Analysis and enhancement of an Individual Based Model strategy to study tuberculosis at a city level

ENGINEERING PHYSICS BACHELOR'S THESIS

Author:

Bernat Puig Camps

Co-advisors:

Daniel López

Cristina Montañola

June 2017

Abstract

Tuberculosis remains a major health problem. In 2015, it is estimated there were 1.8 million TB death and 10.4 million new cases.. To date, it is one of the top 10 causes of death worldwide and it is roughly calculated that one third of the world population has already been infected by *Mycobacterium tuberculosis*. In particular, the Ciutat Vella neighbourhood in Barcelona has a tuberculosis incidence which is comparable to the incidence of countries like Sudan.

In this work, different datasets have been analyzed with the aim of gaining more insight regarding the social factors that affect tuberculosis spreading, especially for the Barcelona case. This work presents three important findings: up to 34 % of the population is TB infected, the most important monitorable sources of infection are households, and that the relatively high TB incidence in the Ciutat Vella district can be explained through immigration from countries that present a remarkably high TB incidence.

Moreover, a previous NetLogo (free and open source simulation tool that incorporates a helpful user-friendly interface) implementation of an Individual Based Model which is used to simulate the tuberculosis dynamics in Ciutat Vella has been translated to the C programming language, presenting substantial improvements both in computation time and scalability.

Finally, the findings in the data analysis performed proved necessary to consider households and other social meeting points in order to properly model tuberculosis dynamics within a city. For this reason, a new version of the model that accounts for these spaces has been developed by generating realistic households by means of an heuristic model that respects the age distribution of the population.

Acknowledgements

First and foremost, I want to thank my co-advisors, Dr. Daniel López and Dr. Cristina Montañola, for their guidance and advice, which habe been extremely helpful during these last months. I also want to thank Dr. Clara Prats for her valuable help.

I want to thank both the BIOCOM-SC group of research at the UPC and the Barcelona Supercomputing Center for having me lend spaces of work in two different campus. Thanks to all the students and teachers with whom I have shared space and time throughout these last months and they have made this project much more enjoyable.

Moreover, I am grateful to the *Agència de Salut Pública de Barcelona* for the data provided that has been key for this study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Tuberculosis disease

Tuberculosis (TB) is an infectious disease caused by the bacillus *Mycobacterium tuberculosis*. It typically affects the lungs (pulmonary TB) but can also affect other sites (extrapulmonary TB). It is estimated that 85% of the cases are pulmonary [14].

The disease is spread between individuals through the air. When sick people with lung TB cough, sneeze or spit, they propel the TB germs into the air. A person needs to inhale only a few of these germs to become infected. After inhaling the air droplets, the bacteria may reach the alveoli in the lungs. If this happens, the bacillus are absorbed by macrophages through phagocytosis, and they start growing and dividing inside them. This is the starting point of the tuberculosis infection. Depending on the immune response of the individual, the initial infection will remain under control or it will evolve towards an active disease, when the person will sicken. Therefore, two stages of the disease can be differentiated: the latent stage (latent tuberculosis infecion or LTBI) and the active stage (active tuberculosis or active TB).

During the latent stage, bacteria are present in the lungs but the bacterial load is sufficiently low to remain under control and thus, the individual does not experiment any symptoms nor infect others. When the immune system of an infected person can no longer control the infection, the individual develops an active disease (i.e. enters the active stage). Only a relatively small portion $(5 - 15\%)$ of the infected individuals will develop an active disease [14]. Notice that people with some degree of immunodeficiency are more susceptible to develop the active disease.

During this stage, the sick individual can infect other people. It may suffer from different symptoms (cough, fever, night sweats, weight loss, etc) although they may be mild for many months. This can lead to delays in seeking care, and results in transmission of the bacteria to others. Without proper treatment up to two thirds of people ill with TB will die.

1.1.1 Diagnosis and treatment

Latent tuberculosis can be diagnosed through a tuberculin skin test (Mantoux test). In case someone has been exposed to the TB bacillus, a skin reaction to the antigens will appear in less than 48 to 72 hours after the test. In order to diagnose an active disease, there are two main procedures: a chest X-ray or a microscopic examination of sputum. A positive in the sputum test, means that the individual is smear positive which is related to a higher infection power [12]. A cultivation of a clinical sample allows to identify antibiotic resistant TB strains. Identifying these resistant strains is of great importance in order to determine the optimal treatment.

Standard tuberculosis treatment usually last for about 6 to 9 months and consists of a combination of 4 antibiotics. Unfortunately, the 6-month period under treatment is difficult to commit to for some, leading to treatment abandonment.

1.1.2 World situation

Tuberculosis is curable and preventable. Nonetheless, it is one of the top 10 causes of death worldwide. In 2015, 10.4 million people fell ill with TB and 1.8 million died from the disease. Although, 95% of TB deaths occur in low- and middle-income countries, it is also a problem for the high developed countries. The World Health Organization (WHO) estimates that one third of the population is infected by the pathogen [14]. From those, a 10% will develop an active TB disease in the following years. Figure 1.1 shows how the incidence of the disease is distributed worldwide.

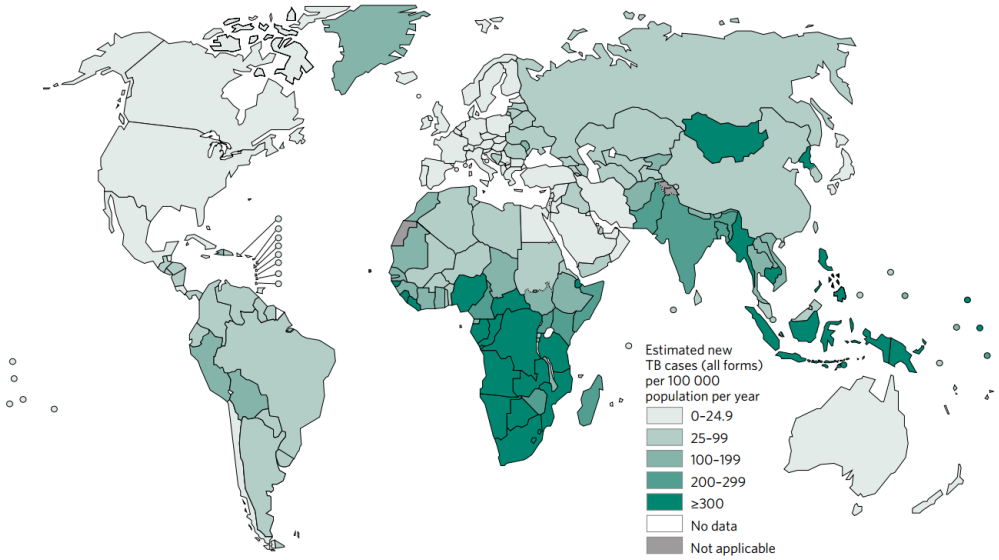

Figure 1.1: Estimated TB incidence rates, 2015. [14]

1.2 Barcelona and tuberculosis

Barcelona is a large city (1 604 555 inhabitants in 1st January of 2015 [4]) with a relatively low tuberculosis (TB) incidence (17.4 TB cases/100 000 inhabitants in 2015 [5]). This is due to the low burden region itself but also because of the high quality public health policies applied to the city regarding tuberculosis. For more than 30 years, the Public Health Agency of Barcelona (ASPB) has been performing a metic-

ulous control of the disease, both by identifying and treating the active cases and by conducting a contact study of such individuals in order to spot further infections [5]. In Figure 1.2, one can observe that, indeed, the Barcelona's TB incidence have been decreasing in the last years. Figure 1.3 shows that the decline of the tuberculosis incidence in Barcelona is faster than Spain's with an average yearly decrease of incidence between 2010 and 2015 of -7.9 %/year in front of a -6.6 %/year.

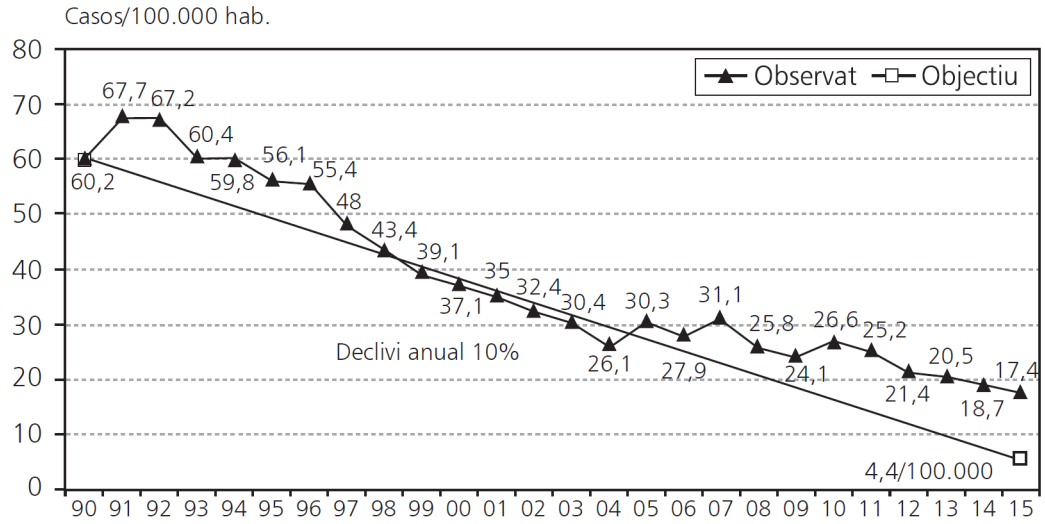

Figure 1.2: Evolution of the TB incidence (new cases/100.000 inhabitants) in Barcelona since 1990. Straight line shows the targeted TB incidence. [5]

Barcelona consists of ten districts with the order of 100 000 inhabitants each and there exist remarkable differences among their population and their life conditions. For instance, while immigrant population represents approximately a 16% of the total of Barcelona, it accounts for 43% of the people living in the Ciutat Vella (CV) district. A significant part of them come from countries stricken by high TB incidences such as Pakistan, the Philipines or Bangladesh [4]. These differences have an impact on the tuberculosis presence in each area. In particular, the Ciutat Vella district presents a huge incidence with respect to the rest of the city (57.3 TB cases/100 000 inhabitants in 2015 [5]). Figure 1.4 shows the TB incidence in each district.

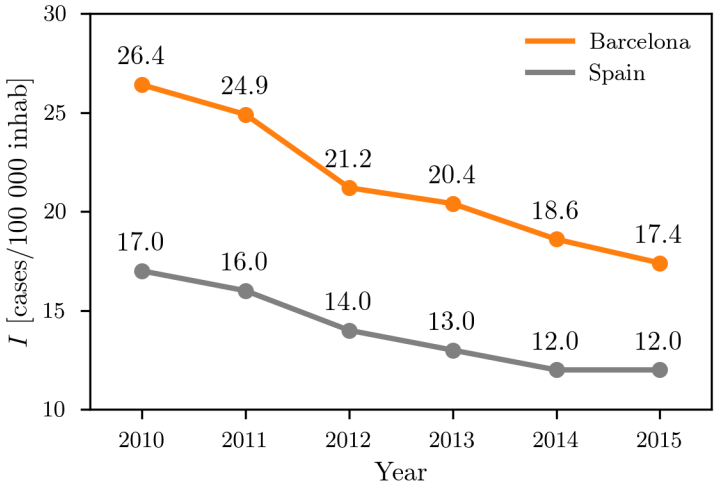

Figure 1.3: Evolution of the TB incidence in Spain and Barcelona between years 2010 and 2014.[14][5]

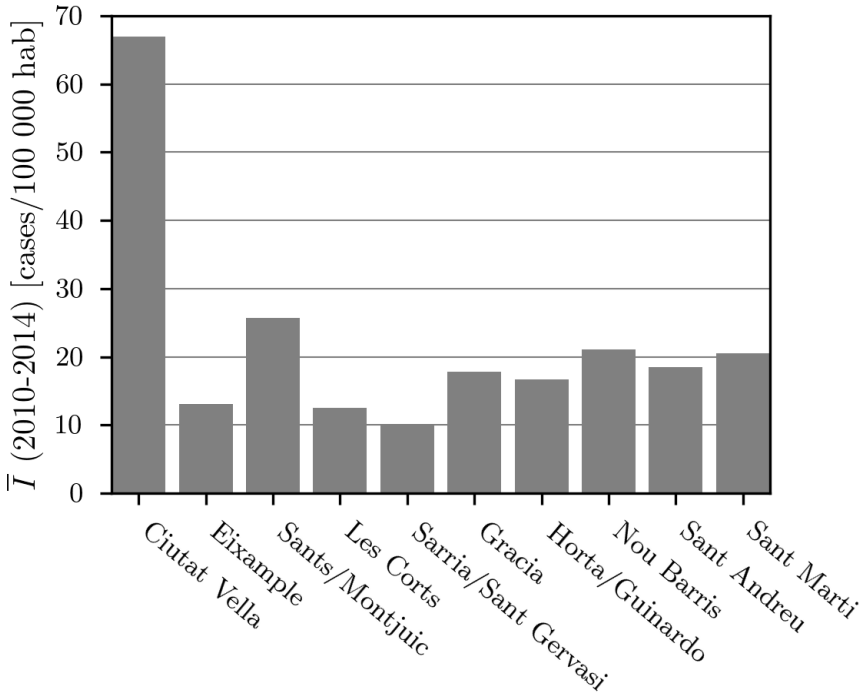

Figure 1.4: Mean TB incidence between years 2010 and 2014 by district. [5]

1.3 Epidemiological models

The ultimate goal when assessing an infectious disease is to control and eradicate it. In order to achieve that, control strategies have to be applied. The better the disease is understood, the more accurate these control strategies will be, thus more efficient. To better understand the dynamics of a given disease epidemiological models are used.

Epidemiology is the study of the factors responsible for the spread of the diseases. To fully understand the mechanisms of transmission of the infectious agents, tools need to be developed and used. Models are a simplification of reality, a conceptual representation of a complex system. Models allow to understand a complex phenomenon by spotting the key aspects of a problem.

Two main strategies can be considered when modelling a certain system: the Top-Down and the Bottom-Up approach. As an example, let us say the flow of a tide has to be simulated; a Top-Down approach would start by understanding the movement of a wave, it would then formulate a differential equation able to predict the position of each water molecule at each time. Instead, a Bottom-Up approach would assign each water molecule a trajectory to then observe the general pattern, the wave.

Therefore, the Top-Down strategy is based on applying a general qualitative theory to the specific in order to generate quantitative conclusions. Mathematicians inclined themselves to this approach and they developed several models that have been extensively used, such as the compartmental models based on differential equations (SEIR, SIR, SIS...). These models consist of different pools of individuals characterized by the status of the disease. The S stands for susceptible; E stands for exposed, people who have been in contact with the disease but are not infectious; I stands for infectious who may transmit the disease and R for recovered, individuals who have overcome the disease.

On the other hand, the Bottom-Up approach is based on modelling the low-level interactions, the behaviour of the elements to observe the dynamics of the global system. Examples of such strategy are the individual-based models (IBM).

1.3.1 Individual-based Models

Historically, the complexity of scientific models was limited by mathematical tractability. Since the only possible approach was differential calculus, they had to be simple enough to be solved mathematically. Thus, people were often limited to model rather simple systems.

With computer simulation, this limitation is removed and hence, more complicated models that account for more system characteristics can be addressed. IBM's are less simplified in one key feature: they represent a system's individual component and its behaviour. Instead of describing the system with variables representing the state of the whole system, the individual agents of the system are modelled.

The IBM consider unique autonomous entities that interact with each other and their environment locally. Each of these entities can have its own characteristics and pursue its objectives. Therefore, these agents present an adaptative behaviour: they adjust their behavior to the current states of themselves, of other agents, and of their environment.[16]

Tuberculosis is a social disease, its incidence strongly depends upon the social characteristics of each individual such as gender, origin, live conditions, etc as it has been seen in the differences among districts in Barcelona. For this reason, seems reasonable to conclude that one of the best ways to simulate the dynamics of the tuberculosis in a city has to be using an IBM.

1.3.2 An IBM of TB calibrated for Ciutat Vella, Barcelona

An IBM which simulates the evolution of tuberculosis in Barcelona was developed by *Prats et al.*[8][12]. It was improved by *Vila* in order to account for gender, origin and age dependencies in the spreading of the disease [17]. The simulator was developed in the NetLogo platform, an individual-based programming lenguage with an integrated modelling environment that includes a user-friendly interface for simulations.

This particular model simulates the evolution of a population confined in a geographical space of variable size. The basic entities of this IBM are individual persons. Each one can be in one of five stages of the disease as it can be seen in Figure 1.5.

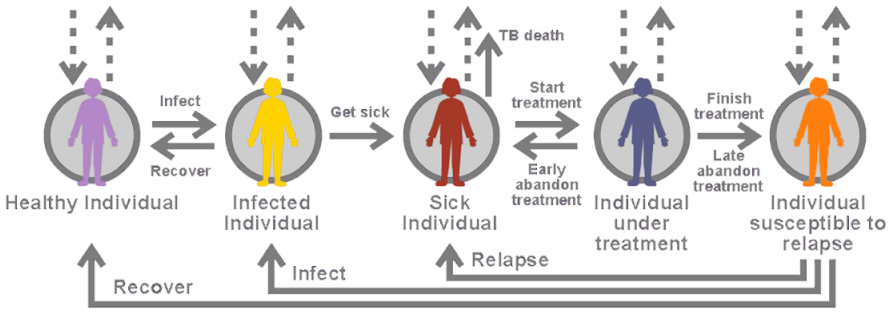

Figure 1.5: State diagram of the IBM model. Output grey dotted arrows refer to deaths.[8]

These agents can be born, grow, and die. They can move and can be infected or infect among other things. This model was calibrated and parameterized for the district of Ciutat Vella in Barcelona. A more detailed explanation of the model can be found in section 3.1.

Although the NetLogo simulator is working, it has some limitations: it requires a rather long computation time, it cannot simulate population of the order of the million and the space is considered in a way that does not account for realistic interactions between agents.

1.4 Aim of the work

This work is part of a project conducted by the Computational Biology and Complex Systems research group (BIOCOM-SC) and the inLAB FIB of the Universitat Politècnica de Catalunya (UPC).

The long term goal of the said project is to develop a counceling tool on tuberculosis control policies for public health agencies. That is, to develop a simulation model able to be parameterized for any city and to conduct simulated experiments in order to evaluate which is the best strategy to control the disease and how much this strategy would cost. This project is done in close collaboration with the Agència de Salut Pública de Barcelona (ASPB), the public health agency of Barcelona.

In particular, the objectives of the work conducted here are the following:

- To analyze real data to evaluate which are the most relevant interactions that need to be considered in the modelling of the tuberculosis dynamics.
- Translate the Netlogo current implementation in a programming language with a lower computational cost, both in time and memory.
- To implement the current TB simulation model with a programming language that allows parallelization and parallelize the code if needed.
- Improve the model towards the inclusion of real interactions between individuals.

Chapter 2

Data analysis

In the context of tuberculosis, and in particular in Barcelona, the amount of available data is remarkable. This opens the door to the possibility of studying in more detail which are the social factors that intervene more intensely in the spreading and prevalence of tuberculosis.

The study performed here tries to answer three questions that have no answer today or, at least, do not have a complete one (in the context of Barcelona):

1. **Which portion of the population is TB infected? What is their distribution?** This is one of the biggest unknown facts regarding tuberculosis as it is nearly impossible to control people who do not present any symptoms. There are estimations by the WHO and other organisms that claim that 1/3 of the worldwide population is infected [14]. Nevertheless, there is not conclusive data, specially for the Barcelona case.

2. **Which are the main features of the interactions between individuals that have to be considered to properly model the TB dynamics?** This is extremely important in order to develop a model that allows to study and test TB control policies.

3. **What is the main factor that leads to such differences in TB incidence between Ciutat Vella and the other districts?** The solution to this problem can lead to a better understanding of the disease and its social factors.

2.1 Data sources

Both in the data analysis and in the parameterization of parts of the model, real TB data and of the city of Barcelona has been used. This section accounts for all the sources used in the work.

2.1.1 Contact tracing study

ASPB has been doing a meticulous work in the control of the TB cases in Barcelona for many years [5]. Among other things, in the recent years they have been performing a contact tracing study of each case and they have yielded us the obtained data.

Methodology

The care of patients with tuberculosis is mainly conducted in the clinical units (CU) in the main hospitals of the city. The CU are coordinated with other centers such as primary attention centers.

When antibiotics are prescribed for an individual in order to treat tuberculosis, the patient is considered to be sick, hence a case. Then, the UC follow the patient and a number of the people that have a considerable degree of periodic relation with him/her, the so called contacts, are surveyed and tested for latent or active TB. These contacts are usually people that share household, family, people who work in the same workplace or kids attending the same school. People who go to the same places regularly are also considered as contacts.

Since it is not possible to know where and when the positive contacts (i.e. contacts with LTBI or active TB) were infected, the case is assumed to be the source of TB in their environment.

Survey dataset description

The data that has been analysed this work comprehends data of the TB cases of Barcelona between years 2005 and 2015 and of the contact tracing study performed between 2010 and 2015.

The former consists of 4346 cases. The relevant information for this study available for each case are age, gender, origin, year of diagnose, district where they live and different risk factors, which include HIV, smoking, alcoholism and diabetes.

The latter consists of 1285 cases and 10685 contacts. The useful data is age, gender, origin and risk factors for both cases and contacts. The year of diagnose, the district, the environment where the contact took place (i.e. family, school, workplace or leisure) and the intensity of the contact (i.e. time case and contact are together) are also available. Information whether the case has pulmonary TB or not is also present. There is also information about the state of the contact (i.e. healthy, infected or sick).

2.1.2 Demographic data of Barcelona

The data about Barcelona demographics of Barcelona, hence the characteristics of the population that lives in the city has been mined from the public database and statistics results of the Ajuntament de Barcelona [6].

In particular, the age distribution, the gender distribution, the nationalities and the household types and sizes have been obtained from this source.

2.1.3 TB world data

The public database of the World Health Organization (WHO) has also been used. In particular, the tuberculosis database which contains the yearly TB incidence for almost all the countries in the world [13].

2.2 Software

All the analysis has been performed using Python. In particular, the pandas package has been of great use. This package allows to migrate Excel files to Python and be treated as vectors or matrices, making really easy do computations with them and find relations or particular conditions.

2.3 Results

2.3.1 Infected individuals

Trying to determine the exact portion of TB infected people of Barcelona population would require conducting a complete study that would involve monitoring a large part of the population. An study of this magnitude would be expensive and hard to justify. Nonetheless, estimations can be done.

As previously said, the contact tracing dataset has information whether the cases have pulmonary TB or non-pulmonary TB. Notice that non-pulmonary TB is not infectious. That is, the contacts of non-pulmonary cases have not been infected by the sick individual. Hence, this cases can be considered as a representative sample of the population, although they may be biased taking into account that they are part of a social environment that do have a TB sick. Therefore, this sample can lead to the best estimations one can perform from real data.

In total, there were 1379 contacts on non-pulmonary TB sick cases from which 899 were healthy, 473 were infected and 7 were sick. This means that up to a **34**% of the population is infected of tuberculosis, close to the WHO estimation [14]. The age, gender and origins distributions of the contacts of non-pulmonary cases are presented in Figure 2.1.

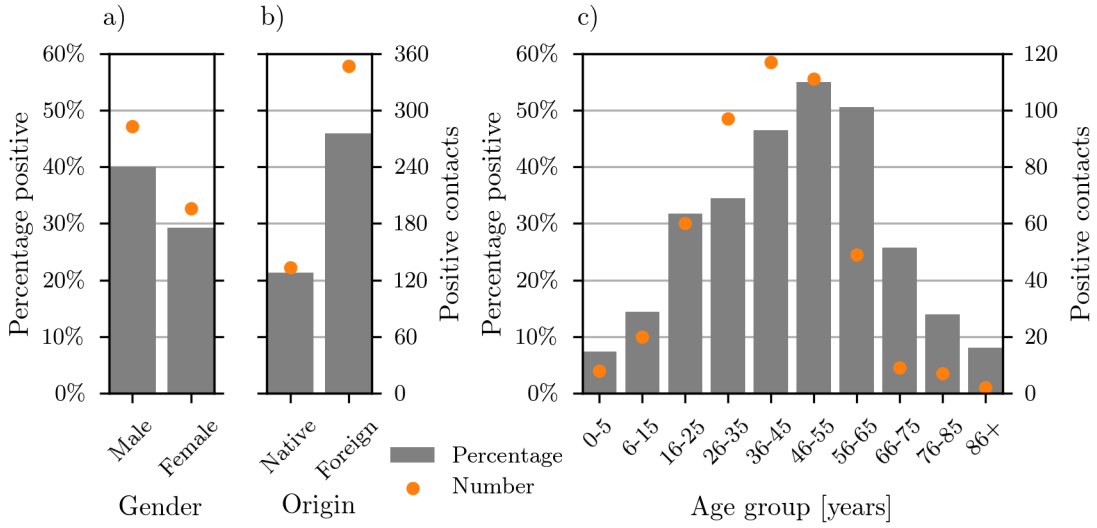

Figure 2.1: Information about the contacts of non-pulmonary TB cases. All values regarding positive contacts (i.e. infected or sick individuals). a) Gender distribution. b) Origin distribution. c) Age distribution

34% of infected people can be considered an upper bound because the sample is probably biased in some degree.

The percentage of men infected is higher of the women. This, rather than be a biological feature is probably a social one. The same can be said regarding the origin distribution.

Until the age of 55 years, the probability of being infected increases. This should be an expected results since the older one is, higher the probability of being exposed to TB infection. Nonetheless, the probability of being infected decreases with age for people older than 55 years. This can be attributed to two phenomena. The first one is that the number of contacts older than 55 years of non-pulmonary TB cases is scarce, thus the percentage presented for these ages might not be representative.

The second one might be related to immigration: foreign people is more likely to be infected but the international immigration in Spain is a relatively recent event and majorly conducted by relatively young individuals.

2.3.2 Social interactions

Determining which are the main social interactions that favour the tuberculosis spreading is limited to the data available. The information in the contact tracing study database that can be related to this matter consists of three types:

1. **Intensity:** The amount of time that the case and the contact spend close by. The study considered 4 categories: >6h/day, <6h/day, >6h/week and <6h/week.

2. **Environment:** The environment in which case and contact meet. The study considered family, workplace, school and leisure (i.e. others).

3. **Cohabitation:** Whether if case and contact share household. That is, they live in the same place.

An interesting plot that can be done using the environment data is presented in Figure 2.2. The result is only presented for Ciutat Vella since considering all the contacts leaded to an unreadable plot. In this plot, different interactions can be identified: workplace interactions are confined in a square between 20 and 65 years (the working age); kids have relation with two age groups: the parents and the grandparents; teenagers relate to others and teachers in the school and old people mainly relate to each other.

The general results obtained for the three aspects are presented in Figure 2.3.

Almost all the subgroups studied and presented in Figure 2.3 have at least over 250 individuals, which means that the statistic results obtained here can be considered as substantially valid.

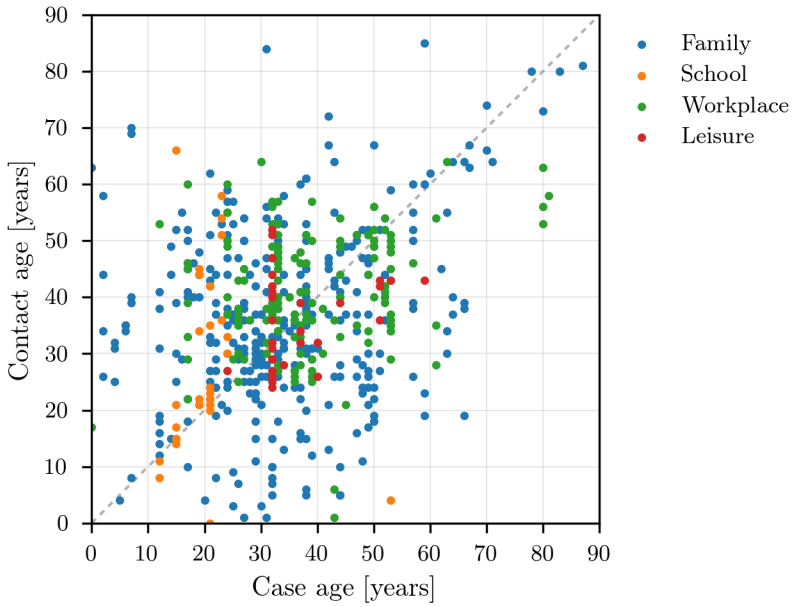

Figure 2.2: Age relation between case and contact sorted by environment. Ciutat Vella. 2010 - 2015.

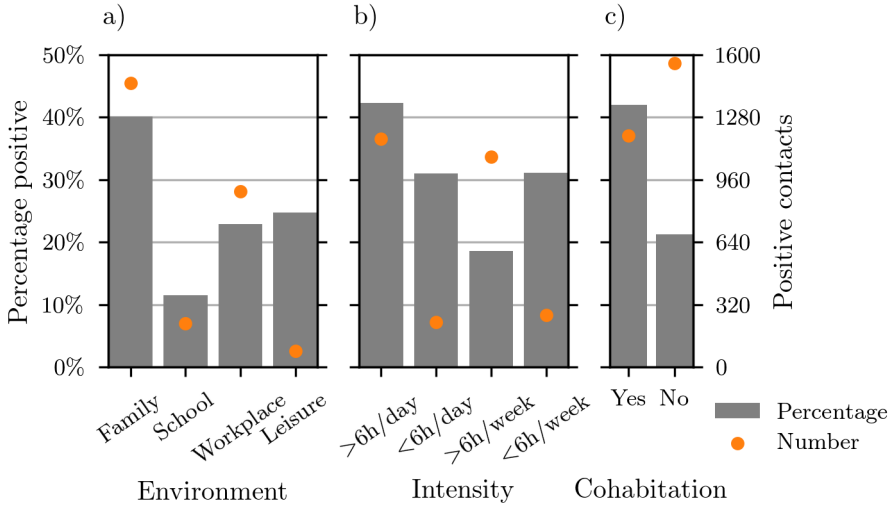

Figure 2.3: Social interaction results. a) Environment. b) Intensity. c) Cohabitation.

The family environment is the one with greater portion of positive cases. This can be attributed to the type of interactions: relations between family members tend to be closer than other types. Thus leading to a higher probability of becoming infected.

Although the more intense contacts are the ones that present higher portion of positive cases, there seems to be no clear relationship between contact intensity and probability of becoming infected.

The cohabitation condition is the one with more clear results: people who share household with a TB sick have a remarkably higher probability of becoming infected.

2.3.3 Ciutat Vella's oddness

Many efforts have been devoted to explaining and quantifying the singularity of Ciutat Vella's incidence from other districts data available but none worked. For instance, Figure 2.4 presents the relation between tuberculosis incidence and the mean surface per inhabitant by district and, as it can be observed, Ciutat Vella (CV) does not fit the rest of the districts. Neither it does work with the mean income per inhabitant by district, which turns out to have a certain correlation with the mean surface per inhabitant.

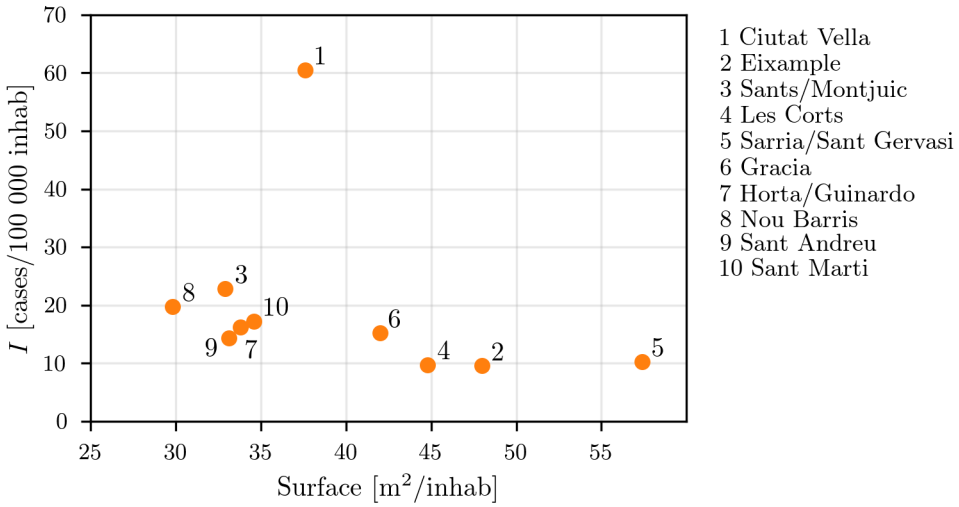

Figure 2.4: Relation between household surface by inhabitant and tuberculosis TB incidence by district. Barcelona. 2014.

Many works have suggested that immigration from countries with a high tuberculosis incidence can have a considerable impact in the total incidence of low burden

countries [1] [11]. In particular, *Ospina et al.* studied the epidemiology of tuberculosis in immigrants in Barcelona and pointed out that a significant number of people from countries with a high tuberculosis incidence do settle in the Ciutat Vella district [15].

This raises the hypothesis that the major factor that leads to the unusually high incidence in CV can be explained through immigration from countries with a high tuberculosis incidence. The aim of this section is to verify that hypothesis and quantify the impact of such factor.

Equivalent tuberculosis incidence by origin

In order to quantify theoretically the impact of the immigration in the overall tuberculosis incidence, the equivalent incidence by origin I_O is proposed.

$$I_O = \frac{\sum_{i=0}^{N} I_{O,i}\ n_i}{\sum_{i=0}^{N} n_i} \tag{2.1}$$

Where N represents the total number of different nationalities considered, $I_{O,i}$ stands for the tuberculosis incidence in the given country and n_i correspond to the number of inhabitants with nationality of the given country.

This magnitude represents the equivalent tuberculosis incidence that a certain population would present if all the people had the same incidence as if they were in the country of which they have nationality.

Figure 2.5 presents the evolution of I_O during the period $2010-2015$. As it can be appreciated, this magnitude decays slowly over time. This matches the expectations since it contains incidences from many countries and not all of them have the disease under control. The other remarkable feature is that Ciutat Vella present a much higher value than the rest of the districts.

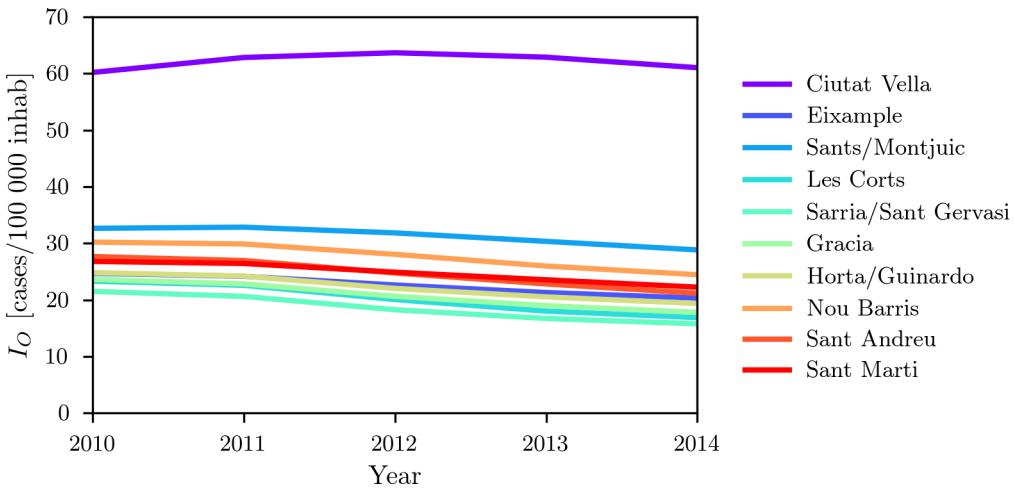

Figure 2.5: Equivalent incidence by origin by district between years 2010 and 2015 in cases per 100 000 inhabitants.

Country	**Inhab.**	**I_O contr.** [%]	**N**	%
Spain	1 342 096	43.63	145	51.8
Pakistan	19 432	14.21	31	11.1
Phillipines	8 530	7.44	3	1.1
Morocco	12 655	3.66	6	2.2
Bolivia	10 014	3.17	10	3.6
Others	211 089	27.87	85	30.2

Table 2.1: Number of inhabitants, contribution to the equivalent incidence by origin, number of cases detected N and percentage of the total cases detected % per nationality. Barcelona. Year 2015.

Moreover, one can evaluate the relation between the actual incidence and the equivalent incidence by origin by districts. As it can be seen in Figure 2.6, the relation between the normalized magnitudes is lineal.

Since each country considered in the magnitude has an associated value $\frac{I_{O,i}\ n_i}{N} \Big/ \sum_{i=0} n_i$, one can evaluate the contribution of a certain nationality to the final value of equivalent incidence and compare it to the real incidence values. The results are presented in Table 2.1.

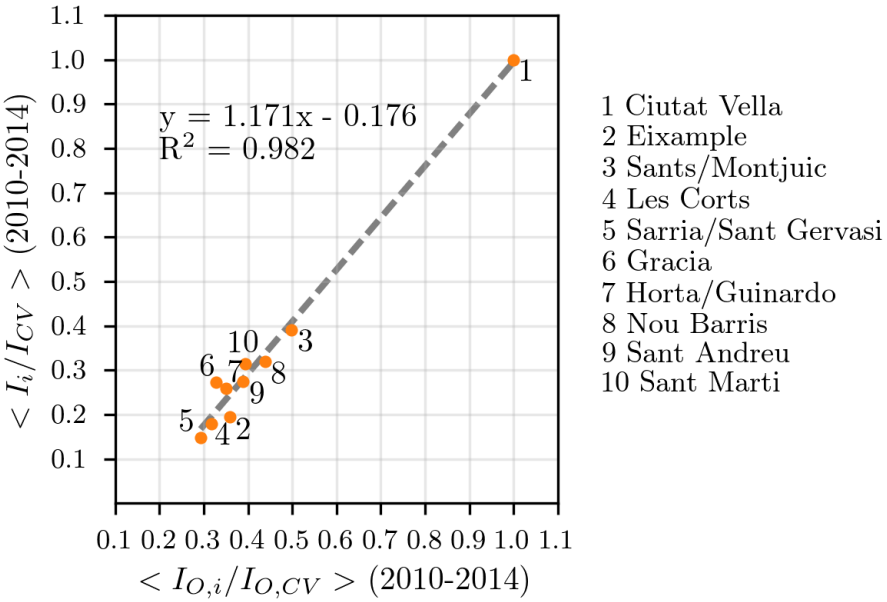

Figure 2.6: Relation between the mean value of the equivalent incidence by origin normalized by its highest value each year and the mean value of the actual incidence normalized by its highest value each year. Barcelona. Years 2010 - 2014.

Country	**Inhab.**	**I_O contr.** [%]	**N**	%
Pakistan	6 648	30.32	10	17.85
Phillipines	4 566	24.83	2	3.87
Spain	57 507	11.65	17	30.35
Bangladesh	2 550	9.69	8	14.28
India	1 350	4.94	6	10.71
Others	27 447	18.54	13	23.21

Table 2.2: Number of inhabitants, contribution to the equivalent incidence by origin, number of cases detected N and percentage of the total cases detected % per nationality. Ciutat Vella. Year 2015.

The same can be done with the Ciutat Vella district. The results are presented in Table 2.2.

The correspondence between the equivalent incidence by origin and the actual incidence percentages is better with a greater sample. This should be expected since the numbers for Ciutat Vella are too low to allow proper statistical results.

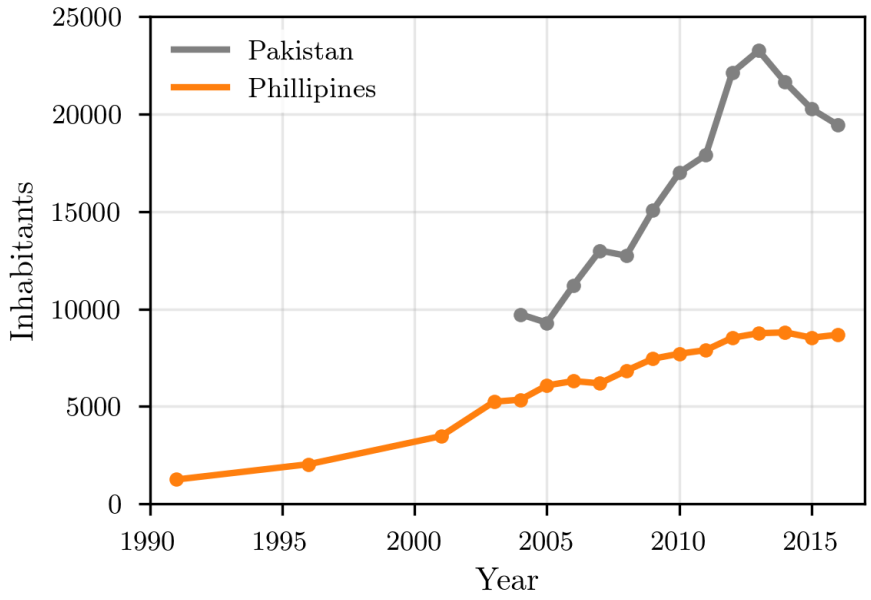

Figure 2.7: Population from Pakistan and Phillipines living in Barcelona. Prior to 1991 there is no available data of this matter and information about people from Pakistan is only present since 2004. [6]

Immigrants from Phillipines would be expected to sicken more than they do in contrast with the people from Pakistan who do get sick representing a percentage close to the expected, at least at the city level. One hypothesis to explain this is that the peak of immigration from Phillipines was more than 7 years before 2015 (time in which an infected individual is more likely to develop an active disease as explained in 3.1) while the peak from Pakistan has been during the past 7 years, thus the people that came from their country infected is still getting sick. Figure 2.7 presents the data related to the people from Pakistan and Phillipines who lives in the city of Barcelona.

2.4 Conclusions

From this study, three main conclusions can be drawn:

1. A considerable portion of the population of Barcelona is currently infected.

2. The foreign individuals that come from countries with relatively high TB incidences do have an impact in the TB incidence of the city. They modify the infected population profile and thus, not only their presence but also their arrival time should be considered when modelling TB dynamics.

3. In the modelling of TB dynamics within a city, space is of vital importance. The space where individuals are has to make sense. Households represent the main controlled place where TB infections take place, thus they must be considered in a realistic way. Workplaces and schools also have a certain impact and should also be taken into account.

Chapter 3

C Simulator

The current NetLogo implementation of the IBM model developed presents important limitations regarding the computation time and the increase of scale (i.e. the simulation of larger systems). Moreover, it does not offer the chance of parallelization and this is a problem since the computation time required in the NetLogo implementation suggests that this might be needed.

In order to meet all the requirements specified in Section 1.4, the C programming language is proposed to develop the new simulator. It is a middle level language which offers a remarkably faster performance and a lighter memory load. Moreover, the BIOCOM-SC in collaboration with the inLAB of the UPC has access to a parallelization library in C.

In this chapter the TB dynamics model will be introduced and explained, along with the particularities of the C simulator developed and its validation.

3.1 ODD description of the model

Individual-based Modelling has been used in multiple fields, from simulation in social science to business, and also biology and epidemiology. As a consequence of this variety, IBM models have had almost as many description procedures as models

have been developed. In order to deal with all the different description systems and standardize the description of IBM models the ODD (Overview, Design concepts, and Details) protocol was defined [9]. This protocol consists of three blocks, which are subdivided into seven elements: Purpose, Entities, State variables and scales, Process overview and scheduling, Design concepts, Initialization, Input data and Submodels; although not all of the elements have to be used in the description.

In this section, the ODD protocol that was developed for the original model by *Gilabert et al.* [8] and updated by *Vila et al.* [17] will be presented.

3.1.1 Overview

Purpose The objective of this IBM is to analyse the evolution of pulmonary tuberculosis incidence in a community. It is fitted to the Ciutat Vella neighbourhood, considering the population to be constant. The adequacy of the simulation results will be checked and compared to the epidemiological data. The final purpose of this IBM, which could be met in a further project is to check through virtual experiments the possible effects of epidemiology control strategies and public health decisions.

Entities, state variables and scales The fundamental entities in the model are persons. It is considered that persons can go through five infection states: healthy, infected (i.e., with a latent TB infection), sick (i.e., with an active TB), under treatment, and recovered. Persons in four out of the five states, all but healthy, are simulated as individuals. We consider that the characteristics of a healthy population remain constant (e.g. native/immigrant distribution and HIV+ percentage, among others). Moreover, a healthy collective is much larger than infected or sick collectives. Therefore, it is not necessary to control healthy individuals one-by-one; they are considered as a property of space (i.e., the number of healthy people in a spatial cell). A healthy person will

acquire individuality once he/she enters the infection cycle. This strategy is an important optimization for drastically reducing the computing time. It was previously tested to provide results comparable to those obtained considering healthy people as individuals [12]

The state variables of the individuals mainly refer to their status in the tuberculosis infection cycle as well as the time spent in such phases and individual diagnostic time when getting sick. Other individual state variables and parameters are age, native/immigrant origin, female/male gender, risk factors (e.g. smoking), diabetes, and possible immunosuppression (mainly HIV infection). Once a person gets infected, the presence (or not) of pulmonary cavitation is also considered. A state diagram of the model is presented in Figure 1.5. The population simulated is 105,123 people, which represents all the people of Ciutat Vella (2012).

The model is partially spatially explicit, i.e., space is considered but it does not mimic the real space of Ciutat Vella. Simulation occurs in a discrete area of 501 x 501 spatial cells. Each spatial cell represents a local abstract space where two persons can meet. The time step is set to 1 day, and the simulation may cover up to a period of 1 or more years.

Process overview and scheduling The model was build in C. The simulation starts with the set-up of the initial configuration, where the population is randomly generated according to the input distributions of parameters and randomly distributed in the 500 x 500 grid. The model assumes discrete time steps of 1 day, as mentioned. Each day, all individuals execute a series of actions, and their variables are updated immediately.

The individual actions may be: to age, to move, to get infected, to get sick, to be diagnosed and to start a treatment, to abandon or to finish the treatment,

to recover, and to die. Some of the actions take place daily for all the individuals in the system (e.g., aging and movement) and the other procedures are daily evaluated when necessary (e.g., the possibility of a sick individual to be diagnosed is daily assessed until it finally occurs).

When an individual dies, a new healthy is generated adding 1 to the number of healthy people in a random spatial cell. At the end of each time step, global variables are updated.

3.1.2 Design concepts

Basic principles The model is based on general knowledge of the natural history of tuberculosis. There are two essential characteristics of TB that must be taken into account in any epidemiological model. On the one hand, an infected individual does not necessarily develop an active disease; on average, only 10 % of infected people become sick. Moreover, a person remains infected for an extended period and may develop active tuberculosis after several years, but the probability of developing the disease decreases with time [7][2]. Infected people are usually not diagnosed, except those detected in punctual screenings or in a contact tracing study. On the other hand, only TB sick can disseminate the infection. The infection rate increases if the patient has TB with cavitation. Once a TB sick is diagnosed, the pharmaceutical treatment takes 6 months [14]. Once the treatment is finished, the possibility of getting sick again because of a TB reactivation remains at 1 % for 2 years. If an infected is detected, it may be treated with preventive drug therapy. This treatment is longer than the one given to persons with active TB. It lasts 9 months and is administered to infected individuals to prevent the development of an active disease once they have been detected during a screening process [14]. There is also a probability of relapse to the infected state that is calculated similarly to the first treatment.

Emergence Emerging phenomena are mainly related to long-term dynamics of the infection at the population level. On the one hand, only non-treated people with active TB can spread the disease. Therefore, diagnosis time is an essential parameter for the prevalence of the disease. On the other hand, infected persons may develop active tuberculosis a few years after the infection. Therefore, global consequences of particular conditions at a precise moment may be detected some years later.

Interaction Local interactions between healthy and ill individuals are explicitly modelled and crucial for the dynamics of the system. They refer to the meeting of two persons favored by the spatial proximity between them and the possibility that one of those individuals with an active TB may infect the other person. The interaction within or between some specific groups (i.e., among immigrant male adults) is of special interest in the context of Ciutat Vella, since it accounts for specific observed patterns of transmission and its consequences on the particular TB dynamics in this neighborhood.

Stochasticity Randomness is introduced at all levels of the simulation. The initial distribution of individual properties is randomly executed according to input distributions. Movement is assumed to be random. Each action is associated with a certain probability and thus executed according to a stochastic number.

Collectives Different collectives may be distinguished, according to the individuals origin (native and immigrant), due to the individuals gender (female and male) and due to the individuals age (by age group). The difference between native and immigrant is the diagnosis time, and depending on the age and gender the probability of sickening will also differ. Due to the social patterns, it is considered that an individual of a specific collective (i.e. 5-year-old native female)

will be more likely to infect another individual from specific characteristics (i.e. 35-year-old native female more likely than 60-year-old immigrant male).

Observation Output data show the yearly evolution of number (or prevalence) of healthy people, infected people, sick people, people under treatment, and persons already treated. It also shows the number of cases per year (or incidence).

3.1.3 Details

Initialization For this particular study, most of the input parameters were taken from official reports. The initial population was fixed at 105,123 individuals. All percentages shown in Table 3.1 were used for calculating the configuration of initial population: rates sick, under treatment and recovered individuals per 100,000 inhabitants; mean diagnosis delay (MDD); mean treatment abandon rate; individuals with risk factors and with HIV infection. Other initial variables are assigned randomly: individuals age, gender and origin (following the percentages shown in Table 3.1), and time spent in the infection state assigned.

Submodels

Age all individuals increase their age by 1 day each time step.

Move all persons can move randomly through the surrounding local space, once a day.

Get infected if there is a number of individuals susceptible to TB (healthy and treated) different from zero in the proximity of a sick individual, meaning one of the 4-neighbouring spatial cells, this sick person may infect one of them with a certain probability. The total of susceptible neighboring individuals is computed and then the infection process is repeated as many times as healthy and treated people have been found. The infection probability depends on the type of TB

District of Ciutat Vella	**Value**	**Units**
Total population	105123	persons
For the infected population **:		
Immigrant	73.57	%
Male	68.02	%
Age distribution:		
0-4	2.20	%
5-14	5.58	%
15-24	12.01	%
25-34	25.21	%
35-44	27.75	%
45-54	18.27	%
55-64	5.75	%
65-74	1.86	%
75-84	1.18	%
85+	0.17	%
For the sick population **:		
Immigrant	71.57	%
Male	69.78	%
Age distribution:		
0-4	2.70	%
5-14	3.15	%
15-24	13.93	%
25-34	30.00	%
35-44	22.25	%
45-54	11.69	%
55-64	7.19	%
65-74	3.82	%
75-84	3.93	%
85+	1.35	%
Total anual mortality	0.83	%
Cavitation forms*	22	%
Diagnosis delay native (median)	42	days
Diagnosis delay foreign (median)	33	days
Diagnosis delay (std)	4	days
Treatment abandon rate*	2.20	%
Cases of TB/HIV+	32	persons
Risk factors*	24.1	%
Diabetes cases	5.60	%

Table 3.1: Official data of Ciutat Vella used in simulations [17]. (*) Percentages with respect to the total number of TB sick people. (**) Percentages with respect to the total number of TB infected people.

		Age of the cases			
		0-20	20-60 male	20-60 female	60-90
Age of the contacts	0-20	30 %	5 %	12 %	5 %
	20-60	65 %	94 %	85 %	45 %
	60-90	5 %	1 %	3 %	50 %

Table 3.2: Distribution of the new-infected age according to the age of the infectious [17].

disease that the sick person has, either smear-positive or smear-negative. A smear-positive case is considered to double the infection probability. The value of the infection probability is closely linked to the spatial and temporal scales, i.e., the probability of infection is inseparable from the spatio-temporal scale. A change in any of these scales entails the revision of its value. Therefore, it is not a real infection probability when a sick individual meets a healthy person, but an effective infection probability given the particular spatio-temporal constraints. In this case (501 x 501 spatial cells and 105123 individuals), the value of this probability was fixed at 49.7 %. Once a person is infected, a newly infected individual is created with the properties assigned according to the characteristics of the infectious. Whether the new person will be set to native or immigrant, male or female, and the age will depend on the characteristics of the infectious individual. Those probabilities are summarized in Tables 3.2 and 3.3. The infection time of the new individual is set at 0 and starts increasing with each time step.

Get sick once infected, the individual may develop active TB according to a particular annual probability that decreases with infection time during the 7 years post-infection [7][2]. It is neglected for the subsequent years (t > 7 years). The probability of developing active TB will also depend on the age, gender and origin of the individual. For children aged 0-15, a factor multiplies the probability.

		Gender and origin of the cases			
		Foreign female	Foreign male	Native female	Native male
Gender and origin of the contacts	Foreign female	37 %	21 %	16 %	13 %
	Foreign male	49 %	65 %	16 %	19 %
	Native female	6 %	3 %	33 %	29 %
	Native male	8 %	11 %	35 %	40 %

Table 3.3: Distribution of the new-infected properties according to the characteristics of the infectious [17].

Since simulation time does not cover periods longer than 10 years, the approximation is good enough. For immunodeficient people, a certain factor multiplies this probability. The same happens if there are other risk factors (smoking, alcoholism) or if the patient has diabetes. The chance of becoming a TB sick individual is evaluated at each time step for all infected persons. Globally, the average of 10 % of infected developing an active disease is satisfied. The possibility of relapse (getting sick again) for recovered patients is also evaluated daily according to the individual relapse probability (see below). Once a person gets sick, the disease time counter starts running until the individual diagnostic time is reached.

Be diagnosed and start treatment each individual has a particular diagnostic time that is randomly assigned when getting sick. These individual times are assumed to be distributed following a normal distribution centered around the mean diagnosis time and standard deviation shown in 3.1. When the sick time counter reaches these values, the individual is diagnosed. Once diagnosed, medical treatment is assumed to start and TB to stop spreading. Individual time under treatment is initially fixed at 0 and then updated at each time step.

Abandon the treatment there is a certain probability that an individual abandons the treatment before finishing it. This possibility is evaluated daily for

each patient under treatment, according to the input abandonment probability. If a person leaves treatment during the initial 15 days post-diagnosis, he/she becomes ill again. If he/she abandons the treatment after 15 to 180 days post-diagnosis, the model will consider him/her to be recovered but with a certain probability of relapse during the following 2 years. This probability is assumed to decrease linearly from the 100 % of a 15-day abandonment to the 1 % of the 180-day treatment period.

Recover when a sick individual is diagnosed and treated for 180 days, he/she becomes recovered and a relapse probability of 1 % is assigned (the chance of getting sick before being considered healthy). 2 years after the diagnosis, the individual is considered to be healthy

Die each individual has a certain probability of dying according to his/her age. These probabilities are fixed using demographic data from Ciutat Vella in 2012. Accordingly, the daily dying probabilities are considered to be $6.88\text{x}10^{-5}$ % for individuals under 10, $5.45\text{x}10^{-4}$ % for individuals between 10 and 65, and $1.22\text{x}10^{-2}$ % for individuals over 65, which is a simplification of the real mortality distribution. Furthermore, TB sick people have a distinct probability of dying from tuberculosis. This probability is evaluated daily for each sick individual, taking into account that 40 % of non-treated TB sick may die in 5 years. Each time an individual dies, a new healthy individual is introduced into the simulation world in a random position with the aim of maintaining a constant population.

3.2 C implementation

In this section, the particularities of the C implementation will be described. The reader is assumed to be familiar with the C language. For further information, refer to the source code.

3.2.1 Details

Space

The space is considered as a 2 D grid (represented by a 2D array). Although each individual can be in one of the points of the grid, only the healthy ones are considered as values in the array (i.e. a 3 in the array means that 3 healthy individual are at that particular point). For the rest of the states, the position is one property of the individual (further explained in the following section).

Individuals

All individuals but healthy are considered separately. That is, the individuals of each state are represented by different structures which contain different properties. This structures are implemented with *struct*'s. The common properties for the not healthy individuals are age (in days), gender (male or female), origin (native or foreign) and position in the grid (i.e. an x and a y value). The particular properties of each state are as follow:

Infected time infected (in days), HIV (positive or negative), diabetes (positive or negative) and smoking (yes or no) which includes can also be interpreted as the risk factors.

Sick time sick (in days), diagnosis time (in days) and smear or cavitation (positive or negative).

Treatment Time under treatment (in days) and smear or cavitation.

Treated Time spent under treatment (in days), time since finishing the treatment (in days), smear or cavitation and probability to relapse before being considered healthy.

3.2.2 Files

Since this is a complex simulator, the source code has been split in different files. This allows to classify the code according to its topic and thus, it is easier to debug, test and maintain the code. The different files are explained below:

IBMheader.h This header file contains all declarations of functions and global variables of the simulator.

IBMparams.c The file contains all global variables and parameters used in the simulator. This file is particularly useful because allow to change easily any parameter without the need to modify the source code. Although some of the parameters has been introduced in previous tables, the others can be found in this file (for instance, the infection probability). Notice the parameters found in it are the same as used in the Netlogo model which is based on.

IBMsetup.c Contains the definition of the functions needed for the initialization of a simulation. That is, the functions that generate the individuals in each state according to the initial distributions. It also initializes and generates the structures where the individuals are embedded in.

IBMdynamics.c Contains the definition of the functions needed for the evolution of the simulation. That is, the functions that represent the dynamics of the model such as infection, movement, diagnosis, etc.

main.c Main file of the program. This contains the function main and the flow of the simulation, it is where the functions in the two previous files are called.

random_functions.c Contains the functions related with (pseudo) random numbers, from the generation of a pseudorandom number according to a certain distribution to the choice of an element between a collection with different probabilities.

list.h Header file that contains the declaration of functions of the file list.c.

list.c Contains the definition of the functions of the double linked lists used. This implementation is independent of the IBM model and could be used in other programs.

Therefore, assuming that compiles with GCC (a commonly used C compiler), the line needed to compile the program from these files would be:

```
gcc main.c IBMparams.c IBMsetup.c IBMdynamics.c random_functions.c
list.c
```

3.3 Evaluation

Once the simulator has been developed, it has to be tested in order to see if it does present the same dynamics as the previous simulator did and mimics the real behaviour of the system as well as which improvements, if any, present with respect to the previous simulator.

3.3.1 Validation

In order to evaluate the dynamics of the system, simulations of 10 years have been performed starting with the parameters shown in Table 3.4. Figure 3.1 presents the results of 10 simulations, their mean and the real data between 2005 and 2015 (the initial parameters were chosen in previous works to reproduce the behaviour within these years [8][17]).

One can observe that one simulation presents a noise level similar to the real data and that the mean between them do follow the same tendency as the real that. In

State	Number
Healthy	100488
Infected	4500
Sick	3
Under treatment	11
Treated	121

Table 3.4: Initial number of individuals in each state for the validation simulations [17].

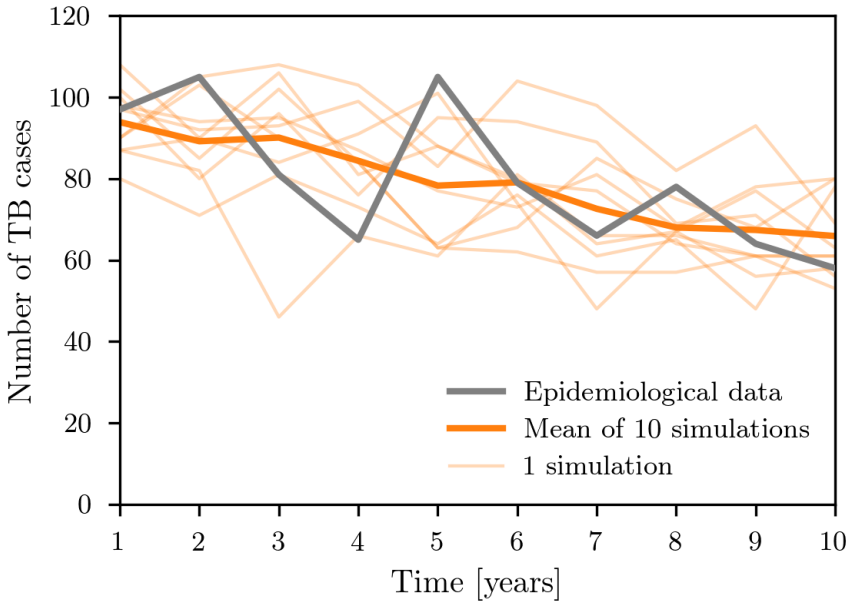

Figure 3.1: Results of ten simulations compared to the real TB incidence of Ciutat Vella between years 2005 and 2015.

conclusion, it can be said that the implementation of the model in the C simulator has been successful.

3.3.2 Performance

The performance of the simulator will be evaluated taking into account the two factors previously mentioned: computation time and scalability. To test these properties, simulations of 1 year have been performed in each simulator with a different number of individuals, corresponding to a district, the size of Barcelona (medium city) and

the size of Bombay (big city). The proportion between states is the same as the one shown in Figure 3.1. The space is enhanced accordingly by multiplying the longitude of the grid by $\sqrt{N_{new}/N_{old}}$ where N is the number of individuals.

The computer in which the simulations have been ran is a MacBook from early 2015 with OS X El Capitan (v10.11.3). It has a processor Intel Core M of 1.1 GHz and a memory of 8 GB 1600 MHz DDR3. The results obtained are presented in Table 3.5.

Individuals	**Simulator**	**Computation time**
105 123	Netlogo C	198 s 6 s
1 600 000*	Netlogo C	30 min 80 s
1 600 000	Netlogo C	- 84 s
12 500 000	Netlogo C	- 12 min 39 s

Table 3.5: Computation time for each simulator with different loads of individuals. Blank spaces corresponds to sizes with which NetLogo simply crashes and cannot reach the end. (*) Same space size as for 105 123 individuals.

3.4 Conclusion

From the evaluation of the C simulator, the following can be concluded:

- The new version of the simulator can reproduce the real data at the same level of the Netlogo simulator. This verification confirms that the same ODD model is reproduced.

- The simulator in C does present a lower computational cost. It can simulate properly populations as much as 100 times bigger or more than Netlogo in a reasonable time and perform smaller simulations more than 30 times faster.

- From the output of bigger simulations, it can be affirmed that there is no need to parallelize the code, at least for the Barcelona case. One year with the population of Barcelona takes slightly more than a minute to be simulated in a not so powerful computer, which is really affordable. However, the parallelization could be needed for higher scenarios.

Chapter 4

Towards an improved model

The results of Chapter 2 have shown the need to consider explicitly the social meeting points to properly model and study the dynamics of tuberculosis within a city as well as the effect of specific control strategies. Thus, this is the next step towards the final goal of the research project.

This chapter focuses on setting up realistic households, which have proved to be the most important factor, that will allow future works to build up the model considering these elements.

4.1 Households

In order to generate the households, an heuristic model based on previously published models (e.g. [10]) that matches marginal distributions of household size and population age structure, and maintains realistic generational age gaps within households, respecting as best as possible the actual mix of age groups. The population's age distribution and the household distribution has been obtained from [6]. A sketch of the heuristic model employed is shown below:

	Household type	Proportion
1	Single person households	29.0%
2	Married couples without children	22.4%
3	Married couples without children under 25 years old (*)	6.8%
4	Married couples with children under 25 years old (*)	21.2%
5	Singles without children under 25 years old (*)	6.2%
6	Singles with children under 25 years old (*)	5.8%
7	Other households	8.6%

Table 4.1: Distribution of household types in Barcelona, 2011 [3]. (*) indicates the possible presence of an aggregated elderly member.

Household size	1	2	3	4	5	6
Proportion	29.0%	33.6%	18.7%	13.6%	3.8%	1.4%

Table 4.2: Distribution of household sizes in Barcelona, 2011 [3].

1. determine the household type by choosing among seven household types from the proper frequency distribution (see Table 4.1).

2. determine the household size by sampling from the proper frequency distribution (see Table 4.2).

3. assign an age to the household head, a_h, by sampling from the population's age distribution and according to type-specific constraints, detailed below.

4. assign an age to the additional members by sampling from the population age structure and taking into account general constraints, detailed hereafter, as well as type specific ones.

General constraints

These constraints are applied throughout the initialization of households.

- All families are composed by a household head, at most one partner, at most one aggregated elderly person (assumed to be a parent of the household head or of the partner), and at most four children.
- The age of the household head must fulfill $a_h \geq 20$.
- The age of the partner, a_p satisfies $a_h - 8 \leq a_p \leq a_h$ and $a_p \geq 20$; thus, by partner it is indicated the younger member of the couple.
- The age of children, a_c, satisfies $a_h - 45 \leq a_c \leq a_p - 16$ if there is a couple in the house, and $a_h - 45 \leq a_c \leq a_h - 16$ otherwise.
- The age of the aggregated elderly, a_e, satisfies $a_p + 16 \leq a_e \leq a_h + 45$ if there is a couple in the house, and $a_h + 16 \leq a_e \leq a_h + 45$ otherwise.

Partners are present only in households of type 2, 3 and 4 (see Table 4.1. Aggregated elderly are present only in households of type 3, 4, 5, 6, 7. The probability of having an aggregated elderly was estimated from census data [3] as about 6.84 %.

Type-specific constraints

- Households of type 3 represent one of the following two subcases:

  a) married couples without children an aggregated elderly person (size 3).

  b) married couples with one or more children at least 25 years old (size ≥ 3) and possibly an aggregated elderly.

  Therefore:

  – if size is > 3 (a subset of case (c)), at least a child is present; therefore, the age of the household head and partner are forced to be ≥ 41 since otherwise they would be too young to have a child of age ≥ 25.

- if the size is 3 and the partner is < 41 years old, the couple is too young to have children of age ≥ 25; therefore, the presence of an elderly aggregated is forced.
- the remaining individuals are forced to be children, with $25 \leq a_c \leq a_p - 16$.

- Households of type 4:
  - the age of the household head is limited to be < 71 since otherwise the couple would be too old to have a child of age < 25.
  - the first other individual besides the couple is a child of age < 25.
  - other children are assigned with age ranges compatible with parents' ages according to general constraints.

- House of type 5 represent one of the following two subtypes:

  a) singles without children with an aggregated elderly (size 2).

  b) singles with one or more children at least 25 years old (size ≥ 2) and possibily an aggregated elderly.

  Therefore:
  - if size is > 2 (a subset of case (b)), at least a child is present; therefore, the age of the household head is forced to be ≥ 41 since otherwise he/she would be too young to have a child of age ≥ 25.
  - if size is 2 and the household head is < 41 years old, the household head is too young to have children of age ≥ 25; therefore, the presence of an elderly aggregate is forced.
  - the remaining individuals are forced to be children, with $25 \leq a_c \leq a_p - 16$.

- Households of type 6:

- the age of the household head is limited to be < 71 since otherwise he/she would be too old to have a child of age < 25.
- the first other individual besides the household head is a child of age < 25.
- other children are assigned with age ranges compatible with parents' ages according to general constraints

- Households of type 6: all members of households of type 6 other than the household head are not labeled as partners, aggregated elderly or children; thus, they are not subject to age constraints.

4.1.1 Validation of household generation

Once the model described before is implemented, validation of its correctly operation is necessary. To do so, the simulated age distribution is compared to the real one used and the distributions of household sizes are also compared. Figures 4.2 and **??** present the obtained results.

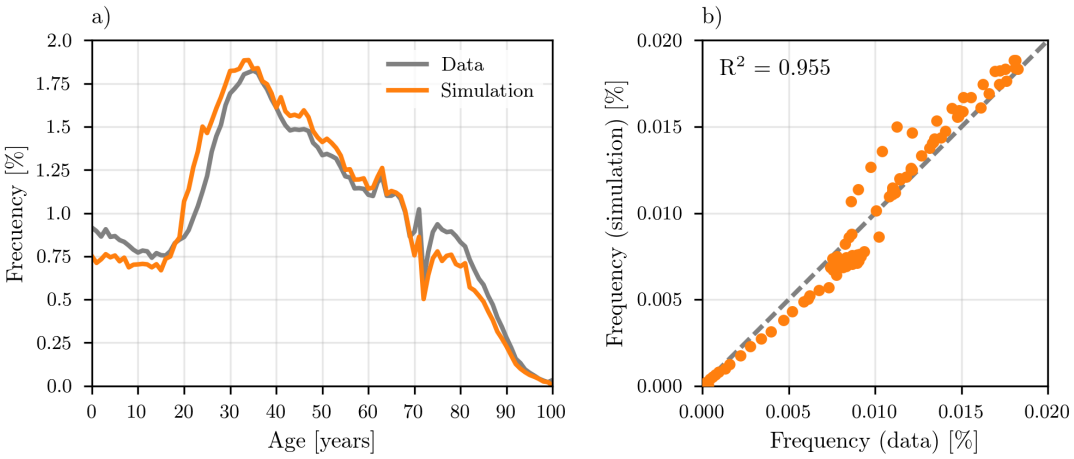

Figure 4.1: Real age distribution of the population of Ciutat Vella, 2011 [3] and distribution obtained after the generation of population through the heuristic model of households.

Although the results do not mimic perfectly the real distributions, they are substantially close to the reality. The results of the age distribution could be improved

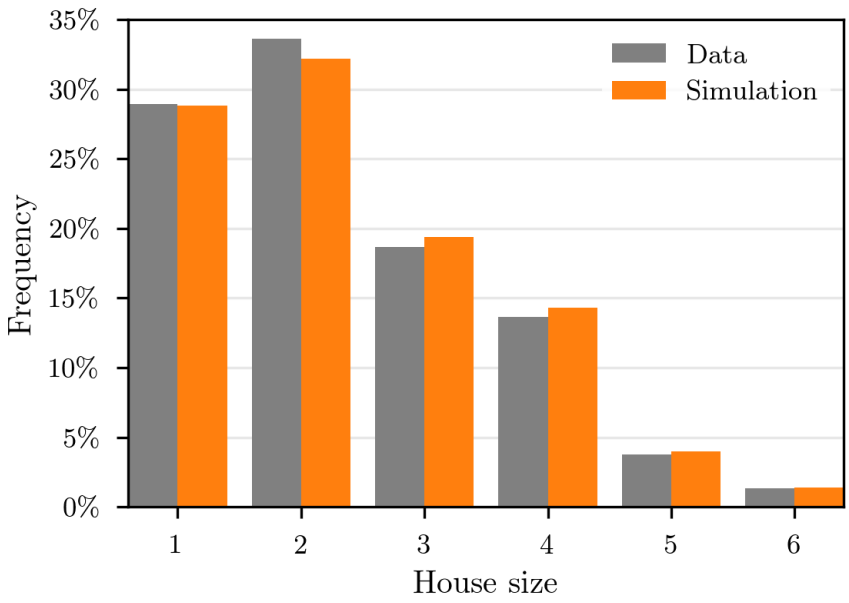

Figure 4.2: Real distribution of household sizes in Barcelona, 2011 [3] and distribution obtained after the generation of the households through the heuristic model.

through iteration, that is, letting the model evolve with population dynamics (e.g. births, deaths, etc) and it will eventually reach the desired distribution. Nonetheless, the goal is achieved: the households are generated realistically and the members within them present feasible familiar structures.

Chapter 5

Conclusion

In this work, a profound data analysis regarding the TB situation in Barcelona has been performed both by analyzing the results of the contact tracing study of the ASPB and exploring the demographic public data of the city. This led to some important conclusions on the subject:

- As estimated by the WHO, a remarkable portion of the population result infected of TB at some point of their lives.
- The immigration from countries stricken by high TB incidence does impact the TB incidence of Barcelona not only in their arrival but also after some years, thus it is important to consider immigrant people and their arrival time when modelling TB dynamics. This suggest that evaluate whether recently arrived immigrants from a high incidence country are infected or not may be an adequate measure with a relatively low cost to improve the TB control.
- Space has to be considered explicitly in the form of social meeting points (e.g. households, schools, workplaces) in order to model the social interactions leading to TB spreading and to be able to test and evaluate TB control policies.

Moreover, a new implementation of the model has been developed in the form of a simulator written in C. This program fulfilled the requirements defined in Section 1.4 and it showed a clear improvement both in scalability and computation time compared to the NetLogo version. Additionally, it has been designed and developed to allow some studies that were necessary such as the evaluation of the impact in the TB incidence of a certain influx of immigrants at a given time.

As the final part of the work and considering the results obtained in Chapter 2, a new simulator has been built to account for the social interactions in the form of social meeting points. As a result of the data analysis, it has been decided to include households representation in the model as a priority. This enhancement allow to realistically represent the distribution of the population in familiar spots. As seen in Section 4.1.1, this objective has been accomplished by means of an heuristic model such households have been generated.

In conclusion, all the objectives defined have been successfully fulfilled. Additionally, it is also a result of this work to avoid the parallelization of the model implementation since the current version is fast enough for a city level.

5.1 Future Work

Future work should include the enhancement of the simulator started here. To do so, schools and workplaces should be modeled too as important places to favour TB spread.The new features will need the adaptation of TB dynamics to the new environment, taking into account that infection will not happen through neighbour cells but in households or the other environments. Moreover, one has to bear in mind that infections within households, schools and workplaces are almost the only ones that can be detected. There exist some kind of casual infections, otherwise the disease will be already controlled. Therefore, a certain probability of infection due to TB sick

in the neighbourhood has to be considered. To deal with these spatial spread points, it may be necessary to give some type of coordinates to the social meeting points generated with the goal to consider things such as buildings with many households or locality at the level of the same street where individuals may have a higher chance to meet.

Beyond that, the simulator will have to be parameterized in order to reproduce the behaviour observed in Barcelona (ideally, with the proper inputs the program should be able to reproduce the behaviour of any city). Only this step can give some insight in the TB dynamics, since some parameters such as the probability to be infected if someone is sick in the same household will be quantified.

The final step of this project will be to implement different TB control strategies in the simulator in order to test them and evaluate their impact in the TB incidence. The control strategies to test will need to be agreed with ASPB or other agencies interested in TB dynamics. That way, a tool able to assess public health agencies on TB control policies would be developed and, hopefully, will contribute to minimize the tuberculosis incidence worldwide.

Bibliography

[1] Robert W Aldridge, Dominik Zenner, Peter J White, Morris C Muzyamba, Miranda Loutet, Poonam Dhavan, Davide Mosca, Andrew C Hayward, and Ibrahim Abubakar. Prevalence of and risk factors for active tuberculosis in migrants screened before entry to the uk: a population-based cross-sectional study. *The Lancet Infectious Diseases*, 16(8):962–970, 2016.

[2] PJ Cardona and J Ruiz-Manzano. On the nature of mycobacterium tuberculosis-latent bacilli. *European Respiratory Journal*, 24(6):1044–1051, 2004.

[3] Ajuntament de Barcelona. Anuari estadístic de la ciutat de barcelona. 2011.

[4] Ajuntament de Barcelona. Anuari estadístic de la ciutat de barcelona. 2016.

[5] Agència de Salut Pública de Barcelona. La tuberculosi a barcelona. 2015.

[6] Departament d'Estadística de Barcelona. Barcelona statistics, accessed June 19, 2017. `http://www.bcn.cat/estadistica/angles/index.htm`.

[7] SH Ferebee. Controlled chemoprophylaxis trials in tuberculosis. a general review. *Bibliotheca tuberculosea*, 26:28, 1970.

[8] Joan Francesc Gilabert. Development of a simulator to evaluate public health control strategies of tuberculosis in big cities. B.S. Thesis, Universitat Politècnica de Catalunya, 2015.

[9] Volker Grimm, Uta Berger, Donald L DeAngelis, J Gary Polhill, Jarl Giske, and Steven F Railsback. The odd protocol: a review and first update. *Ecological modelling*, 221(23):2760–2768, 2010.

[10] Giorgio Guzzetta, Marco Ajelli, Zhenhua Yang, Stefano Merler, Cesare Furlanello, and Denise Kirschner. Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. *Journal of theoretical biology*, 289:197–205, 2011.

[11] Emma S McBryde and Justin T Denholm. Risk of active tuberculosis in immigrants: effects of age, region of origin and time since arrival in a low-exposure setting. *Medical Journal of Australia*, 197(8):458, 2012.

[12] Cristina Montañola-Sales, Joan Francesc Gilabert-Navarro, Josep Casanovas-Garcia, Clara Prats, Daniel López, Joaquim Valls, Pere Joan Cardona, and Cristina Vilaplana. Modeling tuberculosis in barcelona. a solution to speed-up agent-based simulations. In *Winter Simulation Conference (WSC), 2015*, pages 1295–1306. IEEE, 2015.

[13] World Health Organization. Global health observatory data repository, accessed June 19, 2017. `http://www.who.int/gho/database/en/`.

[14] World Health Organization et al. Global tuberculosis report 2016. 2016.

[15] Jesús E Ospina, Àngels Orcau, Joan-Pau Millet, Miriam Ros, Sonia Gil, Joan A Caylà, Barcelona Tuberculosis Immigration Working Group, et al. Epidemiology of tuberculosis in immigrants in a large city with large-scale immigration (1991-2013). *PloS one*, 11(10):e0164736, 2016.

[16] Steven F Railsback and Volker Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2011.

[17] Julia Vila. Analysis and individual-based modelling of the tuberculosis epidemiology in barcelona. the role of age, gender and origin. B.S. Thesis, Universitat Politècnica de Catalunya, 2017.