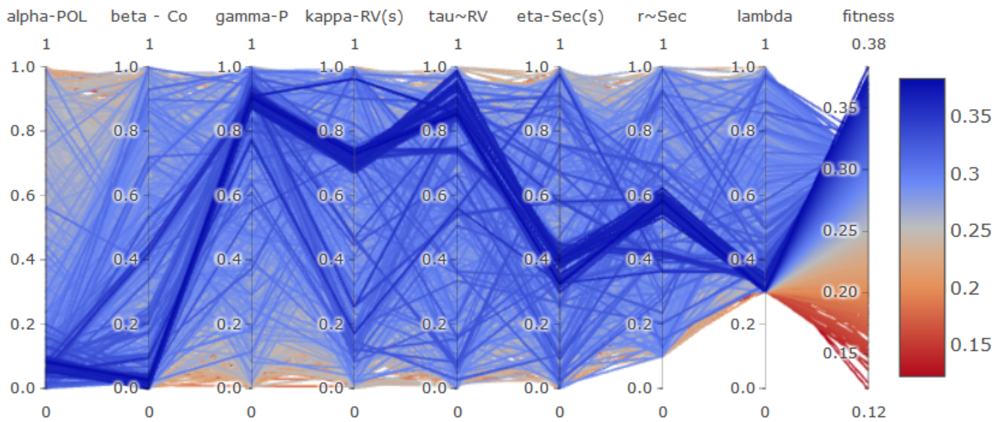


Figure 4.6: Scenario 3. Initial population = 2000

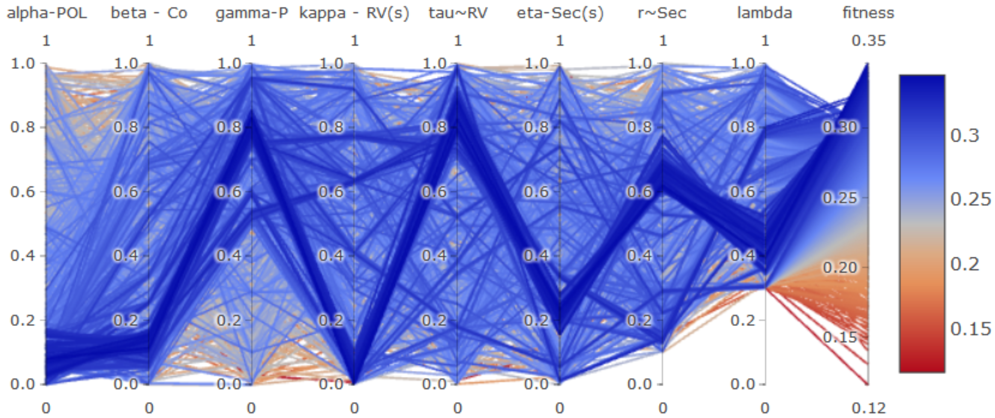


Figure 4.7: Scenario 4. Initial population = 640

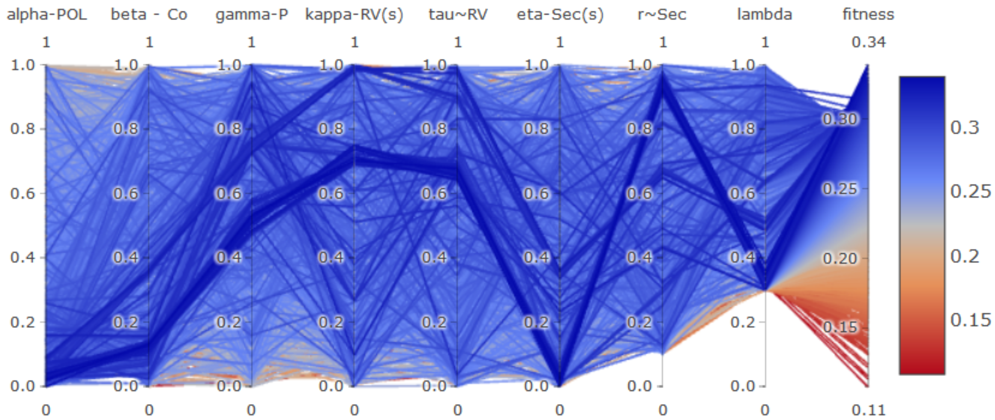


Figure 4.8: Scenario 4. Initial population = 2000

Scenarios 3 and 4, presented above, show fitness between 0.11 and 0.38 but without a clear pattern for the solution set of parameters. Additionally, when we compare the results of the scenario 4 (allocation according SE) with the scenario 3 (completely random), the results show that both distribution of fitness are quite similar. Nevertheless, We can see differences in terms of the size of the initial population, because the maximum fitness is observed when the initial population was 2.000 under scenario 3. This result is expected because a bigger population implies more coverage for the search space, and after the guided search is possible to see the improvements.

Given that the best result was obtained with random allocation of cars and based on the density of dwellings, I decided to execute an additional genetic algorithm with initial population of 1.000, and also I increase the population for crossover and mutation from 128 to 250. Additionally, the variance parameter used in mutation was raised from 0.02 to 0.9 to increase the search space of possible solutions. However, figure 4.9 shows that it was not a successful implementation. Although the results have also a case with high fitness, it is not possible to see a pattern for an specific combination of parameters.

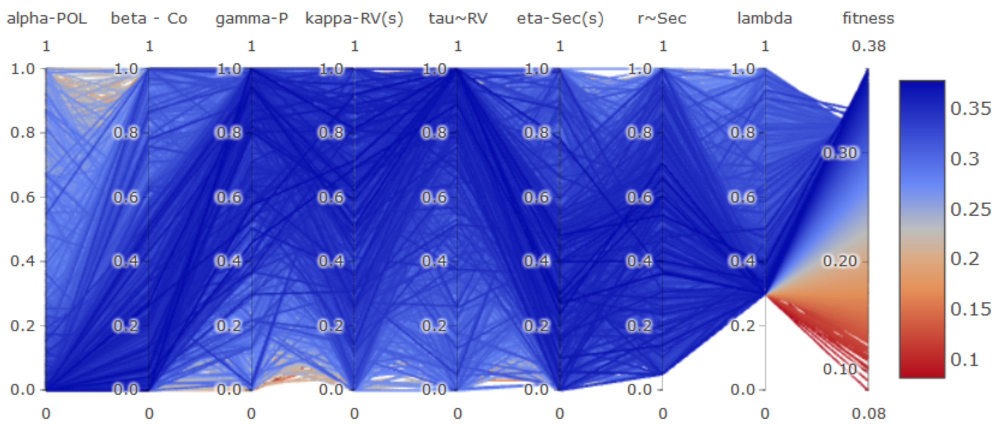


Figure 4.9: Scenario 3. Initial population = 1000

With all the implementations, I chose the better sets of solutions and then I execute 100 times the ABM for each set. I did also this procedure with the worst solution to check the consistency of the results. These sets are:

Scenario	IP in GA	Fitness	α	β	γ	κ	η	r	λ	τ
3	2000	0.38	0.11	0	0.93	0.7	0.36	0.61	0.31	0.9
4	2000	0.34	0.07	0.11	0.57	0.68	0.02	0.97	0.3	0.7
3	1000	0.38	0.17	0.75	0.03	0.9	0.09	0.55	0.34	0.15
3	640	0.35	0.14	0.48	0.6	0.78	0.98	0.4	0.38	0.9
3	1000	0.08	0.8	0	0.05	0.15	0	0.45	0.3	0.85
3	2000	0.12	1	0.3	0.1	0.9	0.37	0.95	0.38	0.83
3	640	0.11	0.6	0.15	0.05	0.1	0.15	0.55	0.35	1

The results of the distribution of fitness are shown in the figure 4.10

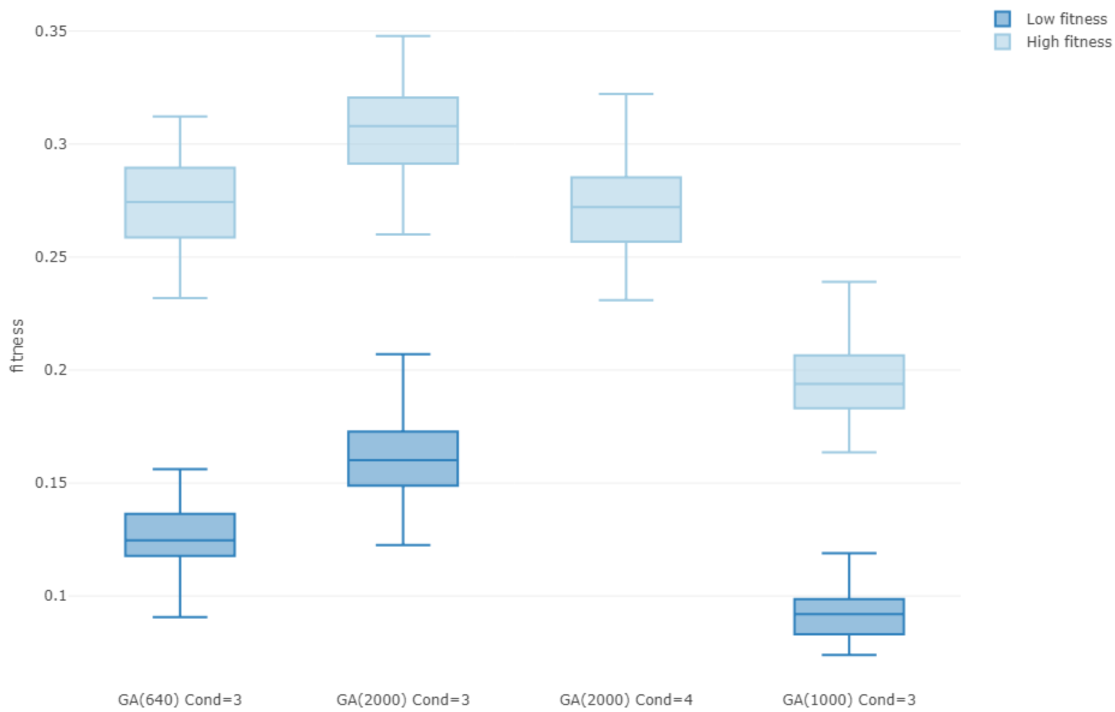


Figure 4.10: Distribution of 100 fitness with different solutions set

Based on this figure, we can see that the guided search with the genetic algorithm worked better with a bigger initial population. However, it was not convenient to increase the size of crossover and mutation. And specifically, it was not right to increase the variance for mutation, because this parameter changes a lot the search space, and then, the mutation loses the ability to guide towards better solutions.

Finally, it is important to compare the results from the GA with a execution of a basic grid. On figure 4.11, we can observe the results generating 70.000 possible solutions.

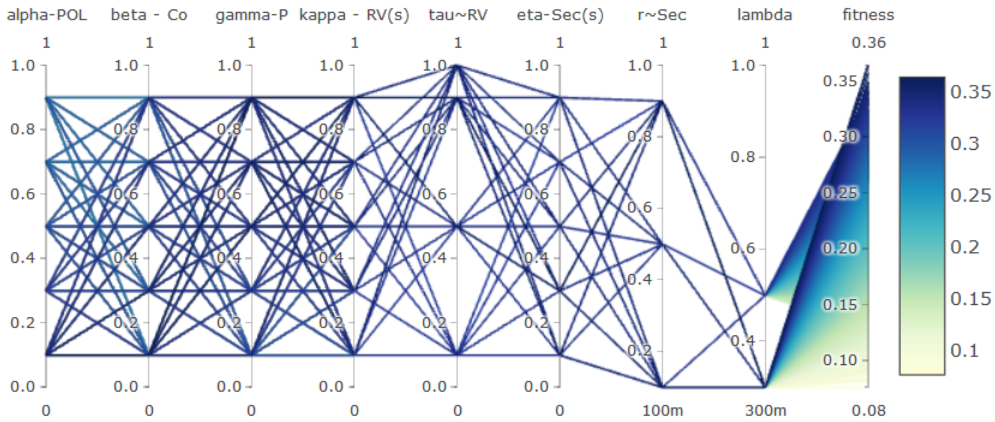


Figure 4.11: Estimation of parameters by grid

Although the grid does not cover the entire space, it was still possible to observe, some patterns, like α it is not a important parameter and alpha is linked with lower values (see figure 4.12).

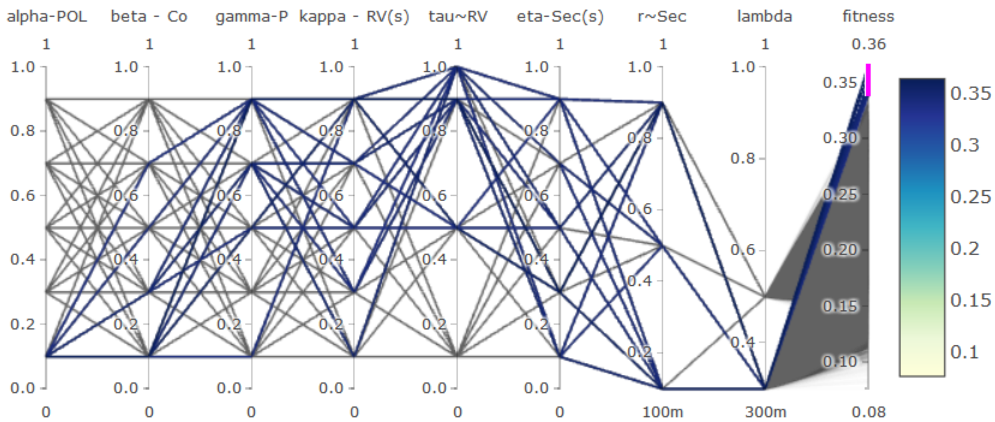


Figure 4.12: Estimation of parameters by grid, filtered by high values of fitness

On the other hand, the main difference of GA with a grid search is basically efficiency. With GA, the ABM was executed 14.544 times for the third scenario with a initial population of 2.000, and it achieves a fitness of 0.38. Grid search was highly demanding in terms of computing, 70.000 ABM models were executed, and the results are not the best. However, it is essential to execute a couple of times the genetic algorithm and some additional variations, to be more secure that the final results is not a local optimum.

4.2 Parameters

In terms of the ABM parameters, the findings are interesting. First in the section 2.3, the count models showed that dwellings have a bigger impact on the model than blocks. This is also reflected in the ABM model, when the results for the scenarios 1 and 2 did not have good fitness. Additionally, if we analyze the best solution:

Scenario	IP in GA	Fitness	α	β	γ	κ	η	r	λ	τ
3	2000	0.38	0.11	0	0.93	0.7	0.36	0.61	0.31	0.9

We can compare these results with the results from the count model in section 2.3. For instance in the zero-inflated poisson regression that includes dwellings, commercial area was quite low and not significant, and from the ABM the parameter *beta* associated to this feature was estimated as zero. Furthermore, it is interesting to see that the parameter related with police station and estimated with the ABM is low ($\alpha = 0.11$). From, the count models we knew that this attribute was a risk factor and it was significant. However, in the ABM model this variable was introduced as a protective factor, but the small number might confirm the findings from section 2.3. A similar conclusion is possible for the parameter γ, this parameter was very significant and it was a protective factor, this behaviour is also confirmed by the ABM.

Regarding the endogenous parameters, the ABM indicates that it occurs repeat victimization at the neighborhood level. Although, I was expecting that the surrounding had more influence over the phenomenon, the low weight of $\tau = 0.9$ indicates that are more relevant the events inside each neighborhood, than the events of the surroundings. About the security attribute, the rate

$r = 0.61$ denotes that almost every time that occurs the first crime, the security increases. However, the value $\eta = 0.36$ indicates that this factor is not so relevant as a "protective" factor, so the security might not be associated with the socio-economic stratum.

Chapter 5

Conclusions

ABM models are very useful specially when the characteristics that we want to evaluate are endogenous. Specially in crime, when the interactions play an important role in the dynamics. However, the level of difficulty increases given all the assumptions required to govern properly the complex system, and then the evaluation of the model not only requires a large understanding of the phenomenon, but also it requires the implementation advance techniques like genetic algorithms.

We can saw that genetic algorithms is a very useful technique that allows to simplify the estimation of parameters in ABM models. Nevertheless, if we formulated assumptions over the model that does not adjust to the reality, there might be cases where genetic algorithms lead wrong conclusions. For instance, in our model we assume on purpose that police station was a factor that helped to prevent crime. In effect, the model shows the lower influence of this variable to prevent crimes, but indeed count models show that this variable was significant and increases the risk for crime. Consequently, the model proposed is not completely wrong, but this convergence to lower values can lead to wrong conclusions.

Nevertheless, It was interesting how the endogenous variables were reflected in the count models and in the ABM results . Thus, it would be interesting evaluate other type of dynamics with this approach. So, it will be a good start point to start with basic endogenous attributes and then build complex models over the simple ones.

After the complete understanding of the simple model proposed, there are many ways to increase the complexity. Although the recommendation would be to include more dynamics in single stages. For instance, the first modification of the model might be to give the attribute of motion to the cars. Additionally, other types of agents can be created like offenders or police to check if the interaction has more impact than one fix police station. With information about offenders, it is possible to associated hot spots, because now we could evaluate the repeat victimization theory, but in the model I did not included how is the dynamic to generate the hot-spots.

Finally, the definition of crime theories help to define the dynamics than govern the complex system, but it is important to delimit the influence of them on each attribute. If one theory has more influence over the phenomenon and we would like to describe it, we have to be careful with the interaction with the others. Given the endogeneity, sometimes is difficult to separate the results and just describe the phenomenon bases on one theory.

Appendix A

Histogram and models with all data

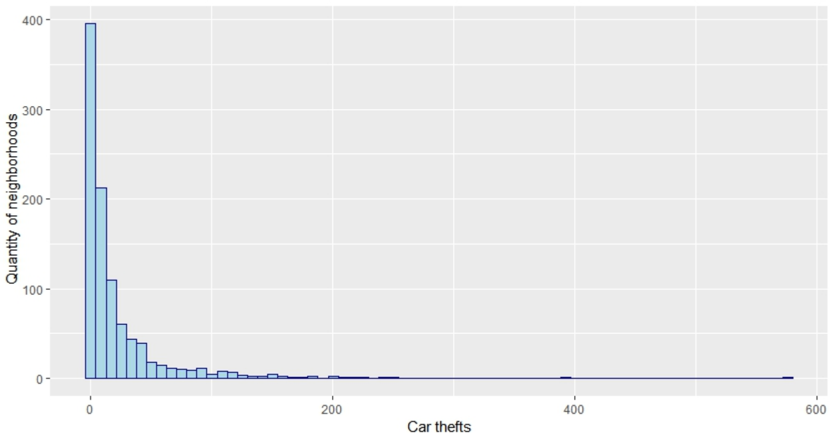


Figure A.1: Histogram car thefts per neighborhood

Table A.1: Poisson and NB models

	Poisson		Negative Binomial	
	Blocks	Dwellings	Blocks	Dwellings
(Intercept)	1.20^{***}	1.80^{***}	0.51^{***}	0.96^{***}
	(0.03)	(0.02)	(0.14)	(0.12)
Police station	0.80^{***}	0.82^{***}	0.86^{***}	0.70^{***}
	(0.02)	(0.02)	(0.13)	(0.13)
Parking place	-1.13^{***}	-4.06^{***}	-1.25^{*}	-5.96^{***}
	(0.07)	(0.09)	(0.51)	(0.72)
Commercial area	0.11	0.93^{***}	-0.31	2.57^{***}
	(0.07)	(0.07)	(0.54)	(0.55)
Average SES	0.36^{***}	0.32^{***}	0.60^{***}	0.55^{***}
	(0.01)	(0.01)	(0.05)	(0.05)
Population block	1.52^{***}		1.60^{***}	
	(0.03)		(0.17)	
Total dwellings		3.31^{***}		5.24^{***}
		(0.04)		(0.61)
AIC	34567.05	34080.77	7348.12	7353.78
BIC	34596.40	34110.12	7382.36	7388.03
Log Likelihood	-17277.52	-17034.38	-3667.06	-3669.89
Deviance	31091.93	30605.65	1145.61	1146.75
Num. obs.	984	984	984	984

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.2: Zero inflated models

	Poisson		Negative Binomial	
	Blocks	Dwellings	Blocks	Dwellings
Count model:				
(Intercept)	1.65^{***}	2.13^{***}	1.02^{***}	1.36^{***}
	(0.03)	(0.02)	(0.17)	(0.15)
Police station	0.76^{***}	0.75^{***}	0.84^{***}	0.69^{***}
	(0.02)	(0.02)	(0.12)	(0.12)
Parking area	-0.91^{***}	-3.62^{***}	-0.93^{*}	-5.12^{***}
	(0.07)	(0.09)	(0.47)	(0.60)
Commercial area	-0.25^{***}	0.55^{***}	-0.42	1.76^{**}
	(0.08)	(0.07)	(0.55)	(0.66)
Average SES	0.31^{***}	0.28^{***}	0.46^{***}	0.46^{***}
	(0.01)	(0.01)	(0.06)	(0.06)
Total of blocks	1.26^{***}		1.40^{***}	
	(0.03)		(0.17)	
Total of dwellings		2.98^{***}		4.48^{***}
		(0.04)		(0.61)
Log(theta)			-0.47^{***}	-0.43^{***}
			(0.05)	(0.05)
Zero model:				
(Intercept)	0.68^{**}	0.34	1.05^{*}	0.52
	(0.23)	(0.20)	(0.53)	(0.39)
Police station	-0.47	-0.61	-0.08	-0.34
	(0.36)	(0.36)	(1.33)	(1.13)
Parking area	1.10	8.90^{***}	-404.89	10.02
	(1.59)	(1.96)	(648.79)	(14.41)
Commercial area	-16.88^{***}	-13.03^{***}	-671.29	-519.89^{*}
	(4.83)	(3.82)	(480.95)	(241.46)
Average SES	-0.49^{***}	-0.46^{***}	-0.60	-0.03
	(0.10)	(0.10)	(0.41)	(0.21)
Total of blocks	-1.79^{***}		0.04	
	(0.37)		(1.03)	
Total of dwellings		-10.31^{***}		-20.16
		(2.13)		(13.03)
AIC	30139.64	29345.38	7285.94	7269.97
Log Likelihood	-15057.82	-14660.69	-3629.97	-3621.98
Num. obs.	984	984	984	984

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Appendix B

Python codes

```
import random as rd
import numpy as np
from scipy import stats

dir = "D:/tesis/Model_neighbors/X4_ABM/"

def cal_pop_fitness(chromosome):
 # Scenarios
 # 1
 # 2
 # 3
 # 4
 if chromosome[7] == 1:
  with open(dir + "_0_1_rd_mz_grid.txt", "r") as file:
       grid_info = eval(file.readline())
  target = sum([i[5] for i in grid_info])
  list_grid = {}
  for i in range(0, len(grid_info)):
       list_grid[grid_info[i][0]] =
           {"CAI": grid_info[i][2],
             "commerce": grid_info[i][3],
             "parking": grid_info[i][4],
```

```
                "target": grid_info[i][5],
                "nb": grid_info[i][6],
                "pt": grid_info[i][7],
                "sg": grid_info[i][8]}
    with open(dir + "_0_1_rd_mz_cars.txt", "r") as file:
        car_info = eval(file.readline())
    list_car = car_info.copy()
elif chromosome[7] == 2:
    with open(dir + "_0_2_mz_gain_grid.txt", "r") as file:
        grid_info = eval(file.readline())

    target = sum([i[5] for i in grid_info])

    list_grid = {}
    for i in range(0, len(grid_info)):
        list_grid[grid_info[i][0]] =
            {"CAI": grid_info[i][2],
                "commerce": grid_info[i][3],
                "parking": grid_info[i][4],
                "target": grid_info[i][5],
                "nb": grid_info[i][6],
                "pt": grid_info[i][7],
                "sg": grid_info[i][8]}
    with open(dir + "_0_2_mz_gain_cars.txt", "r") as file:
        car_info = eval(file.readline())
    list_car = car_info.copy()
elif chromosome[7] == 3:
    with open(dir + "_0_3_dw_rd_grid.txt", "r") as file:
        grid_info = eval(file.readline())



    target = sum([i[5] for i in grid_info])
```

```
    list_grid = {}
    for i in range(0, len(grid_info)):
        list_grid[grid_info[i][0]] =
            {"CAI": grid_info[i][2],
            "commerce": grid_info[i][3],
            "parking": grid_info[i][4],
            "target": grid_info[i][5],
            "nb": grid_info[i][6],
            "pt": grid_info[i][7],
            "sg": grid_info[i][8]}
    with open(dir + "_0_3_dw_rd_cars.txt", "r") as file:
        car_info = eval(file.readline())
    list_car = car_info.copy()
elif chromosome[7] == 4:
    with open(dir + "_0_4_dw_gain_grid.txt", "r") as file:
        grid_info = eval(file.readline())

    target = sum([i[5] for i in grid_info])

    list_grid = {}
    for i in range(0, len(grid_info)):
        list_grid[grid_info[i][0]] =
            {"CAI": grid_info[i][2],
                "commerce": grid_info[i][3],
                "parking": grid_info[i][4],
                "target": grid_info[i][5],
                "nb": grid_info[i][6],
                "pt": grid_info[i][7],
                "sg": grid_info[i][8]}
    with open(dir + "_0_4_dw_gain_cars.txt", "r") as file:
        car_info = eval(file.readline())
```

```
  list_car = car_info.copy()

"""
Global parameters
"""
alpha = chromosome[0]
beta = chromosome[1]
gamma = chromosome[2]
kappa = chromosome[3]
eta = chromosome[4]
t = chromosome[5]
p_lambda = chromosome[6]
spread_nb = chromosome[8]
target2 =
trials = 0

"""
selecting some cars to calculate probability
"""
car_stolen = {}

while target2 < target:
    sample_cars = rd.sample(list_car.keys(), 1)[0]

    grid_trial = list_car[sample_cars]["id_grid"]
    trials += 1

    max_thefts =
        max([list_grid[i]["pt"] for i in list_grid])
    if max_thefts == 0:
        max_t = 1
    else:
```

```
        max_t = max_thefts
    pt_own = list_grid[grid_trial]["pt"] / max_t
    pt_nb = 0
    for nb in list_grid[grid_trial]["nb"]:
        pt_nb += list_grid[nb]["pt"] / (max_t *
            len(list_grid[grid_trial]["nb"]))

    if (alpha + beta + gamma + kappa + eta) > 0:
        delta = ((alpha * list_grid[grid_trial]["CAI"]) +
            (beta * (1 - list_grid[grid_trial]["commerce"])) +
            (gamma * list_grid[grid_trial]["parking"]) +
            (kappa * (1 - ((spread_nb * pt_own) +
            ((1-spread_nb) * pt_nb)))) +
            (eta * list_grid[grid_trial]["sg"])
            ) / (alpha + beta + gamma + kappa + eta)
    else:
        delta = 0
    probability = (1 +
       np.exp(-(list_car[sample_cars]["gain"] -
       delta) / p_lambda)) ** (-1)
    theft = rd.random() < probability
    if theft:
        target2 += 1
        car_stolen[sample_cars] = list_car.pop(sample_cars)
        list_grid[grid_trial]["pt"] += 1
        list_grid[grid_trial]["sg"] += t
        if list_grid[grid_trial]["sg"] > 1:
            list_grid[grid_trial]["sg"] = 1



cor_per = stats.pearsonr(
    [list_grid[i]["target"] for i in list_grid],
```

```
        [list_grid[i]["pt"] for i in list_grid])[0]
 cor_spe = stats.spearmanr(
    [list_grid[i]["target"] for i in list_grid],
        [list_grid[i]["pt"] for i in list_grid])[0]

 return [chromosome, car_stolen,
        list_grid, cor_per,
        cor_spe, trials]


def tournament(population, n_tourn,
    sol_per_pop, fitness_res):

    pop_tournament = []
    fitness_tournament = []
    for i in range(0, (sol_per_pop-1)):
        muestra = rd.sample(fitness_res,
            k=n_tourn)
        max_fitness_idx_tourn =
            np.where(fitness_res == max(muestra))[0][0]
        fitness_tournament.append(max(muestra))
        pop_tournament.append(
            population[max_fitness_idx_tourn])

    max_all_idx_tourn = np.where(
        fitness_res == max(fitness_res))[0][0]
    fitness_tournament.append(max(fitness_res))
    pop_tournament.append(population[max_all_idx_tourn])
    return pop_tournament, fitness_tournament


def crossover(muestra):
```

```
    cutpoint = rd.sample([0, 1, 2, 3], k=1)
    idx_genes = rd.sample([0, 1, 2, 3, 4], k=5)
    if cutpoint == 0:
        cross = np.array(
            [[muestra[0][idx_genes[0]], idx_genes[0]],
             [muestra[0][idx_genes[1]], idx_genes[1]],
             [muestra[1][idx_genes[2]], idx_genes[2]],
             [muestra[1][idx_genes[3]], idx_genes[3]],
             [muestra[1][idx_genes[4]], idx_genes[4]],
             [muestra[0][5], 5],
             [muestra[0][6], 6],
             [muestra[0][7], 7],
             [muestra[0][8], 8]])
    elif cutpoint == 1:
        cross = np.array(
            [[muestra[0][idx_genes[0]], idx_genes[0]],
            [muestra[0][idx_genes[1]], idx_genes[1]],
            [muestra[0][idx_genes[2]], idx_genes[2]],
            [muestra[1][idx_genes[3]], idx_genes[3]],
            [muestra[1][idx_genes[4]], idx_genes[4]],
            [muestra[0][5], 5],
            [muestra[0][6], 6],
            [muestra[0][7], 7],
            [muestra[0][8], 8]])
    elif cutpoint == 2:
        cross = np.array(
            [[muestra[0][idx_genes[0]], idx_genes[0]],
             [muestra[0][idx_genes[1]], idx_genes[1]],
             [muestra[0][idx_genes[2]], idx_genes[2]],
             [muestra[0][idx_genes[3]], idx_genes[3]],
             [muestra[1][idx_genes[4]], idx_genes[4]],
             [muestra[1][5], 5],
```

```
                [muestra[1][6], 6],
                [muestra[1][7], 7],
                [muestra[1][8], 8]])
    else:
        cross = np.array(
                [[muestra[0][idx_genes[0]], idx_genes[0]],
                [muestra[1][idx_genes[1]], idx_genes[1]],
                [muestra[1][idx_genes[2]], idx_genes[2]],
                [muestra[1][idx_genes[3]], idx_genes[3]],
                [muestra[1][idx_genes[4]], idx_genes[4]],
                [muestra[1][5], 5],
                [muestra[1][6], 6],
                [muestra[1][7], 7],
                [muestra[1][8], 8]])
    cross = sorted(cross, key=lambda x: x[1])
    offspring = np.transpose(cross)[0]
    return offspring.tolist()


def crossover2(population, fitness):
    # escogiendo 2 de torunament para hacer crossover
    sample_cr_idx = rd.sample(range(0, len(population)), k=2)
    sample_cr = []
    for i in [0, 1]:
        sample_cr.append(population[sample_cr_idx[i]])
    # aca se calcula el resultado del crosover
    result_crom = crossover(sample_cr)
    # se escoge uno de los papas para sacar de la lista
    sample_pop = rd.sample([0, 1], k=1)
    # se saca de la lista el papa escogido
    population.pop(sample_cr_idx[sample_pop[0]])
    fitness.pop(sample_cr_idx[sample_pop[0]])
```

```
    # se calcula el fitness del hijo
    # result_fit = cal_pop_fitness(result_crom)
    return result_crom



def mutate(muestra, var):
  for i in range(0, len(muestra)):
    idx_mutate = rd.sample([0, 1, 2, 3, 4], k=1)[0]
    muestra[i][idx_mutate] += np.random.normal(0, var)
    if muestra[i][idx_mutate] < 0:
        muestra[i][idx_mutate] = 0
    if rd.random() >= 0.5:
        muestra[i][5] += np.random.normal(0, var)
        if muestra[i][5] < 0.05:
            muestra[i][5] = 0.05
        if muestra[i][5] > 1:
            muestra[i][5] = 1
    if rd.random() >= 0.5:
        muestra[i][8] += np.random.normal(0, var)
        if muestra[i][8] < 0:
            muestra[i][8] = 0
        if muestra[i][8] > 1:
            muestra[i][8] = 1
    if rd.random() >= 0.5:
        muestra[i][6] += np.random.normal(0, var)
        if muestra[i][6] < 0.3:
            muestra[i][6] = 0.3
        if muestra[i][6] > 1:
            muestra[i][6] = 1
  return muestra
```

```
def mutate2(population, fitness, ss_mutate, var):

    sample_mt_idx = rd.sample(range(0, len(population)),
        k=ss_mutate)
    sample_mt = []
    for i in range(0, len(sample_mt_idx)):
        sample_mt.append(population[sample_mt_idx[i]])

    result_mt = mutate(sample_mt, var)

    for index in sorted(sample_mt_idx, reverse=True):
        del population[index]
        del fitness[index]
    return result_mt
```

References

Batty, M. (2005). Agents, cells, and cities: new representational models for simulating multiscale urban dynamics. *Environment and Planning A*, *37*(8), 1373–1394.

Berger, L. M., & Borenstein, D. (2013). An agent-based simulation of car theft: further evidence of the rational choice theory of crime. *Economic analysis of law review. Brasília. Vol. 4, n. 1 (jan./jun. 2013), p. 103-119*.

Bhavnani, R., Donnay, K., Miodownik, D., Mor, M., & Helbing, D. (2014). Group segregation and urban violence. *American Journal of Political Science*, *58*(1), 226–245.

Birks, D., & Elffers, H. (2014). Agent-based assessments of criminological theory. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 19–32). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-1-4614-5690-2_687 doi: 10.1007/978-1-4614-5690-2_687

Birks, D., Townsley, M., & Stewart, A. (2012). Generative explanations of crime: Using simulation to test criminological theory. *Criminology*, *50*(1), 221–254.

Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. Macmillan New York.

Bruinsma, G., & Weisburd, D. (2014). *Encyclopedia of criminology and criminal justice*. Springer Reference.

Bursik Jr, R. J. (1988). Social disorganization and theories of crime and delinquency: Problems and prospects. *Criminology*, *26*(4), 519–552.

Cadaster-Bogota. (2018). *Reference map for bogota. updated on march 2018*. Retrieved from https://www.ideca.gov.co/es/servicios/mapa

-de-referencia

Clarke, R. V., & Cornish, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. *Crime and justice*, *6*, 147–185.

Clarke, R. V., & Felson, M. (2017). Introduction: Criminology, routine activity, and rational choice. In *Routine activity and rational choice* (pp. 1–14). Routledge.

Cohen, L. E., & Felson, M. (2016). Social change and crime rate trends: A routine activity approach (1979). In *Classics in environmental criminology* (pp. 203–232). CRC Press.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment*, *91*(2), 121–136.

de Bogotá, C. d. C. (2017). *Encuesta de percepción y victimización en bogotá-segundo semestre 2017.*

Devia, N., & Weber, R. (2013). Generating crime data using agent-based simulation. *Computers, Environment and Urban Systems*, *42*, 26–41.

Doctor, B. D. (2015). *Fuzzy string matching – a survival skill to tackle unstructured information. http://bigdata-doctor.com/fuzzy-string-matching-survival-skill-tackle-unstructured-information-r/.* Retrieved from http://bigdata-doctor.com/fuzzy-string-matching-survival-skill-tackle-unstructured-information-r/

Farrell, G., & Pease, K. (2014). Repeat victimization. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 4371–4381). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-1-4614-5690-2_128 doi: 10.1007/978-1-4614-5690-2_128

Felson, M., & Clarke, R. V. (1998). Opportunity makes the thief. *Police research series, paper*, *98*, 1–36.

Finanzas-personales. (2018). *Top cinco de colores más buscados para carros.* Retrieved from http://www.finanzaspersonales.co/consumo-inteligente/articulo/top-cinco-colores-mas-buscados-para-carros/50179

Gilbert, N. (2008). *Agent-based models* (No. 153). Sage.

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist*. McGraw-Hill Education (UK).

Groff, E. R. (2007). Simulation for theory testing and experimentation: An example using routine activity theory and street robbery. *Journal of Quantitative Criminology*, *23*(2), 75–103.

Groff, E. R. (2014). Agent-based modeling for understanding patterns of crime. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 32–41). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-1-4614-5690-2_476 doi: 10.1007/978-1-4614-5690-2_476

Groff, E. R., Johnson, S. D., & Thornton, A. (2018). State of the art in agent-based modeling of urban crime: An overview. *Journal of Quantitative Criminology*, 1–39.

Hayslett-McCall, K., Qui, F., Curtin, K. M., Chastain, B., Schubert, J., & Carver, V. (2008). The simulation of the journey to residential burglary. *Artificial crime analysis systems*, 281–300.

Kruse, R., Borgelt, C., Braune, C., Mostaghim, S., & Steinbrecher, M. (2016). *Computational intelligence: a methodological introduction*. Springer.

Lindsey, J. K. (2000). *Applying generalized linear models*. Springer Science & Business Media.

MacDonald, J. M., & Lattimore, P. K. (2010). Count models in criminology. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 683–698). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-0-387-77650-7_32 doi: 10.1007/978-0-387-77650-7_32

Malleson, N., & Evans, A. (2014). Agent-based models to predict crime at places. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 41–48). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-1-4614-5690-2_208 doi: 10.1007/978-1-4614-5690-2_208

Malleson, N., See, L., Evans, A., & Heppenstall, A. (2012). Implementing comprehensive offender behaviour in a realistic agent-based model of burglary. *Simulation*, *88*(1), 50–71.

Miller, J. H., & Page, S. E. (2009). *Complex adaptive systems: An introduction to computational models of social life* (Vol. 17). Princeton university press.

Movilidad-Bogota. (2015). *Movilidad en cifras.* Retrieved from https://www.movilidadbogota.gov.co/web/SIMUR/ARCHIVOS/Movilidad_Cifras_2015_V4_marzo2017.pdf

Norza Céspedes, E., Velásquez, D., Andrés, Y., Castillo Romero, L. F., & Torres Guzmán, G. A. (2013). Automobile theft and strategies against crime: a glance from the academy, the offender, and the police. *Revista Criminalidad*, *55*(2), 49–78.

Ospina Baena, J. H., Peña, C., & Mauricio, M. (2015). Automobile theft under the prospective approach: a current and future characterization of this criminal activity in colombia. *Revista Criminalidad*, *57*(3), 9–25.

Police. (2018). *National police of colombia. criminal statistics.* Retrieved from https://www.policia.gov.co/grupo-informaci%C3%B3n-criminalidad/estadistica-delictiva

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Schutte, S. (2010). Optimization and falsification in empirical agent-based models. *Journal of Artificial Societies and Social Simulation*, *13*(1).

van der Loo, M. (2014). The stringdist package for approximate string matching. *The R Journal*, *6*, 111-122. Retrieved from https://CRAN.R-project.org/package=stringdist

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth ed.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4 (ISBN 0-387-95457-0)

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8). Retrieved from http://www.jstatsoft.org/v27/i08/

Statutory Declaration

1. I hereby declare that the thesis, entitled:
   Empirical Calibration of an Agent-Based Model of Car Theft
   is a result of my own work and that no other than the indicated aids have been used for its completion. Material borrowed directly or indirectly from the works of others is indicated in each individual case by acknowledgement of the source and also the secondary literature used.

2. After completion of the examining process, this work will be given to the library of the University of Konstanz, where it will be accessible to the public for viewing and borrowing. As author of this work, I agree to this procedure.

Konstanz, 23rd July 2019 ______________________

(Date) (Signature)

Erklärung

1. Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema:
   Empirical Calibration of an Agent-Based Model of Car Theft
   selbständig verfasst und keine anderen Hilfsmittel als die angegebenen benutzt habe. Die Stellen, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, habe ich in jedem einzelnen Falle durch Angaben der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

2. Diese Arbeit wird nach Abschluss des Prüfungsverfahrens der Universitätsbibliothek Konstanz übergeben und ist durch Einsicht und Ausleihe somit der Öffentlichkeit zugänglich. Als Urheber der anliegenden Arbeit stimme ich diesem Verfahren zu.

Konstanz, 23.July 2019 ______________________

(Datum) (Unterschrift)