A FRAMEWORK FOR ANALYZING AND OPTIMIZING REGIONAL BIO-EMERGENCY RESPONSE PLANS

Tamara Schneider

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2010

APPROVED:

Armin R. Mikler , Major Professor
Yan Huang, Committee Member
Rada Mihalcea, Committee Member
Robert Renka, Committee Member
Chetan Tiwari, Committee Member
Bill Buckles, Director of Graduate Studies in the Department of Computer Science and Engineering
Ian Parberry, Chair of the Department of Computer Science and Engineering
Costas Tsatsoulis, Dean of the College of Engineering
James D. Meernik, Acting Dean of the Robert B. Toulouse School of Graduate Studies

Schneider, Tamara. A Framework for Analyzing and Optimizing Regional Bio-Emergency Response Plans. Doctor of Philosophy (Computer Science), December 2010, 116 pp., 2 tables, 55 illustrations, bibliography, 104 titles.

The presence of naturally occurring and man-made public health threats necessitate the design and implementation of mitigation strategies, such that adequate response is provided in a timely manner. Since multiple variables, such as geographic properties, resource constraints, and government mandated time-frames must be accounted for, computational methods provide the necessary tools to develop contingency response plans while respecting underlying data and assumptions. A typical response scenario involves the placement of points of dispensing (PODs) in the affected geographic region to supply vaccines or medications to the general public. Computational tools aid in the analysis of such response plans, as well as in the strategic placement of PODs, such that feasible response scenarios can be developed. Due to the sensitivity of bio-emergency response plans, geographic information, such as POD locations, must be kept confidential. The generation of synthetic geographic regions allows for the development of emergency response plans on non-sensitive data, as well as for the study of the effects of single geographic parameters. Further, synthetic representations of geographic regions allow for results to be published and evaluated by the scientific community. This dissertation presents methodology for the analysis of bio-emergency response plans, methods for plan optimization, as well as methodology for the generation of synthetic geographic regions.

ACKNOWLEDGEMENTS

This dissertation was in part funded by NIH grant 1R15LM010804-01. Some of the text and figures have been used in previously published papers and are presented here with permission of WIT Press and Inderscience.

I am very thankful for the support I have received from family, friends and scholastic advisers while pursuing my career. In particular, thanks to my parents, Gangolf and Inge Schneider, who have supported my career goals ever since I decided to attend university. I am very appreciative for the support I received from José and his understanding for my busy schedule and dedication to my work. Thanks to his support, I was able to free additional time to work on this dissertation. My apologies to Tommy, close friends, as well as Cheyenne and Dakota, who during the past months did not always receive the attention they deserved.

I would like to thank the members of my Ph.D. committee, who have helped me on my way to becoming a researcher and a scientist. In particular, I appreciate Dr. Armin Mikler's advice and patience throughout my graduate studies. If it was not for him, I would have not considered applying for the Ph.D program. Thanks to Dr. Chetan Tiwari, who provided valuable feedback and useful hints, particularly in GIS-related matters.

Further, I would like to thank everyone involved in the response analysis project. Specifically, I am grateful for Marty O'Neill's contribution, who read large amounts of traffic data into a database and wrote wrapper methods for easy access to the information necessary for the project. Finally, I am very thankful for the support I have received from the entire CERL research group.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Public health preparedness is defined as the capability of the public health system, communities, and individuals to prevent, protect against, quickly respond to, and recover from health emergencies, particularly those in which scale, timing, or unpredictability threatens to overwhelm routine capabilities [4]. Health emergencies for which routine capabilities, such as childhood vaccination, sheltering, and food dispensing, do not suffice include naturally occurring and terrorist-induced bio-emergencies, such as anthrax attacks. The prevalence of such unpredictable types of threats has led to mandates issued by government agencies requiring U.S. public health departments to prepare for bio-emergencies as part of their emergency preparedness programs. In response to these mandates, many counties have designated regional disaster preparedness committees to devise response plans that suffice government mandated guidelines. A typical response scenario involves the strategic placement of points of dispensing (PODs), at which the population can obtain vaccines or prophylactic medication [26]. These PODs facilitate the large-scale distribution of medication to the general public of an affected region. Response plans are developed by preparedness coordinators and public health experts, who select POD locations and allocate resources based on their knowledge of the region and its demographics. Currently, there are no feasibility studies or analysis tools available that allow for a more strategic planning process. It is crucial to thoroughly analyze the feasibility of these contingency plans, such that in case of an emergency the population of the targeted geographic region can be supplied with the necessary medication within established time-frames that meet Centers for Disease Control and Prevention (CDC) mandated guidelines. For instance, in the event of an anthrax attack, prophylactic medication must be dispensed to the entire population of the affected region within 48 hours [27],[75] following an initial setup phase of 12 hours [105]. Different time-frames apply for

other types of bio-emergencies, such as 72 hours for a smallpox incident (see Figure 1.1) [23], [24], [25].

FIGURE 1.1. Timeline for the event of a bio-emergency involving anthrax

The setup phase allows for authorities to prepare the POD with the necessary resources to carry out POD operation. These resources include medication, as well as appropriately trained personnel. Medications or vaccines are transported from the national stockpile to the affected area, where they are redistributed according to the contingency plan of the local authorities. While some bio-emergencies, such as anthrax, can be addressed by dispensing medication, other incidents would require vaccines be administered (e.g. smallpox). This has an impact on the type of personnel required at the POD locations. Medicine can be dispensed by any personnel familiar with procedures of the corresponding contingency plan. Vaccines, however, must be administered by medical staff, such as doctors or nurses. Further, some individuals may need to lie down after receiving vaccines. Such a vaccination scenario increases the service time significantly, as vaccines cannot be provided on a per household basis, but must be administered to each individual independently. Socioeconomic and cultural diversity may require additional staff present at some of the PODs to overcome language barriers in some regions of the affected area. Further, emergency preparedness coordinators need to take into account limited mobility of parts of the population. Groups with mobility restrictions may include the elderly or individuals who do not own cars. Hence, planning authorities may need to complement their plan by additionally deploying mobile PODs.

Once the setup phase is completed, the population is notified of the incident and the emergency response that has been initiated. Disaster coordinators may choose to broadcast procedures via mass media, such as television, radio and the Internet. Further, specific

Internet or cell phone services could provide additional resources for the population and play a role in circulating information. In order to cater to the needs of individuals without access to such resources, the authorities may need to implement alternative means to disseminate essential information.

During the operation phase of the response scenario, police and security personnel are needed to control traffic flow to the POD locations, as well as the at PODs emerging crowds. The police force, as well as disaster coordinators must also be prepared for the event of a POD breakdown. This can be caused by a variety of factors, such as a sudden lack of resources or acts of terror.

A feasibility study of large-scale contingency response plans requires the integration of different data sources. The placement of the individual PODs is hence critical to the success of this effort and is based, among other things, on the appropriate partitioning of the geographic region into specific catchment areas, thereby defining implicitly the population source for each POD. Further, the capacity of the underlying road network is a critical determinant in meeting the CDC imposed mandates. Thus far, the decision of POD placement and the service capabilities of PODs have been based primarily on qualitative measures and loosely defined quantitative constraints, such as limits on support staff, available parking, and access roads. However, the demographic characteristics of a county, represented by each of its census blocks, together with the available road networks and the corresponding traffic estimates, allow for a quantitative approach manifested in well defined computational methods.

It is imperative to analyze the feasibility of response strategies in respect to the given time limitations. Thus far, tools that facilitate analyses of such response plans are not available. This gives rise to the necessity of developing a new set of tools that aid public health officials in adequately preparing for plausible scenarios, incorporating policies and regulations. Intervention strategies must be designed such that the underlying transportation infrastructure provides sufficient resources to satisfy the emerging traffic demand to and from

the dispensing sites. Therefore, demographic characteristics and distinct behavioral modes of the populations must be taken into account. In particular, different levels of base traffic must be modeled, reflecting a wide range of assumptions upon which the planning process is based. Traffic conditions are likely to differ at distinct times of the day resulting in varied traffic patterns. In addition, movements of the population during a single day can be observed. Hence, traffic distribution must be modeled as a function of the time of the day. This takes into account factors such as commuter traffic, as well as commercial traffic. Each of the PODs defines a catchment area, whose population is assigned to that particular POD. According to public health officials, existing tools focus on the POD design itself and do not take into account emerging traffic in the affected geographic region.

Public Health officials have observed some of the intrinsic problems of existing plan designs during influenza vaccination clinics that were offered at specific public sites. Specifically, the problems of elderly or mobility impaired having to stand in line for hours or being unable to reach their corresponding POD site must be addressed. Further, recent H1N1 vaccination clinics have uncovered limitations of the POD approach that need to be addressed in the planning process. The response to an Anthrax release in the community does not require mass vaccination, but instead the distribution of prophylactic antibiotics. Therefore, strategies must be developed that facilitate the distribution of medication efficiently and within mandated time frames.

1.1. Response Plan Analysis

Regional disaster preparedness committees have devised contingency plans relying on the placement of PODs within a geographic space. In the past, the placement of PODs has been determined manually by public health experts based on the availability of adequate facilities and human judgment. This has led to two problems public health officials have not been able to easily solve with the given resources. First, it must be assessed whether current response plans are feasible in terms of resource and time constraints. This ensures that the response plan will be viable in the event of a bio-emergency. In particular, resources at the POD

locations must meet the expected service demand, and the road network must be able to support the emerging traffic. Second, a feasible placement of PODs with the given resources is to be determined, such that mandated time-frames are met. A current contingency plan might not yield a feasible solution in terms of overall serving-time of the geographic region. Given resource constraints and demographic, as well as geographic properties, a feasible solution can be computed, minimizing the time to serve the entire population.

The PODs will be distributed over the geographic area and may only be reachable by car by the majority of the population. Hence, feasibility studies of response scenarios necessitate the evaluation of traffic emerging within the geographic area served by the PODs, as well as the influx at the POD locations themselves. The feasibility of a response scenario is influenced by different factors necessitating the integration of various data sources pertaining to geography, demography, assumptions, and resource constraints. Pertinent geographic information comprises geographic entities at different granularities and the road network with traffic information. Demographic features of a population can be found within census information, which can be obtained from the U.S. Census website [22] containing data from the last census, which was held in 2000. For the purpose of this research, assumptions refer to hypotheses/expectations defined by public health experts, including the average number of people traveling per car to the POD locations, and the degree of panic within the population. Resource constraints comprise time-frames mandated by government agencies, as well as resource constraints pertinent to regional health departments themselves. The latter also takes into account human resources, as depending on the type of emergency, nurses and other personnel is needed to assist at the POD locations. Resource constraints influence how many POD locations are viable and effectively determine the number of booths that can be operated at each of them. The available road network data-sets for U.S. regions contain most roads within the region, though minor roads might not be included. Nevertheless, minor roads generally contribute to local neighborhood traffic and are deemed less significant and can therefore be neglected. This is due to the fact that neighborhood traffic eventually

leads to traffic on major roads, for which data are available. Using those road segments with available traffic counts, a model can be built estimating the traffic across all road segments in the county. Based on the assumption that most of the traffic of the county will flow towards the POD locations, a higher traffic density may be observed the closer a road segment is located to the POD. Particularly in the vicinity of the PODs, cars will have to enter and leave parking lots or other POD facilities. Hence, catchment areas similar to Voronoi tessellations can be determined (see Figure 1.2) and each catchment area can be analyzed individually, as most of the traffic will be close to the PODs. Integrating the thereby obtained results, traffic along the catchment area borders may be analyzed. Methodology to determine catchment areas to analyze the emerging traffic within geographic regions and estimate demand at the POD locations is discussed in Chapter 3

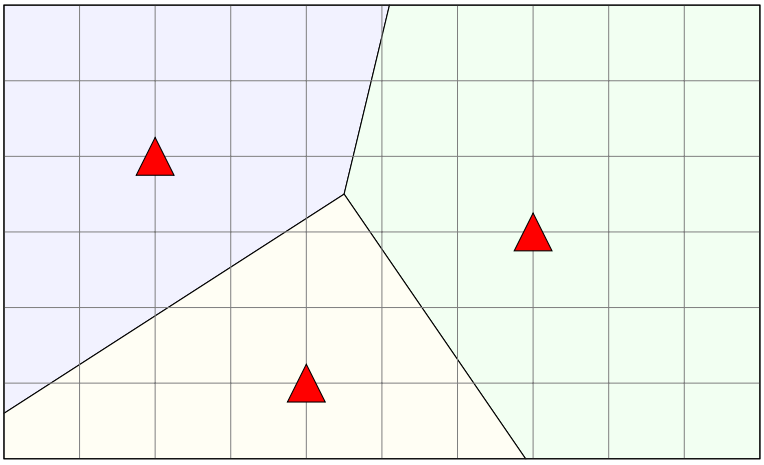


FIGURE 1.2. Partitioning of a geographic region into catchment areas

1.2. Optimizing Response Plans

The analysis of an in-place contingency plan may indicate that the underlying plan will not meet mandated guidelines, thus posing the question: "Does there exist a feasible placement of PODs with the given resource constraints, such that these guidelines are met?" If such a placement exists, the solution shall consist of a set of POD locations constituting such a feasible placement. On the contrary, if the given resources cannot be placed to meet the demand and guidelines, the solution should be an optimal placement that will yield a

best solution to the problem approaching the guidelines as closely as possible. Note that, in reality, the theoretical optima may not be achievable and that there may be more than one best feasible solution. Further, it is of interest to derive an optimal solution without resource constraints and geographical limitations. A resulting solution may not be feasible in reality, but it represents a lower bound for any feasible solution. Computational methods capable of answering these research questions are discussed in Chapter 4, applying methods from discrete and continuous locations science to the placement of PODs. Three specific challenges are of particular interest:

(i) It is to determine whether for a specific geographic region, a partition can be found such that the available resources are distributed to yield a balanced population-resource ratio for all PODs. Thereby, the population is assigned to the individual PODs, such that all PODs finish serving the entire population of their respective catchment areas at approximately the same time. If no such solution exists, then the available resources will not suffice and mandated time-frames will not be met.

(ii) Planning authorities may have identified a set of potential POD locations, but only limited resources are available to operate a sub-set of these locations. Choosing an optimal sub-set of PODs can significantly improve the performance of a response scenario.

(iii) Preparedness coordinators may be interested in the areas covered by the chosen sub-set of PODs within mandated time-frames. Choosing different sub-sets will yield different coverage patterns and distinct overall coverage. Areas that are not assigned to a POD may need to be served by alternate methods, or the overall resources may need to be increased.

1.3. Generation of Synthetic Regions

The analysis and design of bio-emergency response scenarios require the underlying data be kept strictly confidential. In particular, the individual POD locations are only disclosed in the event of an emergency, protecting the PODs from being sabotaged and preventing

high traffic volumes during the set-up phase. Nevertheless, the motivation for generating synthetic geographic environments is manifold. Real data are often subject to confidentiality constraints, which may inhibit research and other specific data utilization. Confidentiality is often achieved by obfuscating underlying data, while maintaining their inherent characteristics. Further, synthetic data give researchers flexibility to explore non-existing regional features, thereby facilitating detailed what-if analyses. Real data rarely reflect temporal dynamics of regional demographics, which limits the ability to predict future demographic or geographic characteristics. Such restrictions can be overcome by providing fully customizable data sources that can be modified to reflect potential future regional features.

The generation of synthetic areas requires the integration of multiple disparate data sources. Information regarding demographics, geography, road networks, and traffic must be combined into a single model. The generation of the layout of synthetic regions involves methodology that is rooted in the field of computational geometry. Specifically, the generation of sub-entities with realistic characteristics requires the placement of features in accordance with realistically occurring shapes. The collocation of demographic features and transportation infrastructure necessitates the use of graph models that facilitate the representation of road features in a region. To enhance the comprehension of features associated with synthetic regions, multi-mode visualization techniques could be employed. Such methods may include the simultaneous display of related features, thereby providing a context for their interpretation.

Once synthetic counties with their corresponding populations have been generated, the results can be integrated into the analysis of bio-emergency contingency plans. This allows for the development of methodology without having to obfuscate underlying data. The results can be published to the scientific community and no confidential information will be disclosed.

1.4. RE-PLAN

Collaboration with county-level disaster preparedness coordinators in an attempt to facilitate the planning process led to the development of quantitative models and computational tools. Analytical results produced by these newly developed tools are immediately integrated into the disaster preparedness planning process, resulting in the modification of existing response policies and the inclusion of new response strategies. As a result of this collaboration RE-PLAN (REsponse PLan ANalyzer) has been developed, which allows for the analysis of response plans involving the placement of PODs. RE-PLAN emphasizes the analysis of emerging traffic and identification of bottlenecks during response plan execution. Some of its components are shown in Figure 1.3. RE-PLAN's framework and methodology for the evaluation of response scenarios were developed in the context of this dissertation. These, however, exclude the actual manipulation of traffic data and implementation of interpolating traffic methods, which were developed as part of contract work for Tarrant County. RE-PLAN's architecture is extensible, allowing for the addition of plan optimization tools in the future. Design and implementation of these additions are National Institutes of Health (NIH) funded, and are in part included in this dissertation.

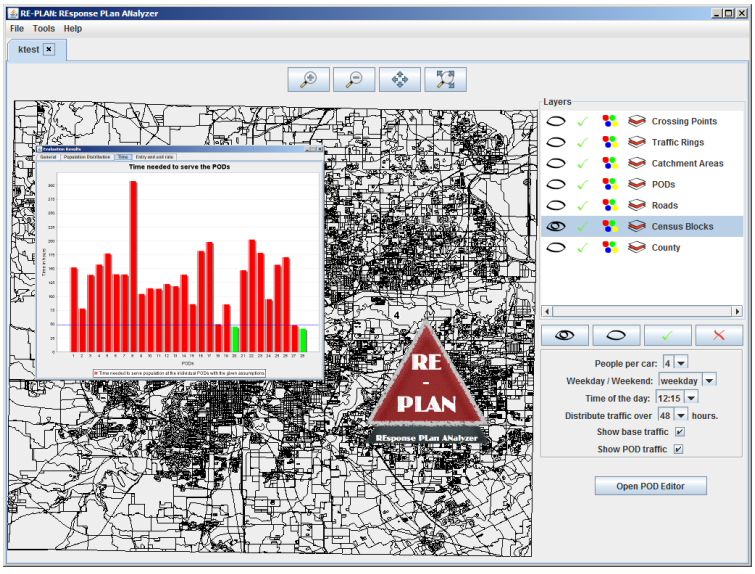


FIGURE 1.3. RE-PLAN

1.5. Overview

This dissertation addresses the question of combining disparate constraints of an emergency response scenario into a multivariate optimization problem. These constraints include government mandated time frames, as well as resource limitations and road capacities. Methodology discussed here provides an evaluation of such response scenarios by identifying bottlenecks in the road network and evaluating the expected demand at specific POD locations. Further, in this dissertation strategies to solve the multivariate optimization problem (demographics, geography, assumptions, constraints) have been developed, such that a feasible solution for the POD placement is determined. Thereby, assumptions and constraints of response plans are simplified such that more complex schemata of the underlying assumptions and constraints can be adjoined in future work. For the optimization task, optimality criteria must be defined. For the purpose of this dissertation, basic criteria, such as population distribution and Euclidean distance are used. These can easily be replaced by more complex functions.

The research described in this dissertation has resulted in new computational techniques to generate synthetic geographic regions with corresponding synthetic populations at census block granularity. These methods facilitate the construction of synthetic regions in the form of synthetic counties with corresponding sub-entities, such as census tracts, block groups, and census blocks. Features in the form of very basic coarse road networks and census populations are assigned to the geographic regions.

This dissertation is structured as follows: Chapter 2 provides a review of literature relevant to the material covered by this dissertation. In Chapters 3 and 4 methodology and corresponding results of the scientific contributions pertinent to response plan analysis and optimization are discussed. This includes the description of RE-PLAN's framework, as well as methods for analyzing and determining feasible response scenarios. Chapter 5 describes methodology to generate synthetic counties with basic road infrastructures and populations

on a census block granularity. Chapter 6 summarizes the findings in the context of this dissertation.

CHAPTER 2

LITERATURE REVIEW

2.1. Response Plan Analysis

In recent years geographic information systems (GIS) have experienced a transformation from mere mapping tools to powerful and valuable tools to improve health care [67]. For instance, GIS have recently been used to examine geographical variation in the need for health services, determine access to health care, show geographic variation in utilization (service-related and socioeconomic factors), and to improve homeland security. For the analysis of contingency plans, it is imperative to account for geographic components, including road infrastructures, as well as point of dispensing (POD) locations and spatial population distributions. The need to integrate computational tools and GIS data into disaster preparedness efforts has been widely recognized. A Montana disaster preparedness exercise demonstrated the usefulness of GIS integration into emergency planning [44]. Existing geographic data, as well as data-sets reflecting the current situation (e.g. affected areas of a chemical spill or locations of disease cases) during the exercise, were used. The lessons learned from that exercise are applicable to many real emergencies. In North Carolina, the use of ArcPad, a mobile GIS software package, replaces the use of paper forms in order to accelerate disaster relief [43]. It aids rapid assessment teams in calculating routes to survey locations and leads to a more timely assessment with more complete data entries. Once the information is collected, it can be uploaded into a database, thereby eliminating data re-entry.

Scientific methods from a variety of fields have been applied to solving problems in public health and emergency preparedness. Wright et al. [103] stress the importance of operations research in homeland security and provide an overview of models being used, specifically discussing their applications in homeland security. Research in the context of homeland

security in general and emergency preparedness and response in particular has been further categorized: location and resource allocation, evacuation models, and disaster planning and response. The latter includes planning procedures that address bio-terrorist attacks using smallpox or anthrax. Response scenarios for the case of such bio-threats have been widely discussed in literature. Miller et al. [69] analyze response plans for attacks with aerosolized smallpox. Specifically, the performance of different public health remediation strategies (ring vaccination, mass vaccination, and quarantine) using the public health infrastructure of San Antonio are evaluated. For the analysis, a combination of a casualty prediction model (CPM) and a health-care complex model (HCM) are used. The CPM represents contacts between individuals, disease transmission, and public health actions, such as vaccination. The output of the CPM serves as input for the HCM, which simulates the use of resources at facilities and patient outcomes. As a result of this simulation, a combination of different strategies, as opposed to following a single strategy, was found most effective. The analysis, however, did not consider the cost of the resources necessary to implement the most effective strategy. Further, Miller et al. state that ring vaccination alone is most likely to fail, as the population in general does not have herd immunity against smallpox. Different emergency responses for airborne anthrax attacks are compared by Wein et al. [101]. The authors use a mathematical model and show that the timely use of oral medications in the exposure region is essential to minimize casualties. Their system consists of an atmospheric dispersion model, an agent-dependent dose-response model, a disease progression model, and spatially distributed two-stage queuing systems for antibiotic distribution and hospital care. Research presented by Kaplan [57] primarily focuses on modeling vaccination policies of large U.S. cities and estimating the number of deaths caused by disease outbreaks. In particular, a traced vaccination policy, which is currently implemented in many U.S. health departments, is compared to a hypothetical mass vaccination policy. For the case of ring vaccination, mass vaccination is only applied, if the disease cannot be contained. For the analysis, homogeneous mixing, rather than nonrandom mixing was assumed. The researchers make use of a smallpox

disease transmission model and show that mass vaccinations yield fewer deaths and faster eradication than traced vaccinations.

Strongly related to the topic of this dissertation is the design of the POD layout, which allows disaster preparedness coordinators to determine the specifics of the POD operation, such as necessary resources. RealOpt, a simulation and decision support system for POD design and planning large-scale emergency dispensing clinics, is presented in by Lee et al. [62] [63]. The system focuses on POD layout, staffing scenarios, and clinics design. Further, it aids in determining POD locations while limiting the maximum travel distance in miles for individuals in the affected area and assigning each jurisdiction (district) at least two PODs. The authors present facility-layout and staffing scenarios for a specific anthrax emergency drill, for which the system was used to estimate necessary staff resources to dispense anthrax medication. Individuals are transported to the POD locations via buses or drive their own cars; these arrivals are modeled via a stochastic process. Upon arrival, individuals are classified into the following categories via a triage: symptomatic, exposed or unexposed. The POD facility itself is modeled as an open-queuing network with the individual stations as the nodes of a directed graph. Service times at the stations are assumed to be independent.

Closely related to the focus of this dissertation is the effort of planning authorities to identify solutions for the fast delivery of medications. An experiment has been conducted in Minneapolis-St. Paul, in which mail carriers deliver anthrax medication [71]. This approach necessitates the stockpiling of antibiotics by mail carriers for themselves and their families. Although this seems to be a feasible solution for the dispensing of medication, it imposes two major concerns: First, it is impractical to keep all mail carriers current on vaccines and medications for a variety of different health events. Second, this approach is not feasible for vaccinations, which in contrast to the simple distribution of medication, need to be administered by trained personnel.

The REsponse PLan ANalyzer's (RE-PLAN) analysis module uses an approach to generate POD catchment areas that is derived from Voronoi tessellations of geographic regions.

Nevertheless, catchment areas could also be defined using different approaches. Similarly, methodology to define locality boundaries is discussed by Coombes [33]. These locality boundaries can yield different partitions depending on which criteria the locality definition is based. Coombes uses synthetic data sets, which provide an abstract representation of the critical information of any boundary set. A variant of Voronoi diagrams, Voronoiesque diagrams, are applied to the redistricting problem by Svec et al. [92]. Voronoiesque diagrams are population weighted Voronoi diagrams. The authors apply a partitioning approach, that makes use of generator points, to New York State to generate 29 legislative districts.

Current evaluation and optimization techniques deployed in RE-PLAN calculate distances based on the Euclidean distance metric. In the future, these distance calculations can be refined by using road distance. In an emergency scenario the population is likely to select short routes to the PODs. A variety of different shortest path algorithms have been applied to transportation networks [76]. These include classical primal and dual algorithms, as well as dynamic approaches. Calculations of the latter are not only based on travel distance, but also take travel time into account. Zhan and Noon [106] address shortest path algorithms specifically on real road networks, as opposed to randomly generated networks. 15 shortest path algorithms are compared on road networks with rural, suburban and urban topologies. If the maximum arc length of a network is within certain bounds, Zhan and Noon recommend Dijkstra's algorithm for solving one-to-one shortest path problems.

Response plan design and analysis exhibit similarities to the construction of evacuation scenarios, and the problem can be mapped onto an evacuation problem. First, depending on the evacuation scenario, affected individuals are trying to either get to a specific point (e.g. exit of a building) or away from a specific point (e.g. hurricane). These points correspond to the POD locations in an emergency response plan. Second, techniques used for the analysis of evacuation scenarios addressing the displacement of large sub-populations can be applied to a large number of evacuees arriving at a POD location. For certain types of evacuation scenarios, such as hurricanes and other natural disasters, a drastic increase in

traffic volume can be expected. Much research has been conducted to address these types of situations. However, as described by Saathoff and Noftinger [85], evacuation in case of *unnatural* disasters is different from evacuation during *natural* disasters, which often can be predicted ahead of time. However, the behavior of individuals is difficult to predict, which was shown for a chemical spill in Arkansas, where an additional 32% of the population decided to evacuate, although instructed to stay. This kind of behavior is also likely to occur for scenarios of mass vaccination or dispensing of medication. Chiu et al. [28] discuss traffic flow in the event of such *no-notice* mass evacuations.

In recent years, researchers have become increasingly aware that interdisciplinary collaboration is important in order to generate good solutions for such complex problems. Shoaf and Rottman [91] stress the importance of this collaboration, while indicating that public health itself is an interdisciplinary field. For early warning systems to be successfully implemented, clinicians must work closely with public health officials [48]. Specifically, when anthrax or anthrax-resembling substances have been distributed through the mail system, overlapping phases of response require collaboration with clinicians. During the initial phase, detection and confirmation of cases or envelopes are essential. Further, trained personnel must be sent to affected sites. During the second phase, full scale investigations and interventions have to be applied. Finally, long-term consequence management and remediation strategies taking into account findings from the prior phases are implemented. The authors consider collaboration with clinicians essential during all phases. In addition to the need for collaboration, the literature emphasizes the need of deploying effective disease surveillance [66]. Further, a distinction between naturally occurring disease outbreaks and bio-terrorist attacks must be carefully evaluated using disease characteristics with respect to geographic area, season, and mode of transmission. Experts consider attacks with the following substances of particular concern: agents of anthrax, plague, brucellosis, smallpox, viral encephalitis and viral hemorrhagic fevers. This is due to the fact that they can be produced easily and inexpensively, are

likely to be fatal, and can be aerosolized. The latter property might be used by terrorists to cover large geographic regions and therefore infect many people.

While not discussed directly in this dissertation, behaviors of individuals and groups play an important role in emergency situations. The computational framework MASSEgress applies theoretical agent-based models for human and social behaviors in emergency situations [77]. Individual behavior includes sensing and decision-making. Social behavior comprises the interaction between individuals, as well as group behaviors, such as competitive, queuing, herding, and leader-following behaviors. Further, social and economic factors influence the resilience to disasters and the abilities of communities to recover from them [34][36]. These findings provide valuable information for emergency planning and may help public health experts to target specific communities that are likely to receive little coverage by standard procedures.

2.2. Optimizing Response Plans

The problem of optimally placing PODs in a geographic region, so as to provide adequate services to the population, resembles problems in the area of discrete location science. Many problems in location science lead back to the Weber problem, introduced by and named after Alfred Weber [99], whose objective is to place a single facility in a plane. Generally, two phases of location science are considered: the classic phase and the contemporary phase. During the classic phase, the Weber problem was extended creating the p-median problem, whose objective is to minimize an aggregate distance function. This was further extended to the network p-median problem, which has been shown to be NP-hard. Specific needs by some of the applications led to the formulation of covering problems (e.g. set covering and maximal covering). The contemporary phase spawned models with more realistic behavior for consumers, producers and the environment. Some of theses models allow for defining multiple objectives, or the use of goal programming methods, which minimize the deviation of a set of standard goals. The need to incorporate different facility types, in contrast to uniform facilities during the classic phase, led to the introduction of hierarchical models [50].

Other literature includes work by Rushton [83], which provides a collection of computer programs solving a variety of facets of location allocation problems.

Location allocation models have been merged with problems introduced by location theory. Beaumont [7] uses location allocation models to operationalize central place theory, which deduces spatial patterns of settlements. The population of an underlying geographic region is often assumed to be equally, but not continuously, distributed. Central place theory delineates market areas, which supply the population using central places. Tewari and Jena [94] evaluate high school decision making in rural India by comparing school locations proposed by authorities to optimal school locations obtained by applying methods of location science. Many of the existing schools do not provide adequate access to school facilities, while the number of children of school age is growing. Taking into account these schools, additional locations are suggested based on average distance to the school locations or by maximizing the population coverage. Church and Eaton [30] present existing formulations of hierarchical service delivery systems. These systems allow for multiple levels of goals (organized by priority) or facilities. Their service types are categorized into parallel service, sequential service, and referral service. Two specific referral hierarchy covering formulations are introduced in detail and the application to rural health service planning in Colombia is presented. Work presented by Casillas [21] discusses the aggregation problem in locational analysis, which assumes dispersed demand be located at a single point. Analyzing the error in the p-median problem shows that data aggregation has little effect on patterns of locations and service centers. Further, the magnitudes of cost-estimate errors are larger for high values of p and the magnitude of the optimality error grows with p, but exhibits no pattern for different aggregation levels.Rushton [84] stresses the importance of carefully selecting the objective function, which often represents the objective of a particular problem insufficiently. Thus, it may take into account consumer-provider relationships, although a service-delivery system is the system to be designed. Rushton suggests that, for improved objective functions, future behavioral patterns should be estimated based on observed data.

Solutions to median and plant location problems minimize the overall resource and distance cost by strategically placing facilities such that the demand of the entire region is met. For bio-emergencies, the facilities correspond to the PODs, while the demand is represented by the corresponding population. The problem has been discussed in the literature and different solutions have been proposed. Canovas et al. [19] introduce additional inequalities for the linear programming formulation of the simple plant location problem (SPLP). Solutions to the SPLP require to choose a subset of candidate plants, and allocate each customer to a plant while minimizing installation and allocation cost. The convex hull of feasible solutions exhibits a polyhedral structure from which additional facet-defining inequalities can be derived. Canovas et al. make use of these inequalities to reduce the number of solutions. Similarly, Sherali and Park [90] make use of the convex hull to formulate a linear programming solution as well as different heuristics to the discrete equal-capacity p-median problem. The problem addresses the placement of p new facilities with uniform capacities on the network. A variation of simulated annealing has been applied to the uncapacitated facility location problem (p-median problem) by Chiyoshi and Galvao [29]. Supply of a single commodity is to be achieved for customers with known demands by selecting a subset of the candidate facility set. The approach that is making use of simulated annealing techniques achieved optimal solutions for most problems. Work presented by Hansen et al. [52] proposes the variable neighborhood decompositions search (VNDS), which is an extension of the variable neighborhood search (VNS). For large problem sizes of the p-median problem the VNDS showed improvements over the VNS while reducing the computing time.

As the time and processing power to obtain these solutions increase with an increasing number of demand nodes and facilities, there have been efforts to reduce this computational cost by using pre-processing techniques [5], or devising heuristic approaches [2],[39]. Avella and Sforza [5] propose reduction tests, which estimate values that a set of variables in the optimal solution will take in advance. This technique yields a reduction of the size of an instance. Specifically, reduction tests for the p-median problem are described and their

impact on the resulting problem size is shown. Lagrangian heuristics have been applied by Agar and Salhi [2] to solve large single-source capacitated plant location problems (CPLPs) and multi-CPLPs, whereas Drezner and Erkut [39] use a genetic algorithm for the p-median problem.

Methods for multiple facilities per site, which correspond to multiple booths or lanes per POD location, are described in by Canavate-Bernal et al. [18], and Ghiani et al. [49]. Canavate-Bernal et al compare three heuristics for the single product capacitated machine siting problem (SPCMSP), which is an extension of the simple plant location problem.The capacitated plant location problem with multiple facilities in the same site (CPLPM), which is an extension of the CPLP, is presented by Ghiani et al. Applications of the CPLPM range from polling stations to warehouses and plants. Poor performance of classic lower and upper bounding procedures is overcome by introducing a Langrangean relaxation and a tailored Langrangian heuristic.

According to Rawls' theory of justice, the quality of a solution is no better than the worst-served entity [81]. Center and covering problems in discrete location science address this issue by taking into account a specific service standard, such as the maximum allowable distance between a facility and its customers. Although this problem has been discussed using a variety of different approaches [79],[65], [58], [3], [11], [13], [20], [12], the resulting solutions might not be feasible within an available budget or its defining characteristic to meet the prescribed service standard. Applying this to the placement of PODs, a non-feasible solution that exceeds the service standard results in a scenario in which underlying resources do not suffice and additional resources are required.

Pelegrin and Canovas [79] propose a seed point algorithm for the continuous k-center problem, which outperforms time-consuming iterative algorithms. The continuous k-center problem requires the selection of k facilities in continuous space, whereby a set of discrete demand points are given. The proposed algorithm uses a two-step approach: It first generates

k seed points and then generates the output partition (assignment of consumers to k facilities) based on these seed points.

The problem of placing congested facilities that maximize the population covered with short queues or waiting times is addressed by Marianov and Serra [65]. They extend earlier work and describe a maximal covering approach, which maximizes the population covered by a limited number of facilities into a set covering approach. The latter achieves full coverage of the population with the least possible number of facilities.

Polynomial approximations for two different versions of the capacitated k-center problem are given by Kuller and Sussmann [58]. Their solution to the k-center problem requires k facilities to be located in a graph while minimizing the maximum distance from a vertex to a facility. Each facility is assigned at most as many vertices as a given capacity allows for.

Recent efforts in discrete location science include generalizations of existing models and the development of hybrid models of typical objective functions. Berman and Krass propose [12] a generalization of the maximal cover location problem. A non-increasing step function determines the degree of coverage based on the distance to the nearest facility. This leads to a partial coverage of the consumers. Berman and Krass propose various integer programming formulations and show that in general, the problem is equivalent to the uncapacitated facility location problem. This has been further generalized by Berman et al. [11] by proposing the ordered gradual covering location problem as an extension of the ordered median location problem (OMP) and the gradual cover location problem (GCLP). The proposed model provides a unifying structure with respect to classical location allocation objectives.

Most work in location science focuses on locating facilities in a network and optimizing an objective function, while assuming static edge lengths. Nevertheless, for specific applications, such as emergency facility locations, edge lengths may not be static. If the edge lengths represent transit times, they can be dynamically expressed as a function of traffic load at a specific time. Bathia et al. [13] propose a model of dynamic distance functions and approximation algorithms for the k-center problem.

Averbakh and Berman [6] group customers together and treat them as discrete points in space. For the analysis of response plans, this can be mapped onto the population of a census block being represented by its geographic centroid. Further, applications in discrete location science consider the placement of ambulance bases in a geographic region, providing coverage within a given amount of time [1],[53], as well as the establishment of rural health facilities in developing countries [80].

Similar to the placement of PODs is the selection of shelter locations for evacuation scenarios. Kongsomsaksakul et al. [59] provide a model for emergency response in the event of flooding. Assumptions for the model are that authorities select the shelter locations and evacuees decide which shelter they will approach and which route they will take. The latter is formulated as a stochastic user equilibrium. The model is implemented by making use of bi-level programming techniques, whereby the authorities are represented by the upper level and evacuees by the lower level. Simulation results show that higher shelter capacities yield a smaller number of selected shelters and a lower total evacuation time.

2.3. Generation of Synthetic Regions

The tasks of protecting confidentiality and data obfuscation have been widely discussed in literature. "Measures of Disclosure Risk and Harm" is the result of requests by the Panel of Confidentiality and Data Access of the Committee on National Statistics of the National Research Council [60]. The author highlights that the definition of disclosure is extremely context dependent. In general, individuals are to be protected from being re-identifiable. This includes perceived re-identification, which constitutes an intruder's belief to have successfully re-identified a record. In particular, statistical agencies for protecting confidentiality may either apply statistical disclosure limitation procedures or restrict access to data itself. Jabine [55] reviews methods and procedures developed and used by U.S. statistical organizations. The general public has grown increasingly weary of participating in surveys, as new analytical techniques are developed and computing power increases, thereby increasing the possibility of revealing formerly protected data. Addressing these issues, the

trade-off between confidentiality and data access is emphasized by Duncan et al. [40]. This embraces individuals who refuse to share personal information with government organizations or corporate agencies due to possible abuse or leakage of information. Factors contributing to this perception include mandatory disclosure by government agencies, data sharing within the government, availability of data to non-government users, and possible compromise of information due to lack of information security. In 2004, the Office for National Statistics in the UK established the Virtual Microdata Laboratory (VML). Its main purpose is to provide access to sensitive micro-data while maintaining the confidentiality and security of the data [82].

Research in the field of computational epidemiology frequently involves the abstraction of real world characteristics. The necessity to collapse vast amounts of data and to impute missing data motivates the use of synthetic representations, which are easily modified to reflect specific geographic or demographic properties. Modeling techniques, such as cellular automata [47], agent-based models [45], and social networks [98] are examples of such synthetic representations that provide the necessary level of abstraction to express specific characteristics. For instance, France is now applying virtual reality techniques to military purposes. These methods are used to aid during design, testing, training, mission preparation, and operations [78]. In contrast to virtual reality with underlying computer-aided design (CAD) models, virtualized realities construct virtual models from images of the real world. Kanade et al. [56] present methods to construct virtualized realities, which allow users to move through virtualized events. These events include a variety of different settings, such as surgeries and basketball games.

Synthetic populations in different forms have been created for a large number of applications. Cirillo et al. [31] discuss the generation of a synthetic population with a model for weekly work participation for Belgium. Information regarding real work participation was collected over a time-frame of four years by conducting surveys. Methodology for the synthetic representation of the entire population of the UK has been proposed by Birkin

et al. [14]. Both households and individuals are considered with the aim of developing a national demographic simulation. Clarke and Wheaton [32] address the problem of data sparseness at the neighborhood level. Although large-scale surveys yield data from many different neighborhood groups, the sample size per group is relatively small. Monte Carlo simulations are used to evaluate different group sizes, and cluster analysis techniques are applied to minimize data sparseness.

For the generation of synthetic geographic regions, it is important to ensure that the synthetic representations maintain defining properties of their real counterparts. The Census Advisory Committee addresses the importance of quality assurance for synthetic data sets and introduces the Longitudinal Employer-Household Dynamics (LEHD) Program, now released as the web-based application On The Map, and the Survey of Income and Program Participation (SIPP) [104]. LEHD provides synthetic work-related information including residence, work places, earnings, and industry. The SIPP creates synthetic values based on a sequential regression and multivariate imputation applied to partially synthetic data with missing values. Melhusin et al. [68] evaluate synthetic household populations that were created using spatial micro-simulation techniques. It is shown that these techniques can obtain reasonable results, if the parameters are adjusted adequately. Motivated by the development of activity-based travel models, Beckman et al. [8] make use of such micro-simulations to create synthetic baseline populations of individuals and households. Iterative proportional fitting is applied to census data and to the Public Use Micro-data Sample (PUMS) to estimate demographic features of individual households in the census entities.

One feature that needs to be generated for synthetic regions is an appropriate road network. Many research efforts have modeled and evaluated properties pertinent to road networks. Espie et al. [42] discuss a model allowing for the estimation of the space-mean speed from time-mean measurements. This is in contrast to traditional approaches, which derive the time-mean from space-mean data. The data for the traffic stream is obtained from single loop detectors. A study suggests that higher population densities coincide with higher

road densities [51]. This becomes even more evident when only considering paved roads. The population and road density coherence is explicable by the lower per capita construction cost for a road in denser populated areas. An additional study has been conducted in which the relationship between road patterns and population density of San Antonio, Texas has been evaluated by using fractal geometry [93]. The main incentives for using this approach are the self-similar attributes that satellite towns around large cities exhibit. Results indicate that there exists a correlation between road and population densities in the research area. For this analysis, however, different classes of roads were not distinguished from each other.

Once a road network has been generated, realistic traffic estimations need to be added. Traffic estimation on real road networks has been examined in a variety of research efforts. The working paper *Road Traffic Data: Collection Methods and Applications* for the Joint Research Center of the European Commission compares and contrasts traditional traffic collection methods against new methods [61]. The latter includes techniques relying on in-vehicle devices, such as cell phones. Examples of traffic flows and speeds obtained by different models are illustrated for a variety of countries. Willumsen [102] discusses different approaches for estimating trip matrices. One of these approaches utilizes a gravity model based on traffic counts. Further, mathematical programming techniques for solving equilibrium assignment problems have been applied to estimate trip matrices. The last method utilizes entropy and information theory considerations. In order to discover traffic flow patterns, density-based algorithms have been used to cluster road segments based on density of their shared traffic [64]. In Ntoutsi et al. [72] base their methodology on a graph-based traffic model used to model traffic movements within the road network. This facilitates the estimation of traffic relationships between the individual road segments. The annual average daily traffic is obtained by making rough estimates based on short-period traffic counts. It has been shown that values for locations without traffic detectors can be forecast by using kringing-based methods [97]. These types of methods can also be used to explore other spatial relationships, such as land values, and household incomes. This has been addressed

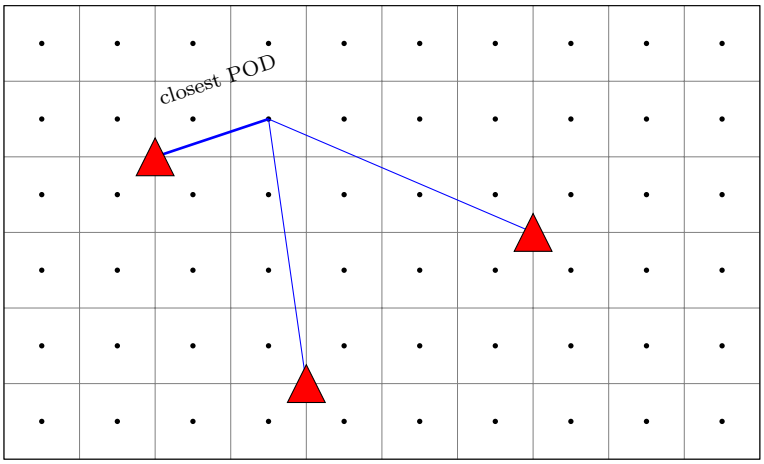
by Borning and Waddell [15], who describe UrbanSim, a model system aimed at linking the planning of land use, transportation, and environmental quality.

CHAPTER 3

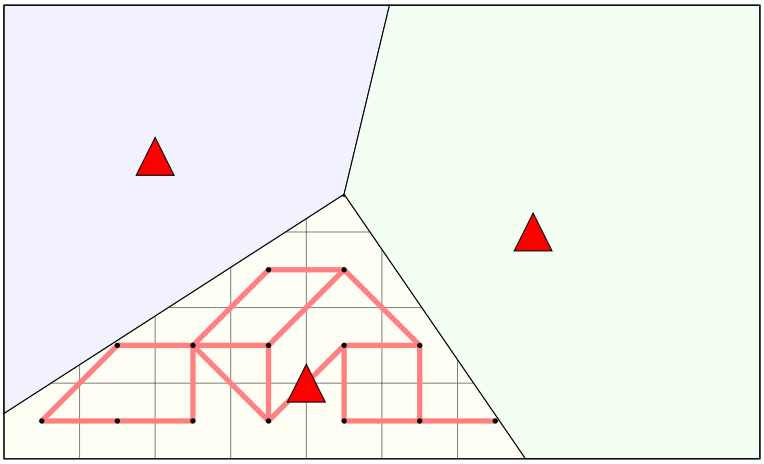
RESPONSE PLAN ANALYSIS

The analysis of existing bio-emergency response plans constitutes a multi-variate optimization problem, as the following data sources have to be taken into account: assumptions and constraints, demographic data, and geographic information. Assumptions are made by public health experts, including the number of individuals traveling in a single car, and the estimated serving time per car at a point of dispensing (POD). The latter influences the POD operation itself, while the number of people per car can have an impact on the traffic conditions. The resources that can be made available for such emergency scenarios include the number of people in the personnel and may limit the number of PODs that can be operated. Additional constraints may be required by government agencies, such as guidelines requiring the response plan to be completed within a limited amount of time. Specifically, the Centers for Disease Control and Prevention (CDC) provide time-frames for supplying the population with prophylactics in the event of anthrax (48 hours [27],[75]) and smallpox (72 hours [23], [24], [25]) attacks. Although the methodology can be applied to different types of emergencies, this dissertation focuses on the event of an anthrax emergency. Demographic information is be used in the form of census data at census block level due to the availability of both geographic and demographic information at this granularity. Geographic information comprises the individual census blocks, the POD locations, as well as the road network with corresponding traffic counts.

Once public health officials have identified specific POD locations, the region's geographic and demographic space is partitioned into catchment areas. This implicitly assigns the entire population to specific PODs, where medication and treatment can be obtained. The granularity of this assignment is currently computed at the census block level, with each census block being represented by its centroid and treated as a single geographic point. This



(a) Assignment of centroids to nearest POD



(b) Catchment area with abstracted road network

FIGURE 3.1. Partitioning of a county into catchment area

abstraction allows for a discrete representation of small geographic regions by a single point. This allows for calculating distances between two points, as the census block regions are represented by their centroids. Each census block is then assigned to the closest POD as defined by an underlying distance metric. This process is illustrated in Figure 3.1(a). Once all catchment areas are defined, the road network within each of them is abstracted into a graph model. Implicitly, the vertices of this graph represent the centroids of surrounding census blocks, whereas the edges represent the cumulative transport capacity between the census blocks. Specific traffic flow within a single census block can be disregarded, as they are assumed to be sufficiently small and the traffic is cumulative within the centroid. An

example of such abstraction is depicted in Figure 3.1(b). This abstraction facilitates the estimation of expected service demands for each POD. It further allows for the analysis of road traffic in the case of a bio-emergency.

Several steps are necessary to analyze a response scenario: the definition of catchment areas, the superimposition of the road network, the construction of a traffic model, and the analysis of the POD capacities. The following sections discuss each of these components in detail.

3.1. RE-PLAN Architecture

RE-PLAN and its underlying methodology have been designed as part of a contract with Tarrant County. Nevertheless, it provides flexibility, such that it can be applied to any real or synthetic geographic region, if the following information is available:

- Non-overlapping, fine-grained sub-divisions of the region: The union of sub-divisions of the geographic space must cover the entire region exactly once. The sub-divisions represent the smallest unit for evaluating response plans and therefore, the entities should be as small as possible. For U.S. counties, such entities are available from the U.S. census in the form of census blocks. Nevertheless, other types of sub-divisions may be used, which also allows for using information from other countries.
- Population counts on a per sub-division basis: For each of the above mentioned entities, population counts are required to ensure that estimation of catchment area populations and traffic volume are as realistic as possible.
- A road infrastructure: This is essential for the bottleneck analysis of response scenarios, as the population will be traveling to the POD locations via the road network. The availability of traffic counts specifically for the underlying region is not required, but yields more accurate estimations of the emerging traffic during a bio-emergency than using a generic model.

RE-PLAN's architectures is designed around the following core components: a geographic information systems (GIS) enabled database, the computational engine, and a graphical

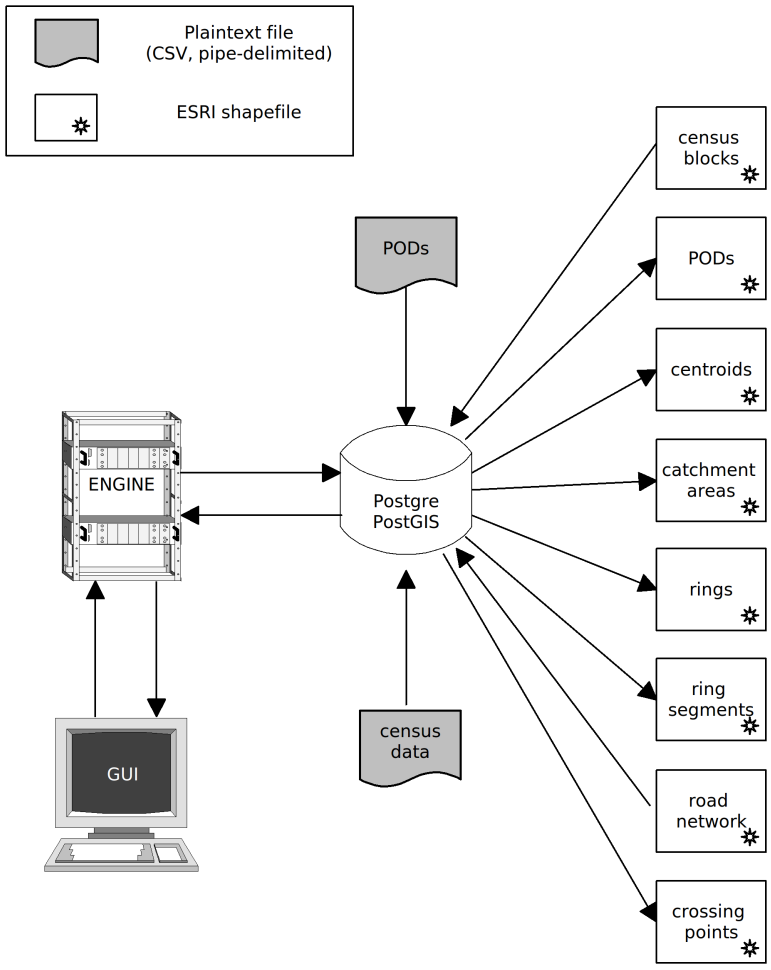


FIGURE 3.2. Structure of RE-PLAN (with courtesy of Inderscience [87])

user interface (GUI). Specifically, the architecture addresses the integration of multiple data sources with diverse formats and data representations. The architecture components and the data flow between the individual parts are shown in Figure 3.2.

RE-PLAN receives input from a variety of data sources in diverse formats, including comma-separated values (CSV) files, pipe-delimited files, and shapefiles as defined by the Environmental Systems Research Institute (ESRI). Input can also be received from a graphical user interface that facilitates interactive data entry for real-time what-if analyses. CSV file formats are most commonly used in the field of public health to represent PODs and data associated with them. They are easily imported into and exported from spreadsheet applications and shapefiles utilized by a variety of GIS tools. Dependencies of the different data

sources are illustrated in Figure 3.3. These dependencies determine the order of computation and show at which stage specific raw data (census blocks, PODs, and road network) need to be provided. Only geographic information regarding the census blocks, PODs and road network need to be obtained from external sources. All other information is dynamically computed as described in the following sections.

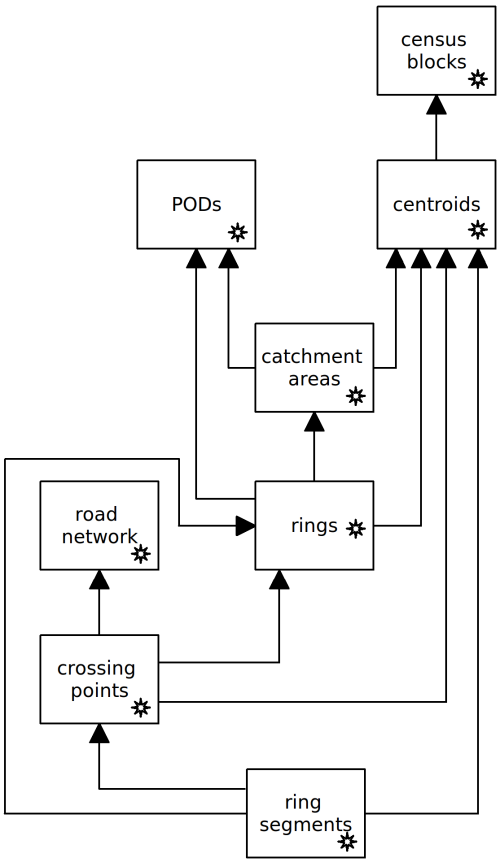


FIGURE 3.3. Dependencies of data components (with courtesy of Inderscience [87])

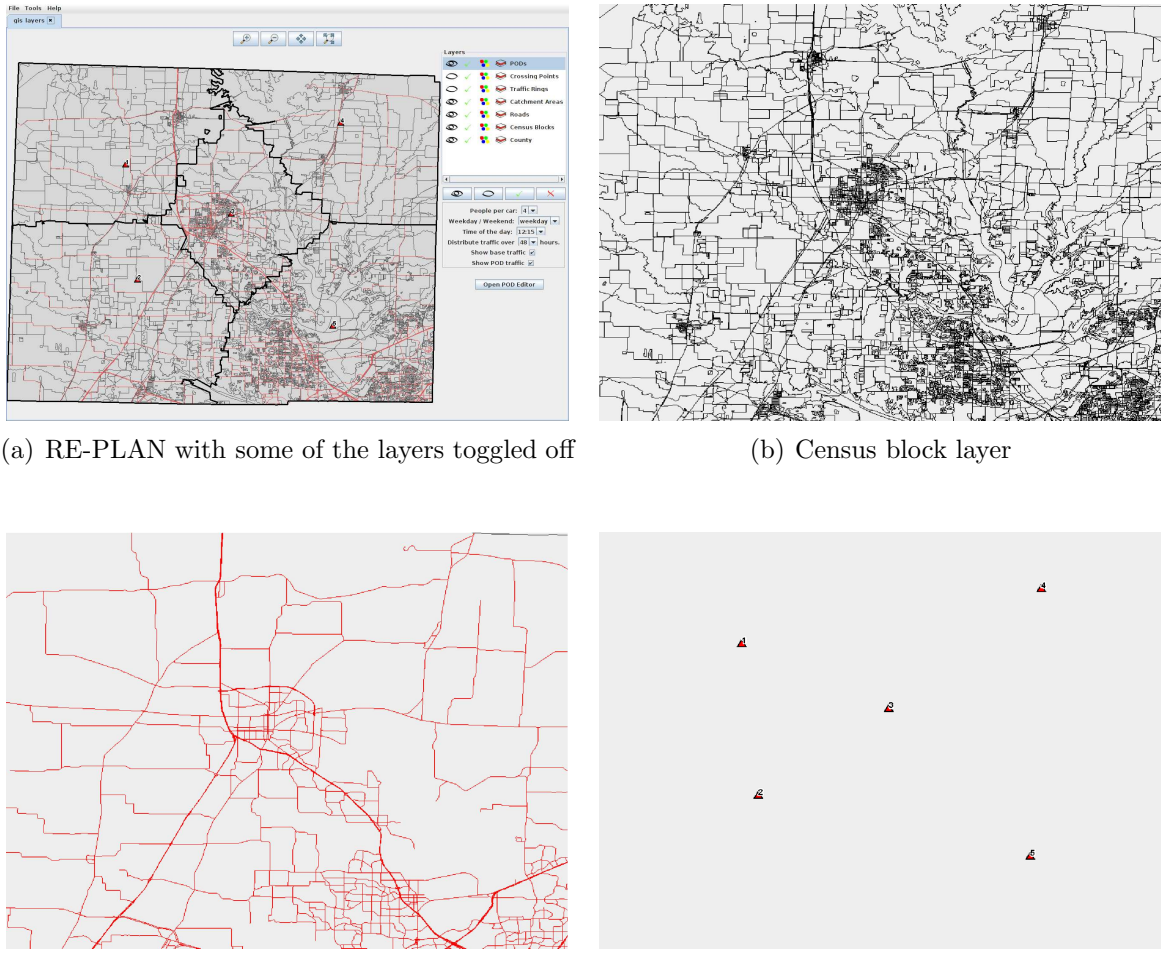
Census data obtained from the US census website "American Fact Finder"[17] are available as pipe-delimited data sets along with the corresponding geographic entities as ESRI shapefiles. Fact Finder provides data sets at levels of different geographic granularities, such as counties and census blocks. For the purpose of this research, a census block granularity was chosen and the corresponding census data provide information on a census block

level. The associated shapefile contains the geographic information represented by geometric shapes, such as polygons and lines. See the "ESRI Shapefile Technical Description White Paper" [41] for detailed information on shapefiles.

The U.S. Census Bureau provides road network information for the United States in the form of topologically integrated geographic encoding and reference system (TIGER©) shapefiles. These files are limited to spatial data and road classifications. Additional information for specific road segments, such as number of lanes and speed limits, may be available from local authorities. For this research, local road networks and traffic information were provided by the North Central Texas Council of Governments (NCTCOG) [70].

The graphical user interface allows users to visually examine existing response scenarios and facilitates an interactive planning process. Upon displaying the geographic region of interest in a main window, users can apply many of the operations found in standard mapping applications, such as zooming and moving. Existing and computed geographic entities are displayed as layers, which can be toggled on and off, thereby displaying specific aspects of response plan features (see Figure 3.4). Existing geographic entities include features that can be retrieved from information sources, such as road networks, census blocks, and POD locations. Computed entities that are calculated during the response plan analysis include POD catchment areas and traffic distributions.

The POD editor provides the necessary functionality to add, remove and modify POD locations. This allows for targeted modifications and an interactive analysis. Functionality to toggle between different POD locations facilitates the experimental optimization of POD placement. Further, PODs can be labeled as *corporate*, which excludes them from the plan analysis. Corporate PODs are facilities, such as hospitals and certain businesses, which provide emergency responses for their employees. RE-PLAN's PODAnalyzer provides information of the expected conditions at the individual POD locations in case of an emergency and is discussed in more detail in Section 3.3.3.



(a) RE-PLAN with some of the layers toggled off

(b) Census block layer

(c) Road layer

(d) POD layer

FIGURE 3.4. GIS layers in RE-PLAN

The input data is converted into database tables and stored in a PostgreSQL database with PostGIS extensions. PostGIS adds necessary support for spatial objects to the PostgreSQL database and is supported by the GIS toolkit GeoTools. Providing a range of geometry functions, such as convex hull, centroid, and distance calculations, PostGIS allows for specific geographic calculations (e.g. calculation of centroids) to be executed within the database. This improves the overall performance by avoiding unnecessary read and write operations and communication between the computational engine and the database. Further, this allows for most computations of the plan analysis to take place on the database server, and therefore, RE-PLAN's execution is partially independent of the performance of

the client machine. The computational engine provides the necessary functionality to export database tables with geographic content into shapefiles that can be imported by other GIS tools.

The computational engine is implemented in Java and utilizes the open-source geospatial library GeoTools. The GeoTools library, described in "Open Source Approaches in Spatial Data Handling" [95], provides build-in support for a variety of GIS-related formats, such as ESRI shapefiles and PostGIS database tables. Its methods are compliant with standards by the Open Source Geospatial Foundation (OSGeo) [74] and implement many Open Geospatial Consortium (OGC) [73] specifications. The following sections provide a description of the methodology implemented in the RE-PLAN engine for the analysis of existing response plans.

3.2. Distance Metrics

For the calculation of the distance between any two entities, such as PODs and census block centroids, different distance metrics can be used. The most intuitive distance metric is the Euclidean distance (ED) between two points. For the purpose of this research, the Euclidean distance was used without taking the underlying road network into consideration. Given the coordinates of two locations (points) s and t, the Euclidean distance is calculated as follows [37]:

$$ED = d(s,t) = \sqrt{(s_x - t_x)^2 + (s_y - t_y)^2}$$

whereby s_x, t_x denote the x-coordinates of s and t; s_y, t_y represent their y-coordinates. The proximity of s and t does not imply the existence of a direct path. That is, service demands emerging from a location l_i in the catchment area for a POD located at a location l_j may affect the entire path of the road network leading to l_j.

Once the road network is included in the graph model, the Euclidean distance metric can be replaced by the road distance (RD) metric. This improves the prior approach in respect to natural and man-made obstacles, such as rivers, mountains, or artificial barriers. This distance metric could be included into RE-PLAN, basing all distance calculations on the road distance. Catchment areas would then show less resemblance to Voronoi tessellations,

as the road distance depends strongly on the road network and its density. In dense portions of the road network it is more likely to find a road distance between two points that is close to their Euclidean distance. Since such distance calculations may necessitate the use of shortest path algorithms, the time complexity of distance calculations would increase. The road distance is generally referred to as the shortest path between two points in the graph [37]. The road distance can be expressed iteratively as

$$RD = \min_j(d(s, l_{p_j,1}) + d(l_{p_j,1}, l_{p_j,2}) + ...$$
$$+d(l_{p_j,n-1}, l_{p_j,n}) + d(l_{p_j,n}, t))$$

where $n+1$ is the number of road segments between s and t on path $p_j \in P$. P contains all paths from s to t. The intersection points between the individual road segments for a path p_j are denoted as $l_{p_j,i}$, $i \in [1, n]$. The function $d(l_{p_j,i}, l_{p_j,i+1})$ with $i \in [0, n]$ takes the two end points of a road segment on path p_j as an argument and returns the road distance stored in the graph model. This yields the shortest road route, as all possible paths in P are evaluated. By setting $s = l_{p_j,0}$, and $t = l_{p_j,n+1}$, the formula can be rewritten as

$$RD = \min_j \left(\sum_{k=0}^{n} d(l_{p_j,k}, l_{p_j,k+1}) \right)$$

This is the sum of all the road segments of the shortest path from s to t. Thereby, results yielded by shortest path algorithms, such as the Bellman-Ford algorithm [10] or Dijkstra's algorithm [38] are obtained.

The last *pure* distance metric is estimated travel time (ETT), for which further information is needed. Based on the road distance, road capacity, speed limits, and base traffic, the time it takes to travel between two points can be estimated. The road distance seems to improve on the Euclidean distance by implicitly including obstacles. However, it does not take into account the actual travel time between two points. Thus, a path with extended road distance, but shorter travel time may be preferable over the path with minimum road distance. Factors, such as speed limits, traffic counts and construction, have to be evaluated carefully and the question arises whether or not GIS-based estimates are sufficiently

accurate. Haynes et al. [54] compare estimated travel times to hospitals with real travel times. One factor to be considered is that cars might not be able to travel at the maximum permissible speed for a particular road segment and may be further slowed at intersections. Therefore, an average speed estimate has to be used. This, however, cannot be generalized to average travel speeds per road segments, as it strongly depends on the underlying traffic, which itself depends on the time of the day. Furthermore, the expected traffic patterns during a bio-emergency are likely to differ from regular traffic. The consideration of this change in traffic volume is essential in order to produce a realistic plan analysis.

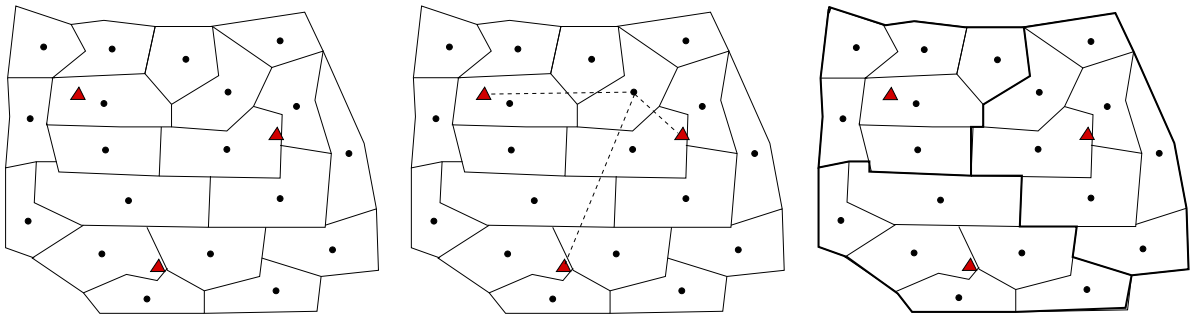
The above discussed distance metrics can be merged into a hybrid model, the mixed distance metric (MD). The individual distance metrics are assigned weights w_{ED}, w_{RD}, and w_{ETT}, with $w_{ED} + w_{RD} + w_{ETT} = 1$. The resulting formula for the mixed distance metric is

$$MD = w_{ED} \times f_{ED}(ED) + w_{RD} \times f_{RD}(RD) + w_{ETT} \times f_{ETT}(ETT)$$

whereby f_{ED}, f_{RD}, and f_{ETT} represent functions that normalize the corresponding distance metric. With $w_i \in [0, 1]$, this formula can be seen as a generalization of the above discussed distance metrics. Such a generalization can reflect weighted factors on which individuals base their decisions to select a POD location. A POD location may be close in Euclidean distance, but a second POD may be closer in terms of road distance. A third POD may be further away, but easily reachable via a major highway and hence has a small estimated travel time. Each of these factors may induce individuals to choose a POD.

3.3. Catchment Areas

Assuming that individuals will obtain vaccines or medication at the POD closest to them, the geographic region is subdivided into catchment areas. These catchment areas are calculated by assigning each census block to the POD nearest to it. Although census blocks represent geographic regions and PODs are single locations, the distance calculation between a POD and a census block yields a single value. This is achieved by representing each census block by its geographic centroid, which is represented by a single point in space.



(a) Census blocks and PODs (b) Distance to all PODs (c) Catchment areas

FIGURE 3.5. Determining catchment areas (with courtesy of WIT Press [88])

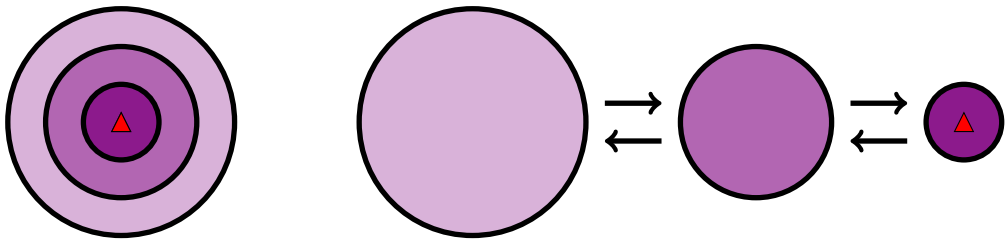
The following describes how census blocks are assigned to the PODs: The distances between the set of census block centroids C and the set of PODs P are calculated and a centroid $c \in C$ representing its corresponding census block is assigned to a POD $p \in P$, if the distance between c and p is smaller than the distance between c and any of the other PODs $p' \in P \setminus p$. This assignment can be expressed as follows:

$$\forall c \in C : pod(c) = min \lim p(\parallel c - p \parallel)$$

where $pod(c)$ takes a census block centroid as an argument and returns the corresponding POD. If a census block centroid has the same distance to multiple PODs, it is uniquely assigned to one of them. This approach is not limited to the Euclidean distance, but allows for the use of other distance metrics, such as minimum road distance, minimum travel time, and a weighted mixed distance metric. The generated catchment areas show strong resemblance to Voronoi tessellations adjusted along census block boundaries. This approach is depicted in Figure 3.5. Figure 3.5(a) shows census blocks with their corresponding centroids. The POD locations are marked by triangles. The process of assigning census blocks to their corresponding closest POD is illustrated in Figure 3.5(b). This leads to the partitioning of the geographic space into catchment areas, as shown in Figure 3.5(c).

3.3.1. Rings of Proximity

For a response scenario to be deemed feasible, the underlying road network of the affected region must be able to route the population to the POD locations within mandated time-frames. Therefore, the analysis of the traffic conditions during a bio-emergency is an essential component of response plan analysis. Individuals in census blocks nearest a specific POD location will be able to access the facility in a shorter time than individuals in census blocks near the outskirts of a catchment area, who will have to travel through many census blocks on their way to the POD. This can be represented by a system of concentric rings that can be drawn along census block borders around each POD. For instance, for a system of three rings or more, the inner rings contains the PODs. Individuals located in the outermost ring will have to traverse multiple interior rings. Consequently, one can expect the traffic density to increase towards the POD location. Figure 3.6 exemplifies this procedure and shows the traffic flow between the individual rings.



(a) Concentric rings surrounding the POD

(b) Traffic flow between the rings towards and away from the POD

FIGURE 3.6. Ring abstraction to determine traffic flow

Traffic on specific road segments is likely to increase the closer to a POD the road segment is located. Assuming different traffic densities as a function of distance, the catchment area of a POD can be represented by concentric rings surrounding the POD. Let k denote the number of rings and p_{max} the point within a specific catchment area that is furthest away from its corresponding POD. Further, let d_{max} denote the distance between p_{max} and the POD p of its corresponding catchment area. Then the census blocks of a catchment area assigned to POD p are classified into sets of rings $\bigcup_{i \in \{1..k\}} R_i$. Ring R_i with $i \in \{1..k\}$ is then

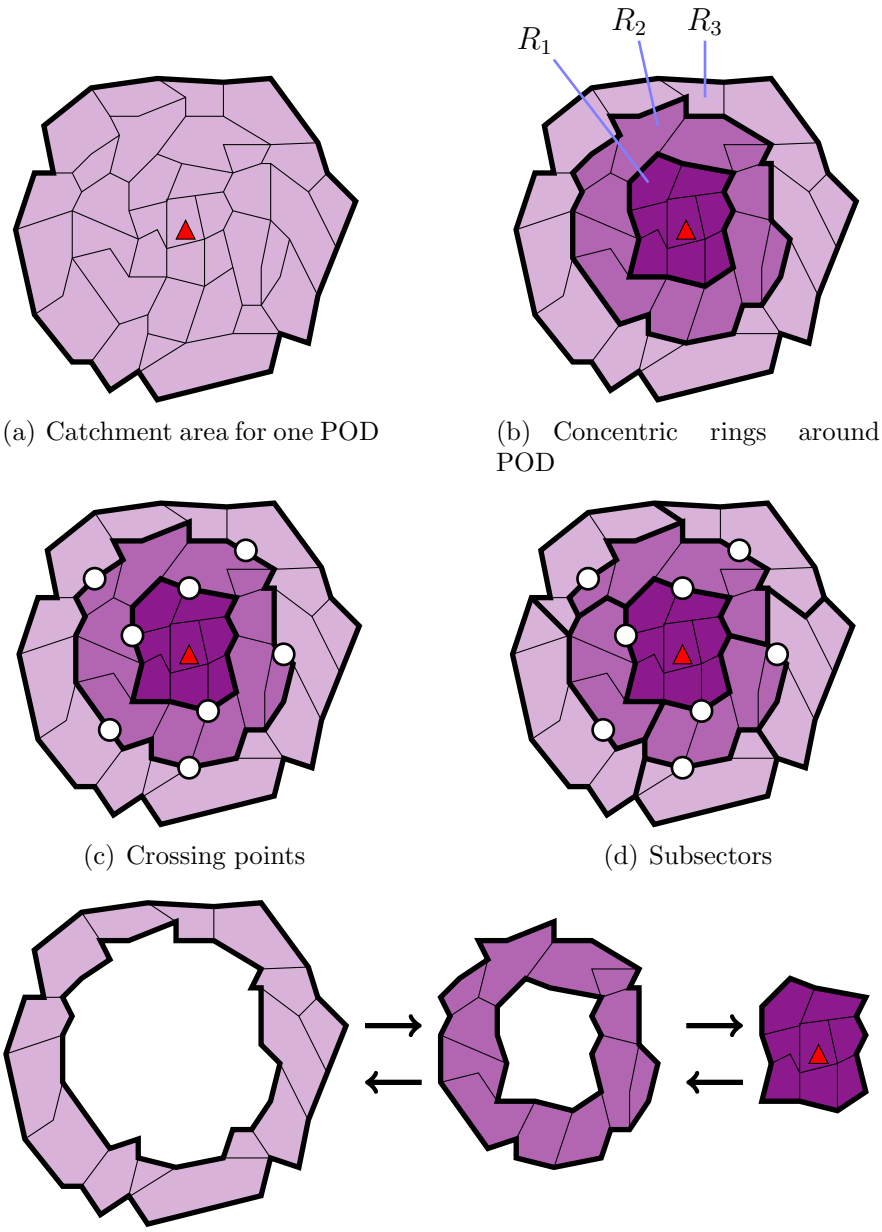
defined as follows:

$$R_i = \left\{ c \in C \mid (i-1)\,\frac{d_{max}}{k} <\| c - p \| \leq i\frac{d_{max}}{k} \right\}$$

Note that census block boundaries are always respected and therefore, the rings are conforming to these boundaries. An example of such a division into rings with $k = 3$ is illustrated in Figure 3.7. In Figure 3.7(a) the catchment area with its census blocks and POD is shown. The resulting subdivision after the ring assignment is depicted in Figure 3.7(b). The figure indicates that in order to travel to the POD from ring R_i with $i \in \{1..k\}$, all R_j with $j \in \{1..i\}$ have to be traversed. This is illustrated in Figure 3.7(e).

Once the rings are computed, road network information is superimposed. The road network data often does not contain minor roads and further, these types of roads are often contained within single census blocks and do not contribute to traffic between census blocks. Hence, minor roads often can not be considered, which does not represent a limitation to the model, as these roads mainly contribute to local neighborhood traffic. For each of the rings, the intersection points of the shared ring boundaries with the road network are computed (Figure 3.7(c)). These points may represent bottlenecks, as traffic must traverse such crossing points in order to reach ring $i-1$ from ring i. Hence, these points provide a measure of traffic influx into the next inner ring. The actual number of cars traversing these crossing points depends on the number of people living in the census blocks of the rings, and the base traffic of the surrounding areas. The latter is addressed in Section 3.3.2.1.

This approach creates small traffic rings for scenarios with a high density of PODs. For scenarios with few PODs, however, the rings become larger and therefore, the crossing points are located further away from the POD locations. The objective of the response plan analysis is to analyze the traffic towards and close to the PODs and therefore, crossing points that are not in the vicinity of the POD locations are of insufficient informative value. This problem has been addressed by utilizing a fixed ring width d_{width}, which describes the distance between two neighboring ring borders. The set of rings for a specific catchment



(a) Catchment area for one POD

(b) Concentric rings around POD

(c) Crossing points

(d) Subsectors

(e) Traffic flow graph

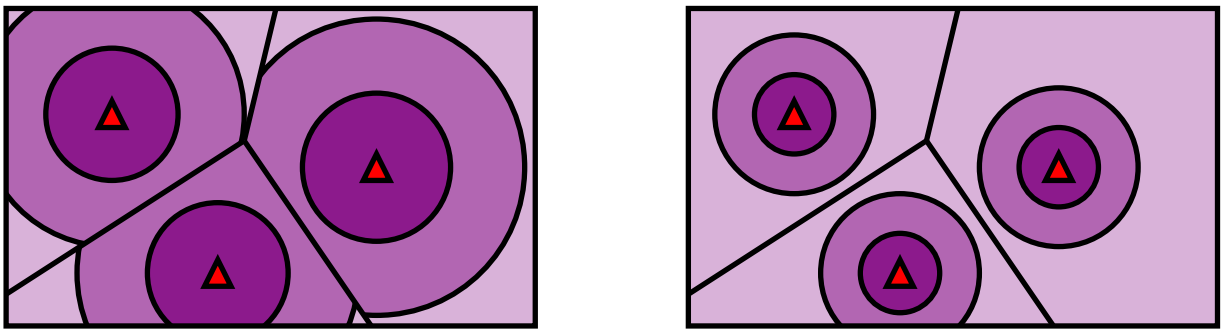
FIGURE 3.7. Subdividing catchment areas

area C is then computed as $\bigcup_{i \in \{1..k\}} R_i$, where

$$R_i = \{c \in C \mid (i-1)\, d_{width} <\| c - p \| \leq i d_{width}\}$$

Figure 3.8 compares both approaches for the same PODs with corresponding catchment areas. In Figure 3.8 (b) all sets of rings have the same sizes, whereas in Figure 3.8 (a) the

ring sizes are determined based on the furthest point in the corresponding catchment are. While the ring widths in Figure 3.8 (a) are the same across all catchment areas, the ring widths in Figure 3.8 (b) vary.



(a) Ring size determined by catchment area size

(b) Fixed ring size

FIGURE 3.8. Concentric rings around PODs with different properties

3.3.2. Ring Sectors

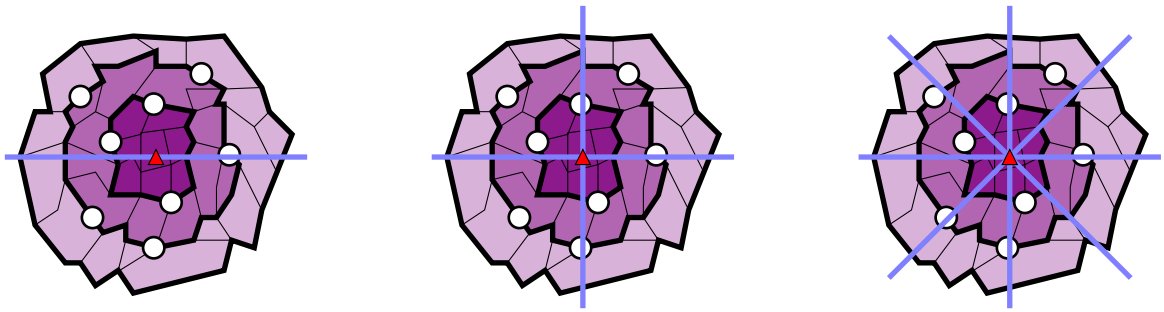
To determine the influx for each of the individual crossing points, the rings are further sub-divided. Analogous to the determination of the catchment areas, all census blocks of a ring a i re assigned to the closest crossing point of ring $i-1$ (Figure 3.7(d)). It should be noted that there is no subdivision for the inner ring R_1, as it contains the POD itself. Starting with the outermost ring R_k, the population of each sub-sector is added to obtain an estimate of the number of people crossing into the next ring R_{k-1}. Parameters, such as the average number of people per car or time of the day, can be specified by public health officials to compare different assumptions and evaluate the resulting scenarios. Once the traffic that is feeding into ring $i-1$ is estimated, the crossing point is assigned this number to act as a "super-centroid" or population source for the next iteration. The closer the rings are to the POD location, the higher the expected traffic feeding into the crossing points. This conforms to the situation one would expect in the case of a real emergency.

Input: List of crossing points L
Set of census block centroids $C_{p,r}$ of POD p and ring r
Output: Ring sectors with mapping of corresponding crossing points and census block centroids
foreach $c_i \in C_{p,r}$ **do**
  crossingPoint$(c_i) = l_{min} \in L : d(c_i, l_{min}) \leq d(c_i, l_j)\ \forall l_j \in L$
end

ALGORITHM 1: Naive assignment

Intuitively, individuals use the crossing points closest to their locations to reach the next inner ring. This idea is developed in the following, starting by directly translating this observation into an algorithm, which then is further refined to address the algorithm's shortcomings. Ultimately, a recursive algorithm is developed, which locally adjusts the granularity of the partition depending on the density of census block centroids and crossing points. The naive assignment method assigns each census block centroid to the closest crossing point l_{min} according to the underlying distance metric (see Algorithm 1 and Figure 3.7(d)). Nevertheless, this may lead to crossing points that have no census blocks assigned to them, if some of the crossing points are not closest to any of the centroids. It is likely to generate *empty* crossing points, if there are many crossing points within a small area. This would indicate that no cars would make use of those particular road segments to traverse into the next inner ring, which is an unlikely assumption. Consequently, this approach must be further refined to reflect more realistic scenarios.

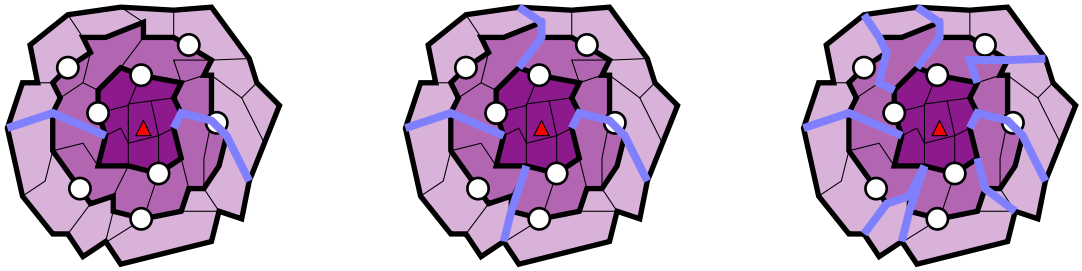
A first refinement of Algorithm 1 starts by drawing a horizontal line through the POD location, which partitions the set of census blocks into two disjoint sets. This is shown in Figure 3.9 (a). For illustration purposes, the crossing points of the individual rings are partitioned without distinguishing on which ring boundary they reside. In reality, crossing points of the respective rings are treated separately. As census blocks are represented by their centroids and are assigned to specific crossing points only based on their centroid location, the horizontal line drawn in Figure 3.9 (a) is implicitly adjusted along census block boundaries. Figure 3.9 (d) shows the horizontal line through the POD location after the adjustment. In the next step, each of the subsets is recursively partitioned by drawing a



(a) Drawing a horizontal line through the POD

(b) Drawing a vertical line through the POD

(c) Adding $y = mx$ and $y = -mx$



(d) Adjusting the horizontal line along census block boundaries

(e) Adjusting the vertical line along census block boundaries

(f) Adjusting $y = mx$ and $y = -mx$ along census block boundaries

FIGURE 3.9. Execution of Algorithm 2

vertical line through the POD, effectively partitioning the geographic region into four sectors as illustrated in Figure 3.9 (b). The implicit adjustment of the vertical line is illustrated in Figure 3.9 (e). The next step yields eight sectors by adding lines $y = mx$ and $y = -mx$ (Figures 3.9 (c) and 3.9 (f)). This approach can be repeated recursively until the desired granularity or a specific stopping criterion is reached. The partitioning method (Algorithm 2) takes a parameter n, which represents the granularity in terms of recursion levels as additional input. The number of sectors generated by this approach is 2^n. This method may lead to sectors, that do not contain any census block centroids or crossing points. Figure 3.9 (f) shows such a scenario. One of the sectors in the fourth quadrant (see Figure 3.10) does not contain a crossing point and hence, the population assigned to the sector would be omitted. Therefore, a recursion step is only completed if each sector contains both crossing points and census block centroids. This also solves the problem of a sector not containing any

centroids, which leads to a crossing point that is not being utilized. For example, Algorithm 2 does not execute the step shown in Figure 3.9 (c) and its execution results in the partition shown in Figure 3.9 (b). If the last step was executed, at least one sector without any crossing points would be created, which causes the algorithm to stop. Using the partitioning method, multiple crossing points may be assigned to specific sectors. The traffic within a sector is then assigned to the individual crossing points proportionally to road capacities at the crossing points.

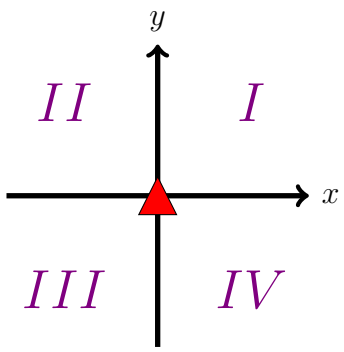


FIGURE 3.10. Numbering of quadrants

Once the initial lines $y = 0$ and $x = 0$ have been used to partition the space into four quadrants with the POD at the origin, each quadrant can be examined individually. In the following, this method is discussed specifically for the first quadrant, which implicitly partitions the third quadrant. Procedures for the second and fourth quadrants are analogous with negative slopes or slopes shifted by 90°. The first partitioning line in the first quadrant in a 45° angle α can be expressed as $y = mx$, or $y = tan(\alpha)x$, alternatively. During the second recursion, lines partitioning the first quadrant into four quarters in angles of 22.5° and 67.5° are added. Figure 3.11 shows the angles that need to be considered at each recursive level.

First horizontal lines are drawn through the PODs. This corresponds to drawing lines through the origin with slope 0°. In general, at recursion level r and for $i \in [0..2^r - 1]$ the following α have to be considered, whereby F is the set of angles generated at the r^{th}

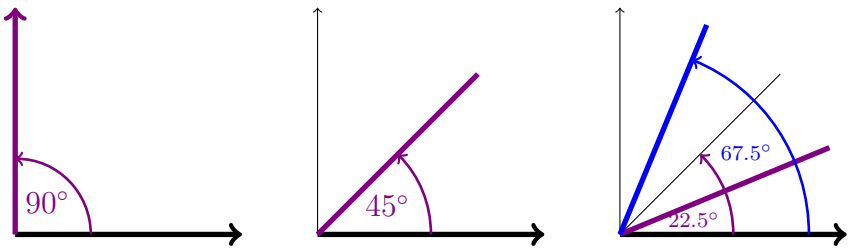


FIGURE 3.11. Angles at recursion steps

recursion level:

$$F_r = \cup \left(\frac{2i+1}{2^r} \times 90° \right)$$

Specifically, these angles can be expressed recursively as

$$\alpha_{i+1,1} = \frac{a_{i+1}}{b_{i+1}} \times 90° = \frac{2a_i - 1}{2b_i} \times 90°$$

and

$$\alpha_{i+1,2} = \frac{a_{i+1}}{b_{i-1}} \times 90° = \frac{2a_i + 1}{2b_i} \times 90°$$

with $a_0 = b_0 = 1$ and $i \in [0..n]$ for a maximum recursion depth n. The resulting straight lines of this approach are denoted as follows:

$$y_{i,1} = tan\left(\alpha_{i,1}\right) x$$

and

$$y_{i,2} = tan\left(\alpha_{i,2}\right) x$$

Figure 3.12 depicts a recursion tree for recursion levels $r = 0, 1, 2$. Each of the lines partitions its sector into two sub-sectors. Hence, in the next step two lines have to be computed to split each of the sub-sectors. This step is repeated recursively. All resulting angles can be listed by executing an in-order traversal on the recursion tree. Such a traversal on the example in Figure 3.12 yields 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, and 157.5°.

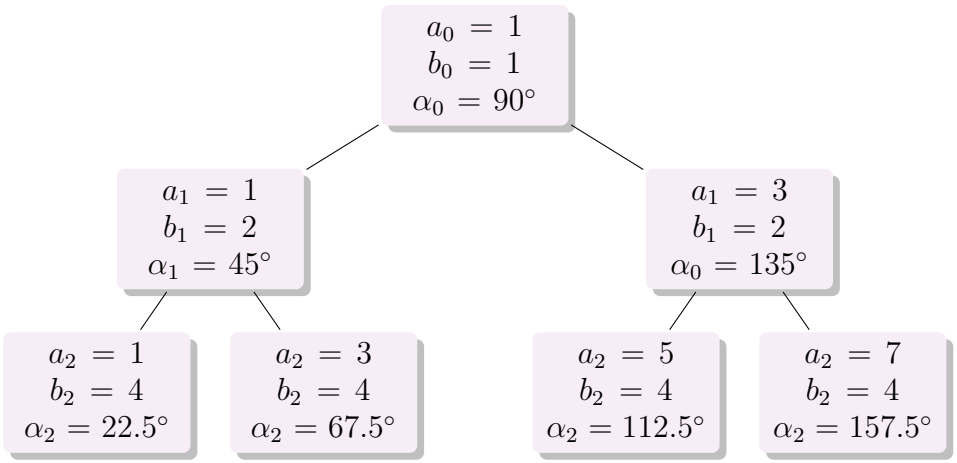


FIGURE 3.12. Recursion tree of angles

Each of these lines partitions the set of centroids C and the set of crossing points CPTS into two sets each: the points above the line and the points below the line. The following inequalities express this partitioning:

$$UpperBlocks_\alpha = \cup_{p \in C} : y(p) > tan\,(\alpha)\, x(p)$$

$$LowerBlocks_\alpha = \cup_{p \in C} : y(p) \leq tan\,(\alpha)\, x(p)$$

$$UpperCpts_\alpha = \cup_{p \in CPTS} : y(p) > tan\,(\alpha)\, x(p)$$

$$LowerCpts_\alpha = \cup_{p \in CPTS} : y(p) \leq tan\,(\alpha)\, x(p)$$

where x(p) returns the x-coordinate of a point and y(p) the corresponding y-coordinate of a point p. Figure 3.13 illustrates the partitioning of a sector into upper and lower sections. Centroids located in the sector $UPPER$ are assigned to the set $UpperBlocks$, whereas centroids located in the sector $LOWER$ are assigned to the set $LowerBlocks$. Crossing points are assigned analogously.

The framework for the partitioning method is summarized in Algorithm 2, whereby P represents the set of all PODs. Sets of all census block centroids and all crossing points are passed to the algorithm. Further, a maximum recursion depth n can be specified. Subsets of crossing points for a specific ring r of a POD p are denoted as $CPTS_{p,r}$. Subsets of census

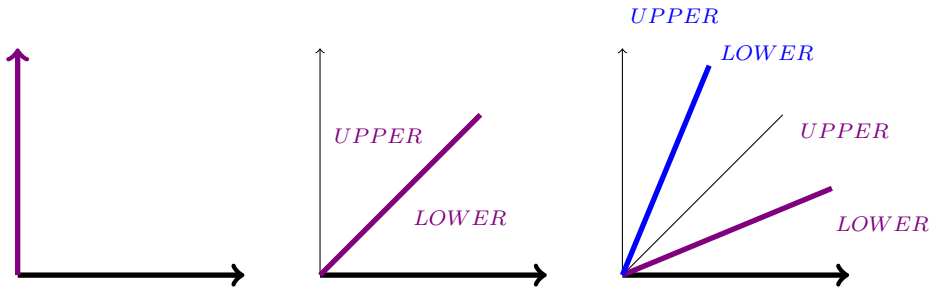


FIGURE 3.13. Partitioning of sectors into upper and lower sections

block centroids are analogously denoted by $C_{p,r}$. In Figure 3.14 for instance, $CPTS_{1,2}$ and $C_{1,2}$ refer to the crossing points and centroids of the second ring surrounding POD 1.

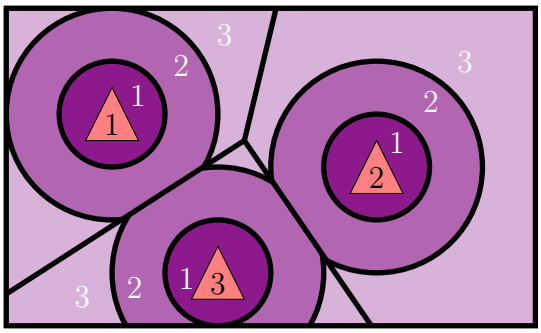


FIGURE 3.14. Numberings of rings

The algorithm iterates through all PODs and partitions the corresponding sets of census block centroids and crossing points into sectors. Sectors are created from the outer ring towards the innermost ring, as the crossing point from a ring become super-centroids for the next inner ring. The sectors for each of the ring may be different depending on centroid and crossing point density. The partitioning process starts by creating sub-sets of crossing points and centroids for each of the four quadrants, whereby the POD is assumed to be located at the origin. The sub-sets are stored in a list and passed to the recursive partitioning method described in Algorithm 3. Each list node contains corresponding centroid and crossing point sets, as well as additional processing information. These sets of each list element are then further partitioned and stored in a new list. If the maximum recursion level has not been reached and all of the new resulting subsets contain census block centroids and crossing points, the method is repeated on the elements of the new list. Otherwise, sectors are

Input: Sets of census block centroids C and crossing points $CPTS$,whereby $C_{p,r}$ and $CPTS_{p,r}$ denote the sub-sets corresponding to POD p and ring r
Maximum recursion depth n
Output: *partitionSets* writes results directly to the database
`/* Iterate through all PODs */`
for $p \leftarrow 1$ **to** $|P|$ **do**
  `/* Iterate through rings except innermost ring */`
  for $r \leftarrow (numRings)$ **to** 2 **do**
    `/* Initialize a and b */`
    $a_1 = 1$;
    $a_2 = 3$;
    $b = 1$;
    `/* Initialize recursion depth */`
    $d = 1$;
    Partition crossing point set into 4 sub-sets based on quadrant they lay in ;
    Partition centroid set into 4 sub-sets based on quadrant they lay in ;
    `/* Store all the sub-sets pairs with corresponding a and b in list nodes */`
    List l ;
    $l.add(\{C_{p,r,I}, CPTS_{p,r,I}, a_1, b\})$;
    $l.add(\{C_{p,r,II}, CPTS_{p,r,II}, a_2, b\})$;
    $l.add(\{C_{p,r,III}, CPTS_{p,r,III}, a_1, b\})$;
    $l.add(\{C_{p,r,IV}, CPTS_{p,r,IV}, a_2, b\})$;
    `/* Call method to recursively partition census blocks and crossing points up to recursion depth` n `*/`
    $partitionSets(l, d, n)$;
  end
  `/* Crossing points become super-centroids for next inner ring */`
  if $r > 2$ **then**
    $C_{p,r-1} = C_{p,r-1} \bigcup CPTS_{p,r}$;
  end
end

ALGORITHM 2: Framework for partitioning method

assigned to each of the nodes of the initial list and the information is written to the database. This terminates the recursive partitioning method. Algorithm 2 then proceeds by calling the partitioning method for the next ring.

The partitioning approach is the principle approach used in RE-PLAN. A stopping criterion is used to prevent the generation of sectors without crossing point or census blocks.

Input: List l with elements containing sets of block centroids, sets of crossing points, seed values of a, b
Recursion depth d
Maximum recursion depth n
Output: Final sector assignment is directly written to the database
`/* Update values for next recursion */`
$d++$;
List *newList* ;
; `/* List new sub-partitions */`
`/* Iterate through all list nodes and create sub-partitions */`
foreach *partition* $\in l$ **do**
  Calculate *UpperBlocks*, *LowerBlocks* ;
  Calculate *UpperCpts*, *LowerCpts* ;
  $b = 2 \times partition.b$;
  $a_1 = 2 \times partition.a - 1$;
  $a_2 = 2 \times partition.a + 1$;
  if *partition in Quadrant I or IV* **then**
    $newList.add(\{UpperBlocks, UpperCpts, a_2, b\})$;
    $newList.add(\{LowerBlocks, UpperCpts, a_1, b\})$;
  else
    $newList.add(\{UpperBlocks, UpperCpts, a_1, b\})$;
    $newList.add(\{LowerBlocks, UpperCpts, a_2, b\})$;
  end
end
if *none of the sub-sets is empty AND* $d \leq n$ **then**
  `/* Call the method recursively */`
  $partitionSets(newList, d, n)$;
else
  `/* Each node in list` l `represents a sector */`
  save sectors to database ;
end

ALGORITHM 3: Partitioning method *partitionSets*

This may yield large sectors and prevents the algorithm from refining the granularity, especially if the census block density greatly varies between different directions of the POD. To avoid this, a hybrid method can be used, which only recurses into sectors where further sub-divisions are permissible. Algorithm 4 shows the necessary modifications to Algorithm 2 to provide the framework for such a refinement. The basic structure of the algorithms are comparable. Algorithm 4 iterates through all PODs and rings in the same order as Algorithm

2. The crossing points become super-centroids for the next inner ring. However, after the sets have been partitioned into sub-sets corresponding to the quadrants, they are not stored in a list, but instead the recursive method for the hybrid partitioning approach (Algorithm 5) is called for each of the centroid - crossing point sub-set pairs independently. Therefore the methods *partitionSets* (Algorithm 3) and *partitionHybrid* (Algorithm 5) fundamentally differ in their operation. Algorithm 3 iterates during each recursion level through all sectors, which are all of the same size, whereas Algorithm 5 only takes into account a single sector. Hence, the latter supports mixed granularities for the resulting partition into sectors.

Input: Sets of census block centroids C and crossing points $CPTS$,whereby $C_{p,r}$ and $CPTS_{p,r}$ denote the sub-sets corresponding to POD p and ring r
Maximum recursion depth n
Output: Ring sectors with mapping of corresponding crossing points and census block centroids

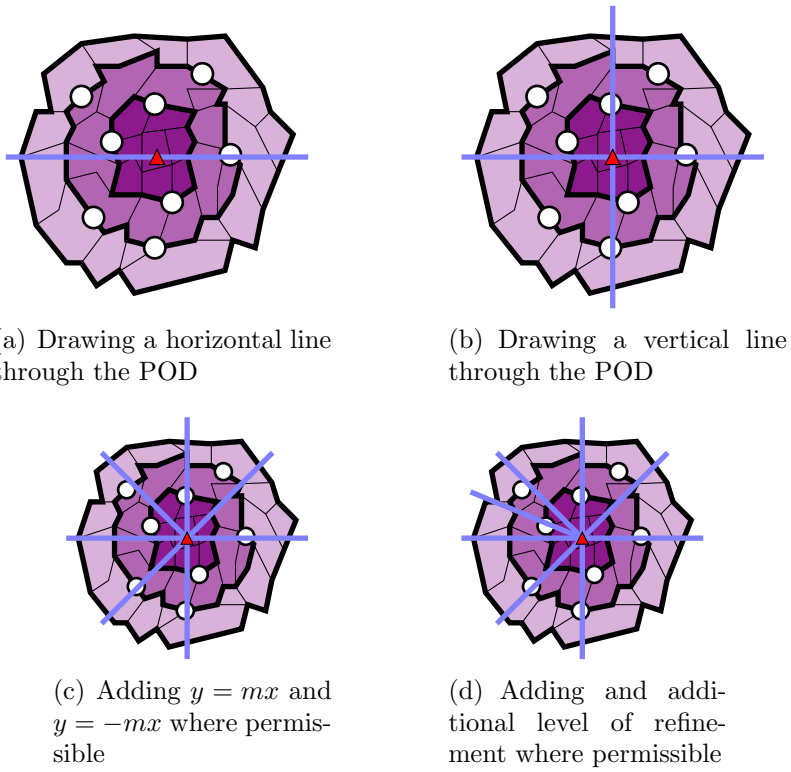
```
/* Iterate through all PODs                                          */
```
for $p \leftarrow 1$ **to** $|P|$ **do**
```
    /* Iterate through rings except outermost ring                   */
```
    for $r \leftarrow (numRings - 1)$ **to** 1 **do**
```
        /* Initialize a and b                                        */
```
        $a_1 = 1$;
        $a_2 = 3$;
        $b = 1$;
```
        /* Initialize recursion depth                                */
```
        $d = 1$;
        Partition sets into 4 sub-sets based on the quadrant they lay in ;
```
        /* For each quadrant (I,II,III,IV) call the method that
           recursively partitions census blocks and crossingpoints   */
```
        $partitionHybrid(C_{p,r,I}, CPTS_{p,r,I}, a_1, b, d, n)$;
        $partitionHybrid(C_{p,r,II}, CPTS_{p,r,II}, a_2, b, d, n)$;
        $partitionHybrid(C_{p,r,III}, CPTS_{p,r,III}, a_1, b, d, n)$;
        $partitionHybrid(C_{p,r,IV}, CPTS_{p,r,IV}, a_2, b, d, n)$;
    end
```
    /* Crossing points become super-centroids for next inner ring    */
```
    if $r > 2$ **then**
        $C_{p,r-1} = C_{p,r-1} \bigcup CPTS_{p,r}$;
    end
end

ALGORITHM 4: Framework for hybrid approach

The method *partitionHybrid* described in Algorithm 5 partitions the sets it receives into two new sets. Only if each set contains elements and the maximum recursion depth has not been reached, the sub-sets are further partitioned recursively. Otherwise, the new partitioning cannot be finalized, and the sector assignment is stored in the database. An example of executing Algorithm 4 along with Algorithm 5 is illustrated in Figure 3.15. The resulting sector assignment shows different sector sizes. The sector assignment across all rings is only used for the purpose of illustration. In reality, sectors are assigned to the individual rings independently.

Input: Set of crossing points and set of census block centroids
seed values of a, b
Recursion depth d
Maximum recursion depth n
Output: Final sector assignment is directly written to the database
Calculate $UpperBlocks$, $LowerBlocks$;
Calculate $UpperCpts$, $LowerCpts$;
`/* Update values for next recursion */`
$d++$;
$a_1 = 2a - 1$;
$a_2 = 2a + 1$;
$b = 2b$;
`/* Only split if both sub-sets contain census blocks and crossing points */`
if *none of the sub-sets is empty AND* $d \leq n$ **then**
  `/* Call the method recursively */`
  if *partition in Quadrant I or IV* **then**
    $partitionHybrid(UpperBlocks, UpperCpts, a_2, b, d, n)$;
    $partitionHybrid(LowerBlocks, LowerCpts, a_1, b, d, n)$;
  else
    $partitionHybrid(UpperBlocks, UpperCpts, a_1, b, d, n)$;
    $partitionHybrid(LowerBlocks, LowerCpts, a_2, b, d, n)$;
  end
else
  `/* Write the sector to the database as it cannot be further partitioned */`
  save sector to database ;
end

ALGORITHM 5: Partitioning method *partitionHybrid*



(a) Drawing a horizontal line through the POD

(b) Drawing a vertical line through the POD

(c) Adding $y = mx$ and $y = -mx$ where permissible

(d) Adding and additional level of refinement where permissible

FIGURE 3.15. Execution of Algorithm 4

The above discussed approaches may lead to sectors with multiple crossing points. The individuals traversing ring borders via specific points are then distributed based on ratios of road capacities of the crossing points within the same sector.

3.3.2.1. Traffic Flow

Once the expected distribution of the population is calculated, the base traffic can be estimated and road network information can be integrated into the model. Base traffic is traffic observed on a daily basis and is not caused by the bio-emergency. The available information includes traffic counts, the number of lanes, and speed limits on a road segment granularity. Traffic counts describe the number of cars crossing a specific road segment at a particular time within a specified time interval. These counts, however, are only available for a minor amount of the road segments of the data set. Hence, traffic information must be interpolated for those segments, for which no traffic counts are available. As mentioned

above, some minor roads are not included in the infrastructure description, and can consequently not be considered as part of the road network. Traffic counts for the major roads in the North Texas region are available in 15 minute intervals. The methodology to estimate traffic counts is sufficiently generic to be applicable to all roads in the road network. Additionally, the North Central Texas Council of Governments (NCTCOG) provided traffic data that contained estimated road capacities for North Texas.

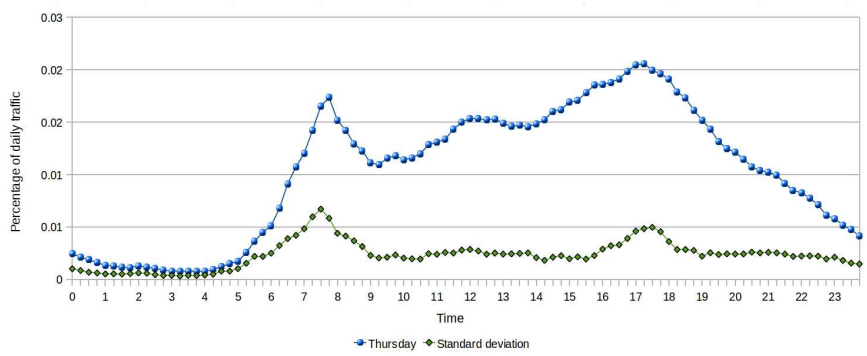
If no road capacities are available, the following approach can be used to estimate these. By assuming the average length of a car l_c to be 18 feet, the space between two cars l_s to be 1 car length per 10 mph, and given the speed limit v_{max} and the number of lanes n, the maximum capacities C_{max} for a road segment can be calculated as

$$C_{max} = \frac{v_{max}}{l_s + l_c} = \frac{v_{max}}{\left(\frac{v_{max}}{10} + 1\right) \times \frac{18}{5280} miles}$$

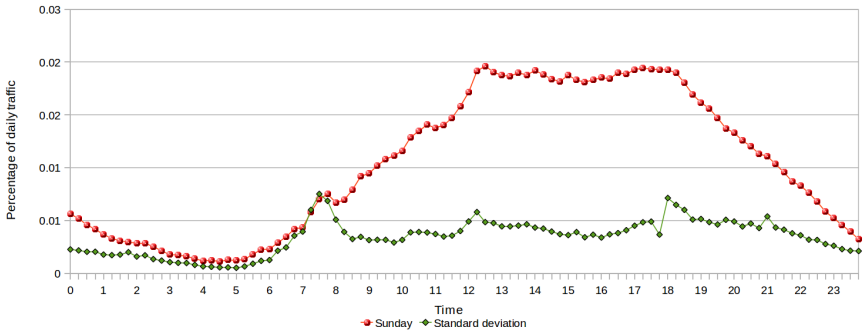
C_{max} is a measure of how many cars can at most traverse a road segment in one hour. Based on the actual traffic counts the proportion of the maximum traffic capacities is calculated. This can be done for different times of the day and allows for the estimation of the roadway traffic for any road segment with the given parameters.

Figure 3.16 shows the traffic count distributions over a day comparing proportions of total daily counts for mid-week and weekend traffic. Specific traffic counts have been obtained from the NCTCOG. Traffic counts of different road segments across different dates have been found to be sufficiently similar to use the results as indicative traffic parameters. Comparing the graphs for the different days of the week, two patterns have been observed that lead to the formulation of two distinct traffic distribution classes, namely for weekdays and weekend days. The two classes are shown in Figure 3.17. For both scenarios, small variations can be observed. The graphs represent the actual proportion $P_c(t)$ of traffic counts averaged over all road segments with available traffic counts for a given time interval, which are calculated as

$$P(t) = \frac{1}{\mid S \mid} \sum_{s \in S} \frac{C^s_{actual}}{C^s_{max}}$$



(a) Traffic counts for Thursdays



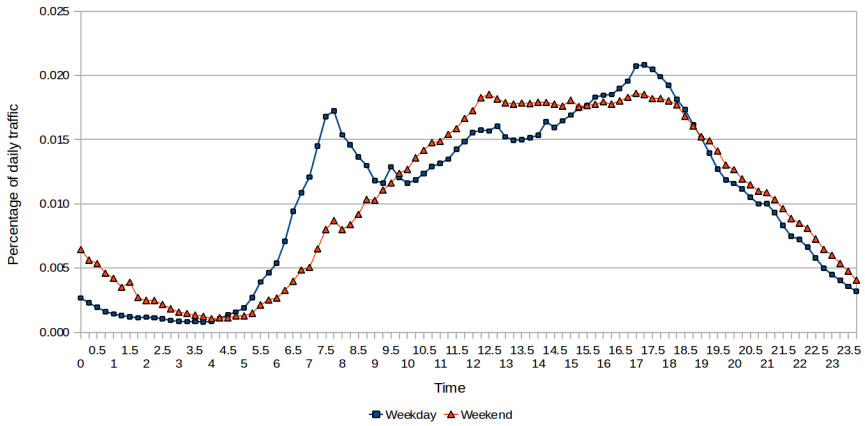
(b) Traffic counts for Sundays

FIGURE 3.16. Sampled traffic distribution (with courtesy of WIT Press [88])

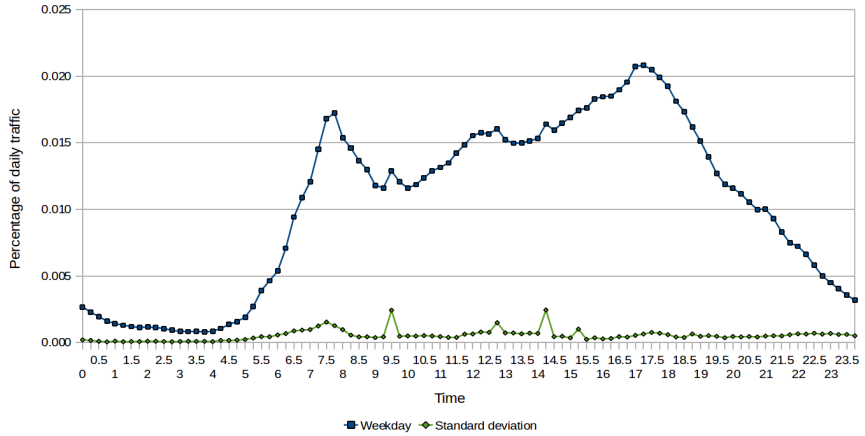
where S is the set of all road segments with available traffic counts.

3.3.3. Traffic Conditions at the POD Locations

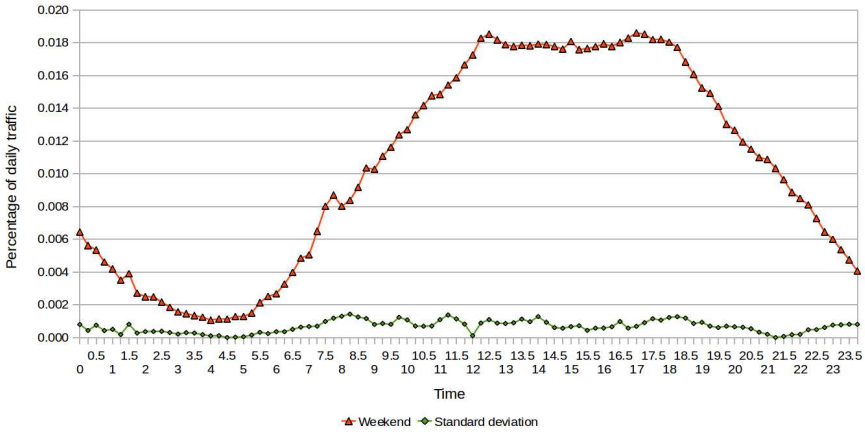
Hitherto, methodologies to analyze emerging traffic within the transportation infrastructure have been presented. These were based on the computation of crossing points that facilitate the estimation of bottlenecks and areas of congestion in the road network. Sufficient road capacities, however, do not guarantee that mandated time-lines for the execution of a specific response can be met, as the POD locations themselves may constitute a bottleneck. Thus, disaster preparedness coordinators are required to carefully balance the number of PODs and human resources allocated to the individual POD locations. An excessive number of booths or lanes at a specific POD location may result in poor utilization, as cars cannot enter and leave the POD location at a rate fast enough to utilize each of the resources



(a) Weekdays and weekend



(b) Weekdays



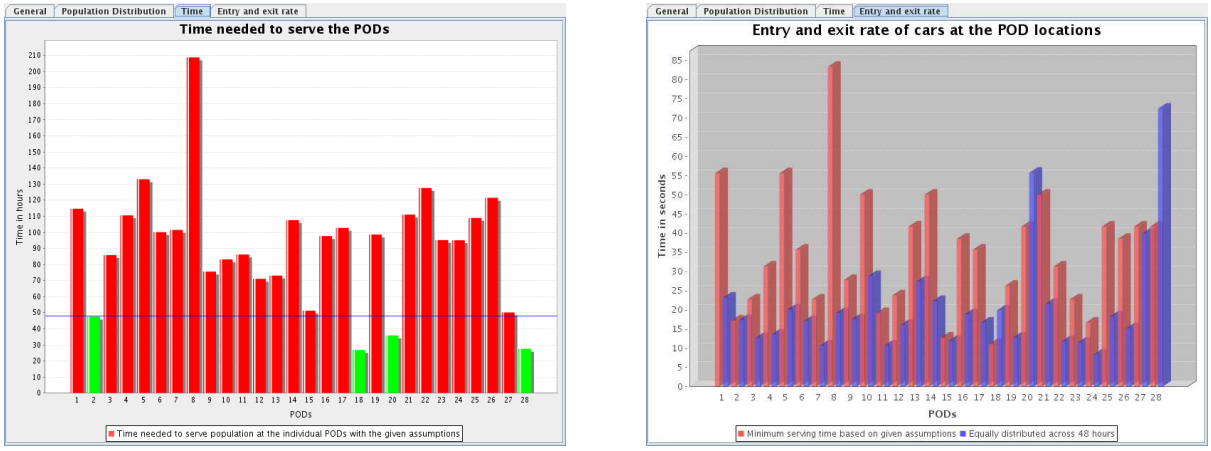
(c) Weekend

FIGURE 3.17. Traffic classes (with courtesy of WIT Press [88])

continuously. Consequently, POD resources may have to be redistributed by introducing additional PODs, thereby dividing the catchment area into multiple smaller ones. Conversely, an insufficient number of booths at a POD location may impede timely plan execution, as despite constant operation, the demand would exceed the available resources. Hence, not only the number of POD locations, but also the number of booths at the locations themselves play a defining role in the feasibility of a response scenario. This poses a trade-off between logistics and staffing on one side, and traffic feasibility at the POD locations on the other side. RE-PLAN's PODAnalyzer aids public health officials to determine the number of booths at each of the POD locations and to identify PODs which in spite of sufficient resources will not be able to serve the demand due to traffic congestion at the POD. Parameters, such as the number of people per car and service time per car, allow for the comparison of the effects and outcomes of different underlying assumptions. The PODAnalyzer gives estimates of the access rate at the individual POD locations and visualizes this information. Further, estimates whether or not a POD location can support the population size of its catchment area within the mandated-time frame are provided. However, the detailed internal design of POD layouts is beyond the scope of this research, but has been addressed by the computational tool RealOpt [62]. RE-PLAN is currently being deployed and evaluated by a Tarrant County's public health department. An example of and evaluation by the PODAnalyzer is depicted in Figure 3.18, showing estimated time requirements for each POD to serve its catchment area, and the estimated and optimal entry and exit rates for cars into and out of the POD locations.

3.4. Results

In the following, RE-PLAN's execution is demonstrated on Denton County. The traffic information has been obtained from the NCTCOG. Demographic data, i.e. number of people for each census block, have been downloaded from the Census 2000 Fact Finder. Specifically, the plan analysis is exemplified on scenarios with three and fifteen PODs, respectively.



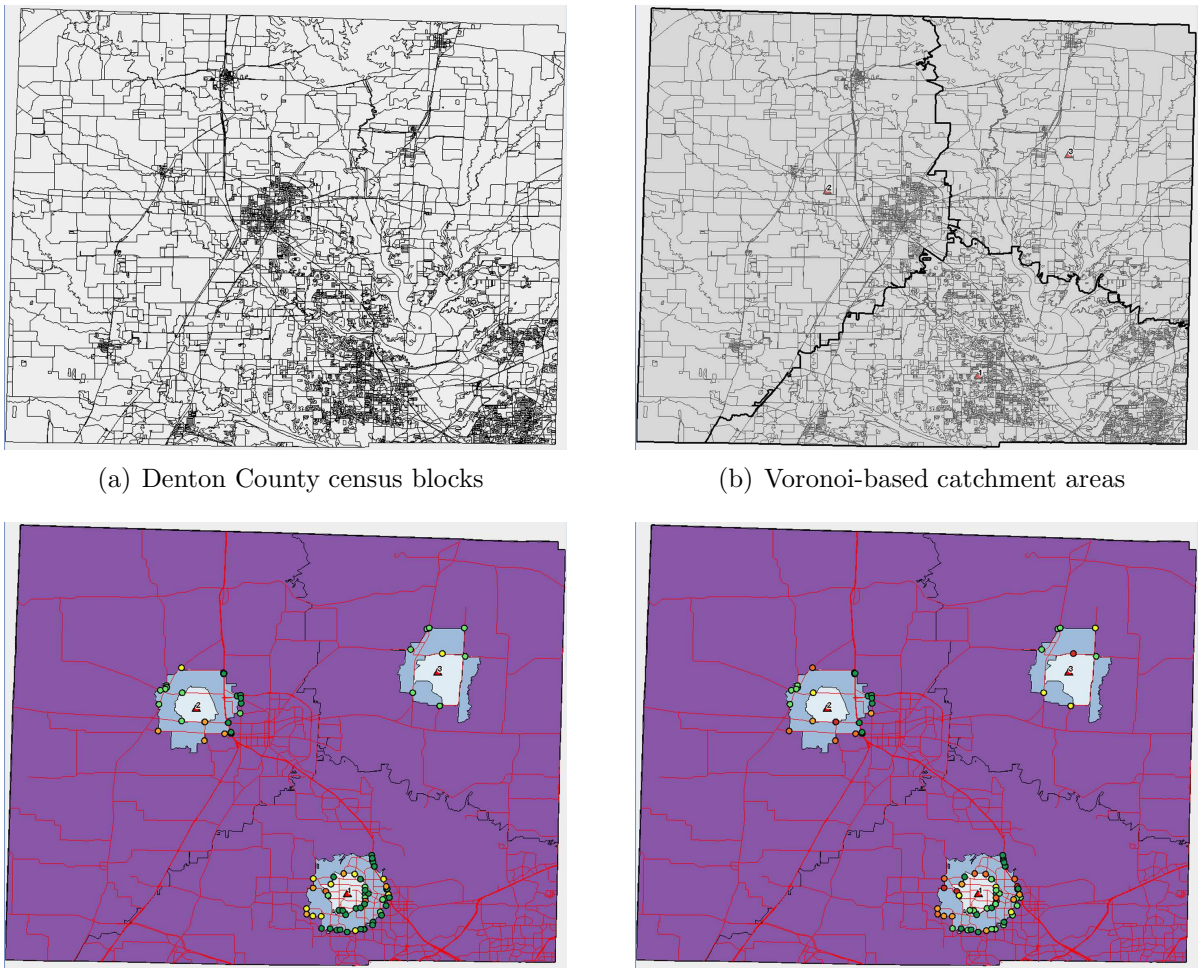
(a) Estimated serving time

(b) Entry and exit rates

FIGURE 3.18. PODAnalyzer (with courtesy of WIT Press [88])

Figure 3.19 shows an example scenario with three PODs placed in Denton County. Starting with the Denton County census blocks (Figure 3.19 (a)) three PODs have been placed within county boundaries. RE-PLAN's execution starts by calculating corresponding catchment areas, which show strong resemblance to Voronoi tessellations (Figure 3.19 (b)) and rings of proximity are determined. Then, crossing points can be computed by overlaying the road network. Depending on the given parameters, such as day of the week and mandated time-frame, traffic at these crossing points can be assessed. Crossing points are colored based on the traffic intensity expected. The color range starts with green (sufficient road capacity) over yellow and orange tones (traffic reaching critical point) to red (bottleneck). Figure 3.19 (c) illustrates such a scenario for a weekday at 8:30am, while an average of two people per car and a 48-hour mandated time-frame are assumed. In Figure 3.19 (d) the parameters are altered to a weekday at 8:00am with only one person per car and a 24-hour mandated time-frame. This change in assumptions yields higher traffic at many of the crossing points and therefore, additional bottlenecks are created.

Figure 3.20 presents the results of a response plan analysis for a fifteen-POD scenario in Denton County. Based on the selected PODs, catchment areas are computed (Figure 3.20 (a)). In Figure 3.20 (b) the generation of crossing points as a result of intersecting the



(a) Denton County census blocks

(b) Voronoi-based catchment areas

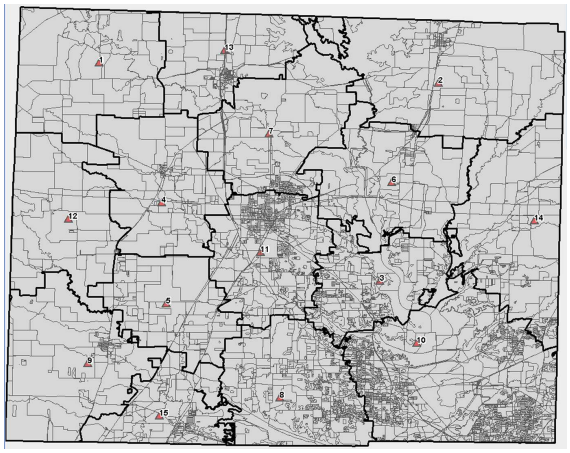
(c) Analysis results (weekday at 8:30am, 2 people per car, 48-hour time-frame)

(d) Analysis results (weekday at 8:00am, 1 person per car, 24-hour time-frame)

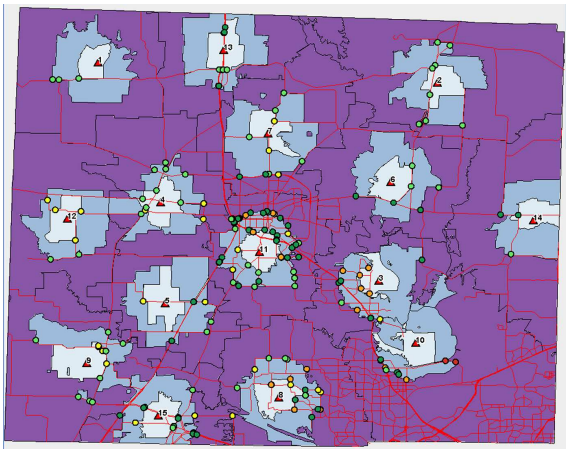
FIGURE 3.19. Three-POD response scenario for Denton County

rings of proximity with the road infrastructure is illustrated. Figures 3.20 (c) and 3.20 (d) compare different assumptions for the response scenario.

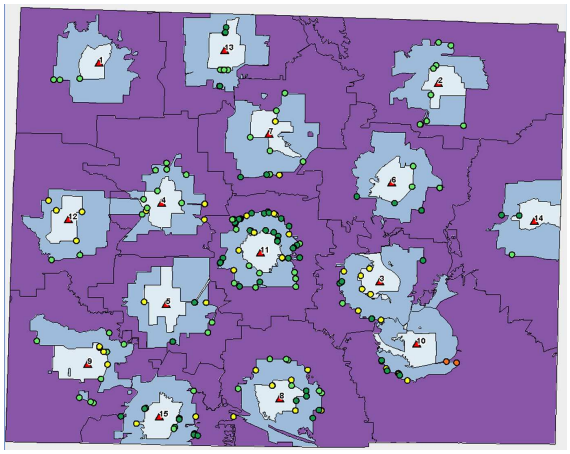
Even if the road network is capable of supporting the anticipated traffic, the PODs themselves might not be able to sustain the emerging demand. Factors contributing to the feasibility at the POD locations are the number of booths operated at a POD, the estimated serving time per car, and the estimated number of cars. The latter is provided by methodology discussed in this dissertation, while the other factors are determined by public health experts and planning committees.



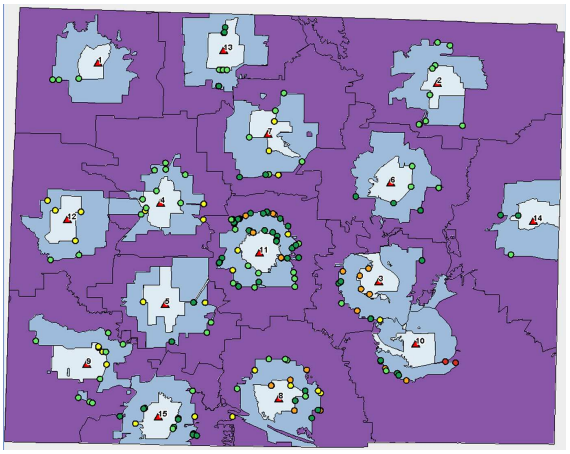
(a) Voronoi-based catchment areas



(b) Intersections between traffic rings and the road network are represented by crossing points



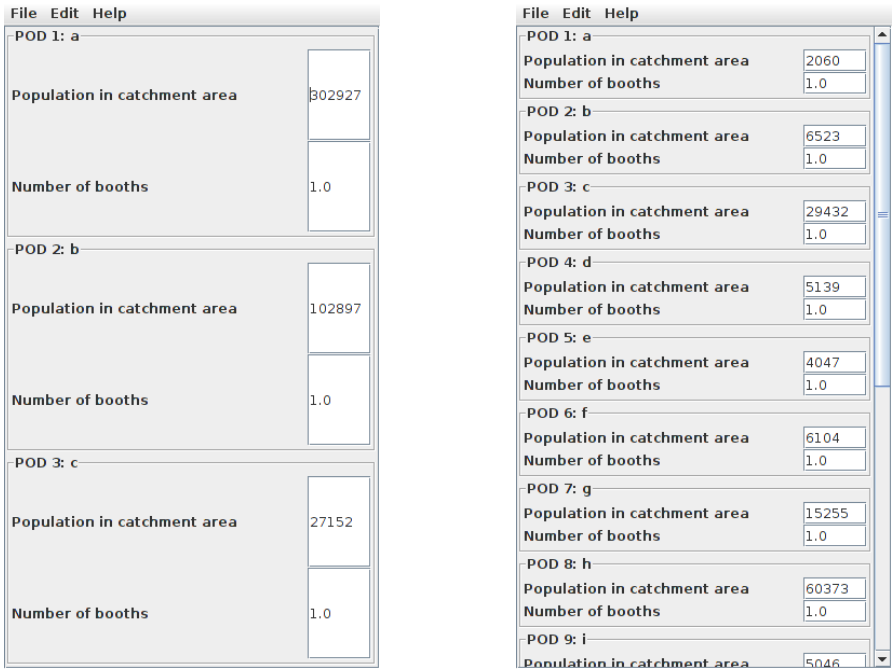
(c) Analysis results(weekday at 12:15pm, 4 people per car, 48-hour time-frame)



(d) Analysis results(weekday at 8:30am, 2 people per car, 24-hour time-frame)

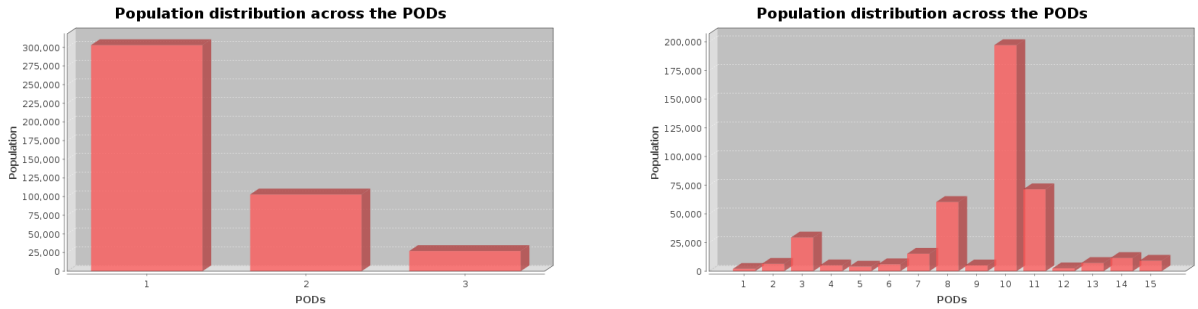
FIGURE 3.20. Fifteen-POD response scenario for Denton County

RE-PLAN's PODAnalyzer allows public health officials to specify parameters, such as people per car, estimated serving time per car, number of booths per POD location, and determines whether the population of a catchment area can be served within a mandated time interval. The length of this interval varies with the type of the underlying emergency, which will determine whether individuals are vaccinated or merely provided medication. The interface of the tool for the above shown scenarios is depicted in Figure 3.21, whereby equally capacitated PODs with a single booth per POD have been assumed.



(a) Estimated serving time (b) Entry and exit rates

FIGURE 3.21. Interface of RE-PLAN's PODAnalyzer

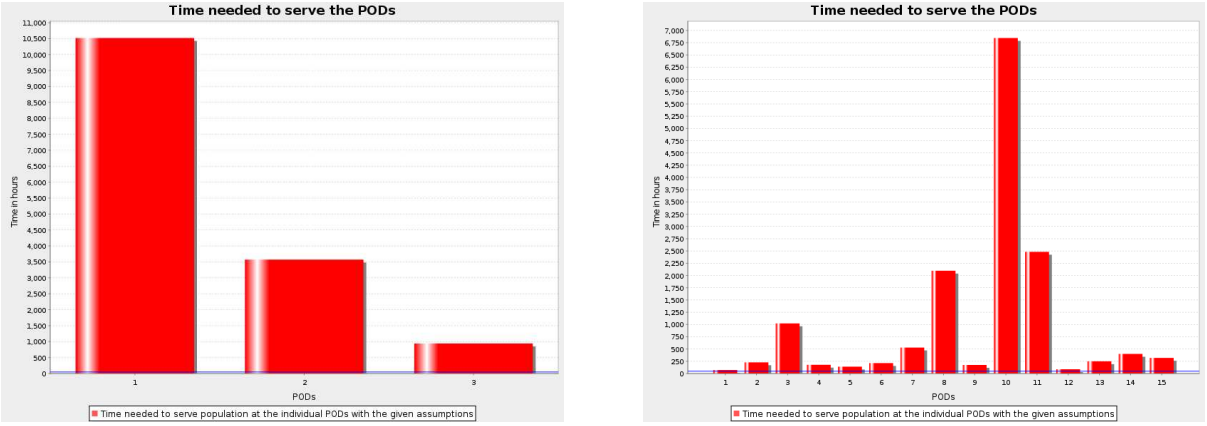


(a) 3 PODs (b) 15 PODs

FIGURE 3.22. PODAnalyzer: Population distribution

The evaluation of the conditions at the POD location provided by the PODAnalyzer includes the visualization of the population distribution (Figure 3.22) and the estimated serving times for the catchment areas (Figure 3.23). Figure 3.22 indicates that for both scenarios the population is unevenly distributed among the catchment areas. PODs with high populations in their catchment areas potentially cause bottlenecks in the road infrastructure. This, however, strongly depends on the available infrastructure and road capacities around

the POD locations. Further, catchment areas with high population counts need more resources (expressed in number of booths) than catchment areas with fewer individuals. As an equal number of booths have been assumed for all PODs, the population distribution directly correlates to the estimated time to serve the population of a POD's catchment area (Figure 3.23). This allows public health experts to determine the resources that are necessary, such that the population assigned to a POD can be served within mandated time-frames.



(a) 3 PODs

(b) 15 PODs

FIGURE 3.23. PODAnalyzer: Performance with respect to mandated time-frames

CHAPTER 4

RESPONSE PLAN OPTIMIZATION

Analyzing an existing emergency response plan might show that the plan is not feasible within the given time-frame. It is noteworthy, that without relaxing some of the constraints, an optimal placement that ensures that the entire population of a region is served within a mandated time-frame, may not exist. The following sections discuss approaches that compute approximations of feasible placements of points of dispensing (PODs).

4.1. Models in Location Science

Location allocation science can be sub-categorized into different fields, including discrete location science and continuous locations science. In what follows, problems found in discrete location science (Sections 4.1.1 and 4.1.2) and continuous location science (Section 4.1.3) are introduced.

4.1.1. The p-Median Approach

The first problem addresses the selection of p POD locations from a given list of candidates, whereby p is a fixed number chosen by public health officials. Given these constraints, the locations are to be chosen, such that geographic region is covered optimally (as close to the mandated time-frame as possible). This approach yields a solution that covers the entire geographic space with the given resources. It does not, however, guarantee that mandated guidelines are met. Nevertheless, it is of importance for public health experts, as it shows the best possible coverage one can achieve with currently available resources. The problem described here can be mapped onto the p-median problem in discrete locations science. Algorithms solving this problem are described in Section 4.3.

The p-median problem [19], [5], [2],[39] takes as input a set I for the demand nodes, a set J for the candidate facilities, and distances d_{ij} with $i \in I$ and $j \in J$ between all demand

nodes and candidate facilities. A common problem to be solved is to choose p facilities out of the candidate facility set J in order to cover the demand arising in all candidate nodes in the set I. Thereby, the overall cost is to be minimized, which is usually denoted as the sum over the distances of the demand nodes to the next facility. For the design of emergency response plans, the facilities in set J correspond to candidate POD locations selected by public health officials. Discrete points of demand can be generated by collapsing small geographic entities into single points, such as assigning the population of census block to their corresponding geographic centroids. Now p PODs can be selected out of the set J serving the population as defined by the centroids in set I. However, an optimal assignment does not imply that the overall solution will meet mandated guidelines. Nevertheless, from this assignment a lower bound in time achievable with the given resources can be derived. As no better solution with the available resources can be derived, the time the response plan pertinent to the derived POD placement requires for the execution is shorter than for any other sub-optimal solution.

4.1.2. The Center and Covering Approach

The goal of emergency response planning is to best allocate resources given a set of constraints as well as mandated guidelines. The latter can be interpreted as a service standard. The approach described above provides coverage to a region, even if the resources do not suffice, thereby relaxing service standards. It is, however, of particular interest to public health officials to device plans that do comply with service standards. This approach reveals shortcomings in the plan, potentially not assigning a POD location to some of the geographic entities. Further, it shows the best coverage that can be achieved with the given resources while not guaranteeing complete coverage. This particular problem is not addressed by the research presented in this dissertation, but in future research, models discussed in Section 4.3 can be extended to address center and covering versions of response scenarios.

Center and covering methods have been discussed in the literature using a variety of different approaches [79],[65], [58], [3] defining a range of different service standards. Such a service standard allows for the inclusion of a time-frame into the computation. Median

and plant location problems find an assignment of demand points to facilities regardless of such service standards. Every demand point is covered, but it is not guaranteed that all demand points meet a service standard. Centering and covering methods, however, ensure that demand points are only assigned to facilities if the given service standard is met. Using this approach, all demand points (individuals at the census block level) assigned to a facility (POD) will meet the time constraint, but not all demand points are assured to be assigned to a facility. This implies that the unassigned demand points cannot receive emergency treatment with the given facilities and service standards. While these methods provide demand points that are critical for the overall success of a response plan and aid public health officials in improving the in-place plans, augmenting such a plan by making use of median and plant location methodologies provides an assignment of the remaining demand points to facilities. The latter assignment does not take into account the service standard, but provides an interim assignment until the response plan has been improved, such that the service standard is met for the entire region.

4.1.3. Continuous Models

While choosing POD locations from a list of preselected locations guarantees the placement of PODs strictly at feasible locations, it is essential to address the placement of PODs without constraints in terms of preselected locations. This allows for the placement of PODs at any point in the geographic space. The resulting assignment of PODs gives an indication of the amount of necessary resources, as well as areas that should be favored for the POD selection. The algorithm presented in Section 4.2 falls into the category of such continuous models.

Continuous models [86],[46],[16],[100] in location allocation modeling do not limit facility placement to a set of candidate locations, but consider the entire geographic space. By leaving the maximum number of facilities unconstrained, a solution meeting a service standard is guaranteed to be found. Nevertheless, this solution is likely to include locations that result in practically infeasible locations, such as lakes, mountains, or private and corporate

properties, whose owners are not willing to grant usage of their property for the response to bio-emergency events. Further, such a solution can result in resources, such as the number of facilities, that cannot be covered with the given budget.

4.2. Applying Continuous Location Science to POD Placement

Input: A set of census blocks B forming a continuous geographic region R with polygonal boundary PB and with population $pop(b)$ and centroid $centroid(b)$ for each $b \in B$; number of PODs k

Output: Partition of the region R into k sub-regions such that the population sizes between any two sub-regions differ by at most b_{max}

Select $p_1, p_2 \in PB : \forall p_i, p_j \in PB, dist(p_1, p_2) \geq dist(p_i, p_j)$ Create list
$L_1 : dist(p_1, b_1) \leq dist(p_1, b_2) \leq \ldots \leq dist(p_1, b_{|B|})$, $b_i \in B, i \in [0..|B|]$ Create list
$L_2 : dist(p_2, b_1) \leq dist(p_2, b_2) \leq \ldots \leq dist(p_2, b_{|B|})$, $b_i \in B, i \in [0..|B|]$ Initialize
$R_1 = R_2 = \emptyset$and$pop(R_1) = pop(R_2) = 0$
Create coefficients $c_1 = \lfloor \frac{k}{2} \rfloor$,$c_2 = k - c$
while $size(L_1) > 0$ **do**
  if $c_2 \times pop(R_1) < c_1 \times pop(R_2)$ **then**
    $b = L_1(0)$
    $R_1.add(b)$
    $L_1.remove(b)$
    $L_2.remove(b)$
  end
  else
    $b = L_2(0)$
    $R_2.add(b)$
    $L_1.remove(b)$
    $L_2.remove(b)$
  end
end
Create Polygons P_1, P_2 containing census blocks in L_1, L_2
if $c_1 > 1$ **then**
  Recursively repeat using $R = P1$, $k = c_1$, and $B = L_1$
end
if $c_2 > 1$ **then**
  Recursively repeat using $R = P2$, $k = c_2$, and $B = L_2$
end

ALGORITHM 6: Placement of k PODs while optimizing population distribution

Continuous Locations Science is one of the sub-areas of Location Allocation Modeling. Continuous models assume demand points at discrete locations and allow for the placement of the service facilities anywhere in the service area. For the placement of PODs, the service

area is a geographic region R, which is partitioned into a set of census blocks B. Each census block $b \in B$ is represented by its geographic centroid $centroid(b)$ and has a population size of $pop(b)$. The population of the entire geographic region R can then be expressed as $pop(R) = \sum_{b_i \in B} pop(b_i)$. Let the distance of a census block b_i to some point p be defined as the distance between p and the centroid corresponding to b_i, i.e. $dist(p, b_i) = dist(p, centroid(b_i))$. For some instance of the problem assume k PODs (service facilities) that are to be placed within the region R, such that the population is equally distributed among all PODs. Algorithm 6 was developed to recursively partition R into k sub-regions (catchment areas) CA, such that $\bigcup_{c \in C} = R$. First, the polygonal boundary PB of R is computed, which is represented as a discrete set of points. Points p_1, $p_2 \in PB$ are chosen, such that $\forall p_i, p_j \in PB : dist(p_1, p_2) \geq dist(p_i, p_j)$. For each of the points p_1 and p_2 corresponding lists L_1 and L_2 are created. Both lists are identical at the beginning and contain all census blocks of the region R represented by their centroids. Using a sorting algorithm, such as Quicksort [89], the census blocks of each list are sorted from shortest distance between centroid and p_1 or p_2 to longest distance. Next, R is partitioned into two regions R_1 and R_2 with $R_1 \cup R_2 = R$ and $R_1 \cap R_2 = \emptyset$. The ratio of the population of the two regions is $c_1 : c_2$ with $c_1 = \lfloor \frac{k}{2} \rfloor$ and $c_2 = k - c_1$. This corresponds to the number of PODs each sub-region is assigned at the current recursion level of the algorithm. Each recursive step terminates after all census blocks $b \in B$ have been assigned to either R_1 or R_2. The census blocks $b \in B$ are sequentially assigned to R_1 or R_2, such that a census block is added to R_i with $pop(R_i) = min\left(pop\left(R_1\right), pop\left(R_2\right)\right)$. This is done by assigning the first census block of the region's R_i corresponding sorted list L_i. Census block b_i is then removed from both lists L_1,L_2. The resulting regions each define new Polygons P_1 and P_2. c_1 and c_2 PODs are then placed into regions R_1 and R_2, respectively. If $c_i \neq 1$ the algorithm is recursively executed for R_i.

Selecting points p_1 and p_2 on the boundary of the geographic region, such that the distance between them is maximized prevents the generation of isolated areas during the execution of the algorithm. If the two points are located closely to each other, it is likely

that islands are created. This is due to the fact that at each step the closest available centroid is assigned to one of the points. The closer the points, the more likely it becomes that they are competing for the same centroid at each of the steps. Although it might seem that one of the lists L_1, L_2 is redundant and they seem to be ordered in reverse order, Lemma 4.1 shows that this is not the case.

LEMMA 4.1. *Lists L_1, L_2 defined by Algorithm 6 are not ordered in reverse order for the general case and therefore, both lists are necessary.*

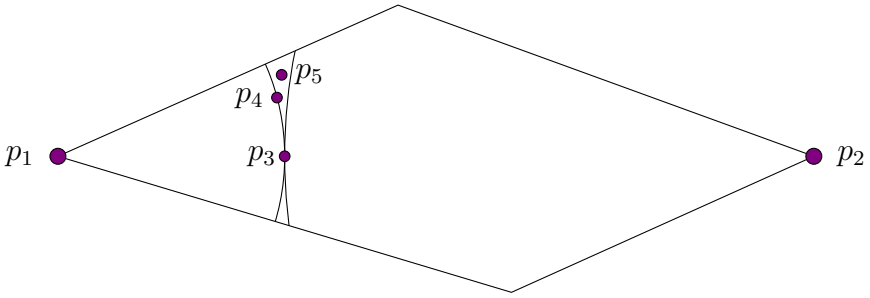


FIGURE 4.1. Proof by counter-example

PROOF. Proof by counter example: Figure 4.1 shows an example, for which L_1 is not the same as L_2 read backwards. Assume p_3 and p_4 are points on a circle of radius r_1 with center p_1. Further, p_3 is a point on a circle of radius r_2 about p_2. p_5 has a distance of $r_2 + \epsilon$ to p_2 for a small $\epsilon > 0$. There are two permissible permutations for L_1, as p_3 and p_4 have a distance of r_1 to P_1:

$$L_1\{p_3, p_4, p_5\} \text{ and } L_1\{p_4, p_3, p_5\}$$

Since p_5 is very close to the circle, there is only one permissible permutation for L_2:

$$L_2\{p_3, p_5, p_4\}$$

None of the permutations of L_1 corresponds to a reversely ordered L_2, and therefore a single list does not suffice to keep track of the ordered distances to nodes p_1 and p_2.

□

4.2.1. Maximum Error

The population of a geographic region is distributed across its census blocks B such that the sum of their population is equal to the region's population p. Further, each point in the geographic region, and therefore its population, is assigned to a unique census block: $\forall b_i, b_j \in B, b_i \neq b_j : interior(b_i) \cap interior(b_j) = \emptyset$ and $\bigcup_{b \in B} b = B$. While a minimum population size of 0 is permissible for census blocks, the maximum occurring value $b_{max} = max_{b \in B}\,(pop(b))$ varies among different geographic regions. Let $B_{max} \subseteq B$ denote the set of all census blocks of population size b_{max}. B_{max} contains at least one element and at most all census blocks of the geographic region.

In the following, assume the number of PODs k to be a power of 2, i.e., $k = 2^h$ for some h. Our results are then be generalized to any arbitrary number of PODs. A catchment area $CA_i, i \in [1..k]$ is defined by POD_i and its population is denoted as $pop(CA_i)$. The maximum observed population difference Δ_{max} between any two catchment areas is defined as follows:

$$\Delta_{max} = max_{i,j \in [1..k]} |(pop(CA_i) - pop(CA_j)|$$

As a consequence of the algorithm, regions cannot be overlapping and each census block is assigned to exactly one of the k PODs. Therefore, the sum of the population in each of the k catchment areas is equal to the population of the entire geographic region:

$$\sum_{i=1}^{k} pop(CA_i) = p$$

LEMMA 4.2. *For the placement of $k = 2$ PODs, a total population size p and largest population size of a census block b_{max}, $\Delta_{max} \in [0, b_{max}]$.*

PROOF. For the placement of $k = 2$ PODs the geographic region is partitioned into two subregions. The lower bound of 0 people difference between both catchment areas is achieved if the population is equally distributed between the PODs (best case). One instance of such a case is a region consisting of only two census blocks with equal population $pop(CA_1) = pop(CA_2) = \frac{p}{2}$. Algorithm 6 then assigns one census block to each of the PODs, distributing

the population equally across the PODs. Consequently, the resulting partitioning results in $\Delta_{max} = 0$.

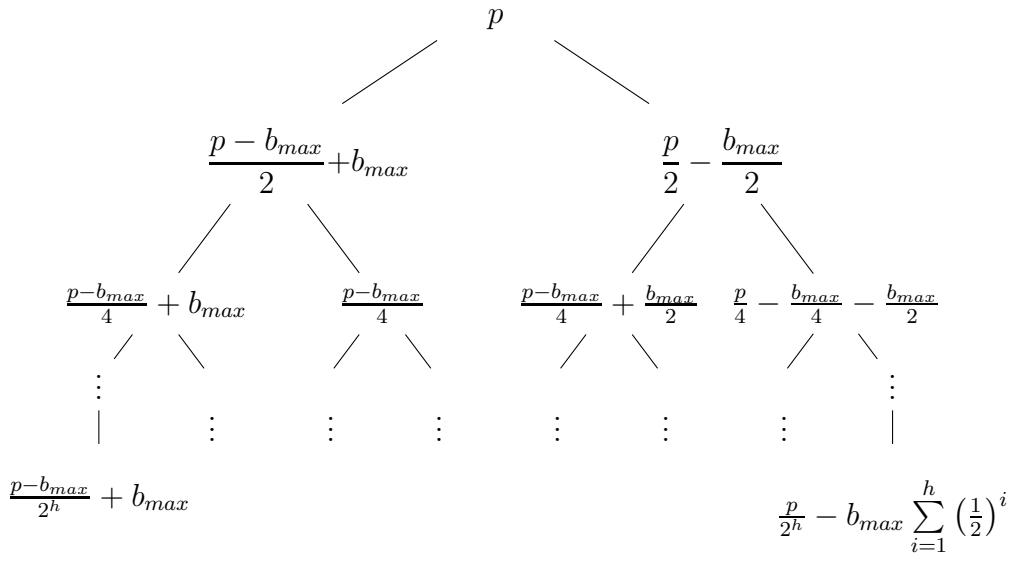


FIGURE 4.2. Maximum and minimum population sizes at each recursion level

Assume that all census blocks except one census block b_{last} have been assigned to the PODs. This results in one of the three following cases:

(i) $pop(CA_1) < pop(CA_2)$: The population of CA_1 compared to the population of CA_2 is smaller and in the range of $[pop(CA_2) - b_{max}, pop(CA_2) - 1]$. The smallest current population size permissible for CA_1 is $pop(CA_2) - b_{max}$, as b_{max} is the largest possible step size and hence the largest difference that can exist between the two catchment areas at any time during the execution of the algorithm. Algorithm 6 assigns b_{last} to CA_1, as it has accumulated a smaller population. After the assignment, the population of CA_1 is in $[pop(CA_2) - b_{max} + pop(b_{last}), pop(CA_2) - 1 + pop(b_{last})]$. As $pop(b_{last}) \in [0, b_{max}]$, the extreme cases yield $[pop(CA_2) - b_{max}, pop(CA_2) - 1]$ and $[pop(CA_2), pop(CA_2) - 1 + b_{max}]$. Therefore, the maximum possible difference between the catchment areas is b_{max}.

(ii) $pop(CA_1) > pop(CA_2)$: This case is analogous to $pop(CA_1) < pop(CA_2)$.

(iii) $pop(CA_1) = pop(CA_2)$: Either of the catchment areas is assigned the last census block b_{last}. Without loss of generality, assume that b_{last} is assigned to CA_1. With $pop(b_{last}) \in [0, b_{max}]$ the population of CA_1 is then in the range $[pop(CA_2), pop(CA_2)+ b_{max}]$. Therefore, the maximum possible difference between the catchment areas is b_{max}.

The overall maximum possible difference between the catchment areas is b_{max}, and consequently, $\Delta_{max} \in [0, b_{max}]$. Having established the maximum difference between two catchment areas, a bound for the maximum difference across the partition must be determined.

□

THEOREM 4.3. *For the placement of $k = 2^h$ PODs, a total population size p and largest population size of a census block b_m, Δ_{max} is bounded and $\Delta_{max} \in [0, 2b_{max}]$.*

PROOF. As stated by Lemma 4.2, the maximum difference of population at each recursion level is b_{max}. During each recursive step, Lemma 4.2 can be applied repeatedly. The resulting possible maximum and minimum populations for different recursion steps are shown in Figure 4.2. The left and right child of each node differ by exactly b_{max} individuals. Note that the sum of the population of all nodes at each level is p. The leftmost leaf node of the tree contains the region with the highest possible population size, whereas the rightmost leaf node of the tree contains the region with the lowest possible population size.

The maximum possible population size pop_{max} and the minimum possible population size pop_{min} for a tree of height h are calculated as follows:

$$pop_{max} = \frac{p - b_{max}}{2^h} + b_{max} = \frac{p}{2^h} + \left(1 - \frac{1}{2^h}\right) b_{max}$$

$$pop_{min} = \frac{p}{2^h} - b_{max} \sum_{i=1}^{h} \left(\frac{1}{2}\right)^i$$

The summation component of pop_{min} is a geometric series [35] and hence, pop_{min} can be expressed as follows:

$$pop_{min} = \frac{p}{2^h} - \left(\frac{\frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^h\right)}{1 - \frac{1}{2}} \right) b_{max} = \frac{p}{2^h} - \left(1 - \left(\frac{1}{2^h}\right)\right) b_{max}$$

The maximum difference in population sizes is calculated by subtracting the minimum population size from the maximum population size of a catchment area:

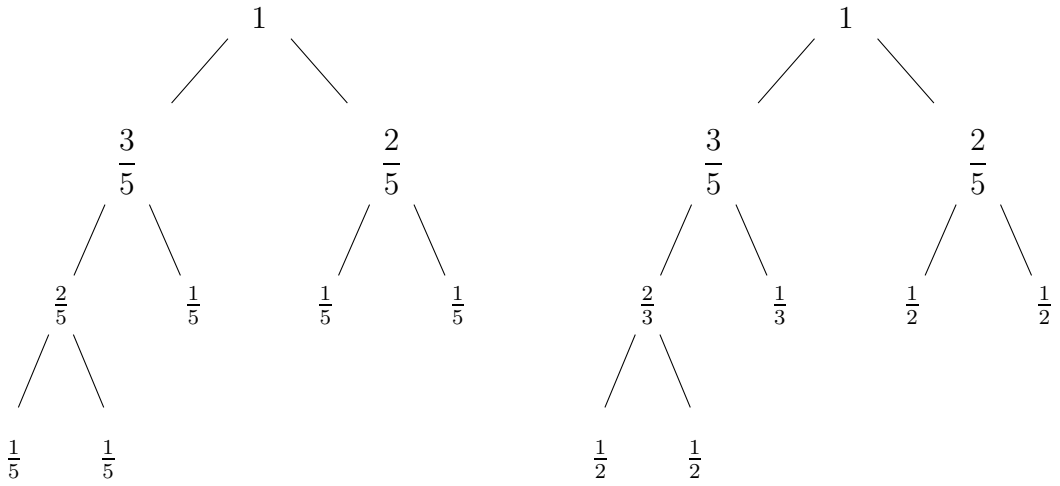
$$\begin{aligned}
\Delta_{max} &= pop_{max} - pop_{min} \\
&= \frac{p}{2^h} + \left(1 - \frac{1}{2^h}\right) b_{max} - \left[\frac{p}{2^h} - \left(1 - \left(\frac{1}{2^h}\right)\right) b_{max}\right] \\
&= \left(1 - \frac{1}{2^h}\right) b_{max} + \left(1 - \frac{1}{2^h}\right) b_{max} \\
&= \left(2 - \frac{2}{2^h}\right) b_{max} = \left(2 - \frac{1}{2^{h-1}}\right) b_{max}
\end{aligned}$$

For a large number of PODs k, approximating infinity, the height of the recursion tree h also approximates infinity. This may be used to calculate the upper bound for the maximum difference in population sizes between any two catchment areas:

$$\lim_{k \to \infty} \left(2 - \frac{2}{2^h}\right) b_{max} = 2b_{max}$$

□

The same bound also holds for an arbitrary number of PODs, which is shown via Lemma 4.4 and Theorem 4.5. Figure 4.3 illustrates such a case for $k = 5$. Figure 3(a) shows the portion of the total population of region R ideally assigned at each step to a particular sub-region. Based on Algorithm 6 during each recursion level, a proportion of a particular region is assigned to its sub-regions. This is shown in Figure 3(b). Multiplying the proportions of any path from the root node to any of the leaf nodes yields $\frac{1}{5}$. Thus, the leftmost and the rightmost paths yield $\frac{3}{5} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{5} = \frac{1}{k}$ and $\frac{2}{5} \times \frac{1}{2} = \frac{1}{5} = \frac{1}{k}$, respectively. Lemma 4.4 shows that this is also true for the general case.



(a) Portion of population (b) Proportions assigned at each level

FIGURE 4.3. Example of $k = 5$ PODs

Assume that at each recursion step, the last census block to be assigned by Algorithm 6 is of size b_{max} and that both sub-regions R_1 and R_2 are of equal population size: $pop(R_1) = pop(R_2)$. Further, assume that the algorithm assigns the census block to the left child. Consequently, the difference in population size between two siblings is always b_{max} Figure 4.4 shows the population sizes assigned to each sub-region at a particular recursion level. The proportions by which the population is split is expressed by the coefficients α_i with $i \in [1..h_1]$ and β_j with $j \in [1..h_2]$ where h_i and h_2 denote heights of the left and right subtrees.

LEMMA 4.4. *For coefficients α_i in any path from the root node to any leaf of a tree of height h, the product of the coefficients yields the reciprocal of the number of PODs:*

$$\prod_{i=1}^{h} \alpha_i = \alpha_1 \times \alpha_2 \times \alpha_3 \times \cdots \times \alpha_{h-1} = \frac{1}{k}$$

PROOF. Let $path_i$ denote a particular path from the root node to a specific leaf and let x_1 denote the number of PODs that is assigned to the sub-region that lies on $path_i$ during the

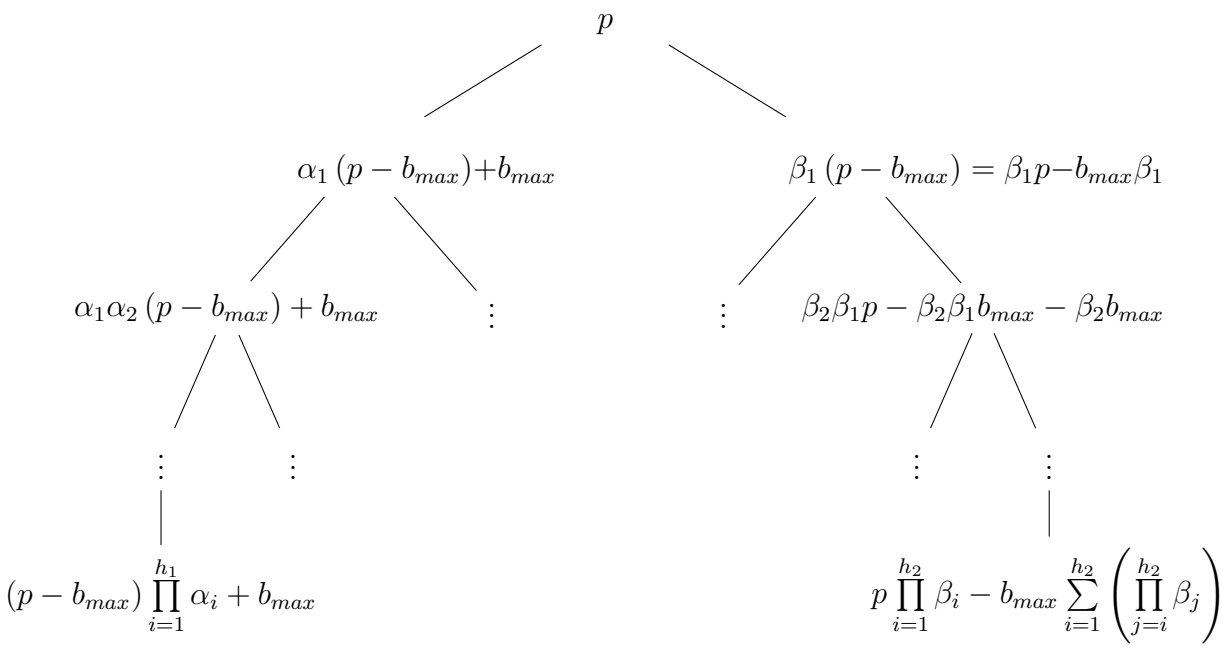


FIGURE 4.4. General case: k PODs while k is not required to be a power of 2

first recursion step. Then, ideally $\frac{x_1}{k}$ of the population is assigned to that sub-region. For the next partitioning step, x_1 PODs have to be assigned. Let x_2 out of the x_1 PODs be assigned to the next sub-region on $path_i$. Consequently, $\frac{x_2}{x_1}$ of the PODs at this level are assigned to it. During the last recursion step, $\frac{x_h}{x_{h-1}}$ are assigned. As a leaf node has been reached, only one POD is assigned, and hence $x_h = 1$. Consequently,

$$\frac{x_1}{k} \times \frac{x_2}{x_1} \times \frac{x_3}{x_2} \times \cdots \times \frac{x_h}{x_{h-1}} = \frac{1}{k}$$

□

THEOREM 4.5. *For the placement of k PODs, a total population size p and largest population size of a census block b_{max}, Δ_{max} is bounded and $\Delta_{max} \in [0, 2b_{max}]$.*

PROOF. During each recursion step, the maximum difference between the sub-regions is b_{max}. In the general case any number k of PODs is permissible. For each split a coefficient denotes the proportion of the population ideally assigned to a particular sub-region. The

goal of the first recursion step in Figure 4.4 is to assign $\alpha_1 p$ to the left sub-region and $\beta_1 p$ to the right sub-region. Assume that at each recursion level the population differs b_{max} between the two sub-regions as shown by Lemma 4.2. The leftmost leaf node in Figure 4.4 shows the highest possible population size for a catchment area, whereas the rightmost leaf shows the smallest possible population size. The population sizes of the leftmost sub-region is calculated as follows:

$$pop_{max} = (p - b_{max}) \prod_{i=1}^{h_1} \alpha_i + b_{max}$$

Applying Lemma 4.4 yields

$$pop_{max} = \frac{1}{k} (p - b_{max}) + b_{max}$$

The population sizes of the rightmost sub-region is calculated as follows:

$$pop_{min} = p \prod_{i=1}^{h_2} \beta_i - b_{max} \sum_{i=1}^{h_2} \left(\prod_{j=i}^{h_2} \beta_j \right)$$

This can be simplified by applying Lemma 4.4:

$$pop_{min} = \frac{1}{k} p - b_{max} \sum_{i=1}^{h_2} \left(\prod_{j=i}^{h_2} \beta_j \right)$$

The right child of a particular node is at most assigned $\frac{1}{2}$ of the population, as the algorithm assigns the larger proportion to the left child. Hence $\beta_j \leq \frac{1}{2}$ with $j \in \{1..h_2\}$. Therefore

$$pop_{min} \geq \frac{1}{k} p - b_{max} \sum_{i=1}^{h_2} \left(\frac{1}{2} \right)^i$$

By applying the geometric series this can be transformed into the following inequality:

$$pop_{min} \geq$$

$$\frac{1}{k} p - b_{max} \frac{\frac{1}{2} \left(1 - \left(\frac{1}{2} \right)^{h_2} \right)}{1 - \frac{1}{2}} = \frac{1}{k} p - b_{max} \left(1 - \left(\frac{1}{2} \right)^{h_2} \right)$$

The maximum difference in populations sizes is calculated by subtracting the minimum population size from the maximum population size of a catchment area:

$$\Delta_{max} \leq$$

$$\frac{1}{k}(p - b_{max}) + b_{max} - \left(\frac{1}{k}p - b_{max}\left(1 - \left(\frac{1}{2}\right)^{h_2}\right)\right)$$

$$= -\frac{1}{k}b_{max} + b_{max} + b_{max}\left(1 - \left(\frac{1}{2}\right)^{h_2}\right)$$

Let h_1 denote the height of the left side of the tree, i.e. the length of the path from the root node to the leftmost leaf. Then $h_1 = \lceil log(k) \rceil$, as Algorithm 6 fills the last tree level, i.e. level h_1, from left to right. Since the recursion tree is not necessarily complete, the height of the right side of the tree can either be $h_2 = h_1 = \lceil log(k) \rceil$ or $h_2 = \lfloor log(k) \rfloor$. If k is a power of 2, it follows that $h_1 = h_2 = \lfloor log(k) \rfloor = \lceil log(k) \rceil = log(k)$. If k is not a power of 2, then let k' denote the largest number smaller than k that is a power of 2. This yields a height $h_2 = log(k')$. Note that if k approaches infinity, k' also approaches infinity.

Case 1: $h_2 = h_1$ (complete tree)

$$h_2 = h_1 = log(k) \Rightarrow k = 2^{h_2} = 2^{h_1}$$

$$\Delta_{max} \leq -\frac{1}{k}b_{max} + b_{max} + b_{max}\left(1 - \left(\frac{1}{2}\right)^{h_2}\right)$$

$$= -\frac{1}{k}b_{max} + b_{max} + b_{max}\left(1 - \left(\frac{1}{2}\right)^{log(k)}\right)$$

Case 2: $h_2 \neq h_1$ (tree not complete)

$$h_2 = h_1 - 1 = log(k') = \lfloor log(k) \rfloor \Rightarrow k' = 2^{h_2} = 2^{h_1 - 1}$$

$$k' \leq k \Leftrightarrow log(k') \leq log(k)$$

$$\left(\frac{1}{2}\right)^{log(k')} \geq \left(\frac{1}{2}\right)^{log(k)}$$

$$1 - \left(\frac{1}{2}\right)^{log(k')} \leq 1 - \left(\frac{1}{2}\right)^{log(k)}$$

$$\Delta_{max} \leq -\frac{1}{k} b_{max} + b_{max} + b_{max} \left(1 - \left(\frac{1}{2}\right)^{h_2}\right)$$

$$= -\frac{1}{k} b_{max} + b_{max} + b_{max} \left(1 - \left(\frac{1}{2}\right)^{log(k')}\right)$$

$$\leq -\frac{1}{k} b_{max} + b_{max} + b_{max} \left(1 - \left(\frac{1}{2}\right)^{log(k)}\right)$$

For both cases Δ_{max} is bound by the same expression. If the number of facilities approaches infinity, the following bounds can be determined:

$$\lim_{k \to \infty} \Delta_{max}$$

$$= \underbrace{-\frac{1}{k} b_{max}}_{\to 0} + b_{max} + b_{max} \underbrace{\left(1 - \left(\frac{1}{2}\right)^{log(k)}\right)}_{\to 1} = 2b_{max}$$

The upper and lower bound approach infinity for large k. Consequently, the maximum possible difference in population size between any two catchment areas is $2b_{max}$.

□

THEOREM 4.6. *The deviation of the population p of any catchment area CA_i from the optimal population size p_{opt} is $\leq b_{max}$.*

PROOF. The optimal population distribution for k PODs would result in $\frac{p}{k}$ people per catchment area.

Case 1: A catchment area has the largest possible population size. Then the deviation from the optimal size is calculated as

$$|pop_{max} - p_{opt}| = \left|\frac{1}{k}(p - b_{max}) + b_{max} - \frac{p}{k}\right| = b_{max} - \frac{b_{max}}{k} < b_{max}$$

Case 2: A catchment area has the smallest possible population size. Then the deviation from the optimal size is calculated as

$$|p_{opt} - p_{min}| \leq \left| \frac{p}{k} - \left[\frac{1}{k}p - b_{max} \left(1 - \left(\frac{1}{2} \right)^{h_2} \right) \right] \right| = b_{max} \left(1 - \left(\frac{1}{2} \right)^{h_2} \right) < b_{max}$$

The largest and smallest possible catchment areas have a maximum deviation of b_{max} from the optimal size and therefore, it is true for all other catchment area sizes.

□

4.2.2. Experimental Results

In what follows, Algorithm 6 has been applied to Denton County, sub-dividing the geographic space into three and fifteen regions, respectively. Figure 4.5 (a) displays the census blocks of Denton County subdivided by Algorithm 6 into three catchment areas. The algorithm has mapped the PODs on the geographic centroids of the catchment areas. In Figure 4.5 (b) RE-PLAN has been applied to these POD locations, yielding Voronoi-based catchment areas, whereas Figure 4.5 (c) shows the results of the response analysis applied on the catchment areas determined by Algorithm 6.

(a) Voronoi-based

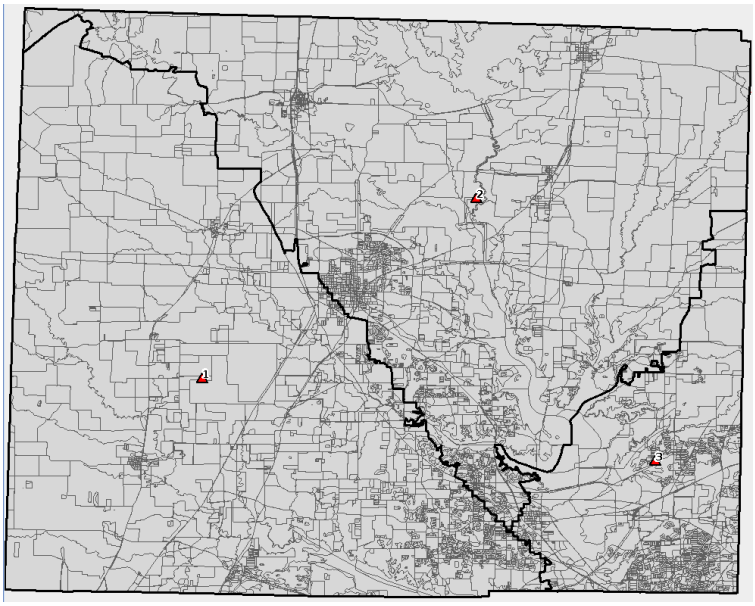
pod	population
1	71580
2	92545
3	268851

(b) Algorithm 6

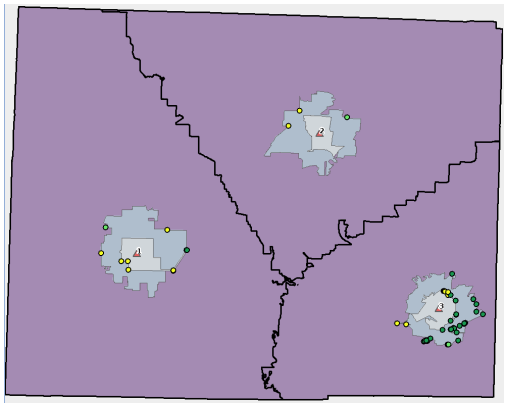
pod	population
1	144689
2	144130
3	144157

TABLE 4.1. Population of catchment areas for $k = 3$ PODs

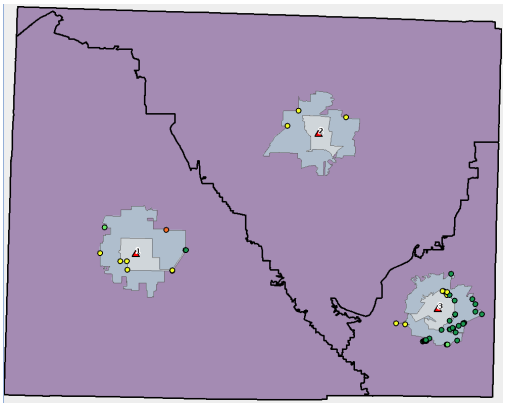
The population distribution across the catchment areas is shown in Figure 4.6. Since Algorithm 6 optimizes with respect to the population distribution, Figure 4.6(b) shows population sizes that are more equally distributed than those obtained by the Voronoi-based plan analysis (Figure 4.6(a)). The actual population sizes of the catchment areas are compared in Table 4.1. The maximum size of a census block in Denton County is 3931, and therefore, Theorem 4.5 guarantees a maximum difference between any two catchment area sizes of 7862 people. The maximum difference observed for Algorithm 6 with $k = 3$



(a) Catchment areas defined by Algorithm 6



(b) Voronoi-based response analysis on PODs determined by Algorithm 6

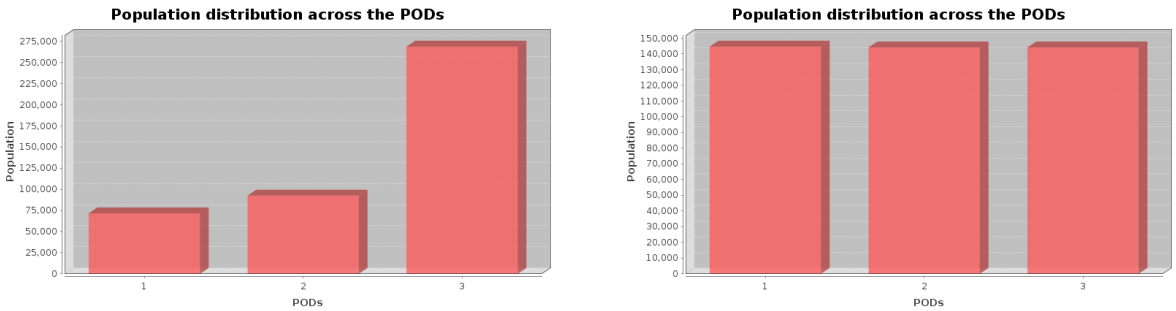


(c) Response analysis on continuous catchment areas determined by Algorithm 6

FIGURE 4.5. Continuous partitioning with $k = 3$ PODs

is 558, which only constitutes 7% of that bound. If for the same set of PODs, however, Voronoi-based catchment areas are calculated, the maximum difference between any two catchment areas results in 197271 people, which is about 25 times the difference allowed by the theoretical bound for Algorithm 6.

Setting $k = 15$, Denton County is partitioned into fifteen catchment areas (Figure 4.7(a)). The results of the response analysis on the two different catchment area types are shown in Figures 4.7(b) and 4.7(c). The corresponding population distributions are compared in



(a) Voronoi-based catchment areas (b) Catchment areas determined by Algorithm 6

FIGURE 4.6. Comparison of population in catchment with $k = 3$ PODs

Figure 4.8. Similarly, to the results for $k = 3$, the algorithm yields a more uniform population distribution. The population sizes per catchment area are summarized in Table 4.2.

(a) Voronoi-based

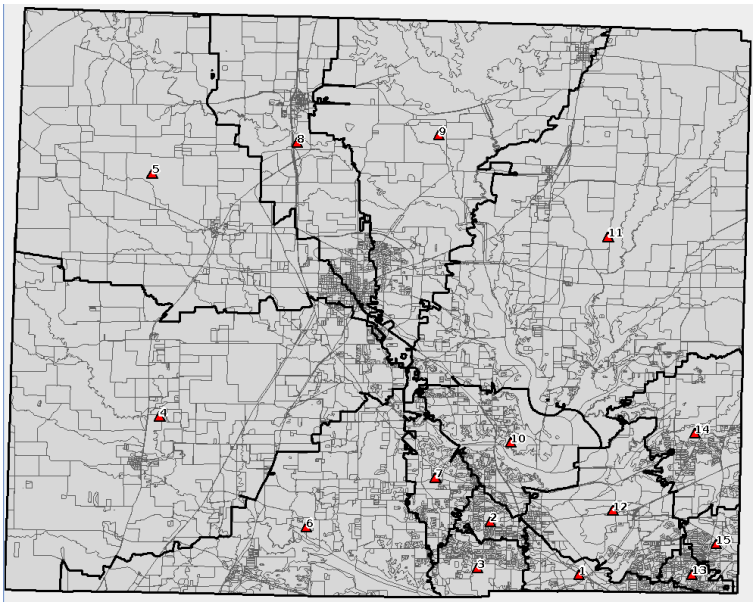
pod	population
1	24771
2	51456
3	34658
4	9164
5	6373
6	19559
7	55711
8	50763
9	26096
10	29217
11	8875
12	5183
13	51747
14	34493
15	24910

(b) Algorithm 6

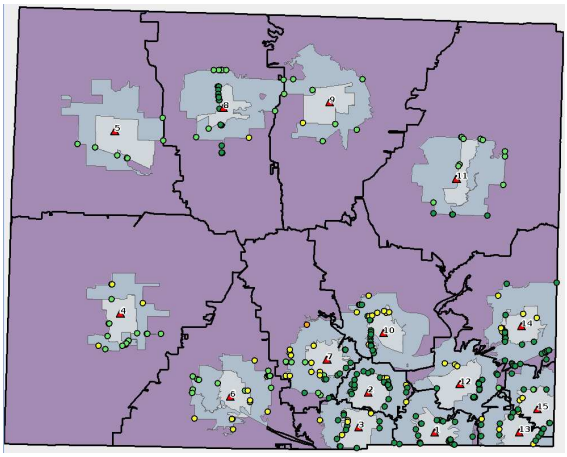
pod	population
1	28803
2	29014
3	28952
4	29175
5	28478
6	28808
7	28802
8	28991
9	28630
10	28836
11	28778
12	29339
13	28928
14	28697
15	28745

TABLE 4.2. Population of catchment areas for $k = 15$ PODs

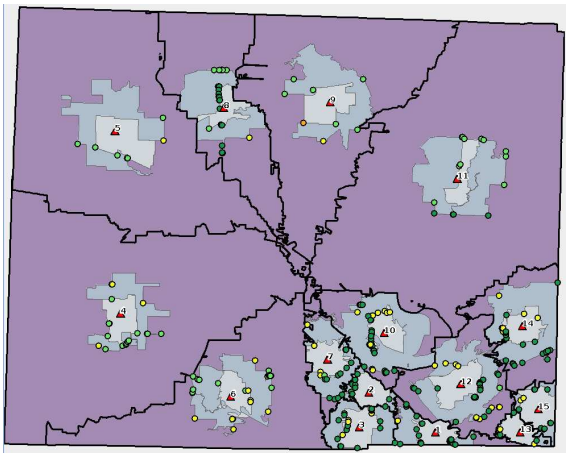
Algorithm 6 yields maximum observed difference in catchment area size of 861. With a theoretical bound of 7862, for this experiment only about 11% of the bound is reached. This stands in contrast to the Voronoi-based approach with a maximum observed difference of 50528, which is over 6 times the theoretical bound for Algorithm 6. The results indicate



(a) Catchment areas defined by Algorithm 6



(b) Voronoi-based response analysis on PODs determined by Algorithm 6



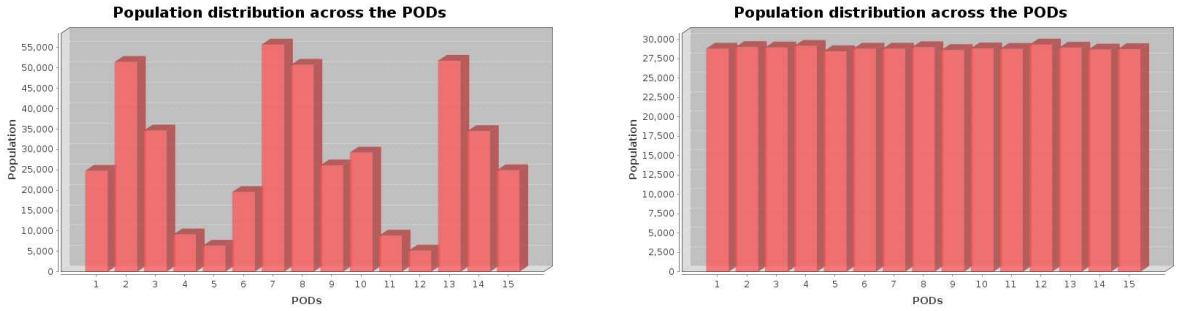
(c) Response analysis on continuous catchment areas determined by Algorithm 6

FIGURE 4.7. Continuous partitioning with $k = 15$ PODs

that catchment area types that are not Voronoi-based may be an alternative that allows for balancing the number of people across all catchment areas.

4.2.3. Variation

A variation of the algorithm that yields multiple results, does not force a decision at each particular step. Algorithm 6 assigns each partition approximately the same number of individuals, if the number of PODs k is even. Nevertheless, if the number of PODs k is odd,



(a) Voronoi-based catchment areas (b) Catchment areas determined by Algorithm 6

FIGURE 4.8. Comparison of population distributions

one partition is assigned k div 2 PODs, while the other partition is assigned (k div 2) + 1 PODs. Consequently, the population is not evenly distributed between the two sub-regions. A variation of Algorithm 6 explores both possible partitions for each occurrence of an odd number. Figure 4.9 illustrates this for $k = 5$ PODs, whereby each node of the tree represents how many PODs are assigned to a particular partition. For $k = 5$, there exist four different partitions.

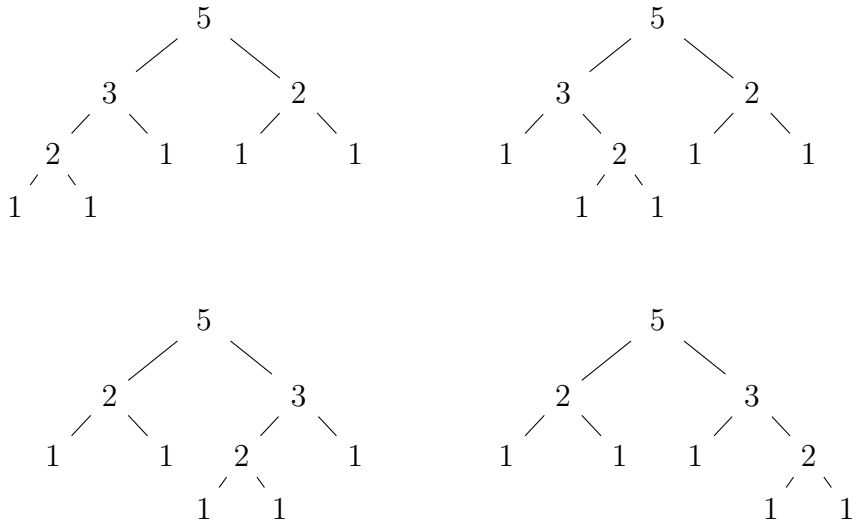


FIGURE 4.9. Variations of POD assignment for $k = 5$ PODs

Input: Set of candidate service locations (PODs) L_n
number of desired PODs k

Output: Subset $L_k \subseteq L$ of candidate locations

```
/* Init set of PODs                                          */
L = L_n ;
for i = n to k + 1 do
    worstValue = 0 ;
    L_remove = null ;
    foreach p ∈ L_i do
        Calculate C_{i-1,p.id} ;
        /* Determine value of removal criterion              */
        value = remCriterion(C_{i-1,p.id}); ;
        if value > worstValue then
            worstValue = value ;
            L_remove = L_{i-1,p.id} ;
        end
    end
    /* Remove a POD based on the removal criterion           */
    L = L_remove;
    /* Final selection L_k                                   */
    return L ;
end
```

ALGORITHM 7: Reverse greedy hill-climing approach

4.3. Applying Discrete Location Science to POD Placement

As opposed to models in Continuous Locations Science, discrete models assume demand points as well as service facilities at discrete locations. For the placement of PODs, the service area is a geographic region R, which is partitioned into a set of census blocks B. The candidate locations for POD placement are stored in a set of size n. Authorities, however, can only provide resources to operate $0 < k \leq n$ facilities. In the following, different experimental approaches to determine a feasible solution are discussed.

4.3.1. Reverse Greedy Hill-Climbing

In this section, a greedy heuristic to determine a solution for the $\binom{n}{k}$ problem is discussed. If resources for all n PODs are available, then catchment areas can be determined as described in Section 3.3. This yields a set of catchment areas CA_n for a scenario utilizing each of the

n PODs. Nevertheless, if only resources for $k = n - 1$ PODs are available, one of the PODs must be removed from the set. This can be done by using a reverse hill-climbing method. Based on a removal criterion, such as the optimal average population size, the lowest ranked POD is removed. This involves computing n sets of catchment areas, each of which reflects the removal of a different POD. $CA_{n-1,id}$ denotes the catchment areas as a result of removing the POD identified by id. $CA_{n-1} = CA_{n-1,x}$ denotes the set of catchment areas obtained by selecting the POD with POD $id = x$ for removal based on the removal criterion. For $k \leq n - 1$ PODs, this method is applied repeatedly, removing one POD per iteration. This yields a sequence of catchment area sets: $CA_n, CA_{n-1}, CA_{n-2}, .., CA_k$. Analogously, the sequence of sets of available PODs at each iteration is expressed as $L_n, L_{n-1}, L_{n-2}, .., L_k$.

Assume that all of the PODs are of equal capacity. More general cases can be included by adding factors for maximum POD capacities to the PODs. The here described approach is summarized in Algortihm 7. For this dissertation, the following removal criteria have been considered:

- Fixed optimum: The "optimal" average population per POD is calculated as the population p of region R divided by the number of PODs k to be placed: $\frac{pop(R)}{k}$
- Variable optimum: The "optimal" average population per POD is calculated at each step as the population p of region R divided by the current number of PODs $i \in [k + 1..n]$: $\frac{pop(R)}{i}$
- Fixed standard deviation: The standard deviation is calculated based on the fixed optimum.
- Variable standard deviation: The standard deviation is re-calculated during each iteration based on the variable optimum at that iterative step.

The reverse greedy hill-climbing approach does not guarantee optimality. Nevertheless, in spite of of having to solve a $\binom{n}{k}$ exponential problem, the presented heuristic generates a solution in polynomial time.

Input: Set of candidate service locations (PODs) L
number of desired PODs k

Output: Subset $L_k \subseteq L$ of candidate locations

`/* Apply continuous partitioning */`
T_k = execute $Algortihm6$;
`/* Initialize result set */`
L_k ;
foreach $t \in T_k$ **do**
    $p = closestPod(L, t)$;
    $L_k.add(p)$;
    $l.remove(p)$;
end
return L_k ;

ALGORITHM 8: Hybrid approach based on continuous method

4.3.2. A Hybrid Approach

The partitioning algorithm as described in Section 4.2 creates k sub-regions of R with approximately equal population sizes. Define T_k as the set containing the catchment area centroids of such a temporary partitioning into k sub-regions. Then assign for each $t \in T_k$ the k closest POD locations in L_n to t. Once a POD has been assigned, it is removed from L_n so that no POD is assigned to multiple catchment areas. Finally, catchment areas are computed as described in Section 3.3.

Section 4.2.3 discusses a variation of the continuous partitioning approach. This variation can be incorporated into the hybrid method by applying to all possible variations. Then, for each scenario, catchment areas are computed (see Section 3.3). Applying the same criterion as in Algorithm 8, the "best" solution according this particular criterion is selected.

4.3.3. Experimental Results

In the following, experimental results obtained for Algorithms 7 and 8 are presented. For the latter, additional experiments were conducted by making use of the variation described in Section 4.2.3.

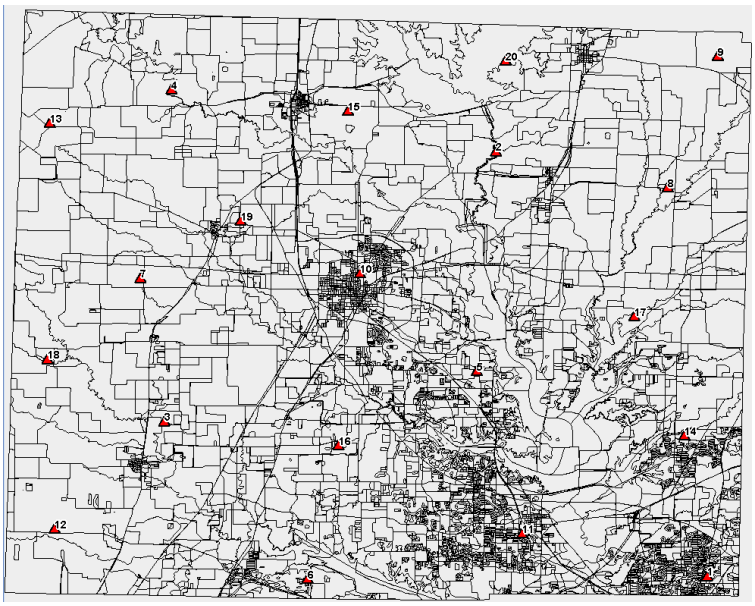


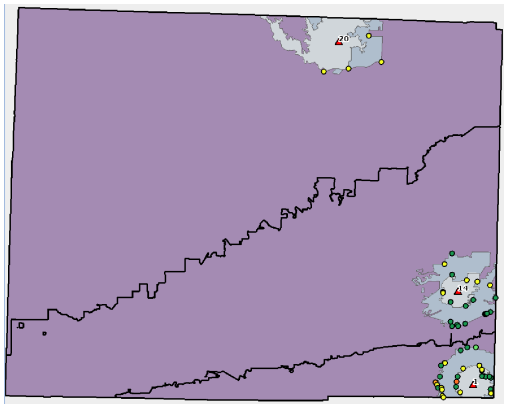
FIGURE 4.10. 20 candidate PODs distributed across Denton County

4.3.3.1. Algorithm 7

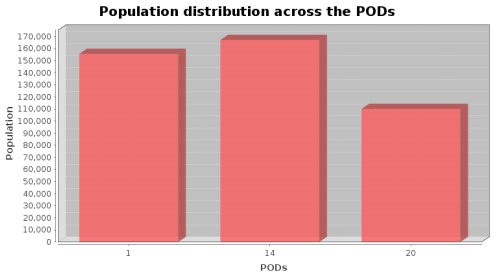
Figure 4.10 shows a set of 20 PODs distributed across Denton County. Algorithm 7 has been applied to this set to select three PODs utilizing each of the four removal criteria. Using a fixed global optimum or a fixed global standard deviation as removal criterion have returned the same subset of PODs. The selected PODs and thereby defined catchment areas are shown in Figure 4.11.

Both of the variable removal criteria yield different sub-divisions of the geographic space. The resulting catchment areas and population distributions are shown in Figures 4.12 and 4.13.

Fifty candidate PODs have been distributed across Denton County (Figure 4.14), out of which fifteen are to be selected. Applying Algorithm 7 with all four removal criteria yields two different results. Utilizing a fixed global optimum, a variable optimum, or a fixed global standard deviation as removal criterion causes the Algorithm to return the same subset of PODs (Figure 4.15) . Only if a variable standard deviation is applied as removal criterion, the resulting subset of PODs is different (Figure 4.16). The latter seems to perform slightly better when visually inspecting the population distribution.

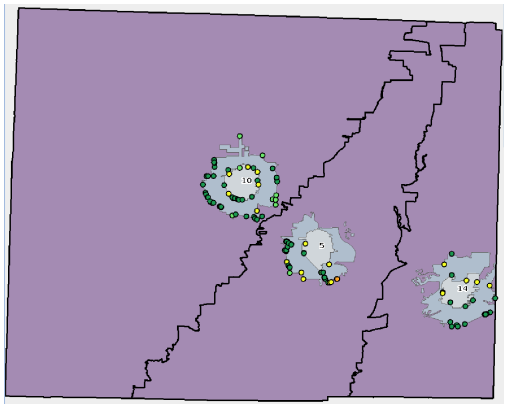


(a) Result of downhill greedy

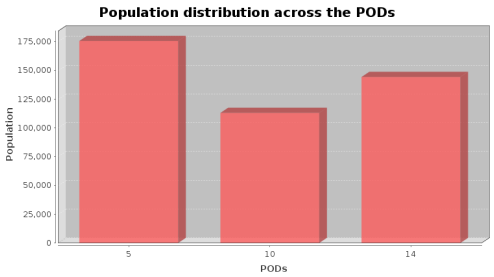


(b) Population distribution

FIGURE 4.11. Algorithm 7: $\binom{20}{3}$ with fixed optimum or fixed standard deviation



(a) Result of downhill greedy

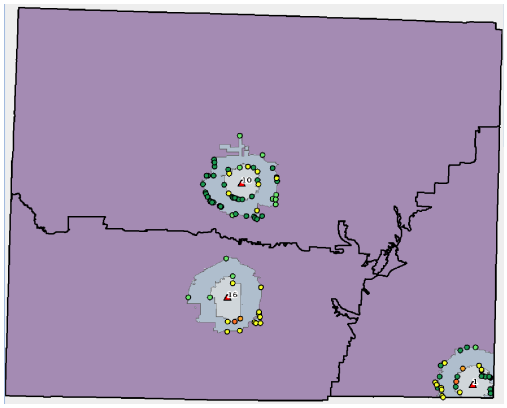


(b) Population distribution

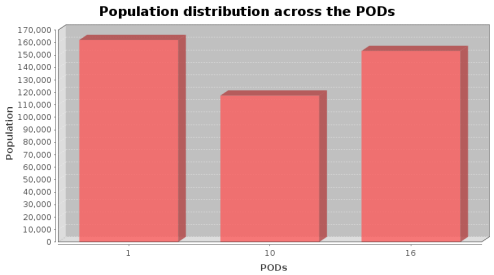
FIGURE 4.12. Algorithm 7: $\binom{20}{3}$ with variable optimum

4.3.3.2. Algorithm 8

Alogrithm 8 has been applied to the set of 20 PODs depicted in Figure 4.10 to select three PODs. This hybrid approach makes use of Algorithm 6 to create subdivisions of the geographic space. One of the 20 candidate PODs is assigned to each of the subdivisions, such that the distance between an assigned POD and the geographic centroid of a specific subdivision is minimized. The partition obtained by Algorithm 6 and the selected PODs based on the partition are illustrated in Figure 4.17 (a). RE-PLAN's response plan analysis module constructs Voronoi-based catchment areas (Figure 4.17 (b)) on the PODs returned



(a) Result of downhill greedy



(b) Population distribution

FIGURE 4.13. Algorithm 7: $\binom{20}{3}$ with variable std dev

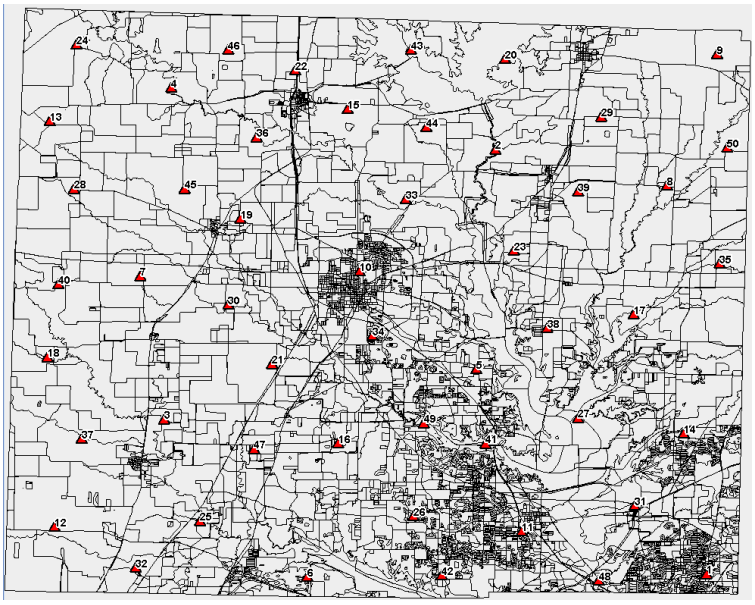
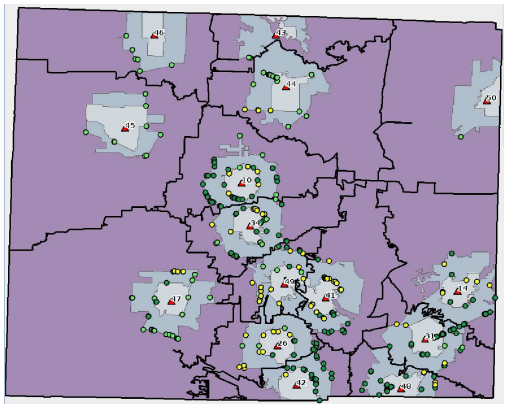


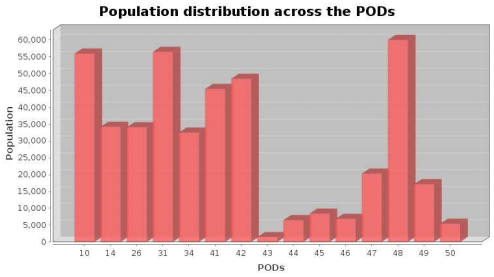
FIGURE 4.14. 50 candidate PODs distributed across Denton County

from Algorithm 8. Figure 4.17 (c) illustrates an overlay of the different types of catchment areas. The result of the response plan analysis is shown in Figure 4.17 (d).

Algorithm 8 has been applied to the set of 50 PODs depicted in Figure 4.14 to select 15 PODs. Figure 4.18 shows the continuous catchment areas constructed by Algorithm 6, as well as the Voronoi-based catchment areas calculated after fifteen PODs have been selected. Specifically, the results of the response plan analysis on the fifteen selected PODs are illustrated in Figure 4.18 (d)

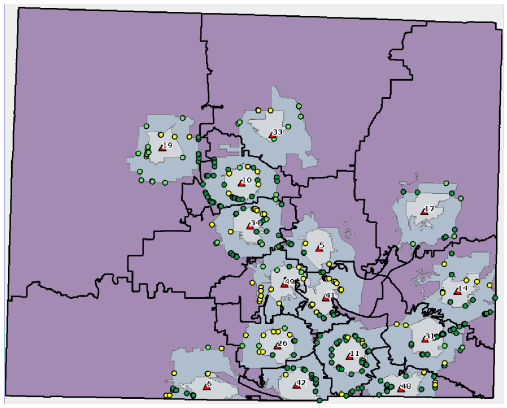


(a) Result of downhill greedy

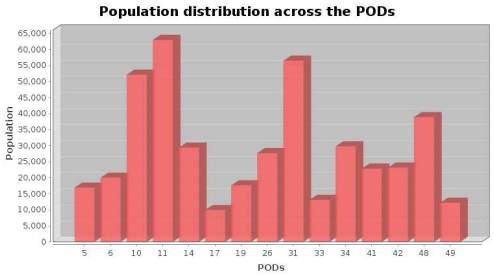


(b) Population distribution

FIGURE 4.15. Algorithm 7: $\binom{50}{15}$ with fixed optimum, variable optimum, or fixed standard deviation



(a) Result of downhill greedy



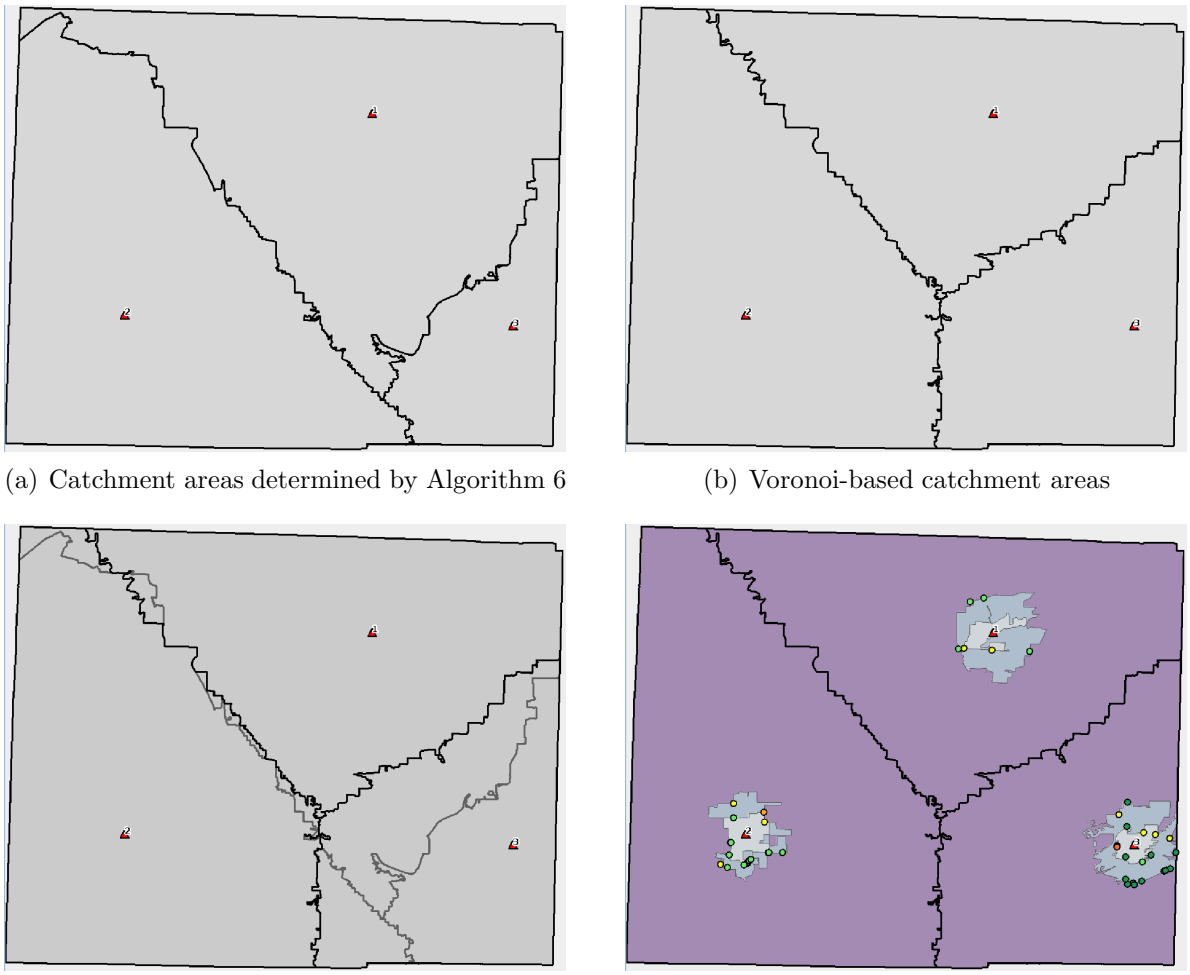
(b) Population distribution

FIGURE 4.16. Algorithm 7: $\binom{50}{15}$ with variable standard deviation

This approach seems to yield a population distribution that is less balanced than any of the other approaches described here (Figure 4.19). Some of the PODs are assigned a large portion of the population, whereas other PODs serve very little population, indicating inefficient use of resources.

4.3.3.3. Algorithm 8 with Variation

By making use of the variation discussed in Section 4.2.3, multiple subdivisions of Denton County can be computed. For the purpose of the experiments discussed here, the *globalaverage* removal criterion has been used and different solutions have been compared



(a) Catchment areas determined by Algorithm 6

(b) Voronoi-based catchment areas

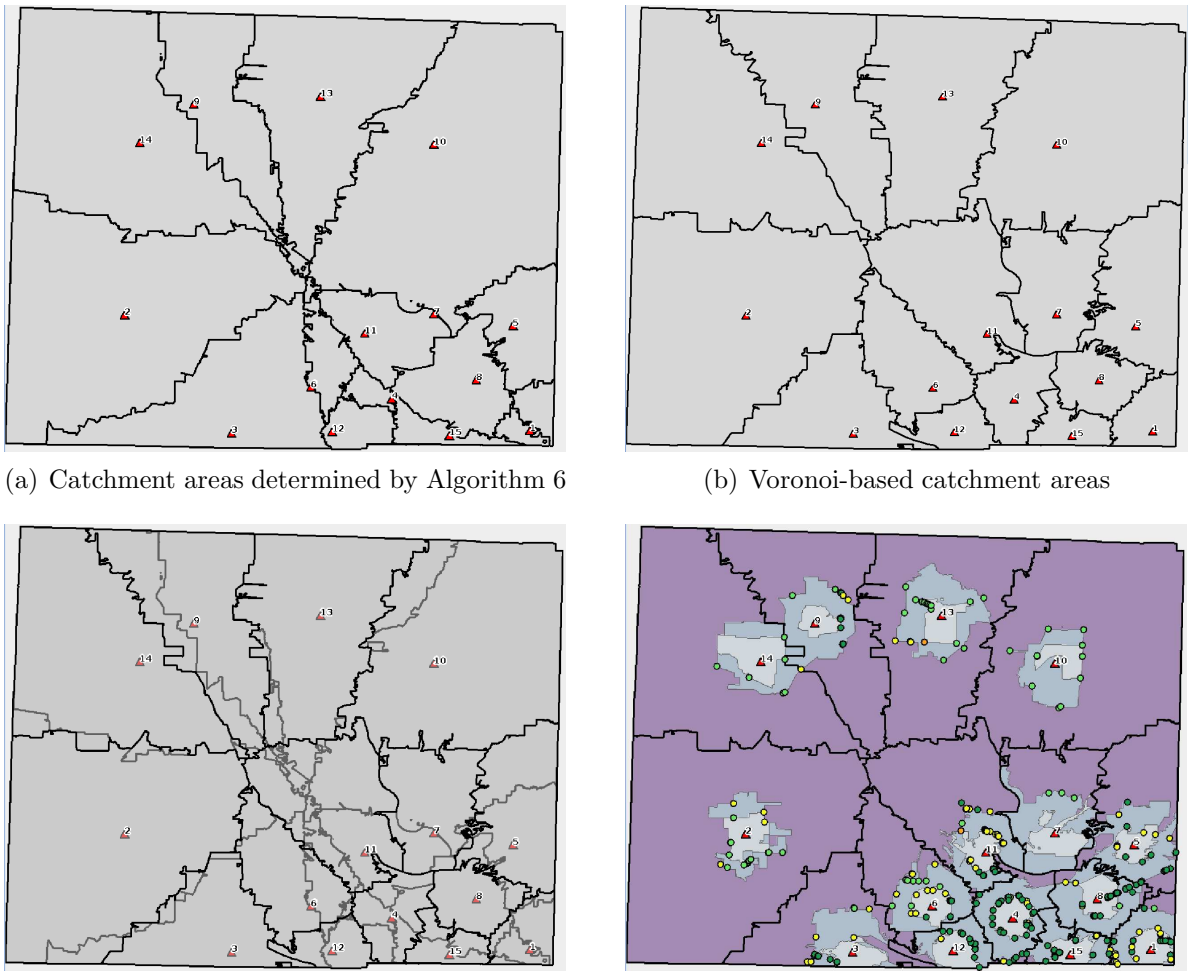
(c) Overlay of different catchment area types

(d) RE-PLAN execution

FIGURE 4.17. Algorithm 8: $\binom{20}{3}$

based on their population distributions across their catchment areas. Alternatively, different metrics, such as standard deviation, can be used. For the case of $k = 3$ PODs, two different subdivisions based on Algorithm 6 can be constructed (Figure 4.20). For three PODs, Scenario 4.20 (a) is selected, which corresponds to the scenario in Figure 4.17.

The variation of Algorithm 6 generates eight different subdivision of the geographic space for fifteen PODs (Figure 4.21). In the experiment, the third option (Figure 4.21 (c)) has been chosen instead of the first option. The results for this option are shown in Figure 4.22. While for the $\binom{20}{3}$-scenario the same subset of PODs is selected as obtained by using Algorithm 8, for the $\binom{50}{15}$-scenario, a different subset is chosen. The population distribution for the



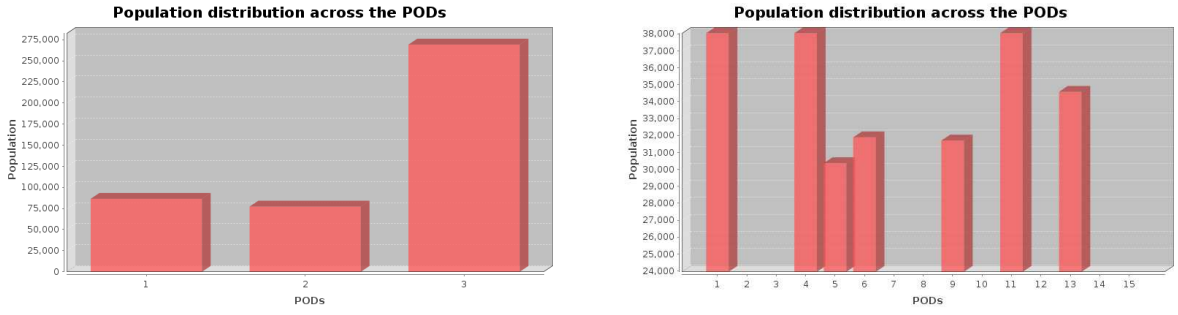
(a) Catchment areas determined by Algorithm 6

(b) Voronoi-based catchment areas

(c) Overlay of different catchment area types

(d) RE-PLAN execution

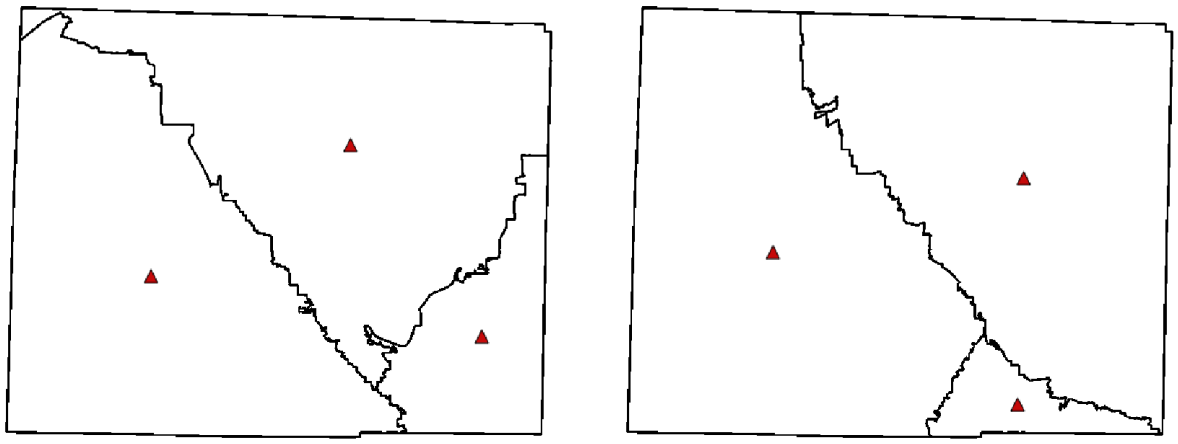
FIGURE 4.18. Algorithm 8: $\binom{50}{15}$



(a) $\binom{20}{3}$

(b) $\binom{50}{15}$

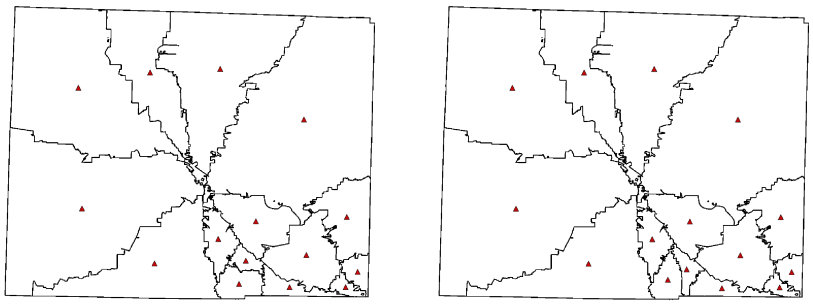
FIGURE 4.19. Population distribution generated by Algorithm 8



(a) Option 1 (b) Option 2

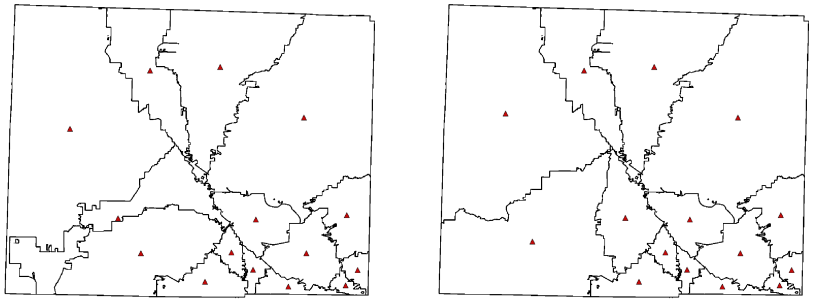
FIGURE 4.20. Partitioning options for $k = 3$ PODs

selected scenarios is shown in Figure 4.23. Although, the population in Figure 4.23 (b) is not equally distributed, it shows improvement with respect to the population distribution in Figure 4.19 (b)



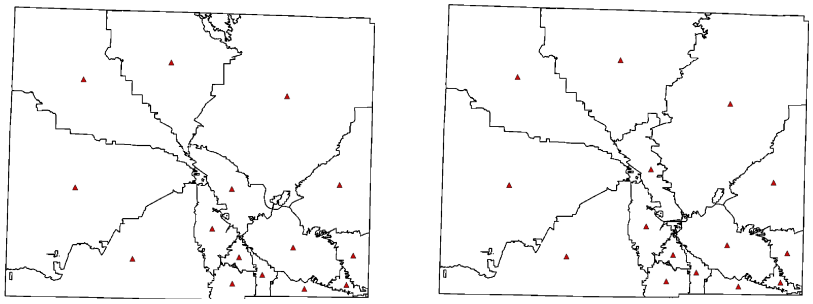
(a) Option 1

(b) Option 2



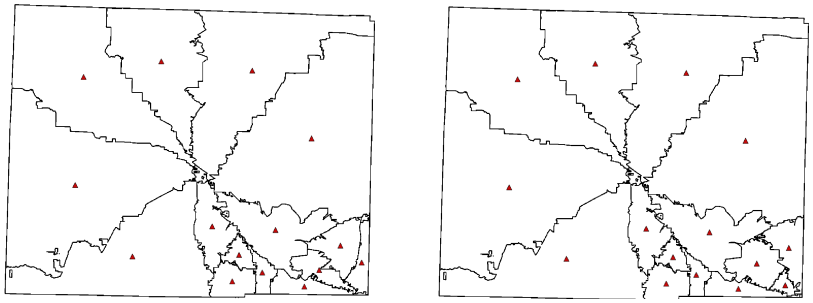
(c) Option 3

(d) Option 4



(e) Option 5

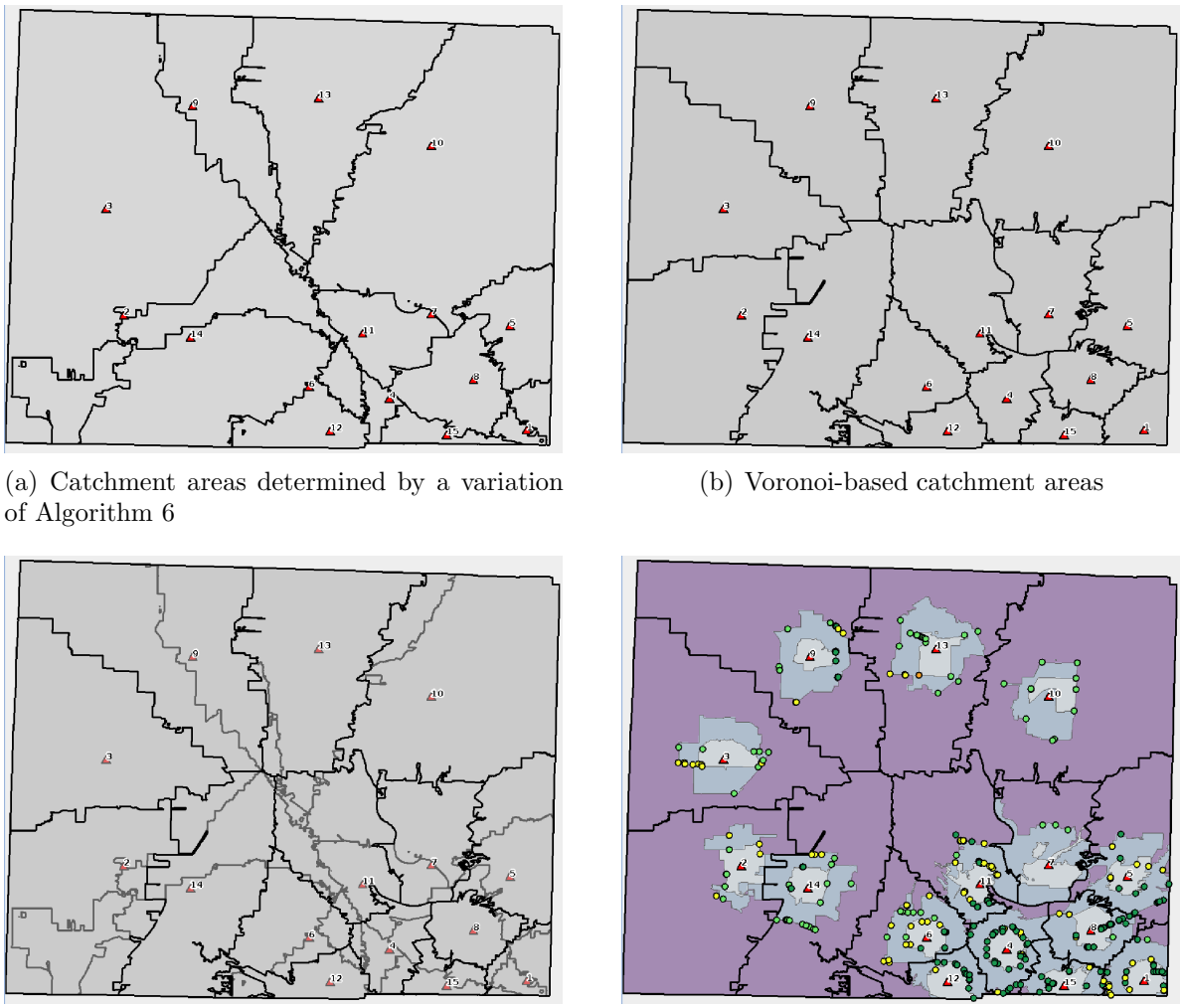
(f) Option 6



(g) Option 7

(h) Option 8

FIGURE 4.21. Options for $k = 15$ PODs



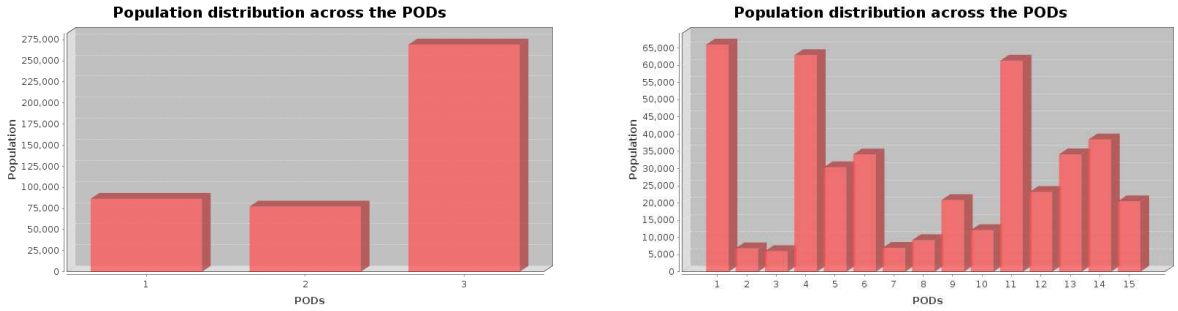
(a) Catchment areas determined by a variation of Algorithm 6

(b) Voronoi-based catchment areas

(c) Overlay of different catchment area types

(d) RE-PLAN execution

FIGURE 4.22. Option 3 (Figure 4.21 (c))



(a) $\binom{20}{3}$

(b) $\binom{50}{15}$

FIGURE 4.23. Population distribution generated by variation of Algorithm 8

CHAPTER 5

SYNTHETIC CITIES

The objective of generating synthetic geographic regions with corresponding synthetic populations consists of two main tasks: the generation of the synthetic geography G with its sub-entities and the generation of a synthetic population P. The synthetic region and the synthetic population can be generated independently, necessitating the definition of a mapping function $M(P, G)$, which assigns the synthetic population to census blocks of the synthetic geography. While many approaches to generate synthetic populations have been discussed in the literature (see Chapter 2.3), the task of generating synthetic geographic regions as means to obfuscate real data is a novel approach. This differs from the generation of geographies for virtual or virtualized worlds.

A tool capable of generating synthetic counties is of special interest for this research. It allows for the comparison of scenarios for different geographies with the same or different demographic properties, as well as the comparison of the resulting changes in the plan analysis due to changes of the demographic properties on a specific geographic area. This can be helpful when using census data that is already outdated, especially for areas exhibiting drastic changes in population size and the demographic distribution. For this dissertation, Census 2000 data sets have been used. The synthetic counties consist of census entities and a road network. The system is extensible such that optional features, such as landmarks, can be added in the future. The individual geographic sub-entities of the county is be described in the following sections.

5.1. Generation of Synthetic Regions: Census Entities

A county consists of a hierarchical system of census entities. It is divided into census tracts, which are further sub-divided into block groups. The smallest entity to be considered

are census blocks. The automatic generation of synthetic counties along with its census entities necessitates the definition of a set of rules based on real world characteristics. Properties and regulations of each census entity (e.g. census block, census tract) are described in the census guide pertinent to the U.S. Census [96]. These properties are summarized in pipe-delineated files and include minimum and maximum number of individual per entity, as well as restrictions in size, if available. From these properties, only approximate rules can be derived, such as minimum and maximum population sizes for census tracts and block groups. This allows for the generation of a synthetic representation of a real county.

Given a population size and the dimensions of a rectangular shaped county, census blocks in the shape of rectangles are created. This is accomplished by partitioning the initial rectangular county recursively into two equally sized sub-areas, assigning each of the new rectangles a portion of the population. The population sizes of the newly generated rectangles depend on a random number. Nevertheless, only valid population distributions are generated respecting minimum and maximum population sizes for census tracts, and block groups, as well as area restrictions for census blocks. These rules were deduced by compiling a list of constraints for the individual entities from the census documentation mentioned above. This simple approach tends to assign a value close the lower bound of permissible population sizes of the individual entities to the rectangles. Hence, a population distribution mimicking the distribution of a real county is incorporated in order to yield a more realistic distribution. This allows for the creation of counties that have properties of a chosen real county. Alternatively, the optimal sizes specified in the census for some of the geographic entities can be used to yield population sizes distributed across the permissible ranges. An example is shown in Figure 1(b). Figure 5.2 shows the population densities of the individual areas, normalized over the area of the census blocks. Patterns emulating metropolitan and rural areas are observable. The granularity shown in the figures is at a census block level. However, during the process of generating the county, census tracts are created first, which then are sub-divided into block groups, which are further partitioned into census blocks. This yields

```
if e.parent.type == county then
    /* generate census tracts                                          */
    repeat
        Draw random number r with min.tract_size ≤ r ≤ e.size ;
        Partition e into e1 of size r and e2 of size e.size − r ;
    until e1 and e2 suffice all additional conditions for census tracts;
else
    if e.parent.type == tract then
        /* generate block groups                                       */
        if e1 and/or e2 are valid census tracts then
            entity.type = tract ;
        end
        PARTITION(e1) and/or PARTITION(e2) ;
        repeat
            Draw random number r with min.group_size ≤ r ≤ e.size ;
            Partition e into e1 of size r and e2 of size e.size − r ;
        until e1 and e2 suffice all additional conditions for block groups;
        if e1 and/or e2 are valid block groups then
            entity.type = group ;
        end
        PARTITION(e1) and/or PARTITION(e2) ;
    else
        /* generate census blocks                                      */
        repeat
            Draw random number r with r ≤ e.size ;
            Partition e into e1 of size r and e2 of size e.size − r ;
        until e1 and e2 suffice all additional conditions for census blocks;
        if e1 and/or e2 are valid census blocks then
            entity.type = block ;
        else
            PARTITION(e1) and/or PARTITION(e2) ;
        end
    end
end
```

ALGORITHM 9: PARTITION(entity e)

a data-set compatible with real world data. The top-down approach was selected over the bottom-up approach starting with the census blocks and grouping them into block groups and census tracts due to the fact that it is an easier problem to subdivide a given geographic

space according to a set of constraints, than finding subsets covering all geographic entities satisfying these constraints. The pseudo-code in Algorithm 9 summarizes the approach described above.

For the sake of simplicity, additional constraints and overlaid population distributions are not included in the pseudo-code and summarized as additional conditions. By making use of GIS software, it is observable that census blocks of size 0 are permissible. Hence, for the census block level, no lower bound on the population size is needed, therefore census blocks are merely restricted based on a minimum size in terms of area.

Emergency planning also takes into account a variety of demographic features, which must be replicated for the synthetic county as well. Therefore, in future research a set of features can be selected to train the system on the original demographic data. Similarly to work presented in [9], ratios of the features can be build. Using this, a synthetic population that mimics the population of the original county is obtainable. In a final step, the population needs to be mapped into the synthetic geographic region.

5.2. Generation of Synthetic Road Networks

The road network consist of two parts: the road structure itself, and traffic counts assigned to the individual roads. The computation of the latter is based on the type of a particular road and its speed limit and can be trained using real world traffic counts and is not part of this dissertation. Superimposing a road network is achieved by using a clustering approach. Clusters of high population density are likely to be inter-connected by major roads, such as highways. Zooming into the individual clusters allows for the recursive repetition of this approach. Nevertheless, with each recursion level, the road size and capacity decreases, generating major roads that are not highways, and ultimately minor roads. During the last recursion level small roads mostly used for local neighborhood traffic can be generated. For some real geographic regions road infrastructure information for minor roads may not be available and can be considered negligible as they provide no contribution to traffic movements on a larger spatial scale.

The generation of a synthetic county must include a corresponding transportation infrastructure in the form of a road network that should strongly resemble its real counterpart. Hence, the underlying algorithm for the generation of a synthetic road network must be based on a set of rules and observations regarding the characteristics of existing transportation infrastructure. In general, roads can be classified into different levels based on capacity and distance. For the generation of a synthetic road network the following classes can be considered:

- Level 1 roads are major arteries with high traffic capacities that connect distant locations, including interstate and state highways.
- Level 2 roads are medium-sized roads that include U.S Routes not being classified as level 1 roads. This class does not include roads that only serve neighborhood traffic.
- Level 3 roads are smaller local roads with neighborhood traffic and are of low capacity.

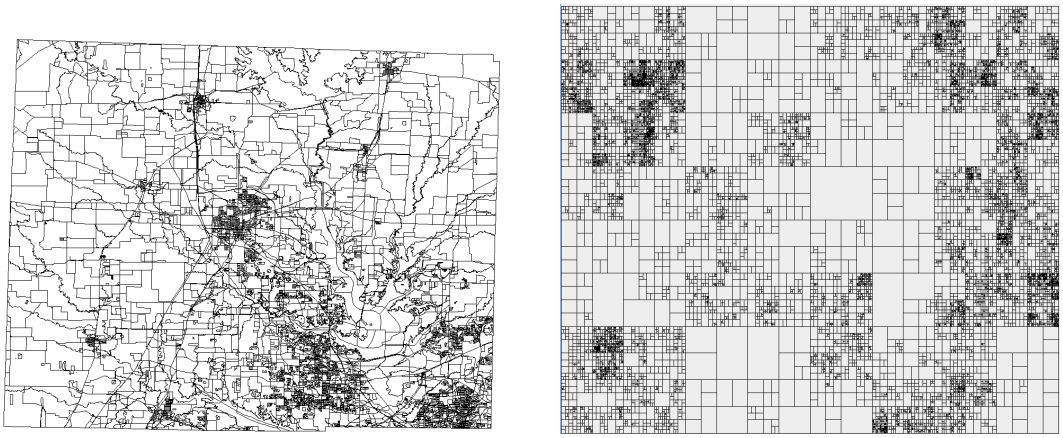
```
for i = 1 to numRoads do
    Choose a starting point within the geographic extension defined by x and y;
    Choose roadL ∈ [minRoadL, maxRoadL] ;
    while roadL > 0 do
        Choose segL ∈ [0, maxSegL];
        if segL > roadL then
            segL = roadL;
        end
        Choose direction of the new point;
        Calculate ending location of segment;
        roadL− = segL;
    end
end
```

ALGORITHM 10: Generate Roads

The number and classification of roads can be further refined based on the application for which the synthetic road network is generated. Input parameters for the synthetic road generation are the spatial extensions of the county x and y, the number of roads to be

generated $numRds$, the minimum length of a road $minRoadL$, the maximum length of a road $maxRoadL$, and the maximum segment length $maxSegL$. For each road, the road length $roadL$ and an initial point p in the geographic area are chosen. From p, the road is *grown* segment by segment until the road has reached the length of $roadL$. This approach is summarized in Algorithm 10. Lower level roads can be constructed analogously.

The incremental generation of synthetic road infrastructures is subject to constraints that depend on the corresponding level of the road to be generated. Such constraints include the angle between road segments, maximum segment length, and maximum and minimum road lengths. For the purpose of this dissertation, parameters for level 1 roads have been determined experimentally. In future work, such parameters can be derived by using real road networks as training sets.



(a) Denton County resolved to the census block level

(b) Synthetic county with dimensions and population sizes comparable to those of Denton

FIGURE 5.1. Comparison of Denton County and a synthetic county with similar dimensions and population sizes

5.3. Additional Components

Optional additional components are excluded from the research presented in this dissertation, but can be added in future research. These include physical landmarks (e.g. mountains, rivers, lakes), human made landmarks (e.g. stadiums, hospitals, malls), and land values. The

addition of landmarks necessitates the generation of terrain, such that mountains and other physical properties can be defined.

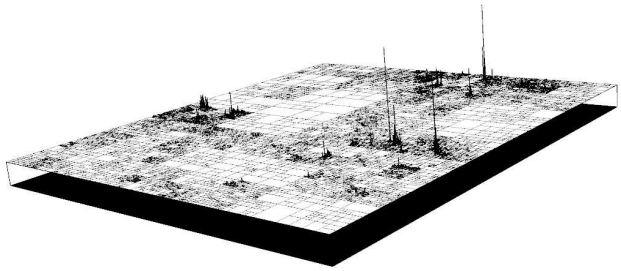


FIGURE 5.2. Population densities of a synthetic county

5.4. Results

Using the approach described above, methodology to generate a rectangular county has been implemented in OpenGL. The input parameters are the dimensions of the county and the population size. Figure 5.1 shows Denton County at the census block level, as well as a synthetic counterpart, also shown at census block granularity. Both representations exhibit patterns of rural and urban areas. There is no unique synthetic counterpart to a geographic region, as the generation makes use of statistical methods.

Figure 5.2 illustrates the population density of the synthetic county expressed by the z-coordinate. The population of the individual census blocks is averaged across the census block area. A basic road network has been constructed for the synthetic county, which can be overlaid onto the synthetic census blocks as shown in Figure 5.3.

Five PODs have been placed into the synthetic geographic region. This allows for the execution of RE-PLAN's response plan analysis module. Figure 5.4 (a) illustrates the resulting catchment areas. The final analysis result is shown in Figure 5.4 (b).

The following figures illustrate the results of applying Algorithm 6 and Algorithm 7 to the synthetic geographic region. The results of Algorithm 6 for $k = 5$ PODs are shown in Figure 5.5. 20 candidate PODs have been selected within the synthetic county and Algorithm 7 has been applied to select five PODs. The candidate PODs, as well as the resulting partitioning are shown in Figure 5.6.

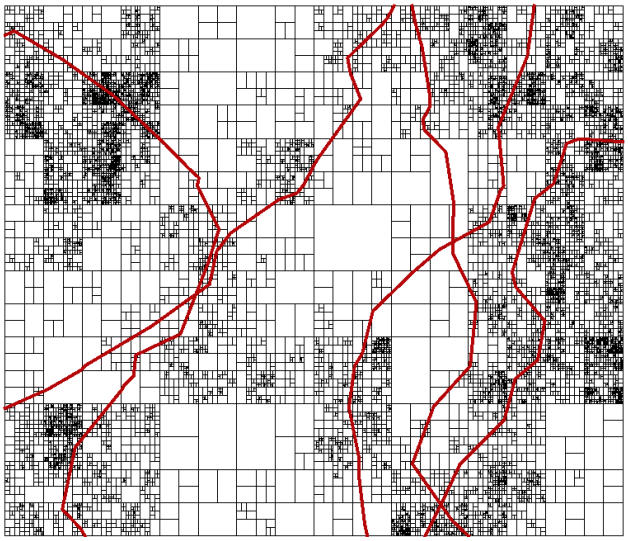
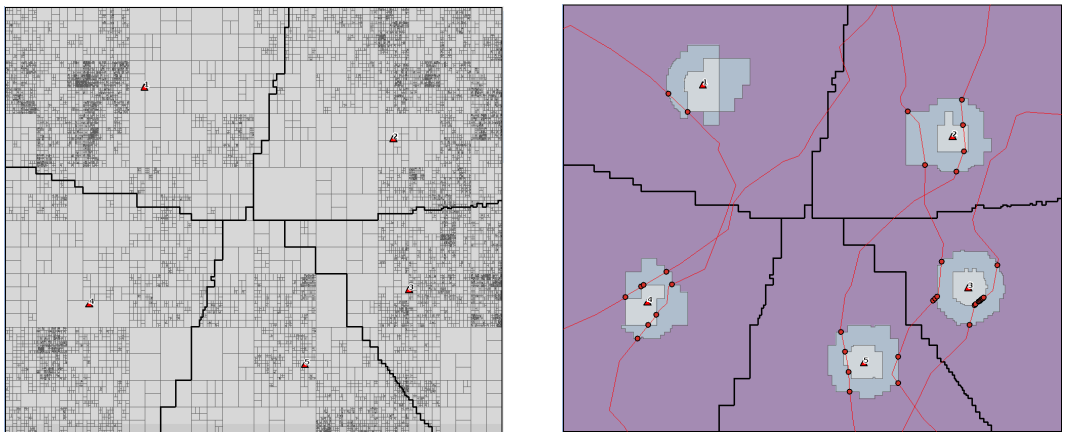


FIGURE 5.3. Synthetic county with synthetic road network



(a) Catchment areas (b) Response plan analysis

FIGURE 5.4. RE-PLAN execution on a synthetic county

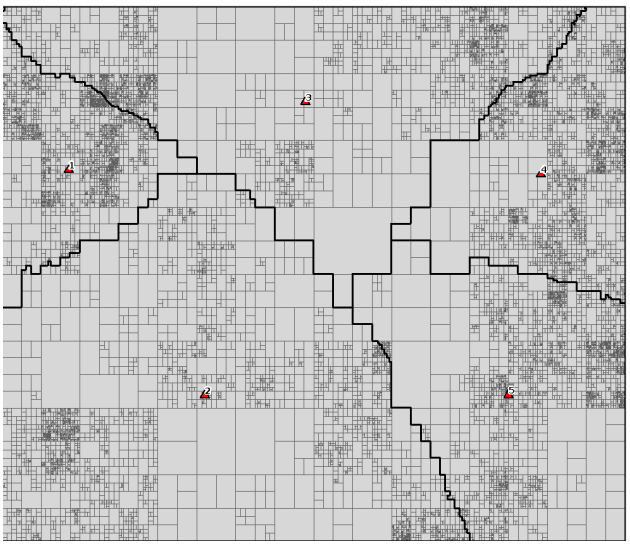
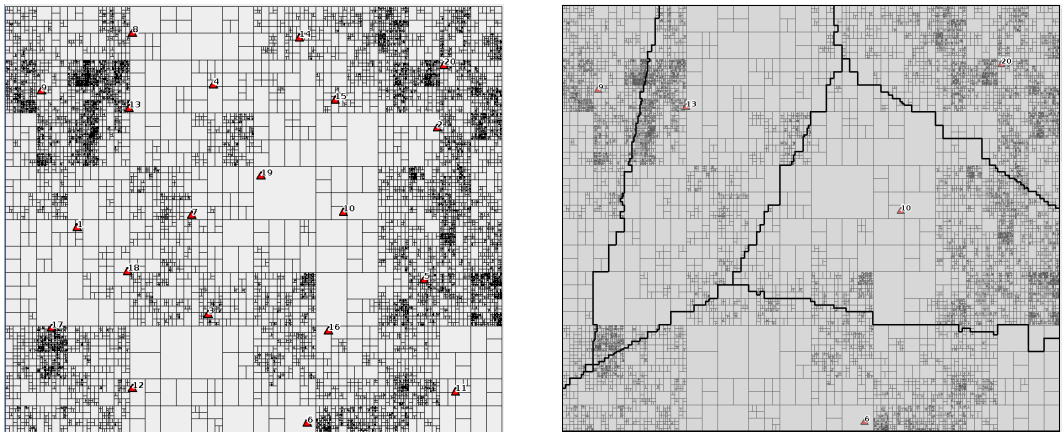


FIGURE 5.5. Continuous partitioning (Algorithm 6) applied to the synthetic county



(a) Catchment areas

(b) Response plan analysis

FIGURE 5.6. Execution of Algorithm 7 on a synthetic county

CHAPTER 6

SUMMARY AND CONCLUSION

Threats of bio-emergencies caused by pandemics or acts of bio-terrorism have led to the design and implementation of specific response plans, whose feasibility is critical for adequate mitigation. As described, a response plan is deemed feasible, if it meets mandated time-lines. Hence, methodology and tools that can assess the feasibility of a response plan in a given geographic region are required. These tools must permit the detailed analysis of such response scenarios while incorporating dynamics in demographics and geographical entities. In this dissertation, the analysis of existing response plans, as well as the determination of a feasible point of dispensing (POD) placement have been presented. Further, confidentiality concerns have been addressed by providing methodology for the generation of synthetic geographic regions.

The research conducted for this dissertation ports computational tools into the context of public health, which allows for faster planning and advanced analysis strategies. Such computational tools can be used to design and implement evacuation and displacement scenarios. Further, they provide public health experts with means to dynamically allocate resources for other natural disasters. One such a tool is REsponse PLan ANalyzer (RE-PLAN), which allows for the analysis and optimization for bioemergency response plans. Since RE-PLAN is being deployed by Tarrant County Public Health, it is tailored to the needs of disaster preparedness coordinators. Further, regular meetings with public health experts allow for feedback to be incorporated during all stages of the design and developmental process. On the county level, it facilitates the comparison of different response scenarios for a particular type of bio-emergency, as well as the comparison of different bio-emergency events. On a larger scale, it serves as a tool for cross-county collaboration, as response plans based on RE-PLAN's analysis results can easily be combined. Since computational

tools facilitate a better understanding of the socioeconomic and cultural distribution of the population, they help to unify the access to public health services. Visualizing placement of public health resources as well as resulting bottlenecks simplifies the response scenario design process.

6.1. Analysis of Existing Response Plans

As part of this research, the GIS-based computational framework RE-PLAN has been developed. The core components of its architecture are a GIS-enabled database, a computational engine, and a graphical user interface. RE-PLAN analyzes response plans by sub-dividing the geographic region into catchment areas, each of which is assigned to a specific POD. For each of the catchment areas, rings of proximity are calculated and intersected with the road network resulting in crossing points that can be used to estimate traffic conditions during plan execution. The results of specific steps of the response plan analysis via RE-PLAN were demonstrated on Denton County and a synthetic geographic region with similar dimensions and population size. Experiments on real data were conducted for Denton County. For this purpose, census blocks, census data, road infrastructure and traffic information were used. The execution of response plan analysis has been demonstrated on Denton County for the placement of different sets of PODs.

The utility of RE-PLAN has become evident during the recent H1N1 vaccination effort, when the effectiveness of existing response plans has exposed limitations and forced their redesign. Public health coordinators indicated that results obtained from RE-PLAN's POD-Analyzer are realistic and that poorly chosen or designed PODs may not be able to support the expected demand in an extensive bio-emergency.

6.2. Feasible POD Placement

The analysis of an existing response plan may suggest that it will not meet government mandated time-frames, hence, posing the question whether there exists a feasible solution with the given resources. In this dissertation, three approaches to determine a POD placement based on available resources are presented. The feasibility of the generated response

scenarios is examined based on their population distribution as well as traffic conditions at crossing points.

While all of the approaches attempt to achieve a fair distribution of the population across the catchment areas, they are based on different areas of location science. The continuous partitioning method can be categorized as an approach from continuous location science, since the POD locations can be placed anywhere in the geographic space. It creates sub-regions within the geographic region for each of the PODs, while maintaining an uniform population distribution. The maximum deviation from the arithmetic average of the population is bounded by the maximum census block size. The second method uses a reverse hill-climbing approach, which first makes use of all PODs along with their corresponding catchment areas. Two of these catchment areas are then merged based on an optimality criterion repeatedly until the desired number of catchment areas remains. Further, a hybrid approach, utilizing methodology of the continuous partitioning method has been proposed to select a subset of a set of given PODs. These methods have been demonstrated on Denton County data.

6.3. Synthetic Geographic Regions

As contingency plans for bio-emergencies often contain sensitive information, methodology for designing and analyzing such plans may not be easily shared or utilized for training purposes. This problem can be addressed by obfuscating real regional information, which can be achieved through the use of synthesized geographic entities. This dissertation has presented two algorithms: one for the generation of synthetic counties and one for the generation of a corresponding road network. Properties of a specific realistic geographic region can be emulated by choosing appropriate parameters. These algorithms have been shown to result in synthetic population distributions and road infrastructures that closely resemble reality. In this dissertation novel methodology for the analysis of bio-emergencies, response plan optimization, as well as the generation of synthetic geographic regions has been proposed.

6.4. Contribution

Part of the contribution of this dissertation is the design of methodology for the analysis of bio-emergency contingency plans. This includes methodology for estimating catchment areas, ring sectors, and the estimation of traffic flow via crossing points. Further, algorithms for determining a feasible placement of PODs have been designed: one algorithm pertinent to continuous location science and two algorithms in the field of discrete location science. For the continuous method, the existence of a maximum bound for the deviation of the optimal population size has been proven. For the generation of synthetic geographic regions methodology to generate census blocks and road networks has been developed. Characteristics to obtain a synthetic region with characteristics of a specific existing region can be specified by selecting appropriate values for the parameters of the algorithms.

6.5. Future Work

Work presented here has answered some of the research questions induced by public health problems and simultaneously created a new range of future research opportunities. RE-PLAN analyzes response scenario within minutes, which is faster than manual analyses of such contingency plans. Nevertheless, additional research could improve the execution time, allowing public health experts to obtain results even faster. This facilitates an improved planning process, as modifications to scenarios could then easily be discussed and evaluated during disaster preparedness meetings. Further, improvements of the implemented optimization methods, as well as the design of new optimization techniques can provide valuable additions to RE-PLAN and additional aid to public health experts in the planning process.

The RE-PLAN methodology itself can be further developed in the future. The crossing points calculation via ring sectors could be substituted by applying geometric optimization techniques. This provides additional flexibility, since clusters of crossing points are not limited to specific ring sectors defined by the angles used to sub-divide the rings. Further, the different distance metrics described theoretically can be added to the system. The minimum road distance is a basis for using the minimum travel time distance. All different distance

metrics can then be combined in the mixed distance metric, which allows for assigning weights to the individual distance components.

While some geographic regions have a sparse public transportation system, the population of other regions, such as New York or Chicago, strongly depends on public transportation. To analyze response scenarios for such regions, methodology for public transit must be added to the RE-PLAN framework. Since social aspects may influence the accessibility of public health resources, certain populations might be undeserved. Including additional census information, such as race, age and income, into RE-PLAN can help to identify areas in need of alternative strategies to provide these parts of the population with adequate emergency responses.

As for the generation of synthetic geographic regions additional features can be generated to provide more realistic synthetic representations of real geographic regions. The system has to be trained on real data in order to generate different levels of roads with properties corresponding to their real counterparts. For these roads, properties, such as traffic counts and road capacities, have to determined. A realistic synthetic population trained on real data can be generated, which by utilizing a mapping function is then assigned to the individual census blocks of the synthetic geographic region. Since methods for the generation of synthetic geographic regions are intended to mimic characteristics of real regions, a set of quality functions has to be defined, measuring quality of a generated solution.

Since RE-PLAN provides an extensible framework, other component could be integrated in the future. This includes the planning evacuation scenarios, as well as other types of Public Health Planning.

BIBLIOGRAPHY

[1] B. Adenso-Diaz and F. Rodriguez. A simple search heuristic for the MCLP: Application to the location of ambulance bases in a rural region. *Omega*, 25(2):181–187, Apr 1997.

[2] M. Agar and S. Salhi. Lagrangean heuristics applied to a variety of large capacitated plant location problems. *Journal of the Operational Research Society*, 49(10):1072–1084, Ooct 1998.

[3] P. Agarwal and M. Sharir. Efficient algorithms for geometric optimization. *ACM Computing Surveys*, 30(4):412–458, Dec 1998.

[4] B. M. Altevogt, A. M. Pope, M. N. Hill, and K. I. Shine. *Research Priorities in Emergency Preparedness and Response for Public Health Systems: A Letter Report*. National Academies Press, Washington, DC, 2008.

[5] P. Avella and A. Sforza. Logical reduction tests for the p–problem. *Annals of Operations Research*, 86(0):105–115, 1999.

[6] I. Averbakh and O. Berman. Location problems with grouped structure of demand: Complexity and algorithms. *Networks*, 31(2):81–92, Mar 1998.

[7] J. R. Beaumont. *Spatial Analysis and Location-Allocation Models*, chapter Location-Allocation Models and Central Place Theory, pages 21–53. Van Nostrand Reinhold, New York, 1987.

[8] R. Beckman, K. Baggerly, and M. McKay. Creating synthetic baseline populations. *Transportation Research Part A*, 30(6):415–429, 1996.

[9] R. J. Beckman, K. A. Baggerly, and M. D. McKay. Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, 30(6):415–429, November 1996.

[10] R. Bellman. On a routing problem. *Quarterly of Applied Mathematics*, 16(1):87–90, 1958.

[11] O. Berman, J. Kalcsics, D. Krass, and S. Nickel. The ordered gradual covering location problem on a network. *Discrete Applied Mathematics*, 157(18):3689–3707, 2009.

[12] O. Berman and D. Krass. The generalized maximal covering location problem. *Computers and Operations Research*, 29(6):563–581, May 2002.

[13] R. Bhatia, S. Guha, S. Khuller, and Y. Sussmann. Facility location with dynamic distance functions. *Journal of Combinatorial Optimization*, 2(3):199–217, 1998.

[14] M. Birkin, A. Turner, and B. Wu. A synthetic demographic model of the uk population: Methods, progress and problems. *Second International Conference on e-Social Science*, pages 692–697, 2006.

[15] A. Borning and P. Waddell. Urbansim: interaction and participation in integrated urban land use, transportation, and environmental modeling. *Proceedings of the 2006 international conference on Digital government research*, pages 133–134, 2006.

[16] J. Brimberg and S. Salhi. A continuous location-allocation problem with zone-dependent fixed cost. *Annals of Operations Research*, 136(1):99–115, 2005.

[17] Bureau of the Census. American factfinder, 2000.

[18] R. Canavate-Bernal, M. Landete-Ruiz, and A. Marin-Perez. On the resolution of the single product capacitated machine siting problem. *The Journal of the Operational Reserach Society*, 51(8):982–992, Aug 2000.

[19] L. Cánovas, M. Landete, and A. Marn. On the facets of the simple plant location packing polytope. *Discrete Applied Mathematics*, 124(1-3):27–53, Dec 2002.

[20] C. Caruso, A. Colorni, and L. Aloi. Dominant, an algorithm for the p-center problem. *European Journal of Operational Research*, 149(1):53–64, 2003.

[21] P. A. Casillas. Data aggregation and the p-median probelm in continuous space. In *Spatial Analysis and Location-Allocation Models*. Van Nostrand Reinhold, New York, 1987.

[22] Census Bureau. Fact Finder, November 2008. [online] `http://www.census.gov/`.

[23] Centers for Disease Control and Prevention. Smallpox response plan and guidelines (version 3.0), Nov. 2002.

[24] Centers for Disease Control and Prevention. Transcript of "smallpox vaccine logistics: Distribution, storage, and security". pages 1–5, Aug 2002.

[25] Centers for Disease Control and Prevention. Smallpox response plan: Executive summary. Aug 2003.

[26] Centers for Disease Control and Prevention. Mass antibiotic dispensing: Streamlining pod design and operations, Apr. 2005.

[27] Centers for Disease Control and Prevention. Key facts about the cities readiness initiative (CRI), Apr. 2008.

[28] Y.-C. Chiu, H. Zheng, J. Villalobos, and B. Gautam. Modeling no-notice mass evacuation using a dynamic traffic flow optimization model. *IIE Transactions*, 39(1):83–94, 2007.

[29] F. Chiyoshi and R. Galvao. A statistical analysis of simulated annealing applied to the p-median problem. *Annals of Operations Research*, 96(1-4):61–74, 2000.

[30] R. L. Church and D. J. Eaton. Hierarchical location analysis using covering objectives. In *Spatial Analysis and Location-Allocation Models*. Van Nostrand Reinhold, New York, 1987.

[31] C. Cirillo, E. Cornelis, and P. L. Toint. Model of weekly working participation for a belgian synthetic population. *European Transport Conference*, Oct 2007.

[32] P. Clarke and B. Wheaton. Addressing data sparseness in contextual population research: Using cluster analysis to create synthetic neighborhoods. *Sociological Methods REsearch*, 35(3):311–351, Feb 2007.

[33] M. Coombes. Defining locality boundaries with synthetic data. *Environment and Planning A*, 32(8):1499–1518, 2000.

[34] K. Cordasco, S. Asch, and J. Golden. Disaster planning and risk communication with vulnerable communities: lessons from hurricane katrina. *American Journal of Public Health*, 97(S1):S109–S115, 2007.

[35] T. Cormen, C. Leiserson, and R. Rivest. The MIT Press, Cambridge, MA, 1990.

[36] S. Cutter and C. Emrich. Moral hazard, social catastrophe: The changing face of vulnerability along the hurricane coasts. *The ANNALS of the American Academy of Political and Social Science*, 604(1):102–112, Mar 2006.

[37] E. Deza and M. M. Deza. *Dictionary of Distances*. Elsevier, 2006.

[38] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[39] Z. Drezner and E. Erkut. An efficient genetic algorithm for the p-median problem. *Annals of Operations Research*, Jan 2003.

[40] G. Duncan, V. D. Wolf, T. Jabine, and M. Straf. Report of the panel on confidentiality and data access. *Journal of Official Statistics*, 9:271–271, 1993.

[41] Environmental Systems Research Institute, Inc. Esri shapefile technical description white paper, 1998.

[42] S. Espié, E. Follin, G. Gallée, D. Ganieux, and J. Hespert. Automatic road networks generation dedicated to night-time driving simulation. *DSC North America 2003 Proceedings*, 2003.

[43] ESRI. Mobile GIS speeds disaster relief. *ESRI*, Mar 2005.

[44] ESRI. Disaster preparedness exercise uses GIS. *ESRI ArcNews*, 28(1), Jul 2006.

[45] S. Eubank. Scalable, efficient epidemiological simulation. In *Proc. of the* 17^{th} *Annual ACM Symposium on Applied Computing (SAC'02)*, Madrid, Spain, 2002.

[46] J. Fernandez, P. Fernandez, and B. Pelegrn. A continuous location model for siting a non-noxious undesirable facility within a geographical region. *European Journal of Operational Research*, 121(2):259–274, Mar 2000.

[47] S. Fu and G. Milne. Epidemic modelling using cellular automata. In *Proc. of the Australian Conference on Artificial Life*, 2003.

[48] J. Gerberding, J. Hughes, and J. Koplan. Bioterrorism preparedness and response: Clinicians and public health agencies as essential partners. *Journal of the American Medical Association*, 287(7), Jan 2002.

[49] G. Ghiani, F. Guerriero, and R. Musmanno. The capacitated plant location problem with multiple facilities in the same site. *Computers and Operations Research*, pages 1903–1912, 2002.

[50] A. Ghosh and G. Rushton. Progress in location allocation modeling. In *Spatial Analysis and Location-Allocation Models*. Van Nostrand Reinhold, New York, 1987.

[51] D. Glover and J. Simon. The effect of population density on infrastructure: The case of road building. *Economic Development and Cultural Change*, 23(3), Apr 1975.

[52] P. Hansen, N. Mladenović, and D. Perez-Britos. Variable neighborhood decomposition search. *Journal of Heuristics*, 7(4), Jul 2001.

[53] S. Harewood. Emergency ambulance deployment in barbados: a multi-objective approach. *Journal of the Operational Research Society*, 53(2), Feb 2002.

[54] R. Haynes, A. P. Jones, V. Sauerzapf, and H. Zhao. Validation of travel times to hospital estimated by GIS. *International Journal of Health Geographics*, 5(40), September 2006.

[55] T. Jabine. Procedures for restricted data access. *Journal of Officail Statistics*, 9(2):537–537, 1993.

[56] T. Kanade, P. Rander, and P. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia*, 4(1):34–47, 1997.

[57] E. H. Kaplan, D. L. Craft, and L. M. Wein. Emergency response to a smallpox attack: the case for mass vaccination. *Proceedings of the National Academy of Sciences*, 99(16):10935–10940, Aug 2002.

[58] S. Khuller and Y. Sussmann. The capacitated k-center problem. *SIAM Journal on Discrete Mathematics*, 13(3):403, 2000.

[59] S. Kongsomsaksakul, C. Yang, and A. Chen. Shelter location-allocation model for flood evacuation planning. *Journal of the Eastern Asia Society for Transportation Studies*, 6:4237–4252, 2005.

[60] D. Lambert. Measures of disclosure risk and harm. *Journal of Official Statistics*, 9:313–331, 1993.

[61] G. Leduc. Road traffic data: Collection methods and applications. *Institute for Prospective Technological Studies, European Commission*, 2008.

[62] E. K. Lee, C.-H. Chen, F. Pietz, and B. Benecke. Modeling and optimizing the public-health infrastructure for emergency response. *Interfaces*, 39(5):476–490, Sep 2009.

[63] E. K. Lee, S. Maheshwary, J. Mason, and W. Glisson. Large-scale dispensing for emergency response to bioterrorism and infectious-disease outbreak. *Interfaces*, 36(6):591–607, Nov 2006.

[64] X. Li, J. Han, J.-G. Lee, and H. Gonzalez. Traffic density-based discovery of hot routes in road networks. In *Proceedings of the 10th international conference on Advances in spatial and temporal databases*, SSTD'07, pages 441–459, Berlin, Heidelberg, 2007. Springer-Verlag.

[65] V. Marianov and D. Serra. Location–allocation of multiple-server service centers with constrained queues or waiting times. *Annals of Operations Research*, 111(1):35–50, 2002.

[66] J. McDade and D. Franz. Bioterrorism as a public health threat. *Emerging Infectious Diseases*, 4(3):488–492, Jan 1998.

[67] S. L. Mclafferty. Gis and health care. *Annuual Reviews of Public Health*, 24(1):25–42, Jan 2003.

[68] T. Melhusin, M. Blake, and S. Day. An evaluation of synthetic household populations for census collection districts created using spatial microsimulation techniques. *26th*

Australia & New Zealand Regional Science Association, International Annual Conference, Gold Coast, Queensland, Australia., 2002.

[69] G. Miller, S. Randolph, and J. E. Patterson. Responding to bioterrorist smallpox in san antonio. *Interfaces*, 36(6):580–590, Nov 2006.

[70] North Central Texas Council of Governments, 2010. [online] `http://www.nctcog.dst.tx.us/`.

[71] L. Neergaard. Mailmen might deliver meds in next anthrax attack. *The Washington Post*, October 2008.

[72] I. Ntoutsi, N. Mitsou, and G. Marketos. Traffic mining in a road-network: How does the traffic flow? *International Journal of Business Intelligence and Data Mining*, 3(1):82–98, 2008.

[73] Open Geospatial Consortium, Inc. OGC website, 1010.

[74] Open Source Geospatial Foundation. OSG website, 2010.

[75] Orange County Health Department. Cities readiness initiative. pages 1–2, Aug 2009.

[76] S. Pallottino and M. Scutella. Equilibrium and advanced transportation modelling. 1998.

[77] X. Pan. *Computational Modeling of Human and Social Behaviors for Emergency Egress Analysis*. PhD dissertation, Stanford University, Department of Civil and Environmental Engineering, June 2006.

[78] J. Papin and P. Hue. French military applications of virtual reality. *NATO*, 2000.

[79] B. Pelegrin and L. Canovas. A new assignment rule to improve seed points algorithms for the continuous k-center problem. *European Journal of Operational Research*, 104(2):366–374, 1998.

[80] S. Rahman and D. Smith. Deployment of rural health facilities in a developing country. *The Journal of the OPperational Research Society*, 50(9):892–902, 1999.

[81] J. Rawls. A theory of justice. *Harvard University Press, Cambridge, MA*, 1971.

[82] F. Ritchie. Secure access to confidential microdata: four years of the virtual microdata laboratory. *Economic and Labour Market Review*, 2:29–34, Jan 2008.

[83] G. Rushton. Computer programs for location-allocation problems. page 330, 1975.

[84] G. Rushton. Selecting the objective function in location allocation analyses. In *Spatial Analysis and Location-Allocation Models*. Van Nostrand Reinhold, New York, 1987.

[85] G. B. Saathoff and J. B. Noftsinger Jr. Mass evacuation and our nation's highways. *Carolina Planning*, 30(1):33–39, Winter 2005.

[86] S. Salhi and M. Gamal. A genetic algorithm based approach for the uncapacitated continuous location–allocation problem. *Annals of Operations Research*, 123(1-4):203–222, 2003.

[87] T. Schneider and A. R. Mikler. Re-plan: A computational tool for response plan analysis. *International Journal of Functional Informatics and Personalised Medicine*, 3:103–121, 2010.

[88] T. Schneider, A. R. Mikler, and M. O'Neill. Computational tools for evaluating bioemergency contingency plans. In *Proceedings of the 2009 International Conference on Disaster Management*, New Forest, England, September 2009.

[89] U. Schoening. *Algorithmik*. Spektrum, 2001.

[90] H. Sherali and T. Park. Discrete equal-capacity p-median problem. *Naval Research Logistics*, 47(2):166–183, 2000.

[91] K. Shoaf and S. Rottman. The role of public health in disaster preparedness, mitigation, response, and recovery. *Prehospital and Disaster Medicine*, 15(4):144–146, 2000.

[92] L. Svec, S. Burden, and A. Dilley. Applying voronoi diagrams to the redistricting problem. *UMAP Journal*, 28(3):313–330, 2007.

[93] J. Tang. Evaluating the relationship between urban road pattern and population using fractal geometry. *UCGIS*, 2003.

[94] V. K. Tewari and S. Jena. High school location decision making in rural india and location-allocation models. In *Spatial Analysis and Location-Allocation Models*. Van Nostrand Reinhold, New York, 1987.

[95] I. Turton. *Open Source Approaches in Spatial Data Handling*, chapter Geo Tools, pages 153–169. Springer, 2008.

[96] U.S. Census Bureau. Appendix A: Census 2000 geographic terms and concepts, 2000.

[97] X. Wang and K. Kockelman. Forecasting network data: Spatial interpolation of traffic counts using texas data. *Journal of the Transportation Research Board*, 2105(13):100–108, 2009.

[98] S. Wassermann and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York, NY, 1994.

[99] A. Weber. *Alfred Weber's theory of the location of industries (Materials for the study of business)*. The University of Chicago Press.

[100] H. Wei, A. Murray, and N. Xiao. Solving the continuous space p-centre problem: planning application issues. *IMA Journal of Management Mathematics*, 17(4):413–425, Jan 2006.

[101] L. M. Wein, D. L. Craft, and E. H. Kaplan. Emergency response to an anthrax attack. *Proc Natl Acad Sci USA*, 100(7):4346–51, Apr 2003.

[102] L. Willumsen. Simplified transport models based on traffic counts. *Transportation*, 10(3):257–278, 1981.

[103] P. D. Wright, M. J. Liberatore, and R. L. Nydick. A survey of operations research models and applications in homeland security. *Interfaces*, 36(6):514–529, Nov 2006.

[104] J. Wu and J. M. Abowd. How to evaluate the usefulness and quality of synthetic data. *Census Advisory Committee of Professional Associations*, 2006.

[105] D. Young. CDC rolls out nerve-agent antidote program, Sept. 2004.

[106] F. Zhan and C. Noon. Shortest path algorithms: an evaluation using real road networks. *Transportation Science*, 32:65–73, Jan 1998.