DANIEL TOSHIHIRO OKANE

CREATING AN INTERACTIVE SUPPORT SYSTEM FOR THE CHOICE OF THE LOCATION OF THE EMERGENCY SUPPLY WAREHOUSES OF CEPDEC-SP

Trabalho de Formatura apresentado à Escola Politécnica da Universidade de São Paulo para a obtenção do Diploma de Engenheira de Produção.

São Paulo

2019

DANIEL TOSHIHIRO OKANE

CREATING AN INTERACTIVE SUPPORT SYSTEM FOR THE CHOICE OF THE LOCATION OF THE EMERGENCY SUPPLY WAREHOUSES OF CEPDEC-SP

Trabalho de Formatura apresentado à Escola Politécnica da Universidade de São Paulo para a obtenção do Diploma de Engenheira de Produção.

Orientador: Prof. Dr. Hugo Tsugunobu Yoshida Yoshisaki

São Paulo

2019

Catalogação-na-publicação

Okane, Daniel Toshihiro
Creating an interactive support system for the choice of the location of the emergency supply warehouses of CEPDEC-SP/ D. T. Okane -- São Paulo, 2019.112 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.LOGISTICA HUMANITARIA 2.SISTEMAS DE APOIO À DECISÃO 3.REGRESSAO 4.PROGRAMAÇÃO LINEAR I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

AKNOWLEDGEMENTS

This thesis concludes my graduation at Polytechnic School of the University of São Paulo. This means I will finally complete my studies and become an engineer. This accomplishment was only possible because of my family and friends. I must thank them for all the support outside the academic world. Without their help, I would not have had the energy and motivation necessary to finish all the projects nor to study for all the tests.

For the development of this thesis, I must thank Professor Hugo Yoshizaki and Professor Irineu Brito for all the support during the short period of time we had to produce the report. We had a small amount of time to accomplish it and their support was key for me to be able to finish it in time.

I must also thank Alfonso Cosenza for his work that was a great base for the studies made on this report. He did an excellent thesis that really served as a guideline to building this next step and improving his previous results.

Finally, yet importantly, I must also thank Major Panzarini and the team from the Department of Protection and Civil Defense from the State of São Paulo for all their aid and thrust. They shared crucial data to the construction of the project that originated this report and also provided a lot of their hours to have discussions to align the goals and tasks.

RESUMO

Todos os anos na época de chuvas de verão, milhares de pessoas são afetadas por enchentes e deslizamentos no Estado de São Paulo. A Coordenadoria Estadual de Proteção e Defesa Civil do Estado de São Paulo (CEPDEC-SP) ajuda as pessoas afetadas por estas catástrofes com um suporte de materiais alimentícios, de higiene pessoal e de habitação. Então para que o auxílio seja efetivo, é importante que o órgão tenha em seus estoques espalhados pelo Estado os volumes necessários para o pronto atendimento dessa demanda. Logo, o planejamento da capacidade e localização desses depósitos vira um aspecto crítico no planejamento estratégico da organização.

Os trabalhos anteriores focaram nos estudos teóricos das localizações ideais para os depósitos de emergência. O objetivo principal era gerar recomendações que iriam otimizar custos e agilidade de resposta aos eventos ocorridos. Mas após a entrega dessas recomendações nada foi implementado. Apesar das interações contínuas e positivas com a equipe de planejamento da Defesa Civil, não se conseguiu gerar tração o suficiente com as autoridades superiores para que o projeto fosse colocado em prática. E como a pequena alteração de circunstâncias pode modificar a solução ideal, a implementação se mostra ainda mais complicada pois a equipe da Defesa Civil não tem as competências necessárias para atualizar sozinha os modelos desenvolvidos.

Para melhorar os resultados, o foco inicial do atual projeto foi em gerar uma ferramenta visual que pudesse gerar essa tração para a implementação dos resultados. Porém, após as interações com a equipe da Defesa Civil, percebeu-se que seria necessário que o sistema fosse também interativo e facilmente atualizável. Então esses 3 eixos se tornaram focos principais no desenvolvimento do protótipo final que foi apresentado para a equipe de planejamento. Os modelos de dados que alimentam o sistema também foram atualizados, mas com uma ênfase menor visto que o objetivo é tornar o sistema modular e simples para que a Defesa Civil atinja uma autonomia mínima para a utilização posterior do produto.

Palavras chave: Logística humanitária, gestão de desastres, pesquisa operacional, visualização de dados, sistemas de apoio à decisão

ABSTRACT

Every year during the raining season, there are floods and landslides that damages the life of thousands of people in São Paulo. The Protection and Civil Defense organism (CEPDEC-SP) helps the people impacted by these events by providing them with basic supplies for eating, cleaning and sleeping. In order to be effective, the CEPDEC has to stock all these items in warehouses that must be able to supply this needs as they appear throughout the entire State. Therefore, the positioning and capacity planning of these warehouses becomes a critical aspect in the effectiveness of the support provided to the affected populations.

Previous works have been done in order to help the planning team to study new configurations to improve the responsiveness and costs of operation of the warehouses where these supplies are located. But even after the delivery of these results, nothing has changed in the configuration of warehouses. The research team has always had positive interactions with the planning team, but the delivery method has not been able to convince the upper management to implement the solutions. In addition, the change of circumstances impacts the optimal solution that no longer is valid. Therefore, without the know-how to update the models by themselves, the CEPDEC team becomes dependent on the work of the other organizations.

In order to improve the results the idea was to create a visualization tool that would be able to provide a better alternative to convince the upper management. After the discussions with the team it was also noted that it was also important for this tool to be interactive and updatable. So these 3 axis became the major concern during the development of the final prototypes. The data that feeds the model was also updated with more recent data, but the focus was on making the system modular so that the CEPDEC team can work independently in the next steps.

Keywords: Humanitarian aid logistics, Disaster management, Operational Research, systems for supporting decision-making

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

CBH-RB	Coordination center for the region of the hydrographic bay of Ribeira de Iguape and South Coastline (Comitê da Bacia Hidrográfica do Ribeira de Iguape e Litoral Sul)
CEPDEC	State Coordination Center for Protection and Civil Defense (Coordenadoria Estadual de Proteção e Defesa Civil)
COMPDEC	City Coordination Center for Protection and Civil defense (Coordenadoria Municipal de Proteção e Defesa Civil)
CPRM	Geological Services from Brazil (Serviço Geológico do Brasil)
CRED	Centre for Research on the Epidemiology of Disasters
EM-DAT	The OFDA/CRED International Disaster Database.
FEMA	Federal Emergency Management Agency
GLM	Generalized Linear Model
IBGE	National Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística).
IGc	Geosciences Institute (Instituto de Geociências da USP)
IFRC	International Federation of Red Cross and Red Crescent Societies
IPT	Technological Research Institute (Instituto de Pesquisas Tecnológicas do Estado de São Paulo)
KDD	Knowledge Discovery in Databases
MIP	Mixed Integer Programming
OFDA	Office of U.S. Foreign Disaster Assistance
REPDEC	Regional Coordination Center for Protecion and Civil Defense (Coordenadoria Regional de Proteção e Defesa Civil)
RMSP	São Paulo Metropolitan Zone

TABLE OF CONTENTS

1 Introduction

This chapter will give an overview to the reader about the goals of the project. In the first section, we present an introduction about the context of the problem and the main objectives of the work developed to solve it. Secondly, we present the historic of previously developed actions and the importance that this project has within this sequence of studies. Then we introduce a differentiation aspect for the project and the improvement requests made by the "clients" (the Protection and Civil Defense coordination team). Finally, the reader will read about the structure of the entire report for all the following chapters.

1.1 Context and objective

There is a saying in Brazilian popular culture that Brazil is blessed by God due to its nature and natural resources. However, the truth is that it is the only country in the American continent ranked in the top 10 countries in the world by the number of people impacted by natural disasters. Estimations says that that the country loses between 250 and 300 billion USD every year due to these events. (United Nations 2015). So in reality the country has a big challenge to face in combating the natural disasters and their impacts in the general population. It is not as blessed by God as the people tend to believe. São Paulo state despite its leadership position in the economic field has not been able to escape this condition. The State and its capital have been suffering with floods and landslides for a long time. And the investments made on infrastructure improvements were not enough to solve the situation (Custódio & Moraes 2002).

The analysis of a series of disasters such as the Rio de Janeiro 2011 raining season shoes that there is a large part of the population affected in the lowest economical percentiles. And in natural disasters in Brazil, a huge proportion of the damages done are to the housing. So many people become unable to continue living at their houses either temporarily or permanently after a disaster. Therefore, the response and allocation of these families under temporary lodging with access to basic supplies is urgent. During the disasters in the Rio de Janeiro State in 2011, studies shows that up to 20% of the inhabitants of a city can be out of housing after a disaster. (World Bank 2012)

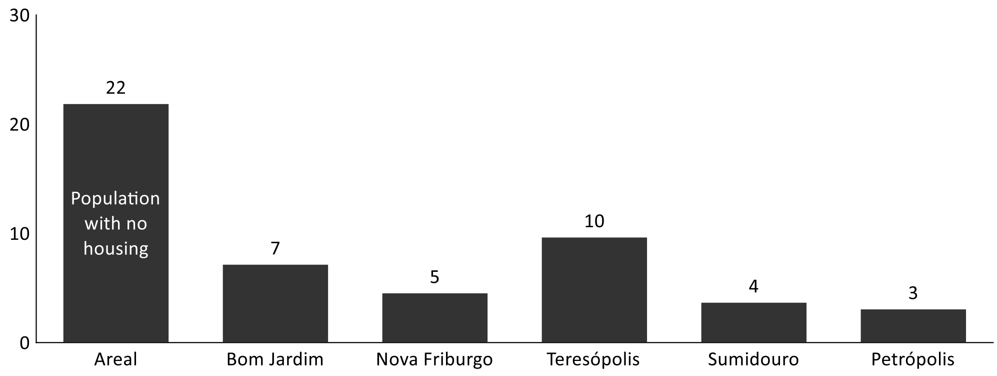


Figure 1- Percentage of population with no housing in 2011 after the tragedy in Rio de Janeiro – World Bank (2012)

Using the historical cases of accidents should provide an accurate view of future risks and help to map the demands for the pre-positioning of resources. (Lei, Wang Chen and Hou 2018). This would improve the response to disasters and also help to understand the regions that should receive investments with a higher degree of priority. In this project the focus will be on the response to the disasters.

We can finally resume the objective of the work into the following question:

“How can we develop a system that will help the public authorities to improve the location of the relief supplies deposit for disaster response?”

The work follows closely what Cosenza (2015) developed for the same institution, the Estate coordination enter for Protection and Civil Defense of São Paulo (CEPDEC-SP). The previous project tried to answer a different question and therefore had another focus. The question studied in this previous report was:

“In the São Paulo State, where should be located the relief supplies deposits for disaster response?”

The two questions can look very similar, but there are subtle differences that create a big gap in the focus of the work. This second question focus on the mathematical approach on identifying the best location for the relief supplies deposits. On the other hand, the first question will be answered with a tangible goal on how to improve the location of the warehouses through the development of an application. Therefore, instead of a simple presentation of the recommendations, the goal is to have something that can be used directly by the Protection and Civil Defense so they can arrive at their own conclusions. A good indicator of success would be if the system can become part of the yearly planning process for the raining season. This will have a series of impacts that will be presented during the chapters 5, 6 and 7 .The details about the requirements of this system becomes critical inside this context and it will be explained in the rest of this chapter and in chapter three where the context is further explained.

All the work happened in a collaboration with the Center for Innovation in Logistics System of the Polytechnic School of the University of São Paulo (CISLOG), the Center of studies and research about disasters of São Paulo State (CEPED-SP) and the CEPDEC-SP team. The first two with the help through the technical expertise and the latter with the databases and clarifications about the use cases.

1.2 Importance and historic

Cosenza (2015) developed a study related to the statewide location of relief supplies warehouses for São Paulo Civil Defense. It focused on developing each of the two steps necessary for the correct establishment of these locations. The first step is calculating the number of people at risk in each municipality in the State of São Paulo and the second using these results as the demand to a location model for the positioning of the warehouses.

We will update these two models, but the focus will be on complementing the work with an additional step. There is a missing point in the visualization of the results that was very simple in the previous works. Since the work done in 2015, there has not been a formal impact in the way the Civil Defense organizes itself. From the discussions with the current team, the PowerPoint presentation used as the deliverable of the 2015 project have been used in the internal presentations and discussion but with no further impact.

So it is important for the new solution to be easy to understand, explain and share. Which implies that having a very user-friendly visualization system is a must have in this case. The presence of these characteristics increases the odds of the users effectively using the proposed system. This will be the major differentiation point compared to the previous iterations that focused on the theoretical results rather than the practical ones.

1.3 Request from the civil defense team

In the kickoff meeting of the project, we had a general discussion with the CEPDEC team (that will be referred to as the client). There was a clear request from the client to go beyond updating the models from the previous iterations. It was important to actually update the concept of the final deliverable.

In the modelling side there was a request to the addition of the capacity as a constraint that was tested and weighed against the more important factors such as the simplicity of the models. The main demand was in the improvement of the system capacity to be operated independently by them. They would like to have a certain freedom to use the tool independently after the project is finished.

The previous projects could not be implemented due to the lack of the necessary resources. These resources are conditioned to the support of the Governor and other high-level stakeholders. So in order to get the resources it is important for the results of the project to be easily understandable by them. So, presenting technical results was not enough, since it did not have the desired impact in their decision making process. Because the risk is to become too difficult to explain when the processes are too complex. So one of the steps will be to simplify the two major steps developed by Alfonso while also making other improvements (regression and location model). A good visualization that is visually appealing and has the right amount of information could be a good answer to these concerns and would allow the use of complex model to transmit simple messages.

Another point discussed was the timing of the current project. Since the arrival of the new Governor, there are ongoing discussions to expand the number of warehouses to a much larger scale. There is a plan for the creation of a series of small centers that integrates multiple State coordination centers staff. So they would have multiple local headquarters through the State with one representative of each unit (Civil Defense, Health, Education, Sports, etc.). And each of these headquarters could have a room used as a warehouse to the Civil Defense needs when necessary. As this is an ongoing discussion it is impossible to provide details on new values for variables such as warehouse locations, capacity and costs. For these metrics the values from the previous interactions during Cosenza (2015) work were kept. As this situation will evolve after the project is finished, the goal is to leave a tool that can be updated by the CEPDEC-SP team independently from the CEPED/CISLOG team. This gain in autonomy would also make the system more flexible and enable the development of a model that could be replicated in other States of Brazil with minor adjustments.

1.4 Structure of the work

The first chapter is the one being read. It will be the first of the nine chapters that compose this dissertation. It consists of the introduction of the general subject and the short debrief on the previous work done by Cosenza (2015).

Chapter two will present the review of the academic literature that discuss the main topics that are necessary to the understanding of the project. The subjects start from the disaster management and go to the regression and location models. Together they will form the base in which the rest of the work will base itself.

Chapter three will present an introduction to the situation of São Paulo state and its CEPDEC regarding the disasters and the response systems that are in place. This will also include a presentation of the available data that will be later used in the models. It will also present a more in-depth review of past projects and current requests from the CEPDEC team.

Chapter four will provide further details on the methodology used on the previous iterations. Then it will present the improvements to the original methodology and the final chosen methodology.

Then in the chapter five, the focus will be on the data processing and the data integration between the different workflows. From the raw data to the linear regression, the treatment and classification of Civil defense data and then finally the location models. All the assumptions and hypothesis taken during the construction of the model will be presented. And most importantly, the decisions made on the simplification of the models in order to arrive at a system that can be used by the Civil Defense team will also be discussed in this chapter.

Finally, in the chapter six the results will be presented and discussed. The focus will remain in the best solutions proposed by the GLM and regression models. Then in chapter seven, the focus will be on the website and interactive system that were built. Moreover, on the chapter eight we will present some conclusions and the next steps that could be made in order to continue this sequence of projects.

The literature review and context explanation in Chapters 2 and 3 are important for the people unfamiliar with the subject. But the reader that wants to go directly to the results can skip the literature review and go directly to the chapter 4 where we present the methodology. The results can be easily seen in the link below and their in-depth description can be found straight on chapter 7.

https://drive.google.com/drive/folders/1LnyvM9gIwUfwwkK2SJjqtkZJ2lib8RgW?usp=sharing

2 Literature review

In this chapter we are going to present the different elements that will impact the project. Each of the subjects will be reviewed with a simple introduction of the subjected. The reader can choose to skip the subjects that he/she is already familiar with. But we recommend the reading of the entire chapter to fully understand the context that will be given in the chapter 3. The reader may also keep in mind to come back to the chapter 2 during the reading of the chapters that will discuss the results and methodology.

2.1 Humanitarian Logistic, Disasters and civil defense organization

2.1.1 General studies

Humanitarian Logistic is an area that studies how to improve the logistic operations that are run to help people under hard circumstances. More often than not, the expertise in the area is needed after the occurrence of a disaster. Disaster can be defined as "a sudden, calamitous event that seriously disrupts the functioning of a community or society and causes human, material, and economic or environmental losses that exceed the community's or society's ability to cope using its own resources. Though often caused by nature, disasters can have human origins" (IFRC 2015)

In general, the disasters occur and then generate a response from the authorities to give support to the impacted people. Often the response from the government must take place with a great urgency as people are put under precarious circumstances. The resource investments made on this stage can be classed in four groups: evaluation, organization, sustaining and reconfiguration. (Thomas 2003)

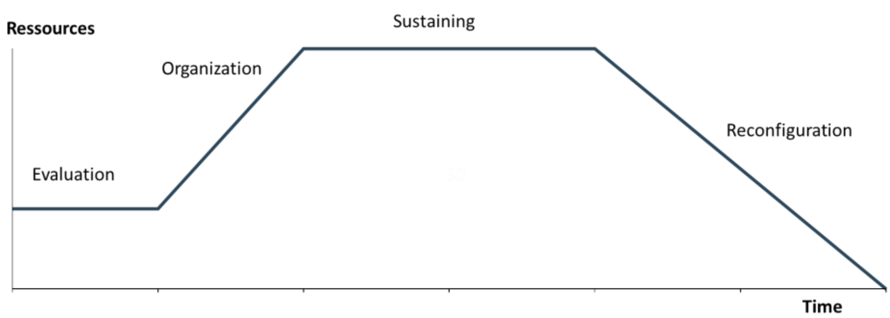


Figure 2- Lifecycle of a humanitarian aid project - adapted from Thomas (2003)

The humanitarian logistics can be applied in all four stages of this lifecycle. In the current project, the focus will be on the organization and sustaining phases. The main concern about locations will refer to the Organization phase while the correct supply allocation will be important in the Sustaining phase.

The evaluation is done on the site as the first and local teams arrive to assess the situation, then once they have done the necessary analysis they will send the supply requests. This will involve the local teams to request the supplies to the central CEPDEC team and then arranging their own teams in order to collect the supplies from one of the warehouses. So the transportation from the CEPDEC warehouses to the affected cities is mostly under the responsibility of the local authorities (according to the interviews with the CEPDEC coordination team).

Vulnerability according to the Red Cross is “the diminished capacity of an individual or group to anticipate, cope with, resist and recover from the impact of a natural or man-made hazard” (IFRC 2015). By the definition, it can be inferred that the socioeconomically situation of a group of people is a defining factor for the definition of the vulnerability status. Poorer people are often more susceptible to obtaining this status as they are surrounded by an infrastructure that is not able to cope with abrupt changes and hazards. Richer neighborhoods have a better capacity to resist to the impacts of hazards.

This can be seen in the report from the World Bank where 91% of the damages to housing infrastructure was made to the poorer segment in the natural disasters of 2011 in Rio de Janeiro State. These disasters can be used as a close proxy to the disasters that happen in the São Paulo State as they were both caused by the rain and the subsequent landslides and floods. (World Bank 2012)

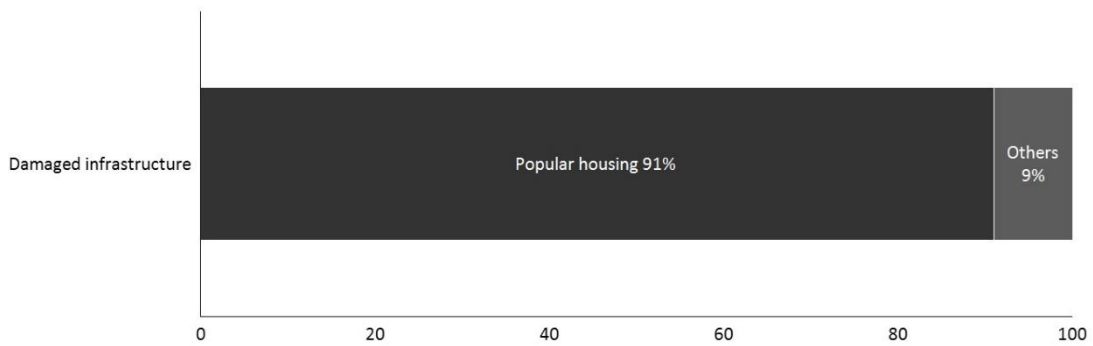


Figure 3- Distribution of damage to housing infrastructure by type of housing - World Bank report (2012)

Therefore, it is important to understand the different level of hazards and different conditions of each region to better quantify the people under risks of disasters. The degree of development of each city is also important as the damages done by hazards affect the entire city and not only one isolated part of it.

The risk of a disaster is a function of the probability of a hazard and the analysis of its impacts (Tominaga, Santoro, and Amaral 2009). It is important to analyze the types of disasters and not only its risks. According to Van Wassenhove (2006), the disasters should be classed using two main binary variables. They are the type of originating hazard and the rapidness of the evolution of the situation. This classification is useful in order to understand which kind of events will have a bigger impact in the supplies from the CEPDEC organization. Slow onset speed often implies that supplies can be gathered from any part of the State while high speed will incur in the need for speedy responsiveness. Natural hazards can mostly be predicted from their historic of occurrences (Lei Wang, Chen and Hou, 2018) while the same is not valid for man-made disasters due to their nature.

The types of originating hazards can be divided into natural or anthropogenic causes. Examples are Earthquakes as natural hazards and terrorist attacks as manmade hazards. The rapidness of the evolution can be classed as sudden-onset and slow-onset. Examples for each of the four types of disasters under this classification can be found in the summary below.

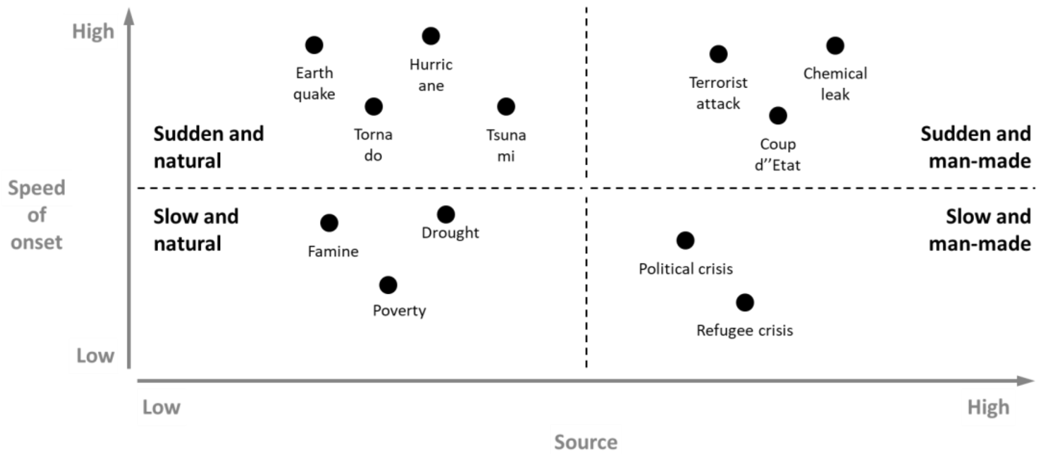


Figure 4- Matrix with examples of disasters per category - adapted from Van Wassenhove (2006)

The natural disasters are different from the man-made ones especially due to the reason a hazard becomes a disaster. The natural disasters depend mostly on the lack of the capacity to absorb natural events while the man-made ones are a damage that society is inflicting in itself. Even if the members generating the damage and those suffering it are often different. It still consists of a part of the society damaging itself as a whole (Albala-Bertrand 2000).

The focus of the Civil Defense in the State of São Paulo is mostly on the natural disasters on the matrix from Wassenhove (2006). Which means that the society must be better prepared to absorb the natural hazard and be prepared to respond quickly to alleviate the consequences of the disaster. Under such circumstances, the positioning of the supplies becomes a critical component of the strategy to be globally prepared to attend the population’s need. This prioritization has been stablished in a national level and can be seen in the Project GIDES. Through the six volumes created, there is a clear movement in the prevention and anticipation side of the planning. (MDR 2013)

The disaster management is a discipline seem as cyclical by most experts. It has very clear phases that are repeated in each cycle. They are split into two, those that happen before the disasters and the ones that happen after. The first one is the preparation to the disaster, which includes the mitigation and the preparedness stages. Once the disaster occurs, there are the stages of response and recovery that take place. It is important to note that the mitigation for the disasters starts during the recovery phase and will only end in the moment the new disaster occurs. (IFRC 2015 and FEMA 2018). For better understanding the cycle, it is important to define clearly the different stages of the disaster management:

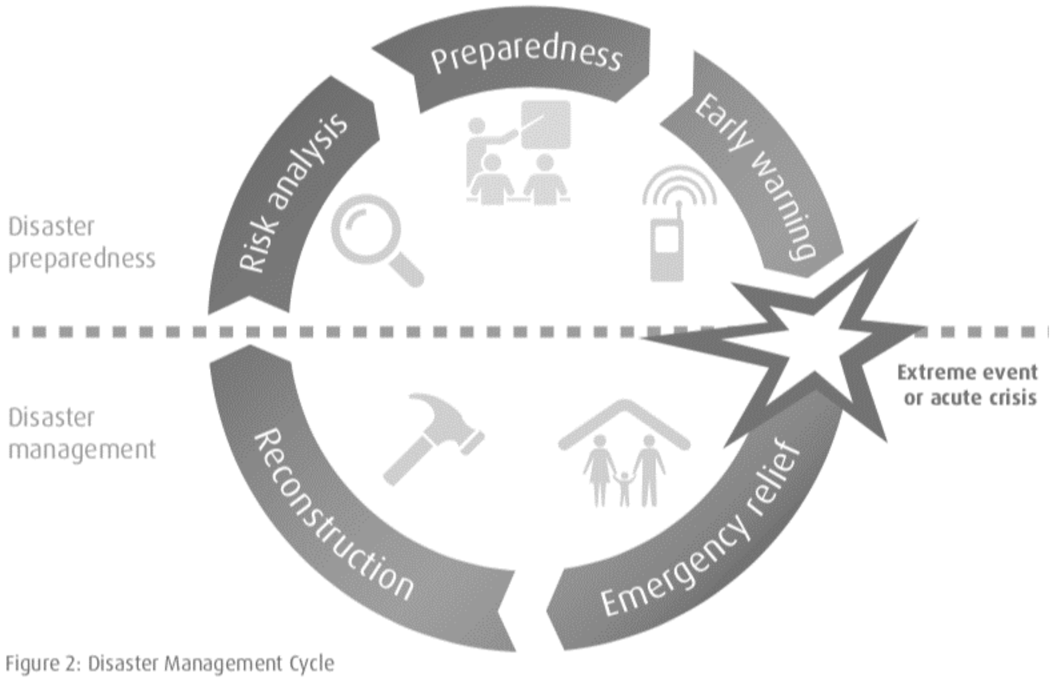


Figure 5- Cycle of a disaster – adapted from Hagenlocher (2016)

- Response: the immediate action after the disaster. The main activities are the assistance to the affected people and the organization of actions to reduce further damages to the society (Brito Jr 2015).
- Recovering: actions taken to restore the normality status. Often with the help of government, the process focuses on re-establishing the functioning of essential system and services (Brito Jr 2015).
- Mitigation: the set of actions of reducing the severity of a disaster or its effect. Includes new laws, programs to raise awareness to the population, etc. (Brito Jr 2015).
- Preparation: is to review what was missing and made possible the past disaster. Then implement new mechanisms to respond to the elements that cannot be mitigated easily (Tomasini and Van Wassenhove 2009). Improving or creating response strategies for each type of disaster is an important step. These actions can greatly reduce the number of victims and damages done for the next events. For a guideline example of what can be done on the individual level check the FEMA preparation guides (FEMA 2018)

Hagenlocher et al did a complete overview of the risks related to natural Hazards in the World Risk report (2016). In their approach there are going to be two major components to the creation of the risk index. The first one is the natural exposure to hazards. The second is the vulnerability of the societal sphere to these issues. In order to determine the vulnerability level, the author utilizes three criteria: susceptibility, lack of coping capacities and lack of adaptive capacities. For the exact method used to calculate all these four indicators, check the Appendix 5. The data sources used by this research was used as a base for possible upgrades to the models developed by Cosenza (2015).

2.1.2 Disasters in Brazil

The country has one of the biggest populations that are impacted by natural disasters. In the period of 1960-2008, there is a record of 5720 deaths and more the 15 million people affected by natural disasters. (Tominaga, Santoro, and Amaral 2009)

The main natural hazards that create disasters in Brazil are floods, landslides and droughts. These are mostly related to intense rainy seasons (during summer in the South and Southeast regions, during winter in Northeast regions) and drought seasons in the rest of the year. During the raining season, there are increased risks of floods, landslides and hailstorms. During the drought, seasons there are risks of forest fires (such as the one affecting the Amazon Forest in the month of August 2019), the lack of water for agricultural purposes and even for human consumption (mostly in the Northeast region of the country). It must be considered that each region in the country is under risk of different types of hazards. The differences of hazards by each region can also be found in the studies developed by Tominaga (2009).

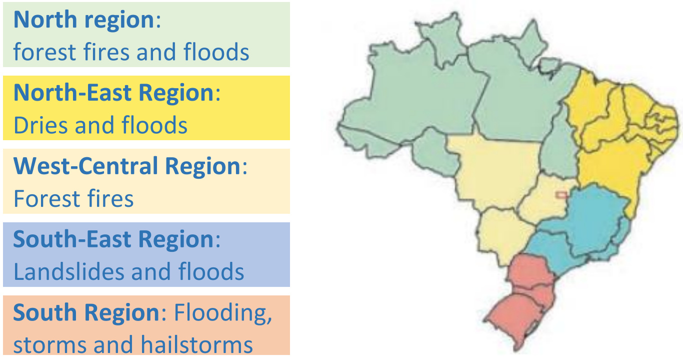


Figure 6- Disaster distribution in Brazil - Adapted from Cosenza (2015)

But despite the different disasters that can happen, there is a high concentration of deaths number of occurrences in just two of these types of hazards. Floods and landslides represent in Brazil:

- 97.4% of mortality from natural disaster
- 76.5% of the occurrences of natural hazards (UNDRR, OFDA and CRED 2018)

2.1.3 Industrial supply chain vs Humanitarian Supply chain

In a typical supply chain, there are activities related to the logistics, warehousing, stock management and provisioning. These four activities tend to follow a demand that is relatively continuous or at least periodical. The goal of the enterprises operating within this context is to maximize profits of the operations, which means reducing the cost the most they can while not influencing the revenue potential of the company (Ballou 2009).

The supply chain system needed for the humanitarian causes changes a lot from this standard model. First, the main goal of the system operators is no longer to maximize the profits. The goal of a humanitarian focused supply chain is rather to minimize the suffering from the affected populations. But the change is not simply on the goals, but also under the operating circumstances. The strongest difference is under the demand behavior. In humanitarian logistics, both the geographical distribution and quantity are hard to be predicted and can have sudden variations. Which means that it is hard to predict where the demand is going to be generated next. Because of the goal to reduce the damages done to the population, the demand generated must be satisfied in the shortest period possible. Pairing this need to be reactive with the limited resources that are often associated to the authorities managing this supply chain and it is easy to understand the biggest challenges of this field. (Van Wassenhove 2006). A simple model of a humanitarian supply chain is shown in the next figure.

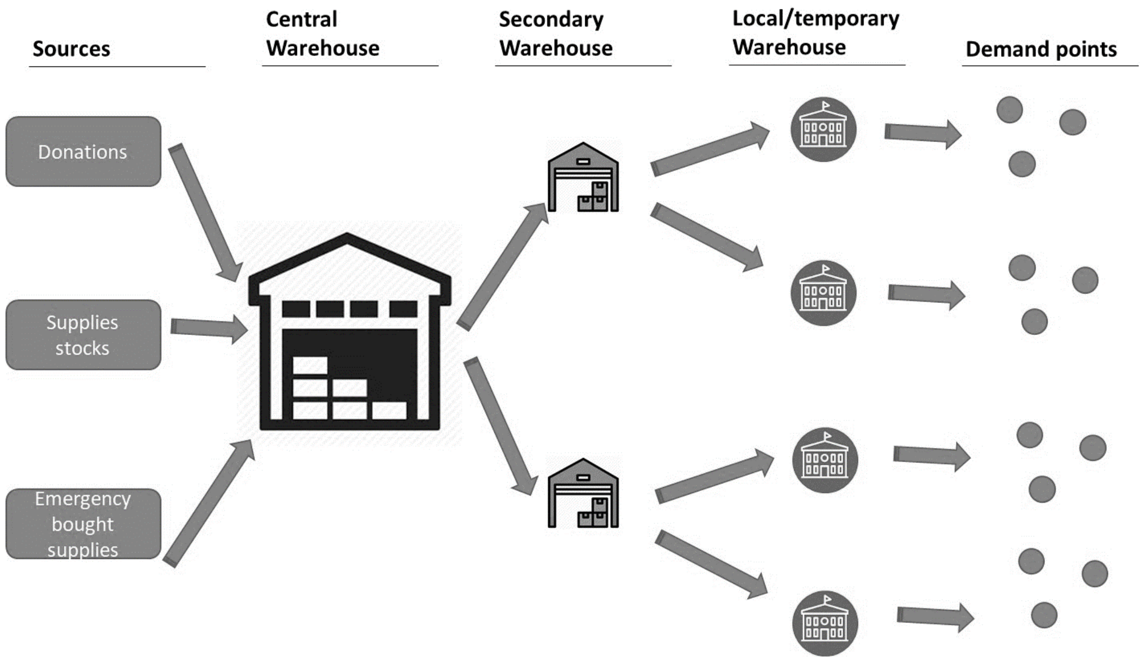


Figure 7- Humanitarian supply chain structure - Adapted from Beanon (2006)

The prediction of the amounts and location of demands is one of the key elements in the optimal utilization of the scarce resources. These activities must be done before the events; therefore, they are in the preparation phase. Once the demand is predicted and the needed stocks are calculated the authorities start the purchase of the needed supplies. They constitute the first shipments that will be made in the case of a disaster, but they are not the only source of supplies in this case. There are the donations that can also be used to provide support to the impacted population. Nevertheless, these supplies are not considered in the preparation phase due to the difficulty in predicting their location and quantities.

Once there is a disaster, the supplies will go from the warehouses to the secondary and local centers. They will take care of the distribution within the deeper capillarity of this network. According to the public authorities of São Paulo State, one of the major costs for their operation is the transport of stocks between the different warehouses in the system. At the same time, the multiple distribution centers can be activated as the supplies in each one might not be enough depending on the scale of the events. In case the stocks are depleted, and the donations do not cover the demands, there is also the possibility of making emergency purchases from local suppliers. The amount of donations received is often associated with media coverage of the events, either locally or internationally. (Nogueira, Gonçalves, and Novaes 2007)

In the first 72h, the only concern of the authorities should be to provide help as fast as possible to those affected. There is going to be little concern with costs, and the biggest barriers will be the lack of resources that was discussed previously. After the initial response, there is an augmentation for the cost-consciousness as the long-term recovery process begin. At this stage the cost effective policies becomes the priority and there starts to be a trade-off between helping people at any cost and using the resources effectively. (Van Wassenhove 2006)

Despite the differences that are considerable, there is an opportunity for sharing improvements between the two types of networks. The humanitarian logistics could start to implement some methods from industrial supply chain. (Van Wassenhove 2006)

2.2 Location models

2.2.1 Generic location models

The objective of a location model is to find the optimal location of the set of warehouses while respecting the constraints chosen. The optimality of the set of warehouses must be evaluated by some criteria, which are often connected to minimizing the costs of operating a system. The iterations of the model and the different solutions with different set of constraints can offer good insights about the problem itself. (Geoffrion 1987; Pidd 1999)

For the constraints, there are multitudes of choices that can be done, the further we constrain the model the closer we get to the reality, but we also make it harder to mathematically find optimal solution. The time available and size of datasets becomes crucial in the decision of which characteristics should be added in the model. There are 5 main criteria to classify a problem and try to group them to find the best resolution methods and sets of parameters. Three of them are very applicable to the situation being studied. (Ballou 2009)

- Goal or driving force: what will be the most influential factor into the model. The costs for retailing, the revenue, the service level or even the response time in emergency cases.
- Number of installations: the easiest models are used to find the best location for a single installation. But in some cases, it might be necessary to find the best location for

multiple units, and in this last case the number of locations can be either fixed or flexible as well..

- Range of choices are the candidate locations limited to a known set or are they open to be placed in a continuous spectrum.

2.2.2 Location models in humanitarian logistic

The growing number of disasters caused by climate change, global warming and terrorism have sparked a similar growth in the research about disaster management (Altay and Green 2006). The field of operations research has also its fair amount of papers published.

Sherali, Carter and Hobeika (1991) had a location model. Instead of calculating the location for relief supply deposits, their model focused on the location for shelters. In their research, they used nonlinear programming.

Other two researches that used location model are O'Kelly and Current (1992) and Srinivasa and Wilhem (1997). They have used the models to calculate the optimal location for nuclear alert sirens (while maximizing the coverage of the alert signal) and the cleaning supplies warehouses for oil leaks in Texas, respectively. In Dekle (2005), there is also another application for the localization of a disaster center for recovery that had a constraint from a maximum distance. In this case, the model used was a Mixed Integer Programming model.

A close model to the scenario being faced was found on Balcik and Beamon (2008). They worked in a problem of maximum coverage with the location of distribution centers for disaster relief supplies. Their model was done with a goal of covering the largest number of people given a delimited response time. The use of estimators for the uncertainty of events and the relationship between different demands for relief supplies and the type of disasters make the further study of this model even more appealing to the understanding of the current problem being studied.

Rawls and Turnquist (2010) used a 2-stage stochastic model and a Larangian heuristic to determine both the location and the quantity of supplies in the installation of warehouses for disaster response. The complexity of the model required the Lagrangian heuristic and the work was further improved in works in 2011 and 2012 that added elements such as service quality.

In a work developed also for the State of São Paulo, Brito Junior (2012 and 2015) is a series of models used to calculate the location of disaster relief supply warehouses in the region of Vale do Paraiba which is heavily impacted by floods and landslides. In the first iteration the 2-stage stochastic model was used, and it considered criteria such as donations, purchased material and considered the limitations of capacity in the warehouses and in the transportation channels. The model was evaluated by an EVPI and VSS systems. Then in 2015 the model was integrated with a MCDA model and it started to take a multi criteria decision analysis. With the mixed integer programming model, the tool became more robust and ready to aid in the decision-making process of the authorities.

Cosenza (2015) used a 2-stage linear stochastic regression to calculate the position and number of warehouses for the Civil Defense system in São Paulo. The model answers to some constraints such as a maximum distance but there are no constraints on the system capacity for the moment. More details about the model will be presented on the Methodology section due to the relationship of the work developed by the author and the current project.

2.3 Regression models

A multivariable regression model is a linear regression model that contains multiple variables as predictors. A linear regression model has a goal to predict a variable Z based on the values of the independent predictors Xi. The General additive regression model can be expressed by a simple function with the k predictors, each multiplied by a βi regression coefficient. The coefficients are expected to change according to the Z variable being predicted. (Hidalgo and Goodman 2012)

$$Z = \beta_0 + \sum_i \beta_i \cdot X_i \quad i = 1 \ldots k$$

One of the most common methods to identify the βi coefficients is to use the least squares estimation. It consists in minimizing the sum of square errors ϵi between the predicted values and the real ones. In order to evaluate the estimator, it is common to analyze the R^2 of the prediction of Z. The R^2 can be interpreted as the proportion of the Z value that can be predicted with the current model. It is used as a test to the fitting of the model with the regressed data. Therefore, it will be used as a quantitative quality indicator

It is necessary to test in simple linear regression and reject the null hypothesis of H_0: β1 =0 for every single relationship between Z and the Xi predictors. Which means to prove that there is a relationship between Z and the predictors. If this test is completed, then there a formal method to prove the variables included in the model with a higher degree of confidence in their relevance. to which there is no useful relationship between Z and the single predictor X.

To test and validate the model there are other tools available as well. One example is the analysis of the residual errors' probability plot. As discussed previously this tool is important in order to assure the reliability of the confidence and prediction intervals estimations for the statistical tests. Therefore, the use of this tool is strongly recommended. And in the cases where the test fails to deliver good results, it does not mean that the model is not reliable. If the number of samples is big enough the statistical significance of the model can still be used and proved. The lower limits for sample dimensions has been estimated to 15. (Lumley et al. 2002)

So, once the method for choosing the variables is known it is necessary to identify the best combination of independent variables. By the utilization of the method (if the number of variables allows the researcher to do this) of adding and subtracting variables one by one it is possible to reach the best model. This is an iterative process where an initial set of variables must be chosen and then at each step it is possible to either add, subtract or do both with different variables (Devore, Farnum, and Doi 2013).

2.4 Generalized linear models

When linear models are not sufficient, generalized linear models can be the solution (GLM). A GLM do not need the dependent variable to be normally distributed, that comes from the nature of the model that uses the probability distribution of variables. This together with the relationship between the dependent variable and the predictor marks the two biggest features of GLMs. F is a generic mathematical function that will be denominated link function. The amount of computation power necessary to solve the optimization of non-linear functions can become a challenge for researchers. (Nelder and Weddeburn 1972)

$$f\big(E(Z)\big) = \beta_0 + \sum_i \beta_i X_i \quad i = 1 \dots k$$

By using the Maximum Like hood Estimation methods, it is possible to generate a function that will measure the fit of different models. With the data input it is then possible to arrive to the model for whom the value of the function is maximized as intended. GLMs can be applied in cases where the dependent variable has a distribution that belongs to the exponential family of functions. In these cases the link function will be connected to this distribution and that will be the necessary condition for its resolution. (Dobson 2002)

When the data presented has a high number of zero values it is necessary to run an adaptation to the models used. The solution to this is the zero-inflated generalized models that treats the zeros in a variable as a different variable. Then with the creation of this new variables it creates its own dedicated regression models that can be used in the final version of the model. A zero-inflated regression model is made by: a GLM for Poisson or negative binomial distribution paired with a binary model, a logistic regression model that determines what to do with the zeroes of the variable and a count model for supporting these models.

2.5 Visualisation methods and data sharing

Even if the analysis is done and the results are ready, the process is not finished yet. As shown by Souza (2004) after the recognition of the patterns it is still important to evaluate and present the data before arriving at the creation of knowledge. This is part of the Knowledge Discovery in Databases (KDD) inside which this work can be grouped. Lima, Oliveira and Gonçalves (2011) have worked in similar cases where KDD was used to improve operations of humanitarian logistics. In their work they discuss how the representations are an important step in order to show the data to the users. The combination of maps and charts is important to transmit the message correctly, because each of these forms of representation have different advantages.

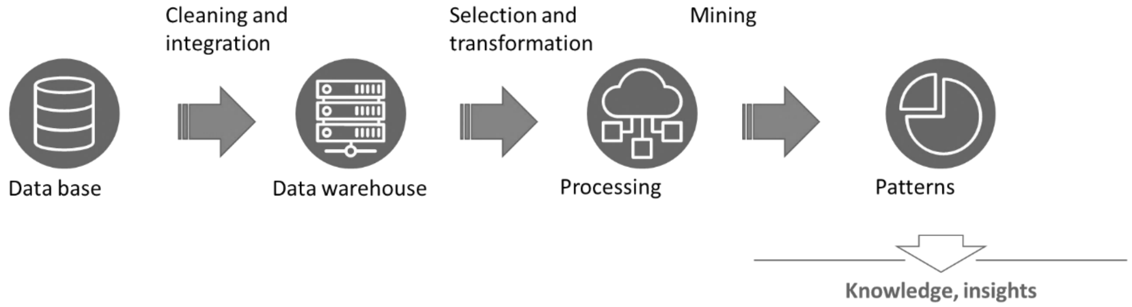


Figure 8- Stages for the KDD Process – Adapted from Lima (2011)

For the creation of maps there are a very wide set of tools and methods that can be used. One of the most known group of tools that are used in geographical based analysis and representation are the GIS software's. A good description of GIS software's can be seen in the following quote:

> "A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. The key word to this technology is Geography – this means that some portion of the data is spatial. In other words, data that is in some way referenced to locations on the earth.
>
> Coupled with this data is usually tabular data known as attribute data. Attribute data can be generally defined as additional information about each of the spatial features. An example of this would be schools. The actual location of the schools is the spatial data. Additional data such as the school name, level of education taught, student capacity would make up the attribute data. The partnership of these two data types enables GIS to be such an effective problem-solving tool through spatial analysis.
>
> GIS is more than just software. People and methods are combined with geospatial software and tools, to enable spatial analysis, manage large datasets, and display information in a map/graphical form." Dempsey (1999)

GIS integration to KDD methods is important, as often KDD do not have major focus on spatial aspects of the data. This is not ideal, as there are potential synergies to be gained with the combination of both, especially when there are interactive maps and automatic map generation. (Andrienko 1999)

But most of the analysis is already done without the need for complex processes that requires the spatial data. This liberates space for the use of other kind of map representations or software's that are more focused on the representation itself than in the processing of the data. As seen in Pickles (1995) the use of GIS at its base is focused on creating the most detailed and correct representations of reality. Abler (1993) is also another promoter of this view of the GIS as a representation of the reality with the biggest amount of details as possible. Due to the scope of the project and the goals of the system, this might not be the best interest as the core mission is really in being able to represent the data and conclusions made to the involved stakeholders.

Another aspect of the system is that the availability of the information is important. In addition, not only any type of information but also the most up to date one is crucial in order to enable authorities to take the correct decisions. Satellite images and studies are a great example on how geographical analysis can be used as a very important tool not only during the response phase of the disaster but after it as well. With the correct representations and analysis, it is possible to better understand the failure points on the system and start to implement corrections to help in this correction (Hagenlocher et al 2016).

After the lecture of this chapter the reader should be ready to understand the project context in the next chapter. Readers should probably come back to chapter two during the reading of the chapters five and six to better understand the methods being used. As they are a bit more technical, they requires a bit more caution during their reading.

3 Diagnostic and problem description

This chapter will introduce the reader to the São Paulo State and to CEPDEC current situation. Then we will discuss more details of the problem at face and the available tools to try to solve them. We will finally talk about the data gathered and received from the public authorities that have collaborated in the project. In the final section, we will introduce the previous projects done in the area while discussing what was left for improvements.

3.1 São Paulo State

São Paulo is the largest State of Brazil in terms of population and GDP. With 44 million estimated inhabitants and a GDP of R$2 trillion (around 33% of Brazilian GDP) it is considered the main engine of the Brazilian economy. These numbers despite being only the 12th biggest State in terms of surface makes the State very dense in relative terms to its other Brazilian counterparts. It has also the 2nd highest HDI level of all the States with a 0,816 grade in 2017. (SEADE 2019)

The economy and population of the State are also very concentrated in the region of the city of São Paulo. The Metropolitan Region of São Paulo (RMSP) has about 21 million inhabitants and with its close neighbor metropolitan zones (Campinas, Santos, São José dos Campos) it exceeds 31 million people. (IBGE Cidades 2019)

The population in this region can be seen in the density map of the state. It is also possible to see that there are other zones with higher densities such as in the regions of Ribeirão Preto and Bauru. The higher density will also be an important factor when there is a discussion of supplies needed to support population affected by disasters. But the population and economy itself do not explain the entire situation. There is a strong inequality between the economic development of the areas near the capital and elsewhere, even in the more populated cities around the state the economic development is no on the same level. (SEADE 2019)

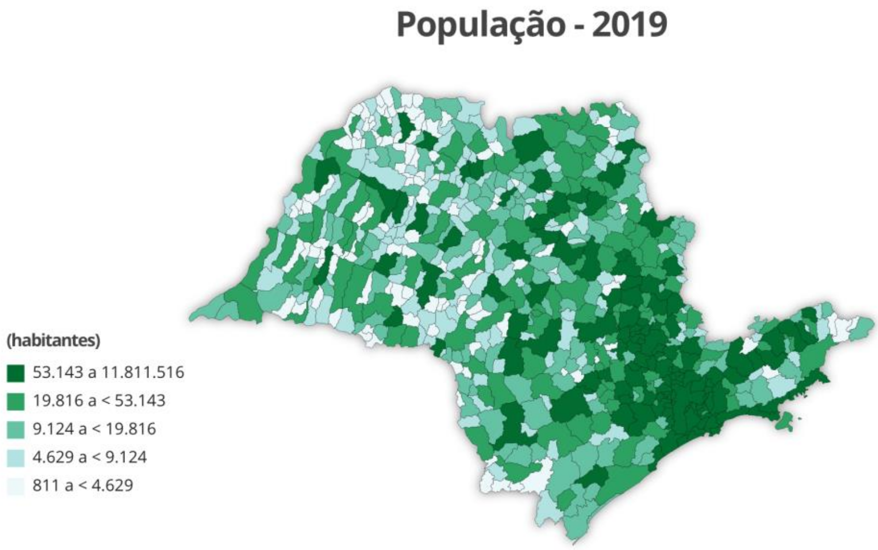


Figure 9- Population by city in the State of São Paulo (SEADE 2019)

The climate is considered subtropical in most regions of the State. The remaining area is considered tropical. In either case the weather is marked by a strong raining season during the summer months. The rain is characterized by its strong intensity and short duration, in the popular culture it is called ‘chuvas de verão’.

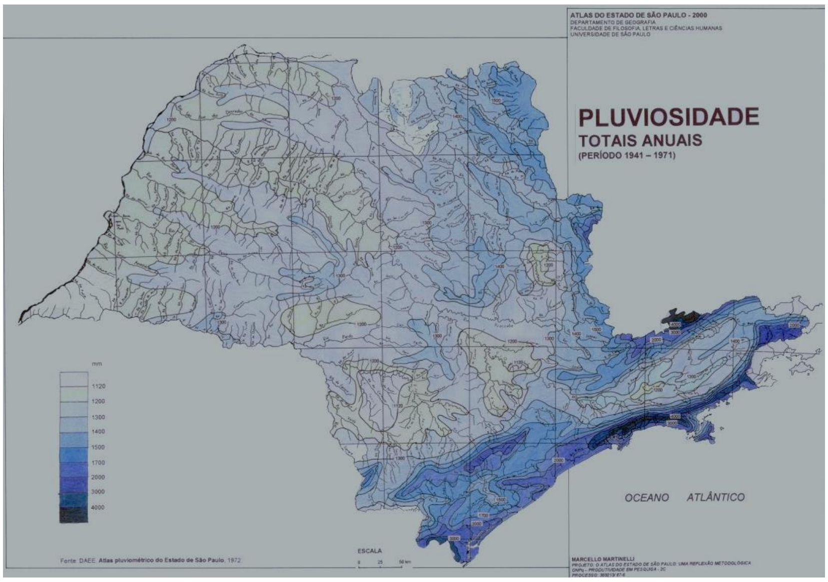


Figure 10- Rain distribution in the State of São Paulo – adapted from Martinelli (2010)

The combination of the rains, the presence of multiple rivers and the mountainous terrain leads to the vulnerability to natural hazards. The main ones are the landslides, mudslides, floods, accelerated erosion (moçorocas in Portuguese) and windstorms. During the drought season there is also another set of risks such as forest fires, but they are less predominant in the state. In Figure 11 it is possible to have an overview of the more common natural hazards in the State. (IG 2008)

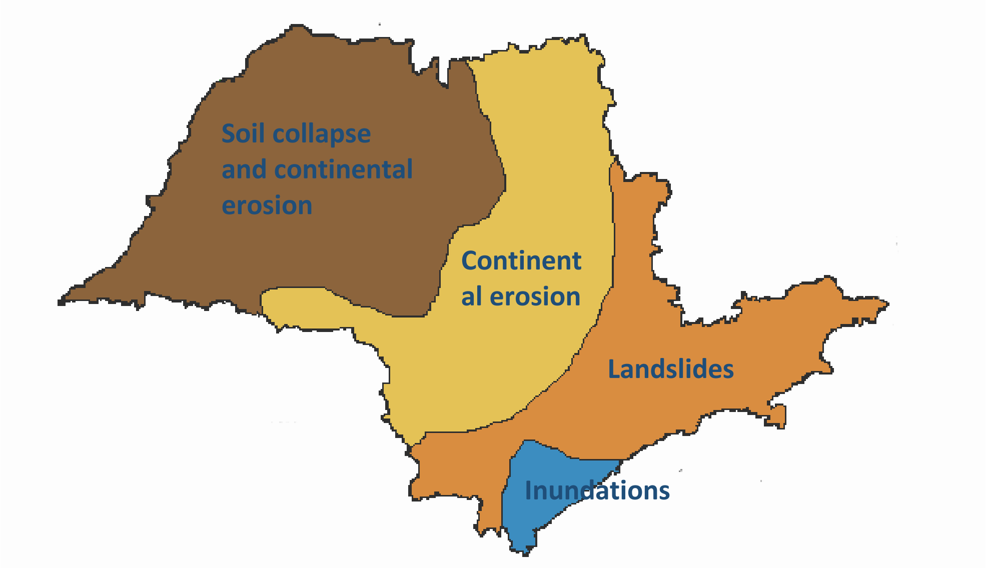


Figure 11- Disaster distribution in São Paulo state (IG,2008 and Cosenza 2015)

3.2 São Paulo State Civil Defense

Civil Defense is the institution in Brazil that is responsible for the protection of the systems in a society during situations that disrupt from normality. The acts of Civil Defense can be divided in actions that happens before and after a crisis. In the group of actions before a crisis it is possible to list the prevention, mitigation and response preparation as the three main responsibilities of the organization.

Colonel Walter Nyakas Júnior leads the Civil Defense of the State of São Paulo during the current administration. The institution follows the values of teamwork and cooperation from all members. The main goals that are even symbolized in their logo are the safety, stability and well fare of all citizens. The logo also shows their main activities of the prevention of disasters and assuring the safety of the population. The blue and orange are the colors chosen to symbolize the tranquility and serenity, and the solidarity and human warmth respectively. The official moto of the institution is "Defesa Civil somos todos nós" which means "Civil Defense is all of us".

The coordination of defense action is the main operational responsibility of the institution. Due to the context of the State weather, the work is organized around combating the raining season and the drought season impacts. Therefore, each year it is the role of the team to update information and reports, elaborate and implement improvement projects and programs as well as evaluating the budget necessary for the yearly operations. In order to support these main activities, it is also important to train human resources, do a review of methods used and make studies of the current capabilities of the system in order to be better prepared in the next season.

Each State in Brazil has its own independent Civil Defense organism. There are guidelines imposed by the National Policy of Civil Defense that try to align all the multiple authorities. Each regional Civil Defense is under responsibility of the State Governor through the State Coordination of Civil Defense (“Coordenadoria Estadual de Proteção e Defesa Civil, CEPDEC”). These are ruled by the law nº 12.608 from April 10th 2012.

In the case of São Paulo it is divided in 19 Regional groups called REPDECS (“Coordenadorias regionais de Defesa Civil”). Moreover, each of these 19 groups have sub organizations at the municipal level, the COMPDECS (“Coordenadorias Municipais de Defesa Civil”). The infrastructure to sustain the entire network of COMPDECS is uneven through the State. As discussed previously the State has a very high level of inequality between its various municipalities. Therefore, in multiple of the poorest regions there is no real facilities nor paid staff. In some cases, the Civil Defense response consists of the work of volunteers and the use of donated/shared space from other government institutions. Despite the military connections of the CEPDEC, the presence of civil and military personnel is a mark of the internal work. Which leads to a lot of REPDEC heads and members being part of fire department, military police or even directly members of the Secretary of the State. The heads of each REPDEC are changed periodically (CEPDEC-SP 2019 website). The Figure 12 contains a quick overview of the organization by region.

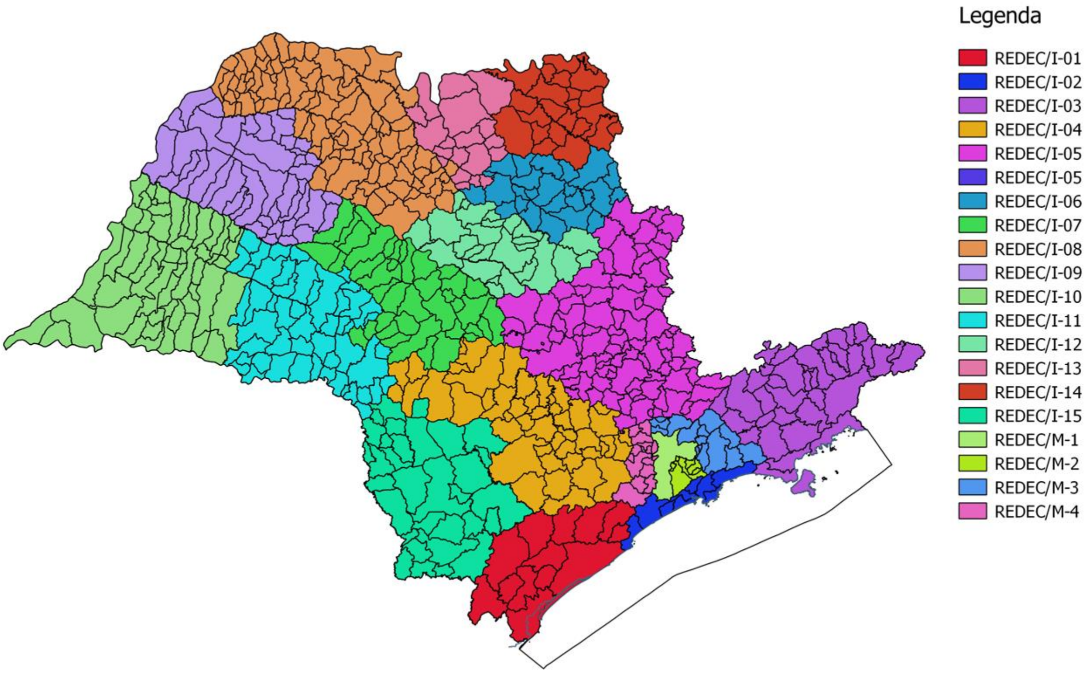


Figure 12- Civil Defense REPDECs (Cosenza 2015)

The current logistic network of CEPDEC-SP is based on a central warehouse in the city of São Paulo at Morumbi and five secondary warehouses in the countryside. The warehouse in São Paulo is the biggest one that centralizes all the purchases and most of the donations, the secondary warehouses have a much lower capacity (at most around 20% of the capacity of the central one. The cities of Apiaí, Aracatuba, Bauru, Caçapava, Presidente Prudente and Registro are the current location of the warehouses. This has changed since the last studies done by Cosenza in 2016 as in that time there were seven warehouses in operation. The five existing ones and two extra in Ribeirão Preto and Araçatuba.

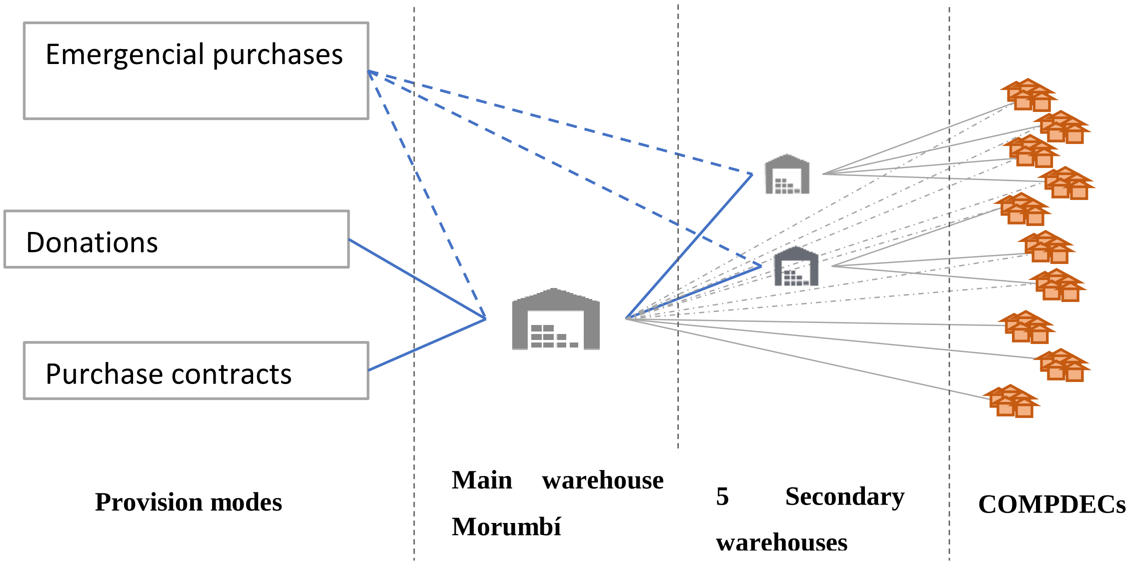


Figure 13 – Updated São Paulo State Civil Defense logistic network structure (adapted Cosenza 2015)

In theory each COMPDEC should receive the material from the closest secondary warehouse when it makes a request. But in reality, there are stock ruptures in the secondary warehouses and then the emergency teams must reach to a more distant warehouse in these circumstances. This impacts negatively the response times to disasters and is a violation of the quality standards expected from the operations.

The central warehouse in the city of São Paulo is in a strategic locations close to the access to the major highways system that are connected to the capital. It has a usable area of approximately 500 square meters for the stock and is being run by only three employees. It is the biggest one in the State and is often used by cities that should be supplied by other warehouses according to the discussions had with the client. With the current system, the CEPDEC department is only responsible for the supplies and maintenance costs of deposits. Because the transport of the supplies from the warehouses to the cities is paid by each municipal authority. Therefore, the supply from the central warehouse to the cities does not impact the costs for CEPDEC, but for a total cost in the system. In the current studies the total costs of all public authorities will be taken into account, into further studies it could be a possibility to analyze only the CEPDEC costs of operation. This information was obtained from a series of interviews and discussions done with the CEPDEC team members.

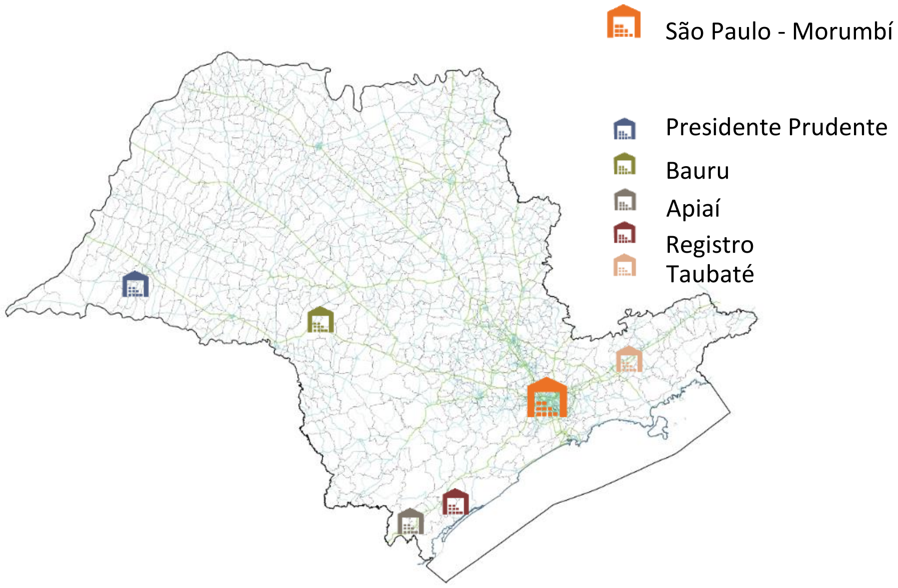


Figure 14- Civil Defense warehouses distribution (adapted Cosenza 2015)

The inventory of the warehouses is mostly going to be constituted of three main categories. The first is the safety equipment and tools for COMPDEC's own personnel and volunteers. The other two categories are relief items for the population that can be categorized internally based on their types. The division is mostly based on the supply kits for the 'Operação estiagem' and the 'Operação Chuvas', basically between the drought season and the raining season kits. The kits for relief have shared subgroups of items. The basic subgroups identified are the following (Cosenza 2015):

- Basic Food Basket: shipped per affected family, contains food to feed four people for a period of 15 days;
- Hygiene Kit: shipped per affected family, contains the basic hygiene tools for a family composed by four people.
- Bedding Kit: shipped per affected person, contains the complete set of mattress and beddings.
- Dressing Kit: shipped per affected person, contains a shirt, a sweater and a pair of tennis shoes.
- Cleaning Kit: shipped per affected family, contains all the tools needed to clean a place after the event.

Figure 15 - Basic food basket and cleaning kit stocked in the central warehouses (REPDEC,2011)

3.3 Types of warehouses

The construction of warehouses will use a PVC canvas warehouse as the base model for the cost projections. This model has thermo-resisting properties and also an estimated lifetime of 10 years of use in service. Each of these warehouses should be able to receive up to a percentage of the stocks from the central warehouse in São Paulo, in the baseline this is going to be 25% from Cosenza (2015). But it will remain as a variable that the user can easily update later on.

Figure 16- Example of Canvas warehouse from Liri Tent

To simplify the optimization, model the possible locations for the installations were reduced to the cities that have a fire department or military police station, over 25 thousand people and cities outside the Great São Paulo area. Each of these simplifications were imagined with the logistical constraints in mind (Cosenza 2015).

The central warehouse is already in the Great São Paulo area and the other two constraints is to assure that the local workforce has the resources to properly operate the warehouses in case the locations are chosen. The final list of candidates has not changed since the work done in 2016 by Cosenza, the same list of candidates can be found in Appendix 3. There was an attempt to reduce the candidate cities to try to improve the execution speed of the models. But as one of the cities (Apiaí) has already a very small population, changing the limits for the cities would impact the possibility of this city as a candidate. So in order to keep the evaluation fair all the criteria remained unchanged.

3.4 Available data

3.4.1 History of interventions

CEPDEC-SP has a few records of the interventions made by the institution. The data is in an Excel format file and contains the supplies requested by the multiple COMPDECs and REPDECs for the emergency operations. There is also another database from them that contains the history of interventions with data such as the number of people affected by each event, number of people forced to leave their houses and the nature of the intervention. It was sent by WhatsApp after multiple exchanges with the client team. This shows that the format of the database is quite fragmented as the individual file sizes were small. The archive required a big processing into consolidating all the different files into a single table.

For the database with the information about people impacted by the disasters, the collection started in 2001. But the constant changes in data format made it possible to use only the information collected from 2006 onwards when the current standards were implemented. The data is aggregated by year and by REPDEC. The data fields available are city of intervention, district and/or street, cause of the interventions, injured people, missing people, deaths, evacuees with alternative housing, homeless evacuees and the nature of the hazard, lightings, floods, landslides, storm, soil collapse. This was the type of data used in the original report and will remain as the main source of information for the interventions.

The database with inventory movements and individual requests has data that starts in September 2012. It has the information on which REPDEC made the request, which city had the demand, the type of material requested, date of request, date of delivery, monetary value of the demanded supplies, warehouse that should respond to the request, warehouse that actually responded to the request. This data was used only as a source for secondary information such as seasonality of the demand, the stock problem issues. As there were 45 different types of items, it is hard to quantify the magnitude of the disasters that are contained in this database. It is for this reason that the data used remained the previously mentioned one.

On this database it is possible to see that there is actually a bigger percentage of the requests fulfilled by the wrong warehouse during the drought season. We also see that there is no growth in the percentage of requests that are being delivered by the wrong warehouse. (2019 data is incomplete). This information is useful in modelling the location problem because it shows the real importance of the warehouse capacity issues. As the % of the volume being delivered by the wrong warehouse during the raining season is small, especially in the interactive model there will be no need to use the capacity as a constraint for the model.

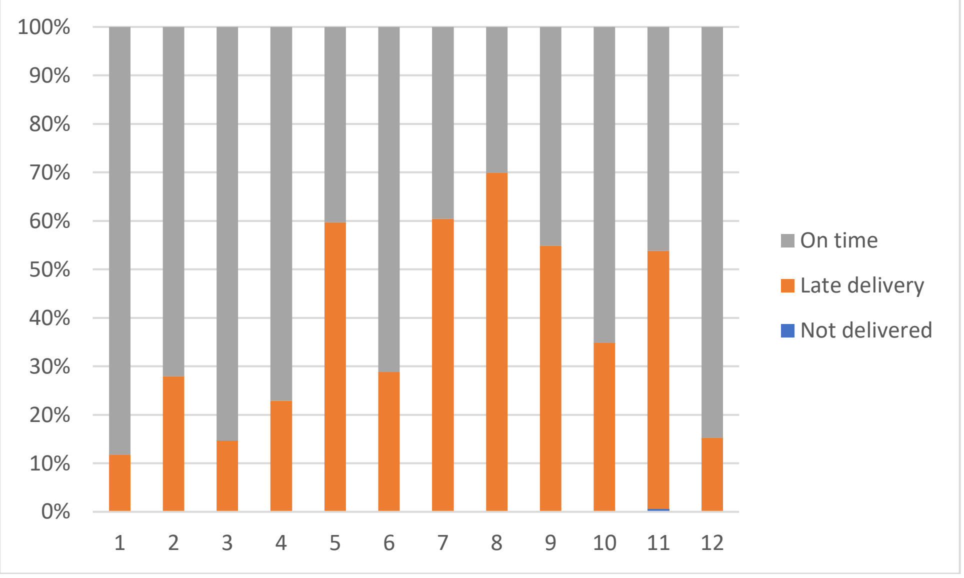


Figure 17- Requests supplied by the correct warehouse by month - CEPDEC data

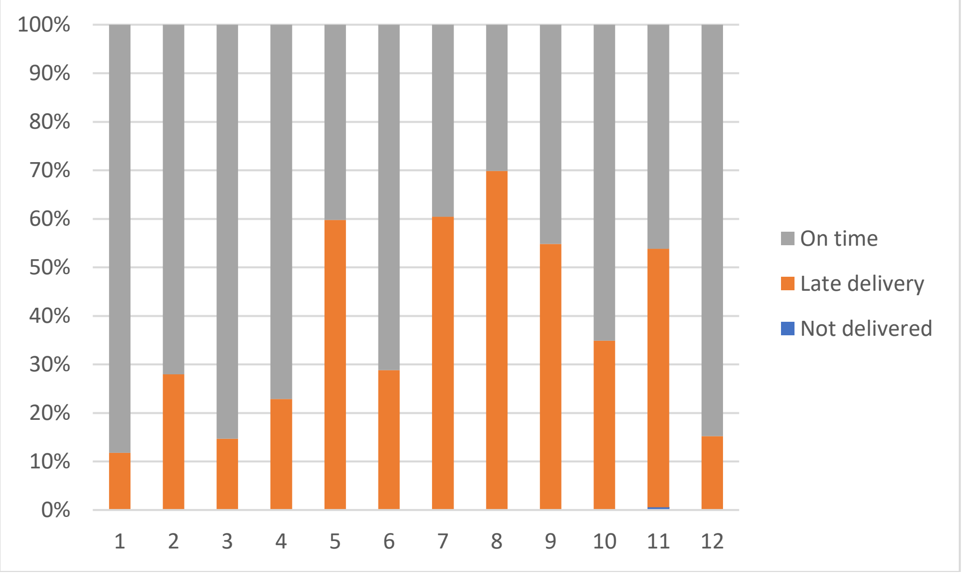


Figure 18- Requests supplied by the correct warehouse by year - CEPDEC data

3.4.2 Risk preparatory plans

The data from the risk preparatory plans can be found on individual documents that reports into the conditions of each city. They can be produced by multiple sources that conducts the analysis. In this report the data from the risk preparatory plans were recovered from Cosenza (2015). An important step would be to update this data if there is time available in order to do this activity. As there are multiple reports that have to be downloaded and read, it becomes a very time-consuming task.

The plans to map risk zones have been developed due to the high frequency of natural disasters. With federal, state and municipal efforts the valuation of these zones and the plans with the mapping of the areas with the biggest risks is something that has evolved during the last few years. The mapping uses the type of hazards as well as the severity and probability of it occurring. There are four identified levels of risk that are going to be mapped:

- R1 - low risk
- R2 - medium risk
- R3 - high risk
- R4 - very high risk

There are multiple institutes that have been involved in the research of the risk levels and the creation of preparatory plans. Some of these institutes are focused on specific regions of the State while others have done research on multiple locations. The four main ones are the following:

- IPT: Institute of technological research (“Instituto de Pesquisa Tecnologica”)
- CBH-RB: Committee of Hydrographic Basins of the “Ribeira de Iguape” and South Littoral.
- CPRM: “Companhia de Pesquisa de Recursos Minerais”, also know Geological service of Brazil
- IGc: Geosciences Institute of the University of São Paulo

The four institutes are combined responsible for 361 out of the 381 risk reports used. This represents risk reports for 245 different cities in the State. Some cities update their risk reports and preparatory plans with a higher frequency and therefore they have multiple risk reports that were made for them/with them. As the date of the reports changes and the structure is also different it is necessary to do a manual aggregation of the data that has led to the mapping of the different zones at risk (Cosenza 2015).

Using the latest risk reports (from the 195 available) and extrapolating the number of houses at risk it is possible to calculate the number of people at risk. It is noticeable that there is a strong correlation between the number of people living in a risk condition to the number of city inhabitants, therefore it is important to normalize the number of people at risk by city inhabitants.

In the available data it was possible to identify 365 cities that needed at least one intervention from the Civil Defense. That means that only 57% of the 645 cities of the state have records of interventions. And from the 645 cities, only 195 have a known number of people living under vulnerable conditions. (Cosenza 2015).

3.5 Previous results and requests made

Cosenza (2015) and Brito Junior (2012) created models and methods to try to calculate the best locations for the warehouses. The method was able to produce results that improved total costs of operation while maintaining the service level to the cities. Some of the proposed solutions were even able to reduce response time while keeping the costs under control. The results were all based on inputs given by the Civil Defense team and all the hypothesis were also discussed with them. In the figure it is possible to see the best solution proposed in 2015.

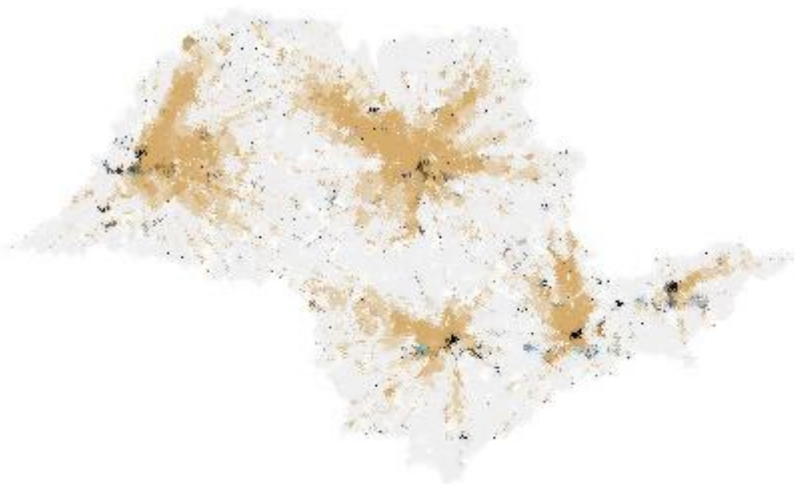


Figure 19- Best solution from location model - Cosenza (2015)

Despite the potential gains, since the presentation of the solution to the Civil Defense administration, the only change to the system was the closure of the warehouse in Ribeirão

Preto. Looking at the numbers of people living under risk conditions in the area, the reduction in costs might be sufficient to justify the lack of responsiveness. The region has fewer people living in risk conditions than other key regions that might have benefits with the extra budget from the closure of this warehouse.

Table 1 - Cost breakdown of the best solution - Cosenza (2015)

	MetCities	MetCities%	MetDem%	MetPop	MetPop%	WHVarCost
'SAO PAULO'	119	18%	65%	24214027	66%	145317
'DRACENA'	163	25%	5%	2344463	6%	13422
'ITAPETININGA'	94	15%	5%	2444161	7%	12146
'MATAO'	229	36%	9%	5797385	16%	23035
'TAUBATE'	40	6%	15%	2006467	5%	34405

Custo de transporte R$	228325
Custo de gestão do estoque R$	20000
Depreciações R$	40000
Custo anual total R$	**288325**

An analysis such as this one indicates that it is possible that a multi-criteria decision-making model should be used. So instead of simply minimizing total costs, it should be a balance with the service levels as well. In this case the final recommendation will probably differ from the current one.

A change since the last studies is the discussion that has been happening to launch a few State government bases in multiple cities of the State. They would be buildings with the presence of members of all the State government ministries. One of the discussions was the possibility to have dedicated stock areas for Civil Defense in these offices. This would completely change the potential candidates for the warehouse locations and would also change the costs structure. Flexibility for the tool developed will grow in importance as the consolidation of this prospects is something that will happen after the conclusion of this project.

The previous system became too complex for the independent operation of the models from the CEPDEC team. So they would be completely dependent on the analysis that had to be done by the CEPED/CISLOG team. This is not a sustainable way of working in the long term and would be a big barrier to the implementation of the system. Also by not being able to understand what was happening and by having limited access to the system inputs they were less confident in the results.

The visualization of the results was also another critical point. The simple presentation on the map was not sufficient to provide all the necessary insights and discussions. And as it was a static result presented in a PowerPoint format it didn’t provide enough to advance the discussion on the necessary evolutions for the current system. In order to achieve that the model must step away from the pure scholar approach and provide an integrated system that can be run by the department team. These are the two main goals that became the focus of the project once their understanding became clear since the early stages.

So after this introduction to the problem the reader can either skip to the results in chapters five, six and seven or go to the chapter four to understand the exact methodology that will be used. Before reading the conclusion it is advised to quickly review the context presented in this chapter. Then the link between the context and the results will be more clear.

4 Methodology

This chapter will discuss the methods and architecture chosen to deal with the problem faced. Readers could start the chapter with a quick visualization of the results in the following link: https://drive.google.com/drive/folders/1LnyvM9gIwUfwwkK2SJjqtkZJ2lib8RgW?usp=sharing

The method chosen will be presented in the first section with a diagram that synthetizes the different sections of the work. Then there will be further discussion on the specific behind the methods chosen as well as some alternatives that were considered but not taken.

4.1 Methodology details

In order to calculate the best architecture for the warehouse locations and then to show the results to the Civil Defense team there will be a couple major steps in the process. The integration of all this step has to be documented in order to create a sustainable system that can become part of the operations of the institution and not a presentation that didn't have concrete impacts in the organization.

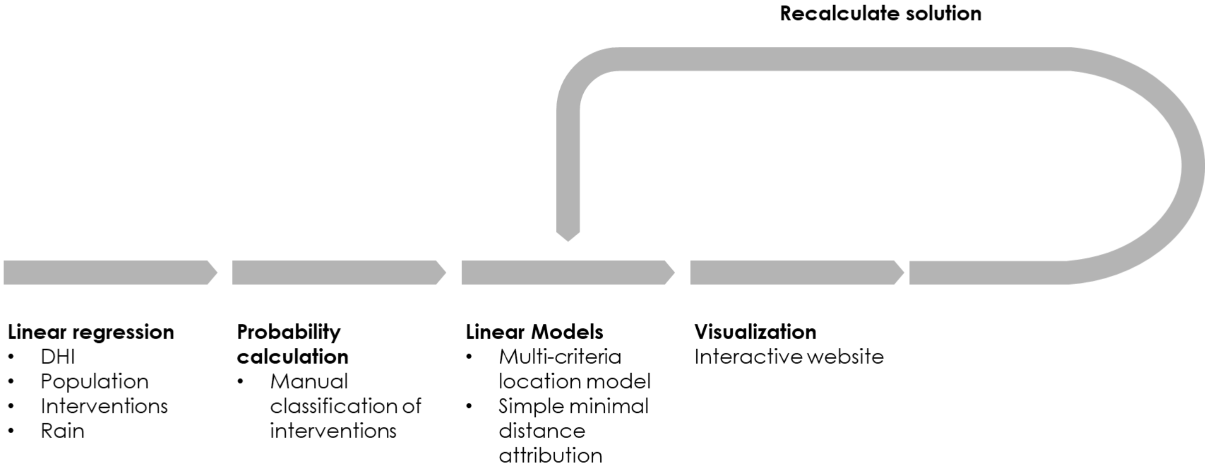


Figure 20a- Architecture of the proposed interactive solution

The first step is building the regression model that is going to be able to estimate the demand for emergency supplies of each city of the state. This will use multiple sources of data that will be tested as predictors to the final model. The basis for the prediction model has already been established in Cosenza (2015), but some alterations will be necessary to his models. The need for this regression comes from the lack of data for all the cities in the State. Therefore, the cities that have known numbers will be used as a learning subset for the regression and data input, and then the results will be extrapolated to the remaining cities. This step is modular from the rest and could be replaced by internal calculations from the CEPDEC team.

In order to use the results from the linear regression model, there will be a creation of multiple scenarios with different magnitudes of disaster. The probabilities of each scenario will be calculated with the historical data available to be then used as the demand in the location model. Then the user will be able to choose which scenario he/she wants to test. With this change, the stochastic model becomes a deterministic one and much more simple. Instead of only minimizing the total operational and installation cost, it will also take into consideration the service levels and response time of the system to the emergency requests.

The choice was to skip the modeling of the disruption of road infrastructure during the disasters. One of the main reasons would be the complexity to model such scenarios in the optimization model. The second reason is because Brazil infrastructure is considered within low risk to be affected by hazards (Hagenlocher 2016).

The capacity of deposits was not taken into consideration in the previous model, but it is an important feature as in the reality, if a deposit is out of stock the demand will be met by another one that might be much more distant. Therefore, the warehouse capacity should be included in the model constraints if there are no further barriers to its implementation. The addition of this will be conditioned to the impacts it will have in the model complexity and the interactivity of the system with the website.

The objective is to have a server running the application with the parameters discussed with the Civil Defense team in advance. The best alternative found to make the website responsive is to record the results of some tests that are ran. And then for the manual tests where the user can choose the parameters the model applied changes. Instead of running an optimization based on the location model, the problem will be for the users to set the locations they want and then the KPIs will be calculated and the relationship between cities and warehouses as well.

This second model is where the biggest number of interactions will happen. And from the discussions with the CEPDEC-SP team and the circumstances of the project, it is expected that this model will be the one used the most by the client. Because of this expectation, this is the page that has received the biggest focus between all the others.

The non-interactive version should even receive a different diagram for its architecture. It does not contains sufficient interactive elements to be considered an interactive webpage. It simply let users choose which of the scenarios they would like to see.

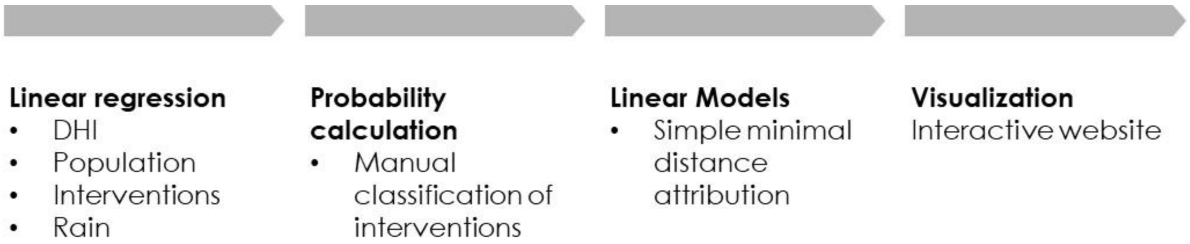


Figure 21b- Architecture of the proposed recommendation based solution

With the methodology in mind, the reader should follow to the next chapters to see the implementation discussed in this chapter. We will start with the data processing in the next chapter, and then discuss the results in the chapter six and the visualization in the chapter seven. In the conclusion, the reader can also find a critical review to the method described here in this chapter and recommendations for improvements that could be made to it.

5 Data processing

This chapter will present in chronological order the steps necessary to arrive in the results that will be presented in the chapter six. The order follows the steps necessary to go from the raw data to the final linear models that gives to the users an answer to the problem. The data starts in the regression where multiple datasets are used in order to arrive at the estimation of people living under risk conditions. Then the interventions dataset is used to calculate the probability of the hazard events to happen in each city. This will be used as the demand input into the location models.

5.1 Regression modelling and use

The regression process is necessary as there are multiple cities with some data of intervention, DHI and population but no calculated number of people living at risk. Therefore, the goal is to help to augment the number of cities with estimated population at risk so that the location problem can have more realistic demand scenarios. The sample with the people living under risk conditions will be used to create the model that will then be applied to the other cities. The data of people living under risk conditions was taken from multiple sources, in this case there were no updates to this dataset that was used from the previous work and the data was taken from Cosenza (2015).

5.1.1 Regression workflow

In the first section of the chapter, the regression process will be described. As discussed previously, the goal of the regression is to determine the number of people living at risk conditions in the cases where direct data is not available. In order to avoid creating a false demand in cities that do not have any historical interventions, the model will be only for the cities that have at least 1 intervention made by the Civil Defense since 2015 when the current data series begins.

$$[People\ living\ at\ risk]_i = f_i(available\ data) \qquad i = 1,2,3,4$$

The initial stage is to search for additional datasets and update the old ones if possible. These sets will be used as possible elements in the regression models. The data used in the first version of the model was used as an initial base, but it was adapted and evolved accordingly.

Table 2- Data sources for linear regression model

Old model data input	**New model data input**
DHI by city	DHI by city
Historic of interventions classified by type	Historic of interventions classified by type
Population	Population
Average density of the city	Average density of the city
-	Pluviometry levels

With a small amount of values to use as a base (cities with complete datasets including the amount of people living at risk), in general it will be important to find a model that works well with the least amount of predictor variables. It does not mean removing predictors just to simplify the model, but to find a balance in having enough significance for the model and also the aiming for simplicity.

Each dataset had to be processed in order to get ready for the use in the regression process. The data gathered had to be aligned to the correct cities through a cleaning process in each case. And for some cases it was necessary to do some assumptions to fill missing values (such as pluviometry levels in each city).

Table 3- Predictor variables for the linear regression model

Predictor variables	Assumption
Number of interventions/ affected population	The high historical number of interventions should indicate higher risk conditions for the population in an area
Human Development Index	Higher HDIs should indicate better capacity to respond to hazards. So, less people would be affected and therefore the expected correlation is negative.
Pluviometry level	The hazards that generate the biggest disasters in the State of São Paulo are often connected to the rain. Regions with higher levels of rain specially during summer should have a higher amount of people living at risk

With the datasets ready, the work done was to replicate the analysis made by Cosenza in 2016. The expectations were to find similar patterns in all the models created. There is a need to create multiple models in order to estimate the number of people living under each class of risk conditions. Which means that there are four models that will predict each of the R1, R2, R3 and R4 affected people.

5.1.1.1 Linear regression, generalized regression and zero inflated models

The expectations for the tests on this model was to not use them as the final one because of the findings of Cosenza (2015). The problems with the usage are connected to residuals distributions that do not follow a normal distribution and therefore makes it harder to test the model fitting quality and its statistical significance. The results of the coefficients were impacted by the big quantity of zeros and differ completely from what the common sense would indicate. The datasets still had the same general characteristics of multiple zeroes.

By using the technique there is a need to fit all the available data under one statistical distribution. This relationship between distributions and the data were unchanged from the previous model where the Poisson distribution was identified as having the best fit. When the number is normalized by 1 million inhabitants the best statistical model to fit it is the negative binomial distribution.

The process was iterative and multiple models were developed and tested. But since there was a starting point with the first model, the convergence of the different tests happened in a faster way. The adaptation of the variable for cases per 1000 inhabitants was found out to be the best one in the previous method and this hypothesis was tested again. To better work with this there is a need for the logarithmic transformation. Due to the nature of the dataset (high frequency of zeros in the dependent variable), the zero inflated linear regression was one of the major focus of the work.

5.1.1.2 The tools used

The implementation of the regression model was made using the R in RStudio. The packages used were the same from previous works:

- Generalized linear model for negative binomial distributed dependent variables: glm.nb from the package MASS.
- Zero-inflated models: zeroinfl function from the pscl package.

5.1.2 Regression model for people living in risk condition R1, R2, R3 and R4

The initial hypothesis for the models that are going to be found can be found on the following table. Basically, it assumes that the previous models will be held true. The difference in the new study is that there is an expectation on the relevance of each independent variable, and specially how the rain factor will impact the other ones. Since the data for known R1, R2, R3 and R4 risk living people was not updated, the fitting of these variables into known distributions can be taken from the previous model. Therefore, the R1 will again be discarded since the relevant tests cannot be made due to the unfitness of the data to the appropriate measures.

Table 4- Expectation for regression models

Risk type	Expectation
R1	With the risk class data not changing, it will remain without any regression calculations due to the lack of fit to the distributions used
R2	Zero inflated with negative binomial model
R3	Zero inflated with negative binomial model
R4	Zero inflated with negative binomial model

From all the populations at risk, the R4 dataset is the one that has the highest correlation to the rains. This is as expected as the people under highest risk are those only impacted by periods of intense rain according to the conditions in the State of São Paulo. But the correlation is still very low at only 0.12. The other variables such as population have correlations in the 70% range. This should indicate that rain will not be a good predictor in the model, but it will have to be tested later with the regression models and will not be discarded just because of the correlation results. In the results below, *chuva* corresponds to the annual rain while *chuva2* corresponds to the rain during the raining season.

```
> cor(r1,chuva2)
[1] -0.0004357691
> cor(r1,chuva)
[1] -0.02415054

> cor(r2,chuva)
[1] 0.05562514
> cor(r2,chuvas2)
[1] 0.09878834

> cor(r3,chuvas)
[1] 0.08842565
> cor(r3,chuvas2)
[1] 0.1239506

> cor(r4,chuva)
[1] 0.05979252
> cor(r4,chuva2)
[1] 0.1056137
```

Figure 22- Results of correlation analysis of Rain and population at risk conditions

The fitting of the model was recovered from the previous iteration of the project. Therefore there was only the need to update the regression models with the new data and new variables. So R1 will not be tested due to the lack of fit of the data to any of the proposed tests. The method consisted in testing the different models with the different datasets until we identified the best performing one for each of R2, R3 and R3.

For R2 we started with tests from R2 with all the variables and the adjustment of R2 affected people by 1000000 inhabitants. With these settings there were errors that were encountered so the decision was to go back directly to the population living at R2 risk. Even with this change the overall performance was similar to the previous iteration. The change is simply a way to try to better fit the data at each time in order to try to apply the models. The addition of the rain did not improve the overall results as shown in the results below, the rain had a bad probability and decreased the global performance. The second result will be the one that will be used later on in the estimation of R2 people in the cities where the data is not directly available.

```
> summary(zeroinfl(r2~idh + intervencoes + chuvas, dist="negbin"))

Call:
zeroinfl(formula = r2 ~ idh + intervencoes + chuvas, dist = "negbin")

Pearson residuals:
     Min       1Q   Median       3Q      Max 
-0.57806 -0.52082 -0.43096 -0.04214  7.05532 

Count model coefficients (negbin with log link):
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -4.9505167  3.4202867  -1.447  0.14779    
idh          12.7188268  4.5382310   2.803  0.00507 ** 
intervencoes  0.1046315  0.0254782   4.107 4.01e-05 ***
chuvas        0.0001240  0.0003443   0.360  0.71860    
Log(theta)   -0.9157526  0.2258244  -4.055 5.01e-05 ***
```

Figure 23- Sample results of zero-inflated regression analysis in R Studio

The same pattern was identified in the analysis for the estimators of R3 and R4. The addition of the rain as a factor into the regression did not help the model to improve its accuracy. So the option was to remove it from the list and remain with the previous variables in the regression model. So after the identification of these patterns, the necessary adaptations were made and then we arrive at the final table of inputs

Table 5- Coefficients for R2, R3 and R4 estimation

ESTIMATED VALUE	INTECEPT	DHI	INTERVENTIONS HISTORY	# AFFECTED PEOPLE	LOG THETA
R2	-4.46263	12.54927	0.10493	-	-0.91928
R3	-4.5392415	13.4272024	-	0.0011361	-1.0970388
R4	-3.1686	10.9091474	-	0.0013703	-0.9030111

Despite a small decrease in confidence in the model compared to the previous study, the alternatives studied presented even worse results. Direct GLM models were tested and performed very poorly. The GLMs were tested with different combinations of variables but they all netted worse confidence in the models. This decrease in performance could come from the different historic for the interventions used or from the different method of classifying the number of affected people. In the end, the correlation found for the variables and the probabilities are still very good, the drop in performance came only from the intercept value. So the decision was to proceed with the current results.

5.2 Historic analyses and scenarios' probability definition

5.2.1 Classification of the interventions

5.2.1.1 Classifications definition

There are four levels of disasters used as a base for the classification of the interventions. These levels are defined based on the number of people living under risk conditions that are affected by a given event. So, the magnitude of a disaster will impact the population that is going to be impacted and suffer its consequences. The higher the magnitude levels imply that even people living under safer conditions will be impacted by the consequences of the current situation.

Disasters of the level n1 will impact people living under R4 conditions, so it is the least impactful disaster on the scale. Then for n2 people living on R3 and R4 will be impacted by the disaster. Then n3 will have people on R2, R3 and R4 and finally n4 have people on R1 until R4 being impacted by the disaster and also people not considered by the existing municipal preparatory plans.

Table 6- Disaster levels used in the construction of the models

	R2	**R3**	**R4**
n1			x
n2		x	x
n3	x	x	x

5.2.1.2 Analyses of the history of interventions

By using the definitions established in the previous section, the interventions in the Civil Defense database were classed. In order to be able to class smaller interventions there was a need for the creation of additional categories. Those categories will be based on 10% of n1 and 50% of n1, which means that only a percentage of people living under the risk condition R4 have been affected by these disasters. Then there will be also other scenarios classified following the n group of events. For exemple: n3 (100% of R4), n4 (R3 and R4), etc. The details of this classification can be seen in the section 5.2.3.

It is noticeable the number of interventions with a rating of 0, which means that a lot of the efforts are made to minor requests in terms of number of impacted people. Interesting to note that the percentage of events of above the n3 scale is very close to events of the n3 level. Which indicates that there is a substantial risk of having a very large number of people affected by disasters if compared to only having the people living under very high risk affected (this can be seen in the section 5.2.3).

5.2.2 Probability of the disasters per level

The frequency of each disaster was calculated in the Section 5.2.1.2 when the evets were classed. With two major assumptions, the probabilities of each scenario will be calculated from the numbers obtained in the previous section. The necessary hypothesis to be able to create this relationship are: to assume that each disaster is represented in the historic as one intervention, to assume that each event occurs in a single day. As a consequence of these assumptions, one single day with disasters in different cities will account for multiple interventions in each of them.

Then it is necessary to divide the number of interventions by category by the duration considered in the historic. The results of this division are the probability of encountering a day where there is an occurrence of a disaster in each city. Another assumption is that there is no major trend of growth or decrease for natural hazards. So, the past data is going to be used to predict the future without having to calculate a growth rate for the probability of natural hazards.

5.2.3 Scenarios and their probabilities

For the six categories previously defined, the affected people used on the demand assumption will be the worst case possible. Which means that all the events between two definitions will be given the impact of the higher level. With this assumption made and the calculations completed, it is possible to arrive at a simple table with the amount of people affected and the probability of the event to happen:

Table 7- Scenarios description and probabilities

	N1	**N2**	**N3**	**N4**	**N5**	**N6**	**Period**
Description	10% of R4	50% of R4	R4	R3+R4	R2+R3+R4	2 (R2+R3+R4)	-
Probability by scenario	24,8%	5,3%	2,1%	6,8%	3,5%	3,5%	2015-2018/08

The scenarios were used together in a stochastic model in the original work from Cosenza (2015). But in this case they will be used independently and the users will simply test different scenarios and the impacts on the solutions.

5.3 Multiple models

There are going to be two models being used in parallel. The first one has the goal to calculate the best architecture for the system based on the scenarios given. The evaluation of the alternatives will be made based on a multi-objective basis, with the minimization of costs and non-coverage level. The second model is going to be much simpler. Given a set of locations for the warehouses it will calculate the attribution of each city to each warehouse. Both of them will share the stage where the demands are generated using the inputs from the regression model discussed previously. So basically, the difference is that the first stage in the multi-objective location model is replaced by a deterministic and fixed scenario before the second stage is applied.

5.3.1 The multi-objective location model

The location model will have two main objectives. To minimize total costs and to maximize the coverage of the network. In order to do it, there is a necessity to work with the minimization of the non-coverage level. In this way the objective function can work with the shared goals. In order to implement this, it is necessary to choose weights to each of these two goals. This is something that will have great impacts in the final results and therefore is something that will have to be studied more in depth. The sensitivity analysis becomes a key process for the presentation of results.

The cost is composed by three terms: transportation costs, annual depreciation for the investment in the warehouse and stock management. The coverage level will simply be a number of cities that are outside the desired range of service (six hours is the base for this but a sensitivity analysis of this limit is also important).

The first iteration of the problem had 169 possible candidate deposits, 645 demand point and six scenarios (Cosenza 2015). In order to improve the performance of the model and enable the generation of multiple simulations to present to the users there was a need for simplifying some of these dimensions if the real time interaction for the model was the goal. After a review of the candidate locations and the scenarios with the Civil Defense team, the numbers were kept the same as their current situation needs to keep these possibilities open. Therefore, the interaction with the higher complexity models was discarded and instead the approach will aim to have two distinct set of goals.

In the first iteration, capacity constraints were not used due to the nature of the model. By that it is implied that the hypothesis taken while constructing the demand generator are against the construction of this constraint. The demand generation model does not include a timed sequence of events; therefore, the stocks would need to be evaluated based on the total demand for a year. But in reality, as this is important there will be adaptations to the stock constraints that will enable to use this feature in the second iteration. Then once the first tests are made we will finally make the conclusion about adding or not the capacity as a constraint in the final model.

5.3.1.1 **Model Description**

5.3.1.1.1 Sets:

I: Candidate locations for warehouses ($i \in I$)

J: Cities in the State that will serve as demand points($j \in J$)

5.3.1.1.2 Decision variables

First stage

X_i: Binary, does the i-th candidate receive or not a deposit

Second stage

A_{ij}: Binary variable, 1 if the warehouse in location i attends the demand of the j-th city

5.3.1.1.3 Parameters:

t_{ij}: Travel time between city i and city j (minutes)

d_{ij}: Distance between cities i and j (km)

q_{ij}: Binary variable, 1 if $t_{ij} > t_{limit}$ and 0 otherwise

$demand_j$: Demands calculated for each city j

$tcost_{ij}$: Cost for transportation of the Demand D_{ij} from i to j

$whcost_i$: Warehouse i installation cost

$whmancost_i$: Warehouse i stock management cost

$whcap_i$: Warehouse i stock capacity

ny: Warehouses lifetime

$whmax$: Maximum number of open warehouses

$whmim$: Minimum number of open warehouses

t_{limit}: The max travel time to attend a city that is considered acceptable

$bigM$: Auxiliary parameter used for an iff operation

k: Auxiliary parameter to control the importance given to the cost component and the coverage component of the goal function

5.3.1.1.4 Objective function

Minimizes the first stage objective function, warehouses fixed costs, and the second stage objective function, considering the transportation costs for each scenario:

$$\min \sum_{i} \left(\frac{whcost_i}{ny} + whmancost_i \right) X_i + [\,Q] + k * E^c\,[R] \qquad (1)$$

$$Q = \sum_{ij} tcost_{ij}\; A_{\;ij} \qquad (2)$$

(3)

$$R = \sum_{ij} demand_j \; * \; A_{\;ij} * q_{ij}$$

5.3.1.1.5 Constraints:

Limits for the number of open warehouses:

$$whmin \le \sum_{i} X_i \le whmax \qquad (4)$$

If and only if (*iff)* bilateral relationship for stating that a deposit is going to be opened if and only if it will supply at least one city:

$$\forall i \;\; \sum_{j} A_{ij} \ge X_i \qquad (5a)$$

$$\forall i \quad X_i\; bigM \ge \sum_{j} A_{ij} \qquad (5b)$$

Each city has to be attended by at least a warehouse in each scenario:

$$\forall\, j \;\; \sum_{j} A_{ij} \ge 1 \qquad (6)$$

The demands of the cities being supplied by the warehouse must respect the warehouse capacity. This assumes that each warehouse must meet the entire demand of the cities it supplies to. This is a constraint that might be turned off in the final deployment of the model.

$$\forall i \ \sum_{j} A_{ij} * demand_j \leq whcap_i \qquad (7)$$

5.3.2 The fixed location model

The fixed location model is much simpler than the multi-objective two stage process defined previously. In this case, with the given locations for the warehouses the goal is simply to assign the cities to the closest warehouse. Warehouse capacities were not considered in this model in order to preserve the simplicity of the model. The simplicity over the fidelity was important in this stage due to the need to create real-time interactions with users while calculating the model answers. Due to its simplicity there was an heuristic used to solve it instead of using Gurobi as a Solver as it was done in the multi-criteria model.

5.3.2.1 **Model Description**

5.3.2.1.1 Sets, decision variables and parameters

Same as the previous model.

5.3.2.1.2 Objective function

Minimizes the first stage objective function, warehouses fixed costs, and the second stage objective function, considering the transportation costs for each scenario:

$$\min \sum_{i} [\,Q] \qquad (1)$$

$$Q = \min \sum_{ij} tcost_{ij} \ A_{ij} \qquad (2)$$

Constraints:

Each city has to be attended by at least a warehouse in each scenario:

$$\forall\, j \ \sum_{j} A_{ij} \geq 1 \qquad (3)$$

5.3.2.2 Model implementation

The model was implemented using Gurobi. The code was adapted from the AIMMS version from the previous project.

5.4 Location Model's parameters description

5.4.1 Scenarios

The six scenarios from Cosenza 2015 were kept as the base for the evaluation. Each of them represents a different magnitude of disaster happening. The six available scenarios are:

Table 8- Scenarios of impacted people from Cosenza (2015)

	I	II	III	IV	V	VI
R1						
R2					100%	200%
R3				100%	100%	200%
R4	10%	50%	100%	100%	100%	200%

As previously discussed, the probability of each disaster/scenario uses the Civil Defense history as the main predictor. It is also important to notice that each intervention was classed individually based on a series of criteria discussed in the chapter 5.

In the new iteration the goal is to create more flexibility in the scenarios built. For example, in years where there is the El Niño phenomenon there is an increased rain in the coastline regions. But it does not make sense to raise the probabilities of the disasters in all the regions for this year, only for some specific cities. So, the goal is to create some manual scenarios that will be aggregated to the base six but also will be run separately.

Table 9- Scenarios elaboration example

Scenario	Impact proposed
A	Level I proposed by Cosenza (2015)
B	Level III proposed by Cosenza (2015)
C	Level V proposed by Cosenza (2015)
D	Level II for cluster regions outside the Ribeira Valley, Level IV for Ribeira Valley
E	Level III for cluster regions outside special zones, Level V for special zones (simulation of a rainy year in the coastline)
F	Level IV for cluster regions outside special zones, Level VI for special zones (heavy rains across the entire state)

In order to allow for a simulation of other combination of scenarios there is a tabulation of regions that were grouped and can be affected by different levels of risks. This was developed with the goal to make the model more flexible. In order to make it a viable interactive model, this was not applicable directly to the interactive model but rather to the parameter setup in Excel that will be explained in chapter six. The form about these attributions can be sent in the Excel file format. Then, the impacts on the Location model solving is a quick process that can be recovered from the Excel file to the python program. The focus was again in transforming the academic research into something that could yield real changes in the organization.

For this, the cities will be grouped in clusters where there are going to be manual coefficients applied in the definition of some scenarios. In order to apply the coefficients in the model, they will be used straight into the probabilities of the scenarios happening in the respective clusters. The clusters will not be entirely based on the REPDECs but also in other characteristics such as the climate, terrain and socioeconomically factors. The exact list of the clusters generated can be found on the Appendix 6. The different coefficients for the simulations will be presented in the results section as this will also be part of the interactive options that the users will have. The clusters are relative to the current warehouse that serves the cities:

The creation of the clusters allows for stress testing new models based on the extra demand current faced by the existing warehouses. Therefore, it is a better classification method than having to decide based on other criteria. The diversity of classifications that could be made would increase the complexity of the work without having necessarily any added value. The final scenarios created are:

Table 10- New scenarios created

DIVERSE SCENARIO ATTRIBUTION IN THE NEW MODEL

CLUSTER	A	B	C	D	E	F
1	I	III	V	II	III	IV
2	I	III	V	II	III	IV
3	I	III	V	II	III	IV
4	I	III	V	II	III	VI
5	I	III	V	II	V	VI
6	I	III	V	IV	V	VI

5.4.1.1 **Demand**

The demand for each city is based on the number of people affected by each scenario and the probability of the scenario to occur. Each person that is impacted by the crisis should receive one complete individual package (with 1 dressing kit and 1 bedding kit) and 25% of a family package (cleaning kit, hygiene kit and basic food basket). So, for this parameter is the main connection between the linear regression model and the location model currently being explained.

5.4.1.2 **Travel time and distance, max travel time**

The goal of the model is to reduce the total costs of managing the entire system while keeping the response to a minimum rate. The rate is a little bit flexible, but in general the goal is to be able to respond to any disaster in any city within eight hours. This number was discussed with the CEPDEC-SP team and the 2-hour preparation time needed in the warehouse leads to a desired limit of 6h of transport time. In order to get to the calculation of time to deliver from the warehouse to the affected cities there will be a simple hypothesis for the average speed that is going to be used to divide the distance between the two points. After discussions with the Civil Defense team, the average speed chosen as a base will be of 50km per hour.

$$t_{ij} = d_{ij}/50$$

A simple approach would be to calculate the distance between cities using only the latitude and longitude. But this could have very negative impacts in the real-world application of the results. In order to have a better estimation of the distances the better option is to use a routing tool. So by using an automated tool and Google Maps API we obtained the distance between all the distances between the cities in the State of São Paulo (CISLOG – 2018).

The city center of each city is going to be used as a reference for the model (apart from the central warehouse in Morumbi region at RMSP). The data used in this model was taken from Cosenza 2015 as there are no major changes to the data. The recovery of the coordinates was made using a Google Maps free API.

5.4.1.3 **Transportation cost**

The transportation cost can be seen as a function of distance and supplies being transported. With data from the National Agency of Cargo transportation & Logistics it is possible to create an equation to calculate the total transportation costs (Cosenza 2015 and NTC 2014). The 2015 coefficients for the equation were the following:

$$tcost_{ij}^{c} = 10.34 + 3.384 \cdot 10^{-1} demand_{j}^{c} + \ 1.191 \cdot 10^{-2} d_{ij} + 3.823 \cdot 10^{-4} demand_{ij}^{c} \colon d_{ij}$$

The transportation costs calculation will be updated using NTC as a source again. The recent annual NTC annual reports do not contain any table where the direct coefficients could be updated. There are certainly changes in the costs due to the fuel price increase that will be applied in a global way. NTC report in 2017 indicates that transport cost evolution has increased 12.13% in the last 24 months. Using this factor we can arrive at the new coefficients with fairly reasonable values.

Therefore, the final equation becomes:

$$tcost_{ij}^{c} = 11.5942 + 3.74334792 * 10^{-1} demand_{j}^{c} * kit\ weight + 1.3354683 \ \cdot 10^{-2} d_{ij} + 4.2867299 \cdot 10^{-4} demand_{j}^{c} \colon d_{ij} * kit\ weight$$

5.4.1.4 **Warehouse costs**

The cost of implementing a new warehouse was provided by CEPDEC SP during the discussions in 2015. It consists of the canvas warehouse model used in the locations, plus the cost of the equipment used to move and prepare the material. The cost of the canvas and equipment was amortized for their lifetimes in order to provide an annual cost to be used in the model. The costs for the management of the have been considered according to Brito Junior

(2015). The final numbers are ~R$100.000 for the installation costs and ~R$4000 monthly costs for the operation of each warehouse.

5.4.1.5 **Number of warehouses**

The current system has six secondary warehouses plus the central warehouse. According to the CEPDEC-SP guidelines there could be an upper limit of 1 warehouse per countryside REPDEC. So, the maximum limit is 16, and while there is no lower limit guideline, with the time constraint it is hard to imagine a solution with less than five warehouses. Therefore the two numbers were chosen as the limitations to the model.

This is a factor that might change in the near future. In the State government plans there might be space for the creation of more warehouses based on the needs of other Organizations within the State institutions. The idea to have multiple units across the state with members of all the units lead by the governor, and in some of them have a Civil Defense dedicated warehouse area. For the current project the modeling will not consider this fact as it would completely change the base hypothesis.

5.4.1.6 **Warehouse capacity**

The real-life capacity of each warehouse depends on each candidate location and each existing one. The Central warehouse is the biggest one in the system and the remaining ones have at most 20% of its capacity. In the past modeling this factor was not used in order to avoid further difficulties into the solving of the model.

In further discussions with the Civil Defense team it was a request of the team to be able to include warehouse capacity in the list of constraints. According to them, this is currently a big issue with stock ruptures in the secondary locations. It creates an unbalance in the system and a growing number of supplies request to the central warehouse, which also means a decrease in the quality of service. When a request arrives in the unassigned warehouse not only there is an increased risk of having a delivery time of over 8h in the individual request, but also an increased risk of rupture in this warehouse. Which means that there is an increased risk that a request made by a city within the 6h limit from the warehouse will need to be answered by another warehouse because its stocks were depleted.

So, this could be added as one of the constraints. To simplify the calculations, the limits of the existing warehouses were used directly, and the prospected ones were all simulated with the same capacity. But the direct use of the raw capacity of the deposit is wrong. The requests from different cities in different disasters do not happen at the same time, so there is the possibility to reallocate stocks from one deposit to the other. There could also be new equipment bought during the year to replenish the stock levels.

After discussions with the Civil Defense team it was clear that the transportation between two secondary deposits is rare, so it is not going to be considered. Another point discussed was the procurement of new supplies instead of having to use only the existing stock in order to meet the demands. Currently the deposits should be prepared to respond to the entire demand of the raining season with only the stock already in the system. This happens because of three major reasons:

- Raining season concentrates a big percentage of the demands in a short period of time
- Raining season begins in December/January when there are two factors that blocks new procurements to be made
    - End of budgetary year for the government
    - Holidays in some of the suppliers
- The lead time is over 1 month

So, the stock capacity necessity will be modeled as the percentage of the annual demand that is used during the raining season. In order to do that we will use the individual records of demands that is provided in the warehouse database.

Table 11 - Current warehouses capacities

Warehouse capacity				
Locations	**Mattresses**	**Blankets**	**Cleaning supplies**	**Food supply kit**
Morumbi	2200	2200	1500	2000
Apiai	450	450	200	150
Taubaté	233	233	300	200
Prudente	100	100	105	50
Bauru	76	76	264	50
Registro	400	400	450	200

We have chosen the food supply kits as the capacity benchmark for the warehouses due to the higher degree of importance of this item. This choice was also based on the different rations os inventory of each type of material in the current warehouses. In order to better simulate the real constraints, the annual capacity will be divided by the percentage that represents the raining season use of the stock (so it will increase). This percentage was calculated using the data from the Civil Defense stock management control. The months were classified and the total demand of items in each season was calculated. Around 60% comes from the raining season.

Now that the data is processed, the results can start to be interpreted. This will be discussed in the chapters six and seven. The difference is that on chapter seven the focus will be entirely on the visualization of the results and the new interactive system. If the models or algorithms used in this section were not clear the reader can return to the chapters three and four to better understand the concepts and plan behind them.

6 Results of the model

In this chapter the user will be briefly introduced to the solutions proposed by Gurobi and how they are presented to the users. Then we discuss about the need to show different results in the same page in order to improve the user experience on advancing discussions about the best cases and the alternatives. This chapter can be seen as the more direct evolution of the work developed in Cosenza (2015).

6.1 Results of the model

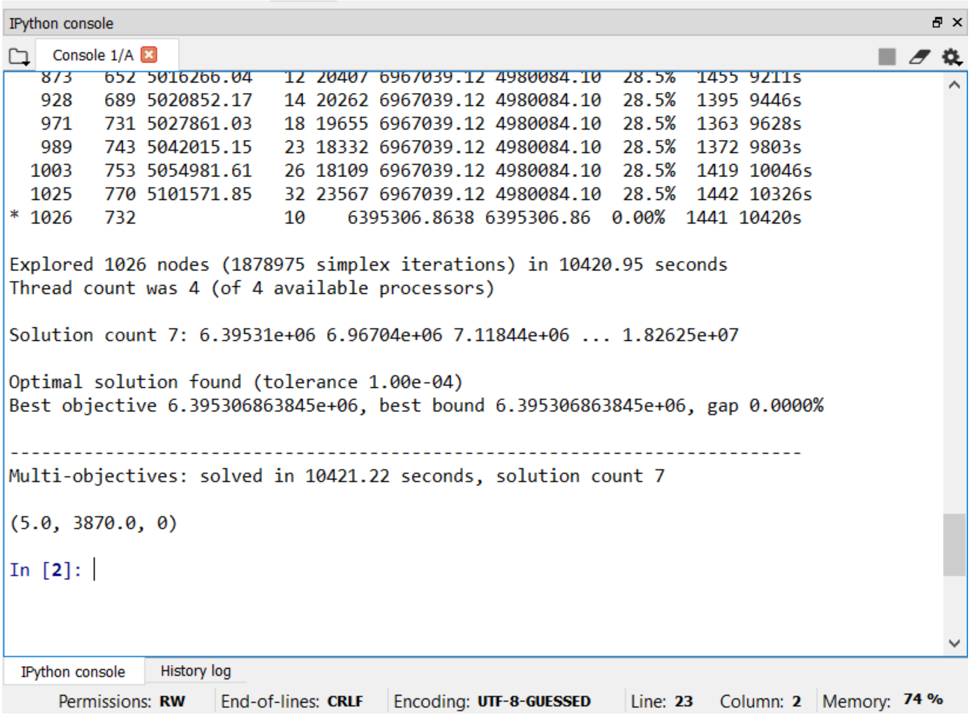
6.1.1 Model solutions

In the end the multi-objective location model optimization is the only one that requires a specialized software in order to be solved. The final choice was to use the Gurobi software due to its integration with Python and the rest of the interface. With pickle library, the recording and recovery of all the solution information is quite easy and accessible. The question of feasibility of the solution was raised only once the tests started. There are two major variables that adds the difficulty and time for the model to find a solution:

- Independent correspondences city-warehouses when running multiple
- Adding warehouse capacity to the model

Each of these variables adds significant time to the finding of the solution, and the combination of both makes it too costly to solve within the available resources. A simple model with none of the possibilities can be solved within two minutes of running time. When the independency was added to the correspondence, the time increased to 10421 seconds until the solution was found, with all the other parameters set equal. When to the simple model the constraint of warehouse capacity was added, the final time to find the solution was of 25257 seconds.

In the end, to simplify the inputs needed to update the system the choice was to avoid both of the conditions and to have a simple demand in the cities. Then the modularity becomes key as the CEPDEC team can provide any changes in the demands by city in order to test the different scenarios. The modularity will be explained in the chapter 7 and will depend on the use of Excel integration with Python.



```
IPython console
Console 1/A
   873    652 5016266.04   12 20407 6967039.12 4980084.10  28.5%  1455 9211s
   928    689 5020852.17   14 20262 6967039.12 4980084.10  28.5%  1395 9446s
   971    731 5027861.03   18 19655 6967039.12 4980084.10  28.5%  1363 9628s
   989    743 5042015.15   23 18332 6967039.12 4980084.10  28.5%  1372 9803s
  1003    753 5054981.61   26 18109 6967039.12 4980084.10  28.5%  1419 10046s
  1025    770 5101571.85   32 23567 6967039.12 4980084.10  28.5%  1442 10326s
* 1026    732              10    6395306.8638 6395306.86  0.00%  1441 10420s

Explored 1026 nodes (1878975 simplex iterations) in 10420.95 seconds
Thread count was 4 (of 4 available processors)

Solution count 7: 6.39531e+06 6.96704e+06 7.11844e+06 ... 1.82625e+07

Optimal solution found (tolerance 1.00e-04)
Best objective 6.395306863845e+06, best bound 6.395306863845e+06, gap 0.0000%

---------------------------------------------------------------------------
Multi-objectives: solved in 10421.22 seconds, solution count 7

(5.0, 3870.0, 0)

In [2]:
IPython console   History log
Permissions: RW   End-of-lines: CRLF   Encoding: UTF-8-GUESSED   Line: 23   Column: 2   Memory: 74 %
```

Figure 24– Example of Gurobi output solution

- Time to run without capacity constraint: 100 seconds
- Time to run with capacity constraint and independence of scenarios: 25.000 seconds

For the fixed location problem, the solution will be simply to minimize the distances between each city and the chosen warehouses. Therefore, the use of a complex operations research model is not needed; a simple attribution model with the database of distances is enough to solve the problem. In addition, the speed of simpler models assures the responsiveness of the website interface after the creation of each scenario that will be tested. As requested by the Civil Defense team all the locations were held open as a possibility, therefore the number of possible combinations is too big to pre-run and store all the possible results. Therefore, the built-in interface will allow the user to make its choices while the backend python server will be running the model in the background after the user inputs are sent to the server.

6.1.2 Solution comparison

An important part of a decision tool is the ability to compare multiple results. In order to do that we will have to keep the different recommendations stored. Otherwise the website would completely lose its responsiveness. And apart from this, it would be important to have an interface that would enable this capability. The full interface will be discussed in-depth in the chapter seven, but an explanation about the comparison tool will remain a topic in this chapter

The goal is for each of the parts to have a design that is directed towards this goal of comparison. One of the major implications is the limitation to show all the information about each model at each time and instead focus on the most important KPIs. In order to establish which are those there was a series of discussions with the Civil Defense team. The final list of KPIs used to compare the different solutions are:

- Costs with costs breakdown by type
- Service level volumes: percentage of requests being responded within the desired timeframe
- Cities being supplied by each warehouse
- Workload on the warehouses

It is noticeable how these KPIs are all linked to some of the elements of the modeling of the Gurobi solution. Therefore, the code must be able not only to output the total costs, but also all the other variables that are important for the final users. This required some alterations of the traditional outputs of a Gurobi/Guropy model and during the construction of these results. So once the final result was available, the same was true for the rest of the KPIs that were necessary.

In order to implement the visual aids the solution used was to use JavaScript integration with HTML to fill relevant information into the appropriate fields designed to be printer friendly. Some of the functions and structural elements used on this were:

- getElementById
- innerHTML
- <div id = “key” > </div>

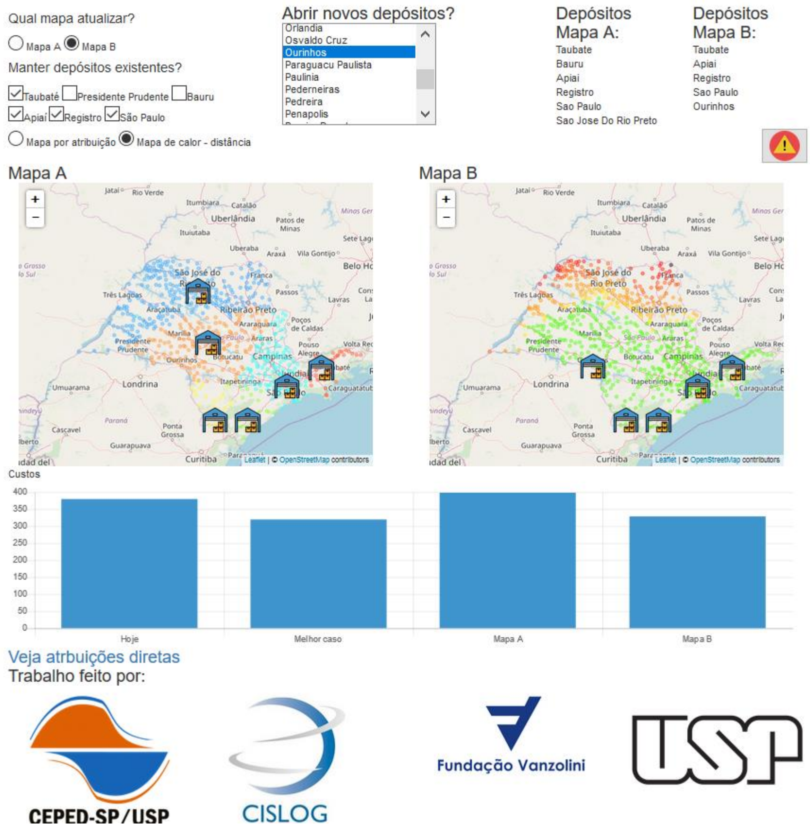


Figure 25- fourth iteration of the website with the comparison elements

In order to arrive at the final version of the model there were multiple iterations with the evolution of the feature. It is important to keep in mind that a managerial tool cannot be imposed but rather built together. Otherwise, there is a big risk of the tool to be discarded completely. The visualization of an interface always generate more comments that will lead to the next iteration and the anticipation of all needs is not reasonable. The focus of managers that will be implied later on the project is something to be taken into account, they must be able to understand quickly what is being presented in a short period of time. The page must be as easy to understand as possible, with an intuitive UI (user interface) and highlighted KPIs dashboard.

6.2 Considerations on the thresholds of the models

There are a few key points to notice as thresholds from the proposed solutions that might impact the real-life application of the model. First of all, the alternatives proposed as locations for the warehouse were all arbitrary and therefore are subject to change.

The goal with using the infrastructure costs is to assure that the model does not over-expand the number of warehouses and because there is no guarantee that in each candidate city there is enough room for the proposed capacity of the warehouses. So the costs are arbitrary and they play a big role in the determination of the optimal solutions. To study and map the conditions of potential warehouses in each city goes beyond the scope of the project but would be ideal if the goal is to improve accuracy in the model. The expected output would be a different cost of opening a warehouse in each city.

The transportation costs are being analyzed in the current optimization model in order to minimize the total costs for the government authorities. But a key aspect is the cost attribution to different authorities. While the construction and maintenance of the warehouses and acquisition of supplies is ruled under Civils' Defense budget, the transportation of the supplies to the cities in need is not. That happens because it is under the responsibility of each city government/each COMPDEC to provide the transportation from the warehouses to the designated cities. So in a different approach the costs of transportations could be removed while maintaining the need to serve every city in the state within 6h of driving distance.

After the generation of the different results and optimal solutions it became clear that simplifications were needed in order to run the interactive model with the web interface. The final module of the site will be presented in the next chapter as well as the content and presentation of the current results in the final website. The development of the interface focused a lot in the deliverable of the interactive model and that is the reason for presenting the two results in different chapters.

7 Visualizations and website interaction

This chapter will present the reader to the biggest contribution of the project towards the final goal. The website built can provide the interactions and insights necessary to gain the support of decision makers about the necessary adaptations to the current system. The chapter starts with the general overview of the site and then goes into the details of the visualizations and reporting tools. The website interactions can be seen on:

https://drive.google.com/drive/folders/1LnyvM9gIwUfwwkK2SJjqtkZJ2lib8RgW?usp=sharing

7.1 Visualization presentation

The final website contains four tabs accessible to explain the project. The definition of the tabs necessary as well as their contents was built in an iterative way with the feedback from the Civil Defense team and the CISLOG/CEPED team. Each of them has a different role and will be used in a unique way that will be thoroughly explained in the following section:

- Home Page: Detailed explanation of the other tabs and the entire methodology of the project. The goal is to substitute a PowerPoint presentation that would be necessary in order to present the project. This way a new user can simply arrive in the homepage of the website and understand the objectives and methodology used. It has a scrolling interface that allows to segment sections that explain in detail each of the other pages

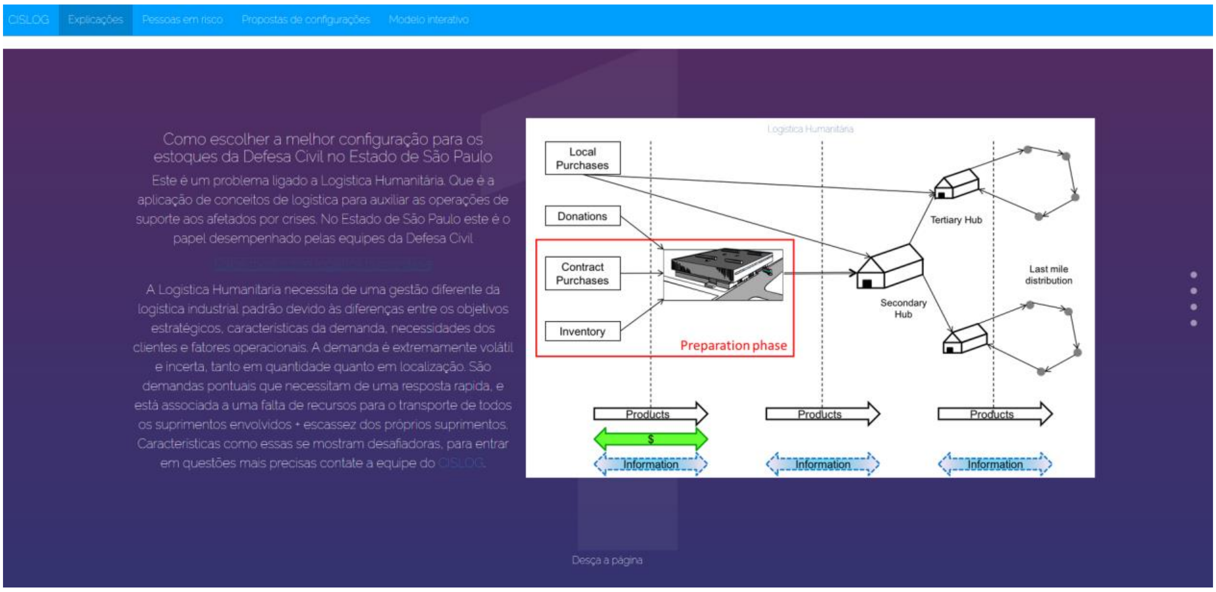


Figure 26- Homepage of the website

- Linear Regression results: Present the results of calculations of people living under risk conditions as well as explanation through maps of the different scenarios built. The idea in this case is to focus on this first step on the methodology so it is clear where the starting point for the regression models is. The visualization and interaction in this page are much simpler than in the following pages where the user inputs are actually needed in order for results to be shown.

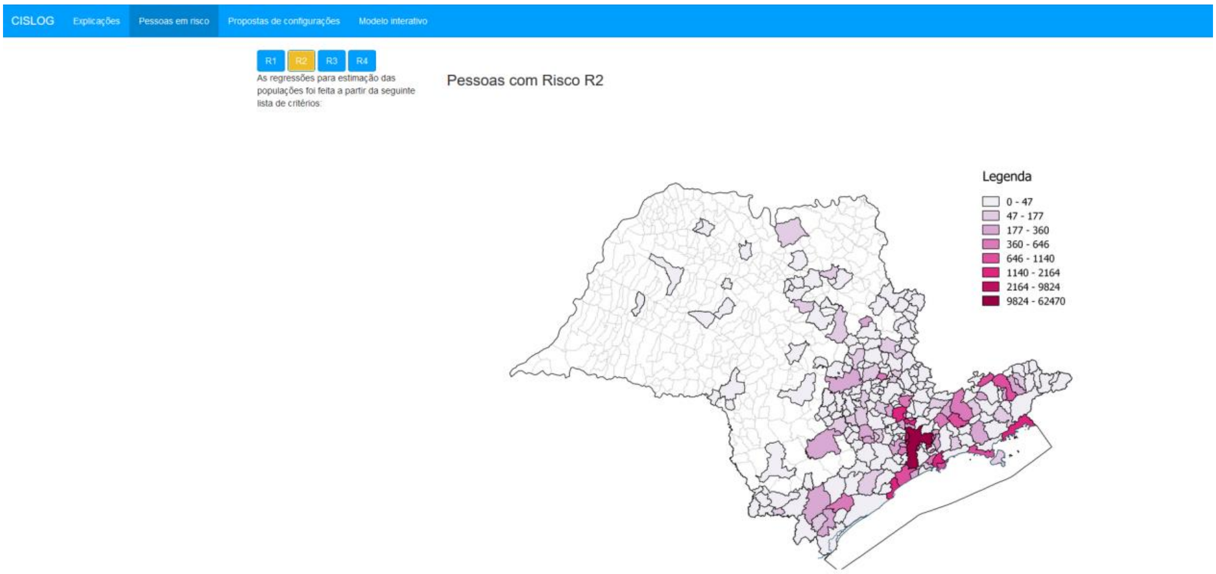


Figure 27- Regression Analysis results page

- Multi-objective model results: presentation of the main scenarios created and their optimal results. These results are pre-run and are the suggestions for the positions for the warehouses. The goal here is to be able to provide suggestion of configurations that could be used by the civil defense team, the same way it was done by Cosenza (2015). For this part of the system there is only a small need for interactions on the website from the user, the person has to simply choose the scenario that he/she wants to see and compare

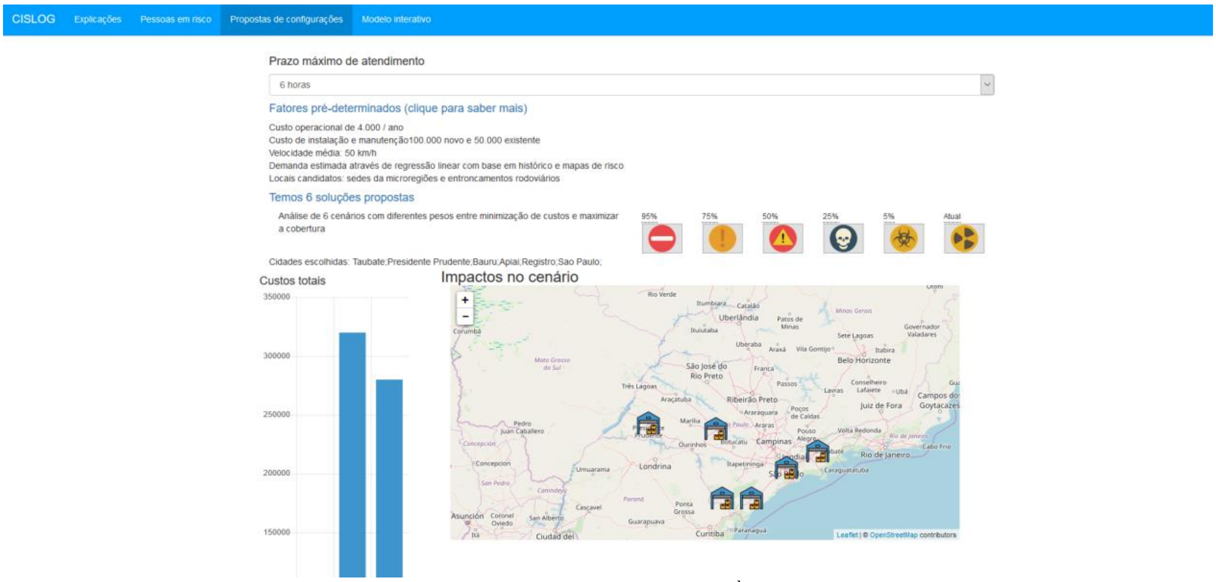


Figure 28- GLM results page, 2nd iteration

- Interactive model: the tab where the users can choose the position of the warehouses and see the costs of the solution. Users can see the comparison between current solution and previous others. This was the page that required the biggest number of iterations as it will probably be the one used the most by their team. This 3rd page is based on multiple assumptions that can easily change overtime, therefore the validity of its conclusions are restricted to a small subset of the real life circumstances that should be updated through the work explained later with the integration with Excel, Python, HTML and JavaScript.

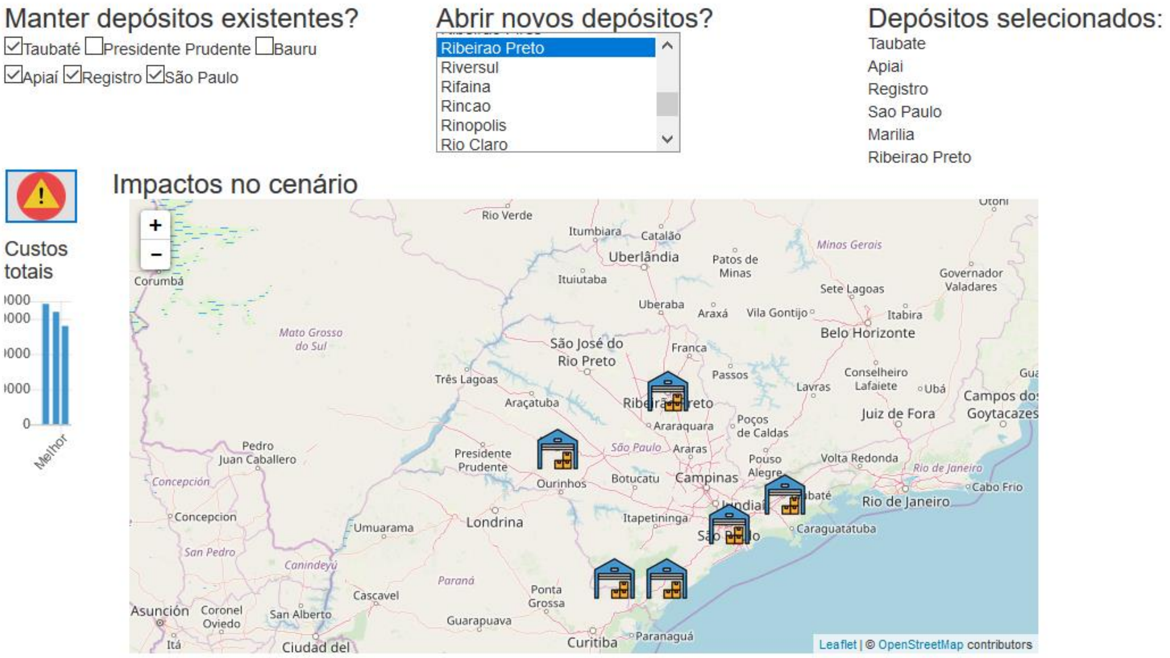


Figure 29- 2nd iteration of the input in the interaction model

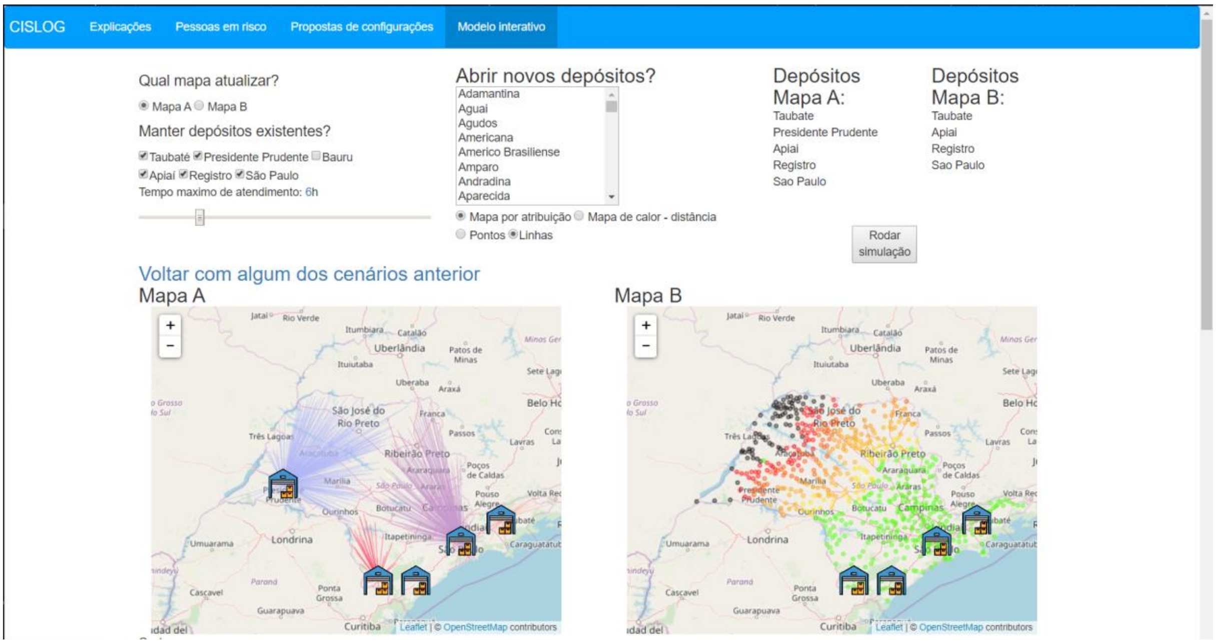


Figure 30- 6th iteration of the interactive model

7.2 Interaction and reporting

For this page beyond the visualization capabilities, it was also important to develop a reporting tool. For the users to be able to take the results and print a version or save it a way that can be used outside this context. In order to do that there is also another section below the interactive maps that focused on the visualization aspect with multiple options such as points or lines to represent both the distance and the attributions of each city to their respective warehouses.

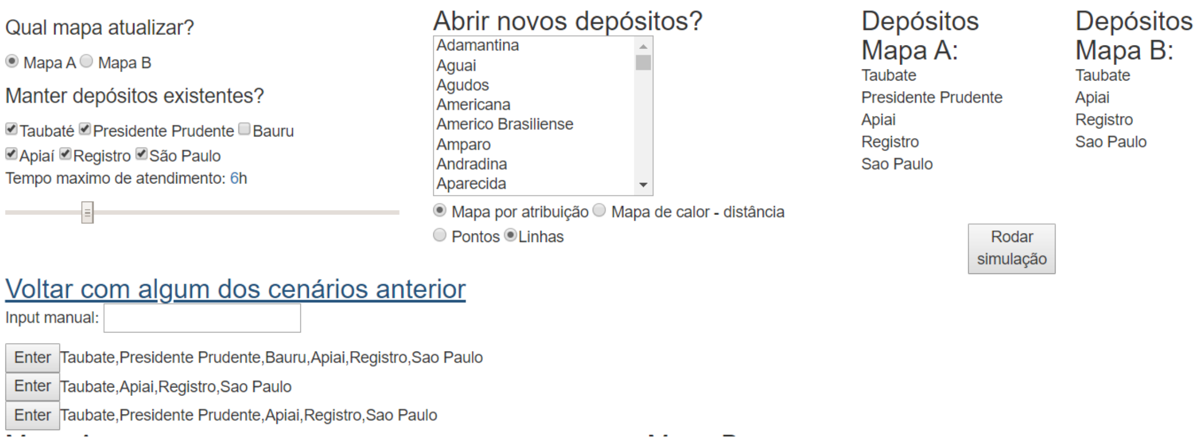


Figure 31- Historic of tests functionality in the website

Adding a section to return easily to previous inputs is important to have a system that can be explored without worrying in memorizing each result. This section could be further improved with the addition to some specific KPIs directly into the historic base but this was not a priority in the development stage.

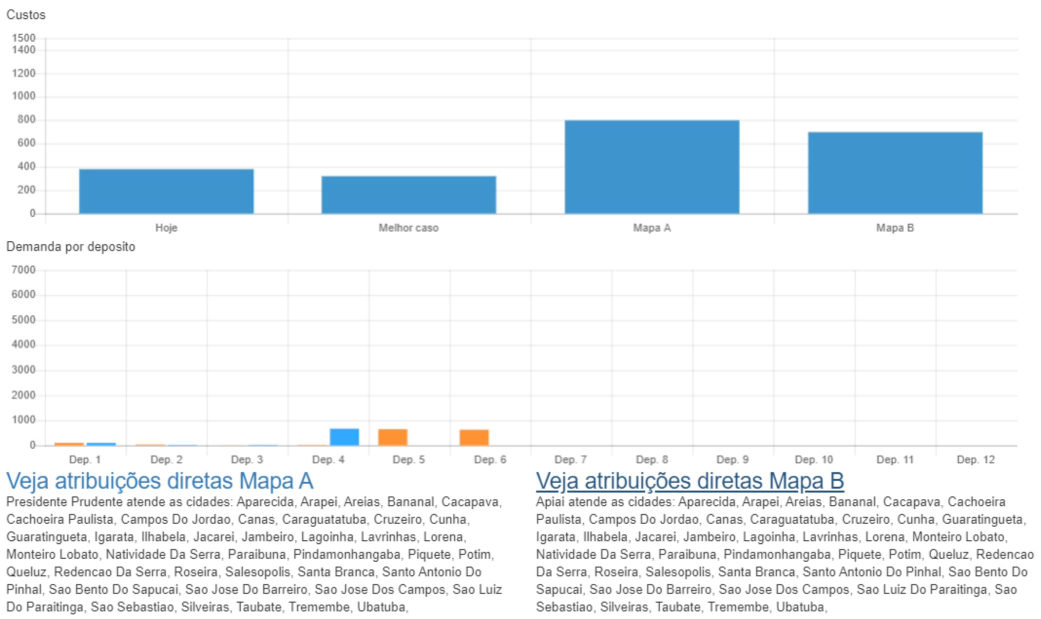


Figure 32- Reporting capabilities in the website (illustrative numbers)

The addition of the costs and capacity demands for each deposit is also another point of attention that was added in the reporting session. With the results represented the way they are, it is possible to print the page in pdf format and send to the decision makers in a simple format.

In order to be able to create a system that is responsive to the user inputs and requests there was a need to implement an heuristics system. This improves the time necessary in order to calculate the results in the interactive model proposed in the chapter five. The choice was to implement a simple algorithm that would use only the smallest distance as an attribution factor. The calculations of the costs arrive only after the attributions are already made. This resulted in a very responsive website that didn't need additional processing time to provide the list of warehouses that serves each city. The full code of implementation of this heuristics can be found on the Appendix 3.

7.3 Updating the model

In order to update the values in the page as well as the scenarios being used as a demand, there is an Excel file that can be used by the users that has then an easy integration with python and HTML. The idea is to give more independence to the CEPDEC team that will run the site and simulations without having to have a in-depth technical knowledge.

Excel is a very popular tool used by many is much more accessible to the general users. It also can be used to implement a more modular system that allows for updates to the website without having to go through the complex system of doing a Regression study, manual classification of events and probability calculation for each type of event. With an Excel model that integrates to the server in Python we open the possibility for the CEPDEC team to input manually the expected demand they want to evaluate in each city and then run the interactive model. Again, this is another step into developing a model that really focus on the interactivity with users and becoming a more user friendly system.

The integration of Excel with Python and HTML is simple. It will consist of simply copying some reference cells from Excel into the appropriate Python files. Then the system will be ready to be used with the newly generated inputs and variables. This could be further simplified with the direct integration of Excel through Python but this would lead to an increased risk of having errors in the integration and therefore the risk of non-usage by the CEPDEC team.

Definiçoes de atendimento		
Velocidade	50	km/h
Tempo max	7	h

Depósitos	
whmax	12
whmin	5
whcost	100000
whmancost	4000
new_wh_ca	300

Kit humanitário	
kit/person	0,25
Kit weight	40,3

Calculo de probabilidades - severidade de eventos						
% populacao afetada	N1	N2	N3	N4	N5	N6
R1						
R2					100%	200%
R3				100%	100%	200%
R4	10%	50%	100%	100%	100%	200%

Construção de eventos - Risco por cluster regional						
Cluster	A	B	C	D	E	F
1	N1	N3	N5	N2	N3	N4
2	N1	N3	N5	N2	N3	N4
3	N1	N3	N5	N2	N3	N4
4	N1	N3	N5	N2	N3	N6
5	N1	N3	N5	N2	N5	N6
6	N1	N3	N5	N4	N5	N6

Resumo de probabilidades por severidade de evento								
Indicador	N1	N2	N3	N4	N5	N6	Período	Dias
Probabildiade	24,8%	5,3%	2,1%	6,8%	3,5%	3,5%	2015-2018/08	2065
Valores antigos	12,4%	3,9%	1,5%	6,6%	3,0%	1,0%	2001-2015	735475

Legenda
Variáveis a escolher
Fórmulas

city	name	Cluster	Demanda esperada manual
3500105	Adamantina	1	10
3500204	Adolfo	2	10
3500303	Aguai	4	10
3500402	Aguas Da Prata	4	10
3500501	Aguas De Lindoia	4	10
3500550	Aguas De Santa Barb	2	10
3500600	Aguas De Sao Pedro	2	10
3500709	Agudos	2	10
3500758	Alambari	4	10
3500808	Alfredo Marcondes	1	10
3500907	Altair	2	10
3501004	Altinopolis	2	10
3501103	Alto Alegre	1	10
3501152	Aluminio	4	10
3501202	Alvares Florence	1	10
3501301	Alvares Machado	1	10
3501400	Alvaro De Carvalho	2	10
3501509	Alvinlandia	2	10
3501608	Americana	4	10
3501707	Americo Brasiliense	2	10
3501806	Americo De Campos	2	10
3501905	Amparo	4	10
3502002	Analandia	2	10
3502101	Andradina	1	10
3502200	Angatuba	6	10
3502309	Anhembi	2	10
3502408	Anhumas	1	10
3502507	Aparecida	3	10
3502606	Aparecida D Oeste	1	10
3502705	Apiai	6	10
3502754	Aracariguama	4	10
3502804	Aracatuba	1	10
3502903	Aracoiaba Da Serra	4	10
3503000	Aramina	2	10
3503109	Arandu	2	10
3503158	Arapei	3	10
3503208	Araraquara	2	10

Figure 33- Manual input system on Excel

	A	B	C	D	E
1	**Usar essa coluna**	**keys**	**key1**	**key2**	**key3**
2	crossover['3500204-3500105']={'d':213,578,'t':4,27156,'q':0}	'd':213,578	'd':213,578	't':4,27156	'q':0
3	crossover['3500303-3500105']={'d':508,806,'t':10,17612,'q':1}	'd':508,806	'd':508,806	't':10,17612	'q':1
4	crossover['3500402-3500105']={'d':538,543,'t':10,77086,'q':1}	'd':538,543	'd':538,543	't':10,77086	'q':1
5	crossover['3500501-3500105']={'d':562,738,'t':11,25476,'q':1}	'd':562,738	'd':562,738	't':11,25476	'q':1
6	crossover['3500550-3500105']={'d':302,912,'t':6,05824,'q':0}	'd':302,912	'd':302,912	't':6,05824	'q':0
7	crossover['3500600-3500105']={'d':425,379,'t':8,50758,'q':1}	'd':425,379	'd':425,379	't':8,50758	'q':1
8	crossover['3500709-3500105']={'d':286,478,'t':5,72956,'q':0}	'd':286,478	'd':286,478	't':5,72956	'q':0
9	crossover['3500758-3500105']={'d':516,407,'t':10,32814,'q':1}	'd':516,407	'd':516,407	't':10,32814	'q':1
10	crossover['3500808-3500105']={'d':89,493,'t':1,78986,'q':0}	'd':89,493,	'd':89,493	't':1,78986	'q':0
11	crossover['3500907-3500105']={'d':286,347,'t':5,72694,'q':0}	'd':286,347	'd':286,347	't':5,72694	'q':0
12	crossover['3501004-3500105']={'d':457,113,'t':9,14226,'q':1}	'd':457,113	'd':457,113	't':9,14226	'q':1
13	crossover['3501103-3500105']={'d':112,414,'t':2,24828,'q':0}	'd':112,414	'd':112,414	't':2,24828	'q':0
14	crossover['3501152-3500105']={'d':542,952,'t':10,85904,'q':1}	'd':542,952	'd':542,952	't':10,85904	'q':1
15	crossover['3501202-3500105']={'d':244,343,'t':4,88686,'q':0}	'd':244,343	'd':244,343	't':4,88686	'q':0
16	crossover['3501301-3500105']={'d':110,567,'t':2,21134,'q':0}	'd':110,567	'd':110,567	't':2,21134	'q':0
17	crossover['3501400-3500105']={'d':174,989,'t':3,49978,'q':0}	'd':174,989	'd':174,989	't':3,49978	'q':0

Figure 34- Python dictionary integration

Possíveis depósitos		
<option value ="	">	</option>
<option value ="3500105">Adamantina</option>		
<option value ="3500303">Aguai</option>		
<option value ="3500709">Agudos</option>		
<option value ="3501608">Americana</option>		
<option value ="3501707">Americo Brasiliense</option>		
<option value ="3501905">Amparo</option>		
<option value ="3502101">Andradina</option>		
<option value ="3502507">Aparecida</option>		
<option value ="3502705">Apiai</option>		
<option value ="3502804">Aracatuba</option>		
<option value ="3503208">Araraquara</option>		
<option value ="3503307">Araras</option>		
<option value ="3503802">Arthur Nogueira</option>		
<option value ="3504008">Assis</option>		
<option value ="3504107">Atibaia</option>		

Possíveis cidades		
<option value ="	**">**	**</option>**
<option value ="3500105">Adamantina</option>		
<option value ="3500204">Adolfo</option>		
<option value ="3500303">Aguai</option>		
<option value ="3500402">Aguas Da Prata</option>		
<option value ="3500501">Aguas De Lindoia</option>		
<option value ="3500550">Aguas De Santa Barbara</option>		
<option value ="3500600">Aguas De Sao Pedro</option>		
<option value ="3500709">Agudos</option>		
<option value ="3500758">Alambari</option>		
<option value ="3500808">Alfredo Marcondes</option>		
<option value ="3500907">Altair</option>		
<option value ="3501004">Altinopolis</option>		
<option value ="3501103">Alto Alegre</option>		
<option value ="3501152">Aluminio</option>		
<option value ="3501202">Alvares Florence</option>		

Figure 35- HTML integration through Excel

Another option studied for updating the decision variables was a direct integration from Python to Excel. The idea was to use the modules from Python to recover the data direct from Excel. But the idea was more prone to errors if there were small changes into file formats and names, and while the correction of these errors is not difficult, they do require more technical knowledge about programming. As copy and pasting was a process that was easier to be done with a how to guide, it was chosen as the final method. A sample of this guide can be found on the Appendixes.

With the website presented, the reader is left with the conclusions in the chapter eight. In the final chapter we will discuss the lessons learned from this project and also the possible improvements that can be made. The goal will be to leave for the next researchers to keep improving the system.

8 Conclusions

In this final chapter we will start with a discussion about the main discoveries and lessons that happened during the execution of the project. Then we will discuss how the project could be further improved to become a more robust and complete system.

8.1 Final comments

In the end, it might feel that the calculation of the optimal placement of warehouses didn't advance during this project. The addition of rain in the regression didn't improve the results and in the location model some simplifications were made to the stochastic part of the methodology. But these setbacks might help in the overall goal of building a system that will help to really improve the operational conditions of warehouses and create real change in the existing configuration.

The modularity and user friendliness are very crucial aspect of building a system that will be used once the development team leaves. And both the simplification of the location model and the failure of adding new variables lead to a simpler system. The engineer must adapt his tools to better fit the end-user expertise and capabilities, sometimes sacrificing some accuracy to gain efficiency. It is easier to calculate the optimal results and then present it to the CEPDEC team, but it will not achieve its goal. In order to help to build a system that will be used, it must be able to be explored and have interactions. This requires going a step further in the development. Having a system that is not understood by the "client" will result in a system that will not reach its potential.

The interactions to assure that the data and model behind the calculations were aligned with the needs of the client is something crucial during the development. And the flexibility to update the entire system in an easy way is also a major concern. The final interaction with the website is therefore an important part of the CEPDEC team experience. But the possibility to update the model with ease and to feel that the model matches their needs is as important feature to guarantee the future use. An example of how to get closer to the final users is to use Excel as a tool to let the user input decisions that will be later integrated in the Python and JavaScript code. This happens because Excel is a tool people are more familiar with than Python or coding HTML.

Only time will tell if the system created will become successful or not. So far the feedback from the team was positive, but it is impossible to judge if the tool will become part of the working life of the operators or if it will remain unused. Even if the system is not used, the project helped to advance the studies a step further to the goal of being an usable system that can be replicated in other regions as well. The modularity of the system becomes crucial as there might be differences in other use cases, the simplification of the integration interface between each stage was another positive point that was not discussed during the previous chapters. It should be a feature that will allow further developments to be made on specific points of the system and not necessarily the system as a whole.

8.2 Further studies

This report has already been developed with advancements proposed in Cosenza (2015). But the system developed so far has taken some simplification hypothesis and could be further improved. The author's suggestions are still valid and could be used but there are focus points that are better candidates for upgrades.

The focus of the works on modelling should be on improvements to the optimization model. Because the errors of the regressions estimations are already quite developed compared to this. There should also be another round of research on new studies that calculate directly the people living under risk conditions. This was not done this time due to the time constraints, but in the next iteration it should be one of the priorities in order to keep the relevance of the model.

In the location model with interactive results, it should be important to study how to further develop the algorithm while not impacting the performance of the system. The types of variables that can be chosen as well as the complexity of the algorithm applied to solve the model could be tested. For example adding the multi-criteria algorithm could be something feasible if the number of possible locations was reduced. Another point could be the addition of heuristics to improve the solving time even with the high number of candidate warehouse locations.

In the online platform there are multiple improvements that are also possible. Some of the more evident ones are:

- Create online visualization that can be shared but that is also secure from undesired access, currently there is no such safety protection

- Make the interactions possible directly from the map using Leaflet functions and avoid having only the HTML interface interactions
- Use of time related visualizations and historical events and evolution to create a notion of flow. This would need a stochastic system and the use of the time variable. But it could be useful to understand the constraints in the system and the type of events that could lead to ruptures

It is important to notice that the current system deals only with the stage of risk analysis and preparedness. The application of technology is possible under all the stages of the disaster management cycle. One further step could be the upgrade of the system for real-time data and the use of the models to be able to help on the response to the events as well. Some use cases of technology that could be imagined as future integrations on early warning reconstruction of infrastructure and others (Hagelocher 2018). With the advances in technology in the recent years we should not abstain from having bigger dreams about possible applications of similar applications to a much larger extend. Having this integrated tool would be something expensive, but that could save thousands of lives in the long term.

Bibliography

Abler, R.A. 1993. Everything in its place: GPS, GIS, and geography in the 1990s. Professional Geographer 45(2), 131-139.

Albala-Bertrand, J. M. 2000. “Complex Emergencies versus Natural Disasters: An Analytical Comparison of Causes and Effects.” Oxford Development Studies 28(2): 187-204.

Altay, Nezih, and Walter G. Green. 2006. “OR/MS Research in Disaster Operations Management.” European Journal of Operational Research 175(1): 475-93. http://linkinghub.elsevier.com/retrieve/pii/S0377221705005047.

Ballou, Ronald H. 2009. Gerenciamento Da Cadeia de Suprimentos - 5.ed.: Logística Empresarial.

Bertha Hidalgo, Melody Goodman (2013) “Multivariate or Multivariable Regression?”, American Journal of Public Health 103, no. 1. pp. 39-40.

Beamon, B. M. (2004) Humanitarian Relief Chains: Issues and Challenges, R 34th International Conference on Computers and Industrial Engineering San Francisco, CA, USA.

Brasil transportes. Edicao anual (2017-2018). Publicacao da Associacao Nacional do Transporte de Cargas e Logistica, NTC&Logistica e editada por DBA&C Associados Ltda.

de Brito Jr, Irineu. 2015. “Localização de Depósitos de Suprimentos de Alívio Para Resposta a Desastres Através de Programação Linear Estocástica E Análise de Decisão Com Múltiplos Critérios.” Escola Politecnica da USP.

Brito Junior, Irineu De, Adriana Leiras, and Hugo Yoshizaki. 2013. “Stochastic Optimization Applied to the Pre Positioning of Disaster Relief Supplies Decisions in Brazil.” In Proceedings of POMS 24rd Annual Conference, 1-10. http://www.pomsmeetings.org/ConfPapers/043/043-0508.pdf.

Chandra, Charu & Kamrani, Ali. (2004). Knowledge Management for Consumer-Focused Product Design. 10.1109/WAC.2004.185546.

Cosenza, Alfonso. 2015. Location of relief supplies warehouses for São Paulo Civil Defense. São Paulo.

Current, John, and Morton O'Kelly. 1992. "Locating Emergency Warning Sirens." Decision Sciences 23(1): 221-34. http://doi.wiley.com/10.1111/j.1540-5915.1992.tb00385.x.

Custódio, Vanderli and Moraes, Antônio Carlos Robert (2002). A persistência das inundações na grande São Paulo. Universidade de São Paulo

Dekle, Jamie et al. 2005. "A Florida County Locates Disaster Recovery Centers." Interfaces 35(2): 133-39. http://pubsonline.informs.org/doi/abs/10.1287/inte.1050.0127.

Devore, J, N Farnum, and J Doi. 2013. Applied Statistics for Engineers and Scientists. Cengage Learning. https://books.google.com.br/books?id=psg_CQAAQBAJ.

Dobson, Annette J. 2002. An Introduction to Generalized Linear Models. 2nd ed. Chapman & Hall/CRC texts in statistical science series.

Garschagen, Matthias & Hagenlocher, Michael & Sabelfeld, Robert & Jin Lee, Yew. (2016). Infrastructure as a risk factor. - Extracted (in part) from a 1999 GIS Lounge post by Caitlin Dempsey

Gennady L. Andrienko and Natalia V. Andrienko. GIS Visualization Support to the C4.5 Classification Model of KDD GMD - German National Research Center for Information TechnologySchloss Birlinghoven, Sankt-Augustin, D-53754 Germany

Geoffrion, Arthur M. 1987. "An Introduction to Structured Modeling." Management Science 33(5): 547-88.

IFRC. 2015. "International Federation of Red Cross and Red Crescent Societies Website." http://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management/

John Pickles. 1995. In, Ground Truth: The Social Implications of Geographic Information Systems. New York: The Guilford Press, pp. 1-30

Lumley, Thomas, Paula Diehr, Scott Emerson, and Lu Chen. 2002. “The Importance of the Normality Assumption in Large Public Health Data Sets.” Annual Reviews of Public Health 23: 151-69.

Martinelli, Marcello. 2010. « Clima do Estado de São Paulo », Confins

Meirim, H. (2007). Logística humanitária e logística Empresarial, mmr brasil. Disponível em: www.mmrbrasil.com.br.

Ministério do Desenvolvimento regional. (2013). Projeto de Fortalecimento da Estratégia Nacional de Gestão Integrada em Riscos de Desastres Naturais - GIDES

Nelder, J. A. and Wedderburn, R. W. 1972, Generalized Linear Models. Journal of the Royal Statistical Society: Series A (General), 135: 370-384

Nogueira, Christiane Wenck, Mirian Buss Gonçalves, and Antônio Galvão Novaes. 2007. “Logística Humanitária E Logística Empresarial: Relações, Conceitos E Desafios.” Congresso Nacional de Pesquisa e Ensino em Transporte: 12.

OFDA/CRED. Feb. 2015. International Disaster Database www.emdat.be Université catholique de Louvain Brussels – Belgium

Pidd, Michael. 1999. “Just Modeling Through: A Rough Guide to Modeling.” Interfaces 29(2): 118-32. http://pubsonline.informs.org/doi/abs/10.1287/inte.29.2.118.

Rawls, Carmen G., and Mark a. Turnquist. 2010. “Pre-Positioning of Emergency Supplies for Disaster Response.” Transportation Research Part B: Methodological 44(4): 521-34. http://dx.doi.org/10.1016/j.trb.2009.08.003.

Rawls, Carmen G., and Mark a. Turnquist. 2012. “Pre-Positioning and Dynamic Delivery Planning for Short-Term Response Following a Natural Disaster.” Socio-Economic Planning Sciences 46(1): 46-54. http://dx.doi.org/10.1016/j.seps.2011.10.002.

S. Lei, J. Wang, C. Chen and Y. Hou..2018. "Mobile Emergency Generator Pre-Positioning and Real-Time Allocation for Resilient Response to Natural Disasters," in IEEE Transactions on Smart Grid, vol. 9, no. 3, pp. 2030-2041

Lima, Fabiana Santos, de Oliveira, Daniel e Gonçalves, Mirian Buss. 2011. A formação de clusters na logística humanitária utilizando mineração de dados. Programa de Pós Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina.

Sherali, Hanif D., Todd B. Carter, and Antoine G. Hobeika. 1991. "A Location-Allocation Model and Model for Evacuation Planning under Hurricane/flood Conditions." Transportation Research Part B: Methodological 25(6): 439-52. http://linkinghub.elsevier.com/retrieve/pii/019126159190037J.

Souza, F.T. (2004) Predição de escorregamentos das encostas do município do rio de janeiro através de técnicas de mineração de dados. tese (doutorado). Programa de pós- graduação em engenharia, universidade federal do rio de janeiro, Rio de Janeiro, 2004.

Srinivasa, Anand V., and Wilbert E. Wilhelm. 1997. "A Procedure for Optimizing Tactical Response in Oil Spill Clean up Operations." European Journal of Operational Research 102(3): 554-74. http://linkinghub.elsevier.com/retrieve/pii/S0377221796002421.

Tomasini, Rolando M, and Luk N Van Wassenhove. 2009. "From Preparedness to Partnerships: Case Study Research on Humanitarian Logistics." International Transactions in Operational Research 16(5): 549-59. http://doi.wiley.com/10.1111/j.1475-3995.2009.00697.x.

Tominaga, Lídia Keiko, Jair Santoro, and Rosangela Amaral. 2009. Desastres Naturais : Conhecer Para Prevenir.

UNDRR (UN office for Disaster Risk Reduction). 2018. Global Assessment Report in Disaster Risk Reduction

United Nations, 2015.The human cost of weather related disasters 1995-2015. UNISDR and CRED

Van Wassenhove, L N. 2006. “Humanitarian Aid Logistics: Supply Chain Management in High Gear†.” Journal of the Operational Research Society 57(5): 475-89.

World Bank. 2012. Avaliação de Perdas e Danos: Inundações e Deslizamentos na Região Serrana do Rio de Janeiro - Janeiro de 2011. Relatório elaborado pelo Banco Mundial com apoio do Governo do Estado do Rio de Janeiro.

APPENDIX 1 – CITIES THAT ARE CANDIDATES FOR THE BASES (COSENZA 2015)

ADAMANTINA
AGUAÍ
AGUDOS
AMERICANA
AMÉRICO BRASILIENSE
AMPARO
ANDRADINA
APARECIDA
APIAÍ
ARAÇATUBA
ARARAQUARA
ARARAS
ARTHUR NOGUEIRA
ASSIS
ATIBAIA
AVARÉ
BARIRI
BARRA BONITA
BARRETOS
BATATAIS
BAURU
BEBEDOURO
BERTIOGA
BIRIGUI
BOITUVA
BOTUCATU
BRAGANÇA PAULISTA
CABREÚVA
CAÇAPAVA
CACHOEIRA PAULISTA
CAJATI
CAMPINAS
CAMPO LIMPO PAULISTA
CAMPOS DO JORDÃO
CÂNDIDO MOTA
CAPÃO BONITO
CAPIVARI
CARAGUATATUBA
CASA BRANCA
CATANDUVA
CERQUILHO
COSMÓPOLIS
CRAVINHOS
CRUZEIRO
CUBATÃO
DESCALVADO
DRACENA
ESPÍRITO SANTO DO PINHAL
FERNANDÓPOLIS
FRANCA
GARÇA
GUAÍRA
GUARARAPES
GUARATINGUETÁ
GUARIBA
GUARUJÁ
HORTOLÂNDIA
IBATÉ
IBITINGA
IBIÚNA
IGARAPAVA
IGUAPE
INDAIATUBA
ITANHAÉM

ITAPETININGA
ITAPEVA
ITAPEVI
ITAPIRA
ITARARÉ
ITATIBA
ITU
ITUPEVA
ITUVERAVA
JABOTICABAL
JACAREÍ
JAGUARIÚNA
JALES
JANDIRA
JARDINÓPOLIS
JOSÉ BONIFÁCIO
JUNDIAÍ
LEME
LENÇÓIS PAULISTA
LIMEIRA
LINS
LORENA
MAIRINQUE
MARÍLIA
MATÃO
MIRANDÓPOLIS
MIRASSOL
MOCOCA
MOGI-GUAÇU
MOGI MIRIM
MONGAGUÁ
MONTE ALTO
MONTE MOR
MORRO AGUDO
NOVA ODESSA
NOVO HORIZONTE
OLÍMPIA
ORLÂNDIA
OSVALDO CRUZ
OURINHOS
PARAGUAÇU PAULISTA
PAULÍNIA
PEDERNEIRAS
PEDREIRA
PENÁPOLIS
PEREIRA BARRETO
PERUÍBE
PIEDADE
PINDAMONHANGABA
PIRACICABA
PIRAJU
PIRASSUNGA
PITANGUEIRAS
PONTAL
PORTO FELIZ
PORTO FERREIRA
PRAIA GRANDE
PRESIDENTE EPITÁCIO
PRESIDENTE PRUDENTE
PRESIDENTE VENCESLAU
PROMISSÃO
RANCHARIA
REGISTRO
RIBEIRÃO PIRES
RIBEIRÃO PRETO
RIO CLARO
SALTO
SALTO DE PIRAPORA

SANTA BÁRBARA D OESTE
SANTA CRUZ DAS PALMEIRAS
SANTA CRUZ DO RIO PARDO
SANTA FÉ DO SUL
SANTA RITA DO PASSA QUATRO
SANTOS
SÃO BERNARDO DO CAMPO
SÃO CARLOS
SÃO JOAQUIM DA BARRA
SÃO JOSÉ DO RIO PARDO
SÃO JOSÉ DO RIO PRETO
SÃO JOSÉ DOS CAMPOS
SÃO MANUEL
SÃO MIGUEL ARCANJO
SÃO PAULO
SÃO PEDRO
SÃO ROQUE
SÃO SEBASTIÃO
SÃO VICENTE
SERRANA
SERTÃOZINHO
SOCORRO
SOROCABA
SUMARÉ
TAQUARITINGA
TATUÍ
TAUBATÉ
TIETÊ
TREMEMBÉ
TUPÃ
UBATUBA
VALINHOS
VARGEM GRANDE DO SUL
VÁRZEA PAULISTA
VINHEDO
VOTORANTIM
VOTUPORANGA

APPENDIX 2 – WORLD RISK INDEX METHODOLOGY

Here the reader can identify all the components for the calculation of the World Risk Index that can be used as a source of inspiration to the type of information that can be used to calculate the amount of people living at risk in the State of São Paulo.

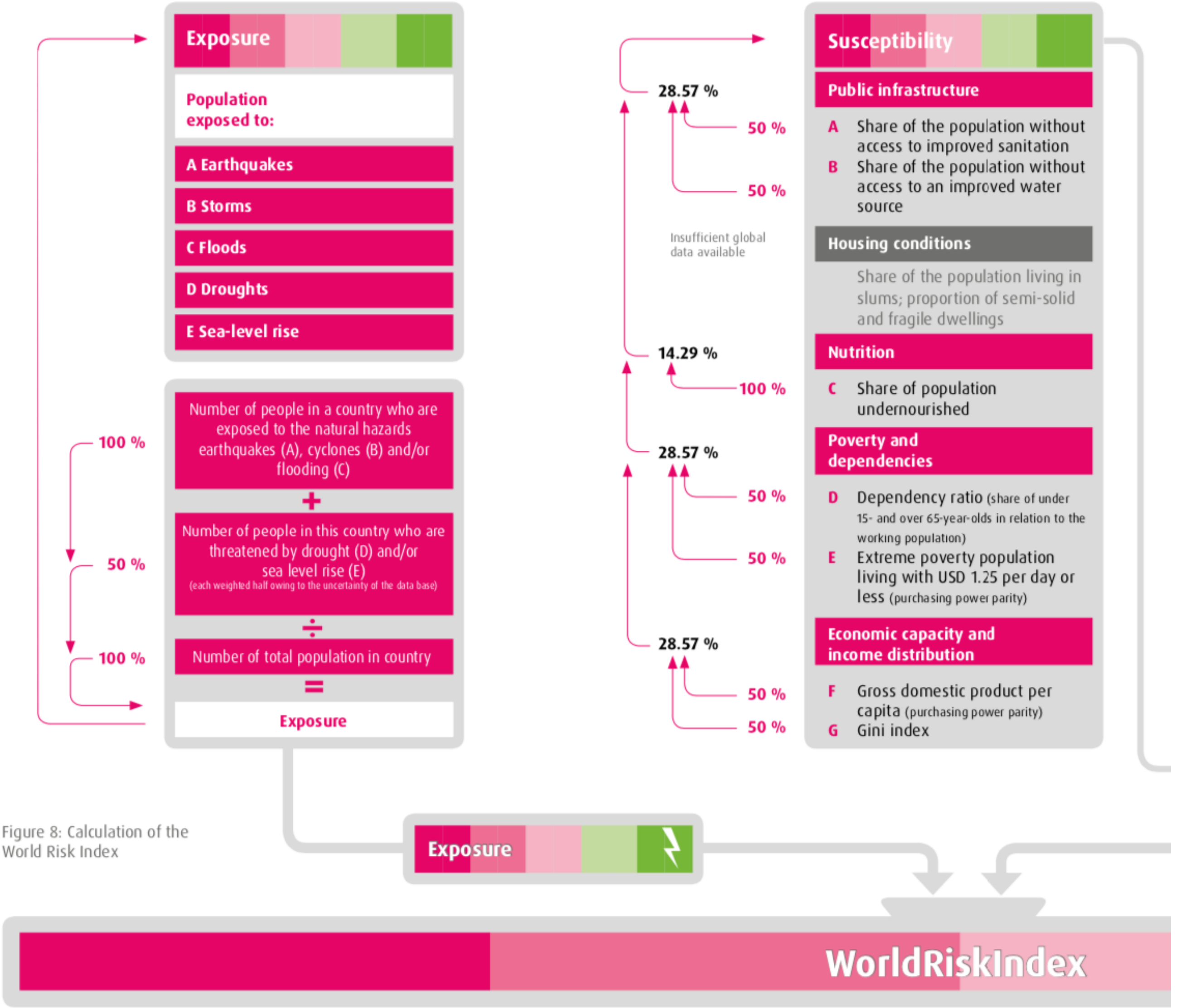
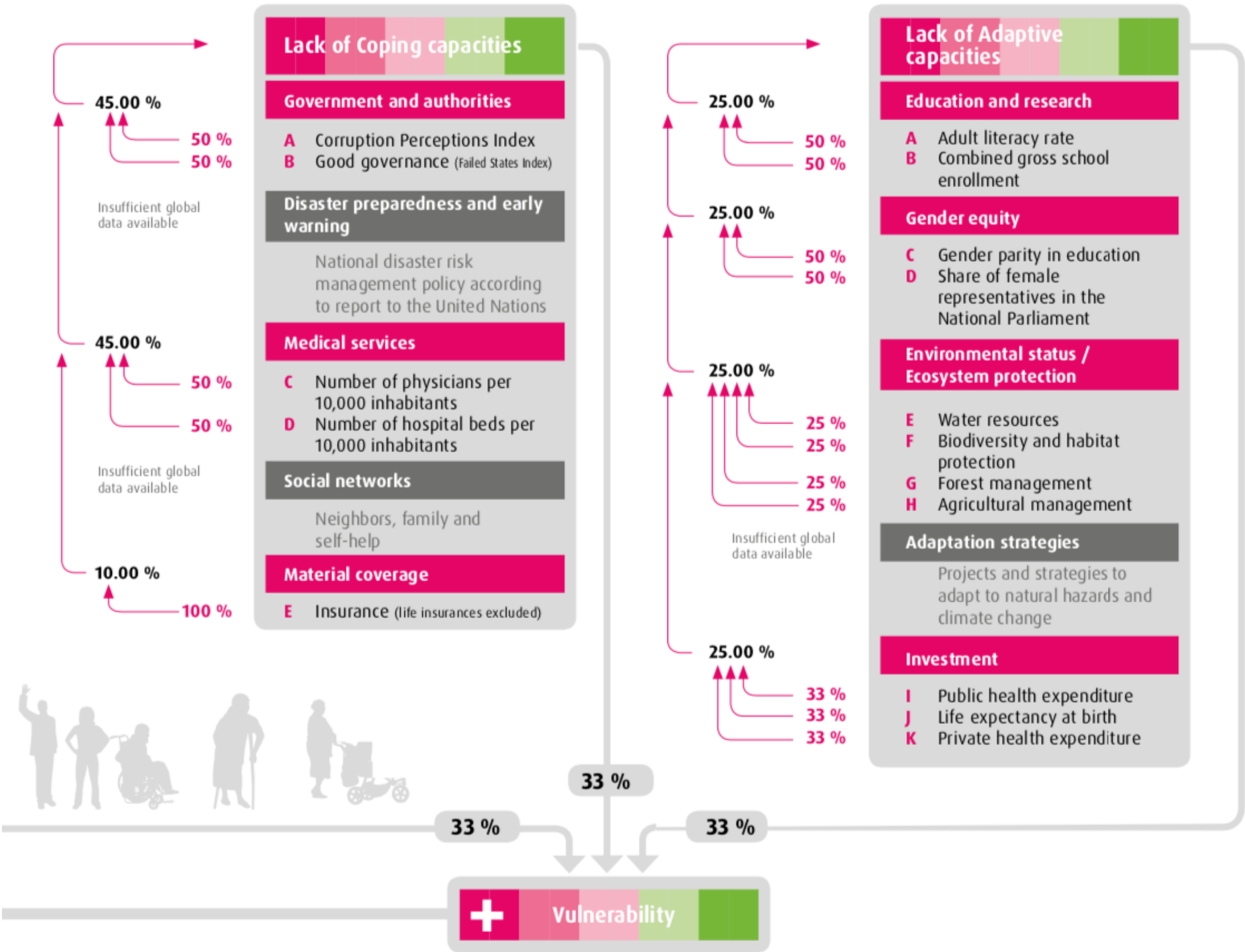


Figure 8: Calculation of the World Risk Index



45.00 %
50 %
50 %
Insufficient global data available
45.00 %
50 %
50 %
Insufficient global data available
10.00 %
100 %
Lack of Coping capacities
Government and authorities
A Corruption Perceptions Index
B Good governance (Failed States Index)
Disaster preparedness and early warning
National disaster risk management policy according to report to the United Nations
Medical services
C Number of physicians per 10,000 inhabitants
D Number of hospital beds per 10,000 inhabitants
Social networks
Neighbors, family and self-help
Material coverage
E Insurance (life insurances excluded)
25.00 %
50 %
50 %
25.00 %
50 %
50 %
25.00 %
25 %
25 %
25 %
25 %
Insufficient global data available
25.00 %
33 %
33 %
33 %
Lack of Adaptive capacities
Education and research
A Adult literacy rate
B Combined gross school enrollment
Gender equity
C Gender parity in education
D Share of female representatives in the National Parliament
Environmental status / Ecosystem protection
E Water resources
F Biodiversity and habitat protection
G Forest management
H Agricultural management
Adaptation strategies
Projects and strategies to adapt to natural hazards and climate change
Investment
I Public health expenditure
J Life expectancy at birth
K Private health expenditure
33 %
33 %
33 %
Vulnerability

APPENDIX 3 – CODE EXAMPLES FOR GLM RESOLUTION AND HEURISTICS FOR SHORTEST DISTANCE ATTRIBUTION

```
X=[]
for i in warehouses:
	X.append(mdist.addVar(vtype = GRB.BINARY,name = "X_%d" %(i)))
	COST_warehouse += X[i] * (wh_list[wh_keys[i]]['whcost']/ny + wh_list[wh_keys[i]]['whcost'])
A=[]
for j in city_list:
	temp1 = []
	for i in warehouses:
		temp1.append(mdist.addVar(vtype = GRB.BINARY,name = "A_%d_%d" %(j,i)))
	A.append(temp1)

for j in city_list:
	for i in warehouses:
		if city_keys[j] != wh_keys[i]:
			key = city_keys[j] + "-" + wh_keys[i]
			COST_transport += A[j][i] * (11.5942+ 0.374334792 * cities[city_keys[j]][c]*weight + 0.013354683 * crossover[key]['d']+ 0.00042867299 * weight* cities[city_keys[j]][c] / crossover[key]['d'])
			COST_nonservice += k * A[j][i] * cities[city_keys[j]][c] * crossover[key]['q']
mdist.update()
mdist.addConstr(quicksum(X[i] for i in warehouses),GRB.LESS_EQUAL,whmax, "Maximo de warehouses")
mdist.addConstr(quicksum(X[i] for i in warehouses),GRB.GREATER_EQUAL,whmin, "Minimo de warehouses")
for i in warehouses:
	#Forçar relação entre fluxos possiveis e abertura de depósitos
	mdist.addConstr(quicksum(A[j][i] for j in city_list),GRB.GREATER_EQUAL,X[i],"Open1_%d" % (i))
```

```
    mdist.addConstr(quicksum(A[j][i] for j in city_list),GRB.LESS_EQUAL, X[i] * bigM,"Open2_%d" % (i))
for j in city_list:
    mdist.addConstr(quicksum(A[j][i] for i in warehouses) ,GRB.EQUAL ,1, "Demand_met_%d" %(j))

COST = COST_warehouse + COST_transport + COST_nonservice
mdist.setObjectiveN (COST_warehouse,0,2)
mdist.setObjectiveN (COST,1,1)
mdist.ModelSense = GRB.MINIMIZE

mdist.optimize()

def closest (chosen_wh_list):
    key_list = []
    for key in cities:
        dist = 10000000000
        bestdistkey = '-'
        bestcode = '-'
        for code in chosen_wh_list:
            searchkey = key + '-' + code
            if crossover[searchkey]['d'] <= dist:
                bestdistkey = searchkey
                dist = crossover[searchkey]['d']
                bestcode = code
        if bestcode == "-"      :
            print((bestdistkey,key,bestcode))
        else:
            key_list.append((bestdistkey,key,bestcode))
    return key_list
def transpcosts(distribution,scenario):
    cost = 0
    for elem in distribution:
        if crossover[elem[0]]['d']!=0:
```

```
                        cost = cost + (11.5942 + 0.374334792 * (cities[elem[1]][scenario]) * weight + 0.013354683 * crossover[elem[0]]['d'] + 0.000428673 * weight * ((cities[elem[1]][scenario]) / crossover[elem[0]]['d']))
        return cost

'''
funcao pega a lista de depósitos ativos e retorna o custo total de operação desses depósitos
'''
def loccosts(wh_locations):
        cost = 0
        for d in wh_locations:
                cost = cost + int(wh_list[d]['whcost']) + int(cities[d]['whmancost'])
        return (cost)
```

APPENDIX 4 – CODE EXAMPLES FOR IMPLEMENTATION OF VISUALIZATION

Code for running the server

```
app = Flask(__name__)

@app.route("/")
def load_main():
	return render_template('index.html')

@app.route("/fase1")
def fase1():
	return render_template('fase1.html')

@app.route("/fase2")
def fase2():
	return render_template('fase2.html')

@app.route("/fase3")
def fase3():
	return render_template('fase3.html')
```

Code for communication server-webpage

```
@app.route('/get_data3', methods = ['POST'])
def get_data3():
	parameters = request.get_json(force=True)
	par = parameters.split(";")
	city_aloc = inter.closest(wh_locations)
	if colour_boolean == "Corresp":
		dots = inter.city_colours(city_aloc,wh_locations)
	else:
		dots = inter.city_dists_colours(city_aloc,wh_locations,max_lt)
	for k in city_aloc:
		city_num.append(k[1])
		wh_num.append(k[2])

	city_name = inter.names(city_num)
```

```
wh_names = inter.names(wh_num)
data_to_send = json.dumps((latlong,names,dots,costs,city_cluster,demanda_wh))
```

Code for map addition

```
if (selectedmap =="A"){
    map = L.map('map',{
    center: [-21.795391, -49.440836],
    zoom: 6
    });
    L.tileLayer('http://{s}.tile.osm.org/{z}/{x}/{y}.png', {
    attribution: '&copy; <a href="http://osm.org/copyright">OpenStreetMap</a>
contributors'
    }).addTo(map);
    }
```

Code for website interaction and updating values in the map and page elements

```
function whchosen(){
    currentones = document.forms[1];
    txt = "";
    select = document.getElementsByTagName('select')[0];
    options = select && select.options;
    opt;
    active_cities = 0
    result
    rerun = document.getElementById("rerun").value
    if (rerun==""){
      result = [];
      for (var i = 0; i < currentones.length; i++) {
        if (currentones[i].checked) {
          result.push(currentones[i].value||currentones[i].text)
          active_cities++
        }
      }
      for (var i=0, iLen=options.length; i<iLen; i++) {
```

```
        opt = options[i];
        if (opt.selected) {
          result.push(opt.value || opt.text);
          active_cities++
        }
    }
}
else {
    result = rerun.split(",")
    document.getElementById("rerun").value = ""
}

var radios = document.getElementsByName('visualch');
var visualmethod

for (var i = 0, length = radios.length; i < length; i++){
     if (radios[i].checked){
      // do whatever you want with the checked radio
      visualmethod = radios[i].value;
     }
};
var radios2 = document.getElementsByName('visualtype');
var visualmethod2
for (var i = 0, length = radios2.length; i < length; i++){
     if (radios2[i].checked){
      // do whatever you want with the checked radio
      visualmethod2 = radios2[i].value;}};
alert(visualmethod2)
update_all(result,visualmethod,visualmethod2)};
```

APPENDIX 5 – EXTRACT OF THE HOW TO USE GUIDE (IN PORTUGUESE)

Passos para atualizar o modelo em casos diferentes.

Mudança de demandas por cidade, ou qualquer um dos parâmetros de riscos, probabilidades

- Atualizar na planilha Excel a tabela de demandas
- Ir para a aba dict1 e copiar todas as celulas da coluna A
- Abrir o arquivo Python dicionario1
- Apagar todas as linhas de 2 a 645
- Colar valores copiados do arquivo Excel
- Salvar e fechar o arquivo
- Rodar arquivo interativo normalmente
- Rodar arquivo de Guropy chamado otimização para conseguir novas soluções ótimas

Mudança na lista de cidades ou do tempo mínimo de atendimento

- Mudar lista de cidades no arquivo Excel
- Repetir passos para mudanças por cidade
- Copiar coluna A da aba dict 2
- Abrir arquivo dicionario2.py
- Apagar todas as linhas a partir da linha 2
- Colar novos valores
- Abrir aba HTML no arquivo Excel
- Copiar valores da coluna A
- Abrir arquivo fase3.html que está na pasta templates dentro da pasta do servidor
- Colar dentro da seção com nome <div id="cidades">

Para mudanças da lista de depósitos

- Repetir os passos para alista de cidades
- Abrir a aba dict 3 no arquivo Excel
- Copiar coluna A
- Abrir arquivo dicionario3
- Deletar linhas a partir da 2
- Colar valores copiados anteriormente

APPENDIX 6 – CLUSTERS OF CITIES FOR SCENARIOS GENERATION

Cluster 1: Adamantina, Alfredo Marcondes, Alto Alegre, Alvares Florence, Alvares Machado, Andradina, Anhumas, Aparecida D Oeste, Aracatuba, Arco Iris, Aspasia, Assis, Auriflama, Bastos, Bauru, Bento De Abreu, Bilac, Birigui, Bora, Brauna, Buritama, Caiabu, Caiua, Candido Mota, Cardoso, Castilho, Clementina, Coroados, Cruzalia, Dirce Reis, Dolcinopolis, Dracena, Emilianopolis, Estrela D Oeste, Estrela Do Norte, Euclides Da Cunha Paulista, Fernandopolis, Flora Rica, Floreal, Florida Paulista, Florinea, Gabriel Monteiro, Gastao Vidigal, General Salgado, Guaracai, Guarani D Oeste, Guararapes, Guzolandia, Herculandia, Iacri, Iepe, Ilha Solteira, Indiana, Indiapora, Inubia Paulista, Irapuru, Itapura, Jales, Joao Ramalho, Junqueiropolis, Lavinia, Lourdes, Lucelia, Luiziania, Lutecia, Macedonia, Magda, Maraba Paulista, Maracai, Mariapolis, Marinopolis, Martinopolis, Meridiano, Mesopolis, Mira Estrela, Mirandopolis, Mirante Do Paranapanema, Moncoes, Monte Castelo, Murutinga Do Sul, Nantes, Narandiba, Nhandeara, Nova Canaa Paulista, Nova Castilho, Nova Guataporanga, Nova Independencia, Nova Luzitania, Oscar Bressane, Osvaldo Cruz, Ouroeste, Ouro Verde, Palmeira D Oeste, Palmital, Panorama, Paraguacu Paulista, Paranapua, Parapua, Parisi, Pauliceia, Pedranopolis, Pedrinhas Paulista, Pereira Barreto, Piacatu, Piquerobi, Pirapozinho, Platina, Pontalinda, Populina, Pracinha, Presidente Bernardes, Presidente Epitacio, Presidente Venceslau, Quata, Queiroz, Quintana, Rancharia, Regente Feijo, Ribeirao Dos Indios, Rinopolis, Riolandia, Rosana, Rubiacea, Rubineia, Sagres, Salmourao, Sandovalina, Santa Albertina, Santa Clara D Oeste, Santa Fe Do Sul, Santa Mercedes, Santana Da Ponte Pensa, Santa Rita D Oeste, Santa Salete, Santo Anastacio, Santo Antonio Do Aracangua, Santo Expedito, Santopolis Do Aguapei, Sao Francisco, Sao Joao Das Duas Pontes, Sao Joao De Iracema, Sao Joao Do Pau D Alho, Sebastianopolis Do Sul, Sud Menucci, Suzanopolis, Taciba, Tarabai, Taruma, Teodoro Sampaio, Tres Fronteiras, Tupa, Tupi Paulista, Turiuba, Turmalina, Urania, Valentim Gentil, Valparaiso, Vitoria Brasil, Votuporanga, Zacarias

Cluster 2: Adolfo, Aguas De Santa Barbara, Aguas De Sao Pedro, Agudos, Altair, Altinopolis, Alvaro De Carvalho, Alvinlandia, Americo Brasiliense, Americo De Campos, Analandia, Anhembi, Aramina, Arandu, Araraquara, Arealva, Areiopolis, Ariranha, Avai, Avanhandava, Avare, Bady Bassit, Balbinos, Balsamo, Barbosa, Bariri, Barra Bonita, Barretos, Barrinha, Batatais, Bebedouro, Bernardino De Campos, Boa Esperanca Do Sul, Bocaina, Bofete, Boraceia, Borborema, Borebi, Botucatu, Brejo Alegre, Brodowski, Brotas, Buritizal, Cabralia Paulista, Cafelandia, Cajobi, Cajuru, Campos Novos Paulista, Candido Rodrigues, Canitar, Cassia Dos Coqueiros, Catanduva, Catigua, Cedral, Cerqueira Cesar, Charqueada, Colina, Colombia, Conchas, Corumbatai, Cosmorama, Cravinhos, Cristais Paulista, Descalvado, Dobrada, Dois Corregos, Dourado, Duartina, Dumont, Echapora, Elisiario, Embauba, Espirito Santo Do Turvo, Fartura, Fernando Prestes, Fernao, Franca, Galia, Garca, Gaviao Peixoto, Getulina, Glicerio, Guaicara, Guaimbe, Guaira, Guapiacu, Guara, Guaraci, Guaranta, Guarei, Guariba, Guatapara, Iacanga, Iaras, Ibate, Ibira, Ibirarema, Ibitinga, Icem, Igaracu Do Tiete, Igarapava, Ipaussu, Ipeuna, Ipigua, Ipua, Irapua, Itai, Itajobi, Itaju, Itapolis, Itapui, Itatinga, Itirapina, Itirapua, Ituverava, Jaborandi, Jaboticabal, Jaci, Jardinopolis, Jau, Jeriquara, Jose Bonifacio, Julio Mesquita, Lencois Paulista, Lins, Lucianopolis, Luis Antonio, Lupercio, Macatuba, Macaubal, Manduri, Marapoama, Marilia, Matao, Mendonca, Miguelopolis, Mineiros Do Tiete, Mirassol, Mirassolandia, Monte Alto, Monte Aprazivel, Monte Azul Paulista, Morro Agudo, Motuca, Neves Paulista, Nipoa, Nova Alianca, Nova Europa, Nova Granada, Novais, Novo Horizonte, Nuporanga, Ocaucu, Oleo, Olimpia, Onda Verde, Oriente, Orindiuva, Orlandia, Ourinhos, Palestina, Palmares Paulista, Paraiso, Paranapanema, Pardinho, Patrocinio Paulista, Paulistania, Paulo De Faria, Pederneiras, Pedregulho, Penapolis, Pereiras, Pindorama, Piraju, Pirajui, Pirangi, Pirassununga, Piratininga, Pitangueiras, Planalto, Poloni, Pompeia, Pongai, Pontal, Pontes Gestal, Porangaba, Porto Ferreira, Potirendaba, Pradopolis, Pratania, Presidente Alves, Presidente Prudente, Promissao, Reginopolis, Restinga, Ribeirao Bonito, Ribeirao Corrente, Ribeirao Do Sul, Ribeirao Preto, Rifaina, Rincao, Sabino, Sales, Sales Oliveira, Salto Grande, Santa Adelia, Santa Cruz Da Esperanca, Santa Cruz Das Palmeiras, Santa Cruz Do Rio Pardo, Santa Ernestina, Santa Lucia, Santa Maria Da Serra, Santa Rita Do Passa Quatro, Santa Rosa Do Viterbo, Santo Antonio Da Alegria, Sao Carlos, Sao Joaquim Da Barra, Sao Jose Da Bela Vista, Sao Jose Do Rio Preto, Sao Manuel, Sao Pedro, Sao Pedro Do Turvo, Sao Simao, Sarutaia, Serra Azul, Serrana, Sertaozinho, Severinia, Tabapua, Tabatinga, Taiacu, Taiuva, Tambau, Tanabi, Taquaral, Taquaritinga, Tejupa, Terra Roxa, Timburi, Torre De Pedra, Torrinha, Trabiju, Ubarana, Ubirajara, Uchoa, Uniao Paulista, Uru, Urupes, Vera Cruz, Viradouro, Vista Alegre Do Alto, Chavantes

Cluster 3: Aparecida, Arapei, Areias, Bananal, Cacapava, Cachoeira Paulista, Campos Do Jordao, Canas, Caraguatatuba, Cruzeiro, Cunha, Guaratingueta, Igarata, Ilhabela, Jacarei, Jambeiro, Lagoinha, Lavrinhas, Lorena, Monteiro Lobato, Natividade Da Serra, Paraibuna, Pindamonhangaba, Piquete, Potim, Queluz, Redencao Da Serra, Roseira, Salesopolis, Santa Branca, Santo Antonio Do Pinhal, Sao Bento Do Sapucai, Sao Jose Do Barreiro, Sao Jose Dos Campos, Sao Luiz Do Paraitinga, Sao Paulo, Sao Sebastiao, Silveiras, Tremembe, Ubatuba

Cluster 4: Aguai, Aguas Da Prata, Aguas De Lindoia, Alambari, Aluminio, Americana, Amparo, Aracariguama, Aracoiaba Da Serra, Araras, Arthur Nogueira, Aruja, Atibaia, Barueri, Bertioga, Biritiba Mirim, Boituva, Bom Jesus Dos Perdoes, Braganca Paulista, Cabreuva, Caconde, Caieiras, Cajamar, Campinas, Campo Limpo Paulista, Capela Do Alto, Capivari, Carapicuiba, Casa Branca, Cerquilho, Cesario Lange, Conchal, Cordeiropolis, Cosmopolis, Cotia, Cubatao, Diadema, Divinolandia, Elias Fausto, Embu Das Artes, Embu-Guacu, Engenheiro Coelho, Espirito Santo Do Pinhal, Ferraz De Vasconcelos, Francisco Morato, Franco Da Rocha, Guararema, Guaruja, Guarulhos, Holambra, Hortolandia, Ibiuna, Indaiatuba, Ipero, Iracemapolis, Itanhaem, Itapecerica Da Serra, Itapevi, Itapira, Itaquaquecetuba, Itatiba, Itobi, Itu, Itupeva, Jaguariuna, Jandira, Jarinu, Joanopolis, Jumirim, Jundiai, Juquitiba, Laranjal Paulista, Leme, Limeira, Lindoia, Louveira, Mairinque, Mairipora, Maua, Mococa, Mogi Das Cruzes, Mogi-Guacu, Mogi Mirim, Mombuca, Mongagua, Monte Alegre Do Sul, Monte Mor, Morungaba, Nazare Paulista, Nova Odessa, Osasco, Pacaembu, Paulinia, Pedra Bela, Pedreira, Pinhalzinho, Piracaia, Piracicaba, Pirapora Do Bom Jesus, Poa, Porto Feliz, Praia Grande, Quadra, Rafard, Ribeirao Pires, Rio Claro, Rio Das Pedras, Rio Grande Da Serra, Saltinho, Salto, Salto De Pirapora, Santa Barbara D' Oeste, Santa Cruz Da Conceicao, Santa Gertrudes, Santa Isabel, Santana De Parnaiba, Santo Andre, Santo Antonio De Posse, Santo Antonio Do Jardim, Santos, Sao Bernardo Do Campo, Sao Caetano Do Sul, Sao Joao Da Boa Vista, Sao Jose Do Rio Pardo, Sao Lourenco Da Serra, Sao Roque, Sao Sebastiao Da Grama, Sao Vicente, Serra Negra, Socorro, Sorocaba, Sumare, Suzano, Taboao Da Serra, Tapiratiba, Tatui, Taubate, Tiete, Tuiuti, Valinhos, Vargem, Vargem Grande Do Sul, Vargem Grande Paulista, Varzea Paulista, Vinhedo, Votorantim, Estiva Gerbi

Cluster 5: Barao De Antonina, Barra Do Chapeu, Barra Do Turvo, Bom Sucesso Do Itarare, Buri, Campina Do Monte Alegre, Capao Bonito, Coronel Macedo, Guapiara, Iporanga, Itabera, Itaoca, Itapeva, Itapirapua Paulista, Itaporanga, Itarare, Nova Campina, Registro, Ribeira, Ribeirao Branco, Ribeirao Grande, Riversul, Taguai, Taquarituba, Taquarivai

Cluster 6: Angatuba, Apiai, Cajati, Cananeia, Eldorado, Iguape, Ilha Comprida, Itapetininga, Itariri, Jacupiranga, Juquia, Miracatu, Pariquera-Acu, Pedro De Toledo, Peruibe, Piedade, Pilar Do Sul, Sao Miguel Arcanjo, Sarapui, Sete Barras, Tapirai