

NISTIR 8039

Nationwide Public Safety Broadband Network Deployment: Network Parameter Sensitivity Analysis

Richard Rouil
Antonio Izquierdo
Camillo Gentile
David Griffith
Nada Golmie

This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8039>

NISTIR 8039

Nationwide Public Safety Broadband Network Deployment: Network Parameter Sensitivity Analysis

Richard Rouil
Antonio Izquierdo
Camillo Gentile
David Griffith
Nada Golmie
*Wireless Networks Division
Communications Technology Laboratory*

This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8039>

February 2015



U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Willie May, Acting Under Secretary of Commerce for Standards and Technology and Acting Director

Acknowledgments

The work presented in this document was sponsored by the Department of Homeland Security, Office for Interoperability and Compatibility (OIC).

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Abstract

The legislation included in the middle class tax relief and job creation act of 2012 [1] established the First Responder Network Authority (FirstNet) for the purpose of deploying and running a nationwide Long Term Evolution (LTE) network for Public Safety called the National Public Safety Broadband Network (NPSBN). This network has unique characteristics that distinguish it from other commercial networks, such as user applications, coverage, and reliability. In this document, we present the modeling method developed by the National Institute of Standards and Technology (NIST) to evaluate the performance of LTE networks for Public Safety. By using sampling techniques and a flexible design, we are able to quickly investigate multiple assumptions and report their impact at a national scale. The results presented in this document focus on the impact of coverage objectives, reliability, and high power user equipment (UE).

Table of Contents

1.	Introduction.....	1
2.	Radio Frequency Network Planning	1
2.1.	Inputs and Assumptions	2
2.1.1.	Geo data	2
2.1.2.	Propagation Models	3
2.1.3.	User Distribution	8
2.1.4.	Traffic Model	16
2.1.5.	Network Configuration	17
2.2.	Performance Analysis	17
2.2.1.	Site Placement/Site Selection.....	17
2.2.2.	Monte Carlo Simulations	19
2.2.3.	Network Analysis	20
2.2.4.	Sample Results.....	21
3.	Tackling Nationwide Scale Modeling.....	22
3.1.	Partitioning the United States Area	23
3.2.	Classification of the Analysis Areas	26
3.2.1.	Input Characteristics.....	27
3.2.2.	Subdivisions	28
3.2.3.	Classification	29
3.3.	Sampling and Analysis	36
3.4.	Extrapolation	38
3.4.1.	Class Extrapolation	38
3.4.2.	Nationwide Aggregation.....	38
4.	Sensitivity Analyses	39
4.1.	Nationwide Coverage	40
4.1.1.	Assumptions	40
4.1.2.	Results	42
4.2.	Impact of High Power devices	44
5.	Conclusion	46
6.	References	47

1. Introduction

The evolution of the communication networks used by Public Safety users toward a broadband wireless technology such as Long Term Evolution (LTE) (as mandated in [1]) has the potential to provide users with better coverage, while offering additional capacity and enabling the use of new applications that make their work safer and more efficient. Designing such a network presents several challenges due to the uniqueness of the deployment (there is no previous nationwide Public Safety network to build upon), the requirements (e.g., it must provide reliable coverage in rural areas and inside buildings, while supporting loads ranging from day to day traffic all the way up to large scale disasters), and the scale, which is national.

The objective of this document is to describe the work carried out to facilitate the deployment of the National Public Safety Broadband Network (NPSBN) by providing insights into the performance of LTE networks for Public Safety under various scenarios. While analyzing a few selected areas across the nation may provide insights on specific aspects or situations, their relevance may be limited when looking at a nationwide deployment. Because of the scaling problems associated with trying to analyze the entire area of the United States, we developed a modeling tool to select, configure, and automatically perform the analysis of representative areas and then extrapolate the results nationwide. This tool makes use of commercial off-the-shelf tools, and extends their core functionalities to handle the particularities of the task at hand.

The rest of the document is organized as follows: Section 2 describes the modeling method to perform a Radio Frequency (RF) planning of a given geographical area. It highlights the type of input information needed for accurate modeling and defines the performance metrics used in the nationwide modeling. Section 3 depicts the approach used to address the scaling issues related to the analysis of the entire nation. Section 4 presents the results of sensitivity analyses and Section 5 provides concluding remarks. This document updates and expends a previous publication by the authors [2].

2. Radio Frequency Network Planning

RF Planning tools facilitate network designs and optimizations by modeling the behavior of the networks. They allow network operators to adapt their network configurations by testing potential scenarios without disrupting their current network. This section describes the inputs used to represent the RF conditions and assumptions about the network operations because they impact the accuracy of the models. In addition, it defines the performance metrics used in the calculation of the coverage status. While the work was performed using InfoVista Planet, other tools may provide similar functionalities; by using the same inputs, assumptions and performance metrics it shall be possible to compare the results of different network plans or RF tools.

2.1. Inputs and Assumptions

2.1.1. Geo data

Computerized geographical/geospatial data, commonly known as geodata, describes the area to cover in terms of parameters such as the elevation, clutter (i.e., land cover), or buildings. The data provides information about obstacles that will affect how the signal propagates from the eNodeB to the user equipment (UE) and vice versa. Table 1 provides a description of the different types of geodata that can be provided.

Table 1: Types of geodata

Geodata type	Description
Elevation	Provides elevation data throughout the map as a function of geospatial coordinates (e.g., latitude and longitude), and is typically given in units of meters above sea level. This will describe natural elements such as mountains or plains.
Clutter	Describes how the land is covered, such as forest, roads, or airports. Each clutter type impacts the signal propagation differently and as such a clutter loss can be assigned to each type.
Clutter height	Provides the height of the clutter as a function of geospatial coordinates, for example the building height. Without this information, a fixed height per clutter type must be assumed.
Building morphologies	Defines the location and shape of the buildings. This information further enhances the accuracy of the RF predictions in urban areas. However, it is not sufficient to model precise in-building coverage, which requires detailed knowledge of interior floor plans and the materials used in constructing walls and ceilings.

The accuracy of the model is affected by both the type of geodata provided as well as its resolution (i.e., the spacing between grid points). A higher resolution (tighter spacing) will usually lead to more accurate predictions, especially in areas where the signal encounters many obstacles, e.g., urban areas. This is illustrated in Figure 1 for downtown Chicago. With a 30 m resolution, it is not possible to distinguish the buildings, whereas a 5 m resolution provides sufficient information to identify streets and buildings.

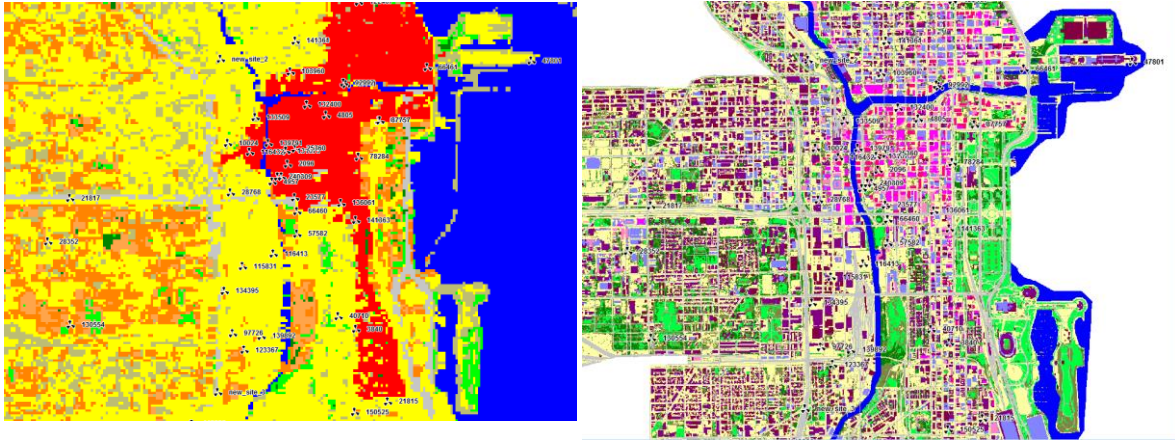


Figure 1 (a) Clutter maps of Chicago – 30 m resolution, (b) 5 m resolution

2.1.2. Propagation Models

Propagation models are essential for wireless network planning. Their purpose is to characterize the RF channel between a transmitter and receiver, specifically how the channel distorts a transmitted signal on its path to the receiver. There are a number of environmental factors which affect an RF channel. One of the most important factors is the terrain of the environment. As explained in Section 2.1.1, the terrain is useful to describe for example whether the environment is flat (e.g., the Great Plains) or whether there are mountains (e.g., Rocky Mountains). Terrain is specified by elevation geodata in terms of a grid-based terrain map; each grid point has an associated elevation value. Figure 2 shows the terrain map for the United States.

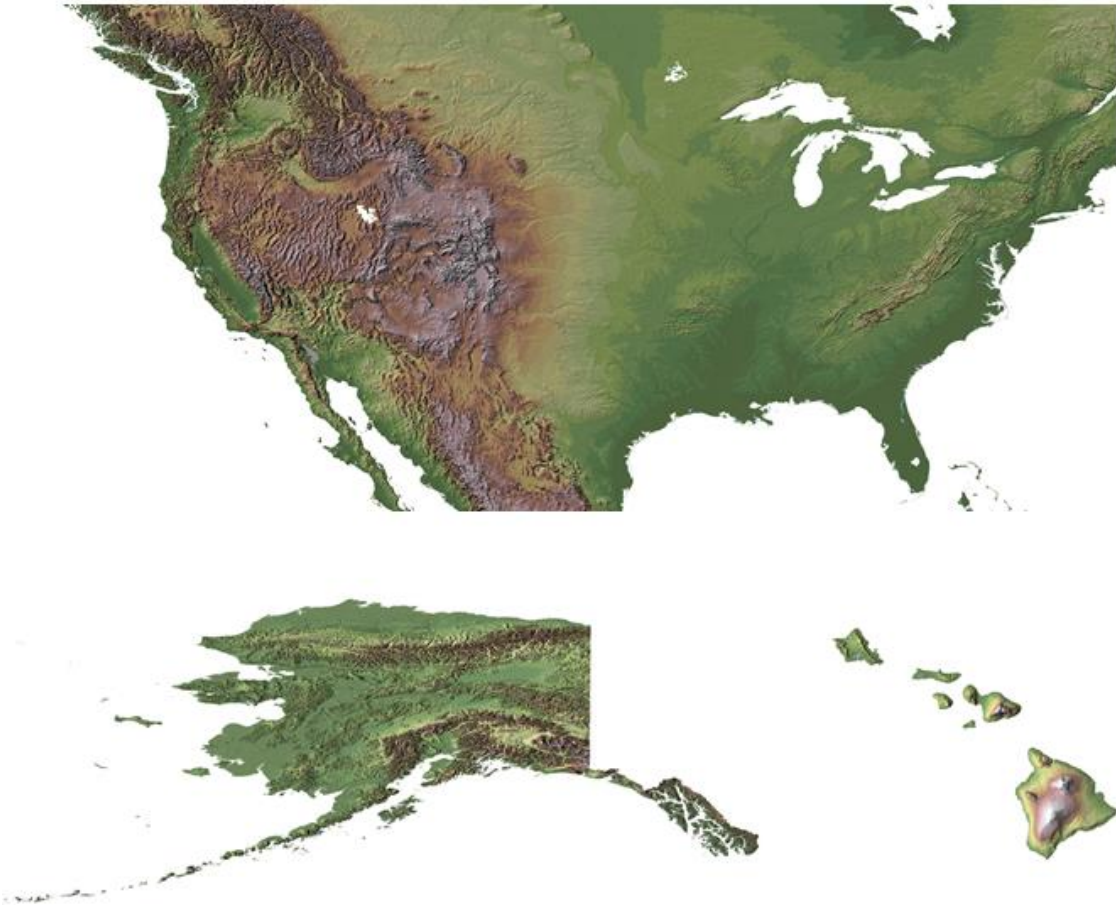


Figure 2: Shaded relief image of the United States

If the direct path from the transmitter to the receiver is unobstructed, the transmitted power will arrive with an attenuation equivalent to that of free space. If the terrain, however, obstructs, or “shadows”, the direct path, additional power will be absorbed. Besides the effects of shadowing, terrain can also cause fading by reflecting and diffracting radiated power, say off hills or a mountain, so that it arrives at the receiver through other paths in addition to the direct path. In some cases, the multiple signals arriving at the receiver combine constructively, boosting the signal strength; in other cases, the signals interfere with each other, reducing the signal strength. The combination of these effects (attenuation over distance, shadowing, and fading) will result in an overall signal degradation known as pathloss. Given the fixed location of a transmitter, pathloss values are generated from the terrain map; each grid point will then have an associated pathloss value.

The other most important environmental factor which affects an RF channel is the clutter in the environment. As explained in Section 2.1.1, like the terrain, the clutter is defined through a grid-based clutter map; each grid point is associated to a discrete clutter class. Figure 3 shows a clutter map for the United States. Typical clutter classes are *urban*, *suburban*, *rural*, *industrial*, *water*, etc. Just as the terrain contributes to the pathloss experienced at a receiver, the clutter causes additional shadowing and fading effects. The purpose of a clutter map is to characterize the incremental clutter pathloss, or simply the

clutter loss, in addition to the terrain. For example, consider the *urban* clutter class: Urban environments are typically characterized by the presence of tall buildings, which shadow, reflect, and diffract power, hence complementing the effects of the terrain on the pathloss. Another example is the *suburban* class. In it, there will typically be lower, residential buildings and so the clutter loss will be less than the typical *urban* class loss, conversely, the suburban clutter loss is typically greater than the loss associated with the *rural* class, which is characterized by sparse buildings and many trees.

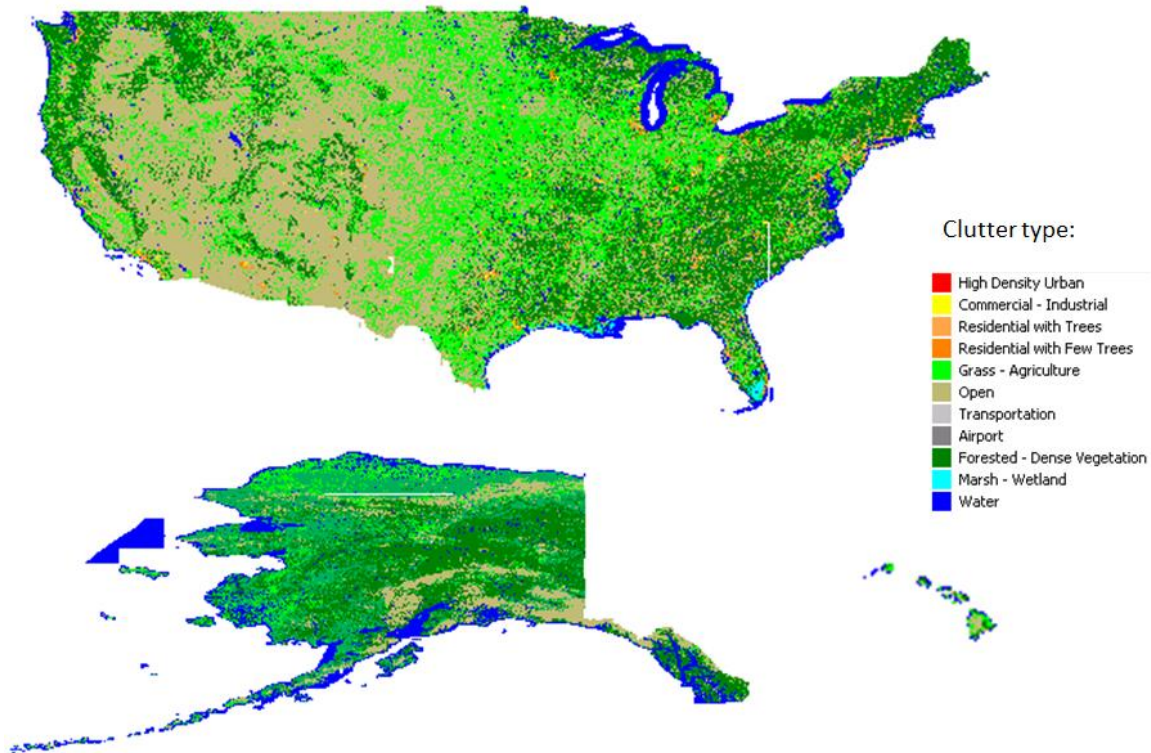


Figure 3: Clutter map of United States

The amount of clutter loss per class will vary from environment to environment. For example, the kinds of trees (height, shape, density, etc.) found on the East Coast will be different than those found on the West Coast. Likewise, the construction materials used for housing in various parts of the country will also differ. As a result, the clutter losses for the same *suburban* class will, in turn, vary. In order to determine the clutter loss for a specific area, it is customary to perform a *drive test* in that area. In a drive test, the received power from a fixed transmitter is collected throughout the area; knowing the transmitted power, the pathloss can be calculated. The measured pathloss values are used to tune the clutter loss per area. A pathloss model resulting from the tuning process is known as tuned model. In order to design their networks, commercial carriers obtain tuned pathloss models, often through secondary vendors. The number of tuned models to achieve accuracy similar to what commercial providers use numbers in the hundreds, with the specific amount being dependent on the specific provider.

2.1.2.1. Propagation Model Mapping

During the course of our work, we were able to obtain 26 tuned models from a secondary vendor. They were generated in four separate regions of the country with varying numbers of models per region: Arkansas (AR) (5 models), Chicago (CHI) (3 models), Northern California (NCA) (11 models), and Washington-Baltimore (WABA) (7 models). Consider Arkansas, in which the 5 models were generated from measurements in the forested area of the state (AR_FOREST), the mixed East and West areas of the state (AR_MIX_E and AR_MIX_W respectively), the mountainous North and South (AR_MTN_N_S), and the suburban area of the East (AR_SU_E). Because the tuned models are recommended for application only in those *tuned* areas in which they were generated, this leaves a void in the rest of the country. Rather than use generic, untuned models to fill that void, we mapped the 26 models to the *untuned* areas. The mapping was determined through the terrain and clutter properties of the analyzed area. We first consider the terrain properties of an area, which we define through the cumulative distribution function (CDF) of the elevations in its terrain map¹. A similarity metric is computed between a tuned (T) area and an untuned (U) area by comparing the aforementioned terrain CDFs of the respective tuned and untuned areas. Specifically, the similarity metric is given through the Kolmogorov-Smirnov (K-S) test [3]. The test yields the maximum distance, $d(T, U)$, between the two CDFs across all values of the random variable. We define the terrain similarity metric to be:

$$f_{TERRAIN}(T, U) = 1 - d(T, U)$$

The metric value ranges between 0 and 1, the former indicating the worst possible fit ($d(T, U) = 1$ (i.e., the CDFs have no values in common)) and the latter indicating the best possible fit ($d(T, U) = 0$ (i.e., the CDFs are equal at all values)).

Similar to the terrain properties, the clutter properties of an area are defined by the distribution of the clutter throughout the area. Because the clutter classes have no numeric value, as opposed to the terrain elevation, a CDF is not computed. Rather, an area is defined by a clutter vector, $\mathbf{v}$, indexed according to the clutter class. The indexed vector entry is the ratio of the area in the class to the whole area. The clutter similarity metric between a tuned (T) model and an untuned (U) area is then given by Pearson's correlation coefficient [3] of the respective clutter vectors $\mathbf{v}_T$ and $\mathbf{v}_U$ as:

$$f_{CLUTTER}(T, U) = \frac{E[(\mathbf{v}_T - \eta_{\mathbf{v}_T})(\mathbf{v}_U - \eta_{\mathbf{v}_U})]}{\sqrt{E[(\mathbf{v}_T - \eta_{\mathbf{v}_T})^2]E[(\mathbf{v}_U - \eta_{\mathbf{v}_U})^2]}}$$

where $\eta_{\mathbf{x}}$ indicates the mean of the vector $\mathbf{x}$. The clutter similarity metric ranges between 0 and 1, the former indicating the worst possible fit (the two areas have no clutter classes in common) and the latter indicating the best possible fit (the two areas have the same distribution of clutter classes).

When comparing a tuned model to an untuned model, both the terrain and clutter properties are considered jointly. This is accomplished by weighting the terrain and clutter similarity metrics equally through a joint similarity metric:

$$f_{JOINT}(T, U) = 0.5 \cdot f_{TERRAIN}(T, U) + 0.5 \cdot f_{CLUTTER}(T, U)$$

Finally, in order to determine the mapping for an untuned area, the untuned area is compared against the 26 tuned models. The tuned model that produces the largest joint

¹ The CDF $F_X(x)$ gives the fraction of a set of values X that is less than or equal to x .

similarity metric is selected. Figure 4 shows the largest joint similarity metric for all subdivisions in the United States. Note that two-thirds of the area has a similarity metric in excess of 0.75, which indicates that there will be a very good fit in many subdivisions. Poorer fits tend to be concentrated in the coastal regions, including the Florida Keys and the shores of the Great Lakes. Figure 5 shows the mapping of the 26 untuned models to the full set of subdivisions in the US.

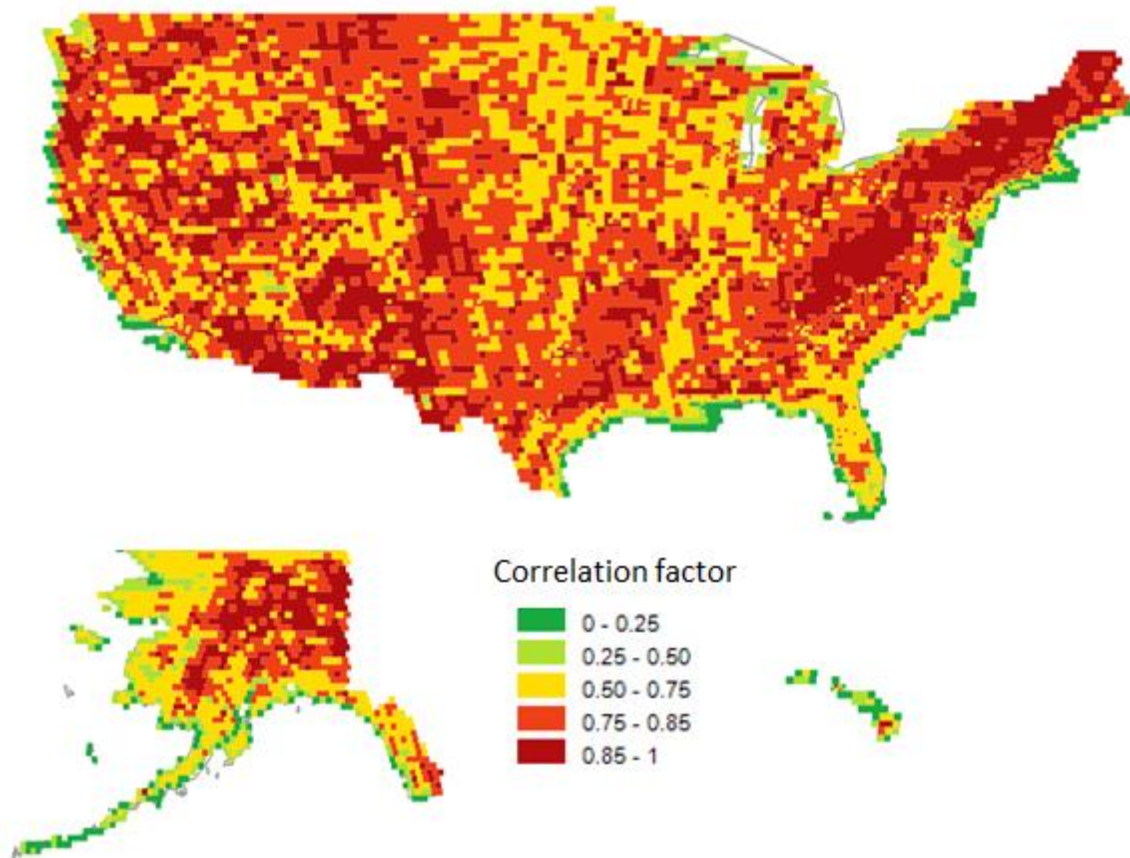


Figure 4: Correlation factor

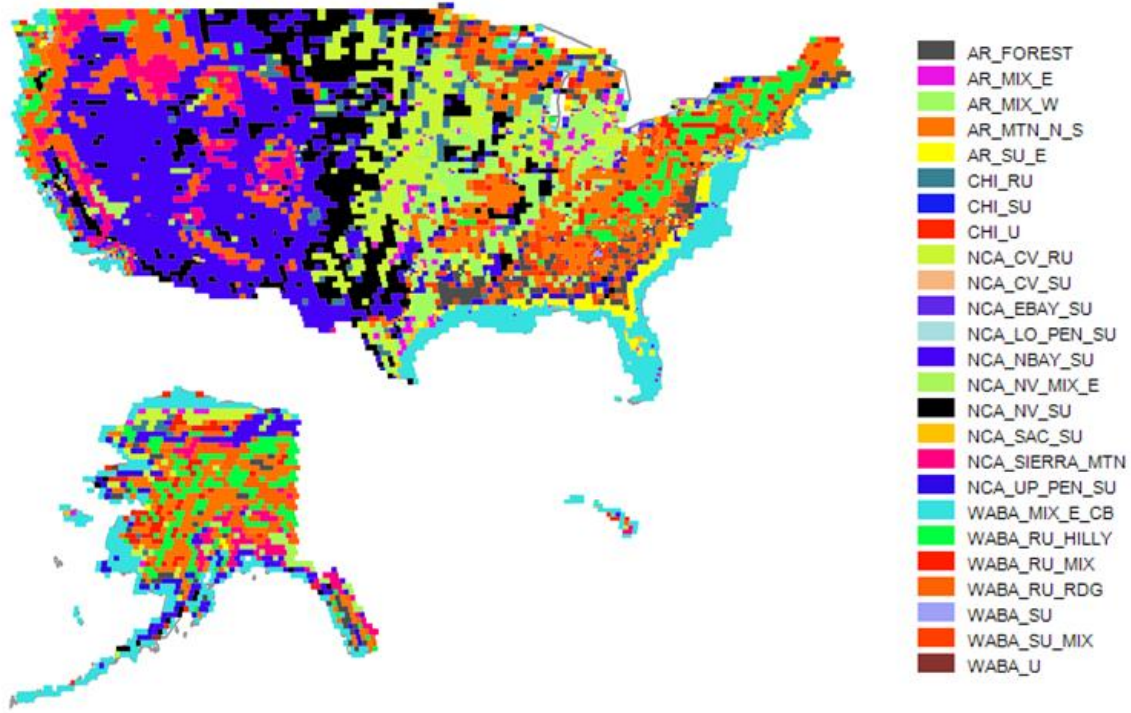


Figure 5: Mapping of the propagation models

2.1.3. User Distribution

The location of the users in the network is another important aspect to consider. When coupled with the traffic models, it determines the distribution of the demand in the network and therefore where the cellular base station sites should be located.

2.1.3.1. Estimation of Public Safety users

The purpose of the NPSBN is to provide a reliable network to Public Safety users. In this document, we assume that it includes law enforcement officers, firefighters, and Emergency Medical Services (EMS) employees at the federal, state, and local jurisdictions. There are an estimated 2.73 million Public Safety employees nationwide and Table 2 provides the breakdown per category.

Table 2: Summary of Public Safety user count

Categories	NIST Estimate
Law Enforcement	1.3 million
Fire	1.2 million
EMS	227 thousand
Total	2.73 million

2.1.3.1.1. Law Enforcement Agencies

A good source to determine the number and distribution of law enforcement officers is the census data collected by the Bureau of Justice Statistics at the federal [4], state, and local [5] levels.

2.1.3.1.1.1. Federal Law Enforcement users

The report on Federal Law Enforcement Officers from 2008 [4] presents data from 73 federal law enforcement agencies. According to the report, there were approximately 120 000 full-time law enforcement officers in September 2008. The raw data used in that report is not publicly available so it is limited to a breakdown of the number of officers per state. Employment information from OPM [6] is an alternative to refine the geographical distribution of federal law enforcement users as it provides the county of employment. Eight out of the 690 occupations listed by OPM identify law enforcement functions representing 110 000 employees. Those occupations are as follows: “General inspection, investigation, enforcement and compliance”, “Park Ranger”, “Criminal investigation”, “United State Marshals”, “Police”, “Correctional officers”, “Border patrol enforcement”, and “Customs and border protection”. Table 3 shows the nationwide employee count per occupation while Figure 6 shows the distribution of those users per county.

Table 3: Federal employment per occupation

Occupation	Employment
1811-CRIMINAL INVESTIGATION	23 627
1896-BORDER PATROL ENFORCEMENT SERIES	21 207
1895-CUSTOMS AND BORDER PROTECTION	20 141
0007-CORRECTIONAL OFFICER	17 586
0083-POLICE	14 323
1801-GENERAL INSPECTION, INVESTIGATION, ENFORCEMENT, AND COMPLIANCE SERIES	6455
0025-PARK RANGER	5113
0082-UNITED STATES MARSHAL	898
TOTAL	109 350

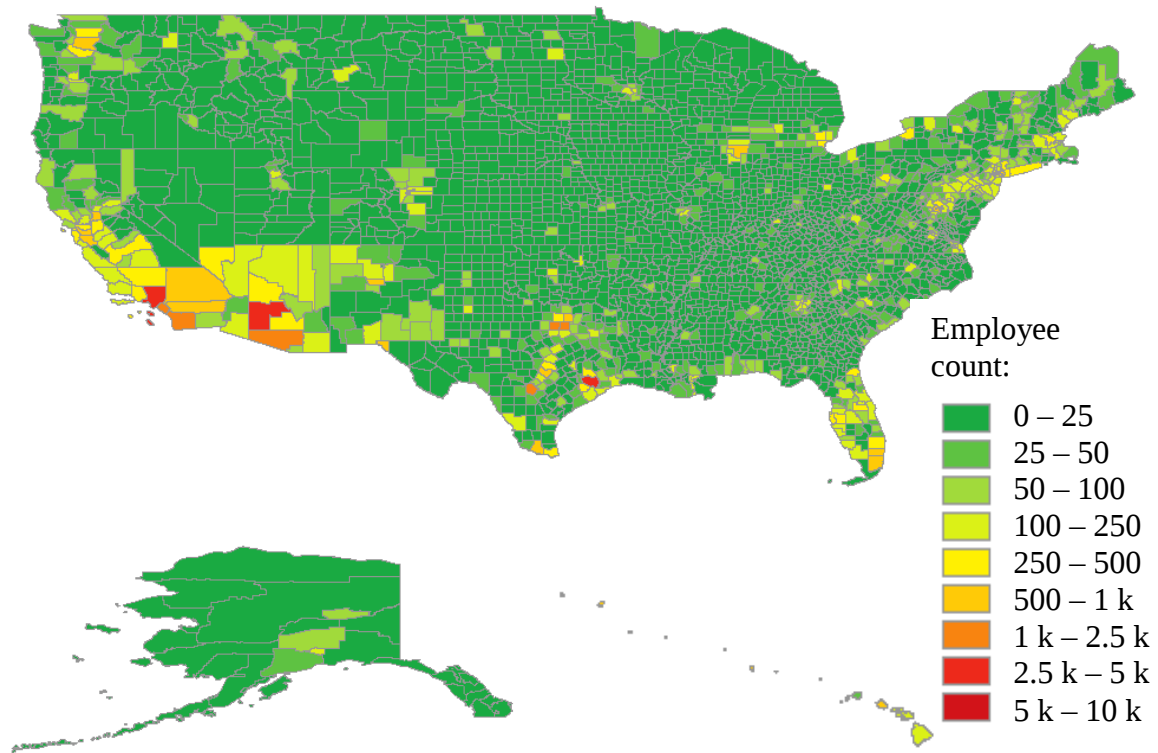


Figure 6: Federal law enforcement employment per county

2.1.3.1.1.2. State law enforcement

The Bureau of Justice Statistics collects data about state and local law enforcement every 4 years; the latest available data is from 2008 [5]. Unlike the Census of Federal Law Enforcement, the raw data is available and allows access to the information reported by each responding agency. The employment information for the states is extracted by querying all agencies where the type of agency is “Primary state law enforcement agency”. Figure 7 shows the number of full time and part time employees for sworn officers as well as civilians reported by each state.

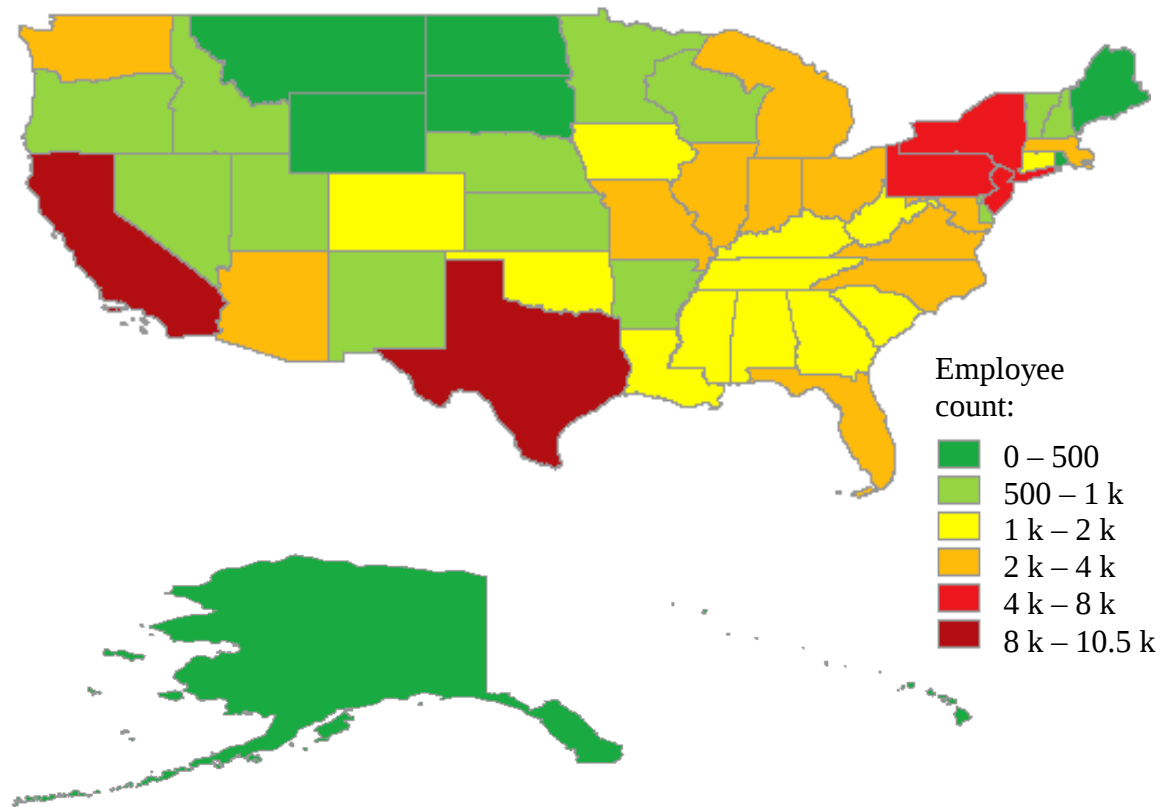


Figure 7: State-level law enforcement employment (officers + civilians)

2.1.3.1.1.3. Local Law Enforcement

Law enforcement data at the local levels, including Tribes, is also extracted from the Census of State and Local law enforcement agencies that was published in 2008 [5]. The address of the agency, including the county name, is used to generate a distribution of the employee count per county, as shown in Figure 8. The census contains information from 17 935 local agencies reporting 748 482 sworn officers and 391 678 civilian employees.

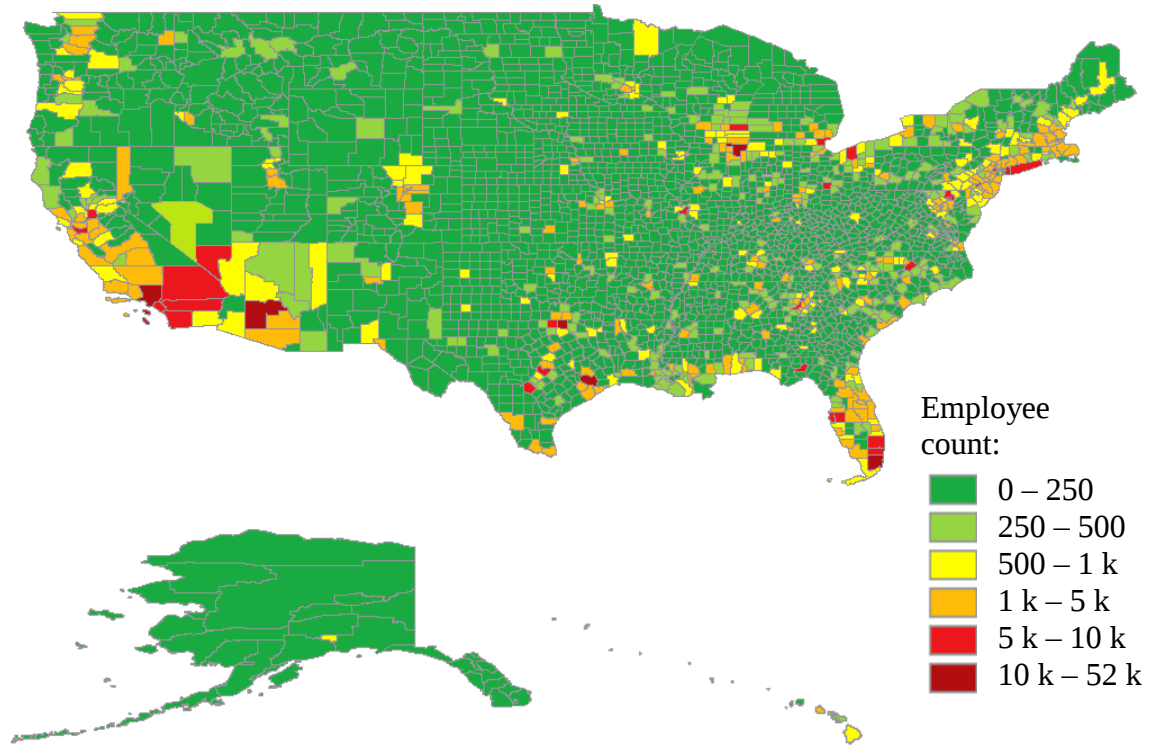


Figure 8: Local-level law enforcement employment (officers + civilians)

2.1.3.1.2. Firefighters

The Government Employment data from the U.S. Census Bureau is not sufficient to determine the number of firefighters because there are many firefighters who are either volunteers or paid per call. More accurate information is published via the National Fire Department Census [7], made available by the U.S. Fire Administration. This data contains staffing information from over 26 000 fire departments nationwide. Because it would be difficult to know the jurisdiction of each department, the analysis aggregated the number of active firefighters (career, volunteer, and paid per call) located within each county. We note that the data covered 3093 counties out of 3153 in the U.S. Nationwide, the total number of active firefighters is 1.05 million and the distribution of firefighters per county is shown in Figure 9.

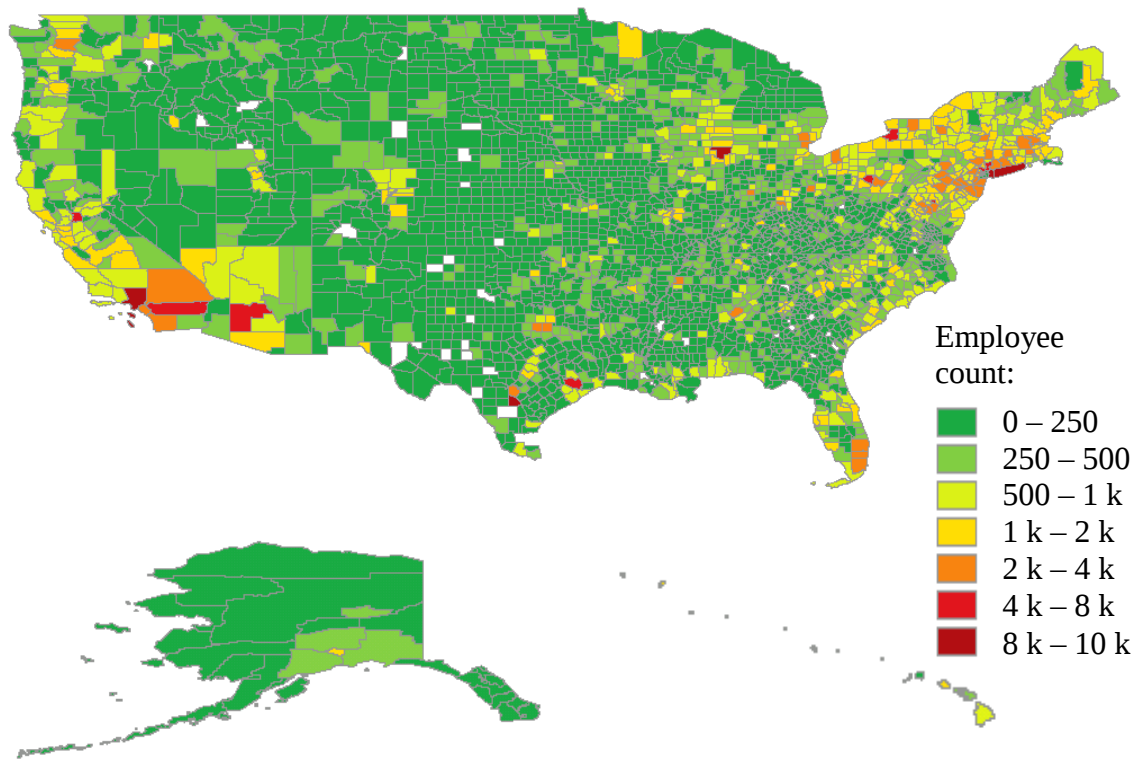


Figure 9: Distribution of active firefighters per county

2.1.3.1.3. Emergency Medical Services

Statewide employment information for Emergency Medical Technicians and Paramedics is available on the Bureau of Labor’s statistics’ website [8] [9] and is shown in Figure 10. Data at a more refined level was not found.

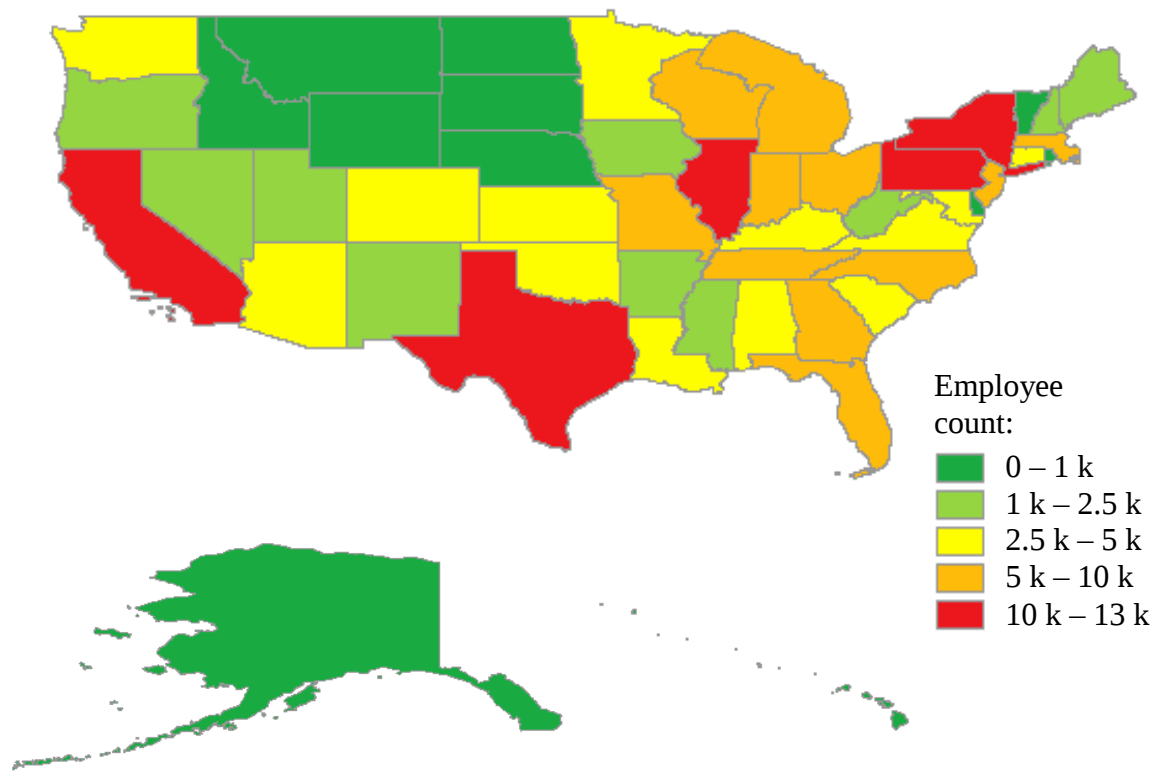


Figure 10: Employment of Emergency Medical Service personnel per state

2.1.3.2. *Traffic map generation for analysis*

By combining the employment data collected for law enforcement, firefighters, and EMS in Federal, State, and Local governments, it is possible to create a nationwide distribution of Public Safety users. When only state level data is available, users in that state are distributed proportionally to the county population in order to provide an estimation of the employee count for each county (e.g., a county whose population comprises half the state's population would receive half of the pool of users for that state). Furthermore, it is unlikely that users are distributed uniformly within a county and this is especially true due to disparities between urban and rural areas (i.e., more Public Safety personnel will be concentrated in urban areas). When the population is a coverage target of the region being analyzed, the users are distributed according to the population density in the county. In our analyses, the population distribution is based on LandScan data developed by Oakridge National Laboratory [10].

For analysis where population is not part of the target coverage, a different user distribution is used as appropriate. For example, when the target is to cover the highways, users are distributed solely along those roads.

2.1.3.3. *User mobility*

The users of wireless networks are mobile by nature. The NPSBN is no different and it is expected that the users will move throughout the day. It was shown in [11] that the network design needs to take transient users into consideration; otherwise, coverage is not guaranteed at all times. Another limitation of using night time information (e.g., the

census data) is that some areas that contain no residences yet are heavily populated during work hours, such as commercial centers and industrial parks, would appear not to have any users in them. In our work the user mobility issue was solved by creating a peak traffic map with the highest user density from both day and night time user distributions. An example of the resulting density map is shown in Figure 11. While this overestimates the number of users, it simplifies the modeling approach and guarantees that the network deployment will provide coverage at any time of the day.

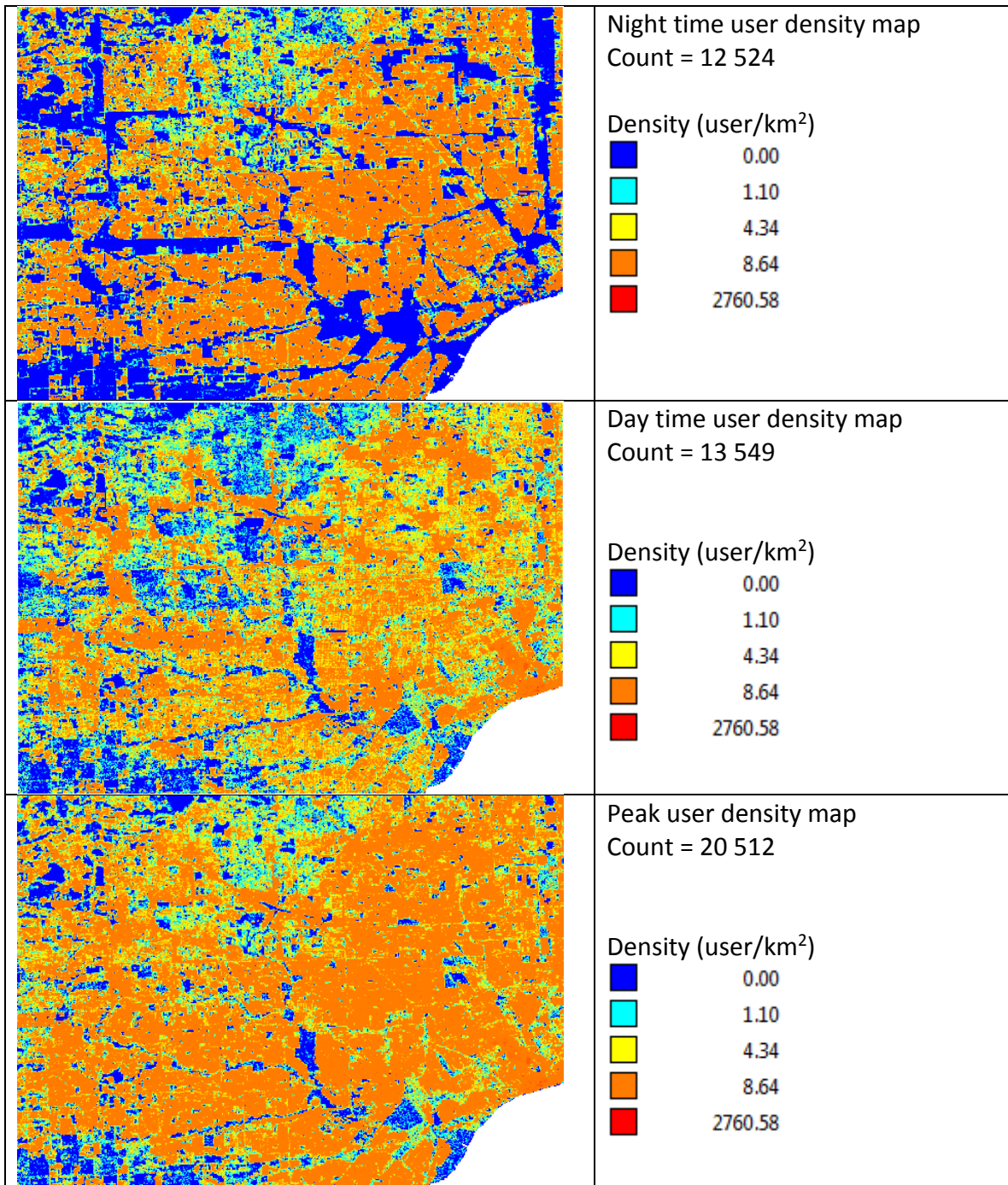


Figure 11: User distributions around Detroit, MI

2.1.4. Traffic Model

The traffic model defines the network utilization patterns of the users. It specifies how many users are using each application, how often, as well as the data rate requirements for each application. While there has been an ongoing effort to characterize the user traffic [12] [13], it is still difficult to find accurate and detailed reports because there is no historical data available, and the range of possible applications (current and future) and configurations (e.g., video resolutions) is vast. An additional issue is that, given the novelty of the NPSBN, practitioners often do not know how the network will be used or how it will change current Public Safety operational practices. Furthermore, Public Safety users respond to different types of incidents differently, therefore requiring different traffic estimations and plans for each type and size of these situations: There are small scale, day to day activities that occur millions of times every year (e.g., motor vehicle incidents, traffic stops, fires) and large scale, infrequent incidents (e.g., natural disasters, active shooters). For the first type of incident, the network is typically operational, while the latter incident types may involve site unavailability. For the modeling of the NPSBN, the day to day traffic shown in Table 4 is a modified version of the incident traffic based on the 2007 collapse of the Interstate 35 bridge in Minneapolis, Minnesota [14]. The traffic is composed of 7 applications and contains a mix of voice, video, and data applications. It is also assumed that 1/3 of the Public Safety users are active at any given time to simulate 8 hour shifts.

Table 4: Day to day Public Safety traffic

Type of device	PS users carrying device (%)	Uplink data rate (kbit/s)	Downlink data rate (kbit/s)	Time device transmits (%)	Time device receives (%)
Mobile Video Camera	25	256	12	5	2.5
Data File Transfer CAD/GIS	87	50	300	7.5	2.5
VoIP	100	27	27	2.5	7.5
Secure File Transfer	12	93	93	2.5	2.5
EMS Patient Tracking	6	30	50	5	2.5
EMS Data Transfer	6	20	25	12.5	2.5
EMS Internet Access	6	10	90	5	2.5

2.1.5. Network Configuration

Assumptions also have to be made regarding the network configuration, including the site locations, antenna configuration (antenna type, azimuth, tilt), transmit power for both the eNodeBs and the UEs, multiple-input and multiple-output (MIMO) configuration, etc. The list includes hundreds of parameters, though not all of them have the same impact on the network performance. The results shown in Section 4 are accompanied by the configuration used for our analyses and highlight some key parameters.

2.2. Performance Analysis

Once the necessary inputs have been collected and the assumptions have been set, it is possible to perform an RF analysis following the process illustrated in Figure 12. Since there is no existing network to evaluate, the first step of the analysis is to find a set of sites that would meet the desired coverage. The next step consists of running Monte Carlo simulations to compute the number of subscribers served and the load for each sector. Finally a network analysis is performed to verify the area and population coverage. The following sections describe each step in more details using an area in the vicinity of Detroit, MI as an example.

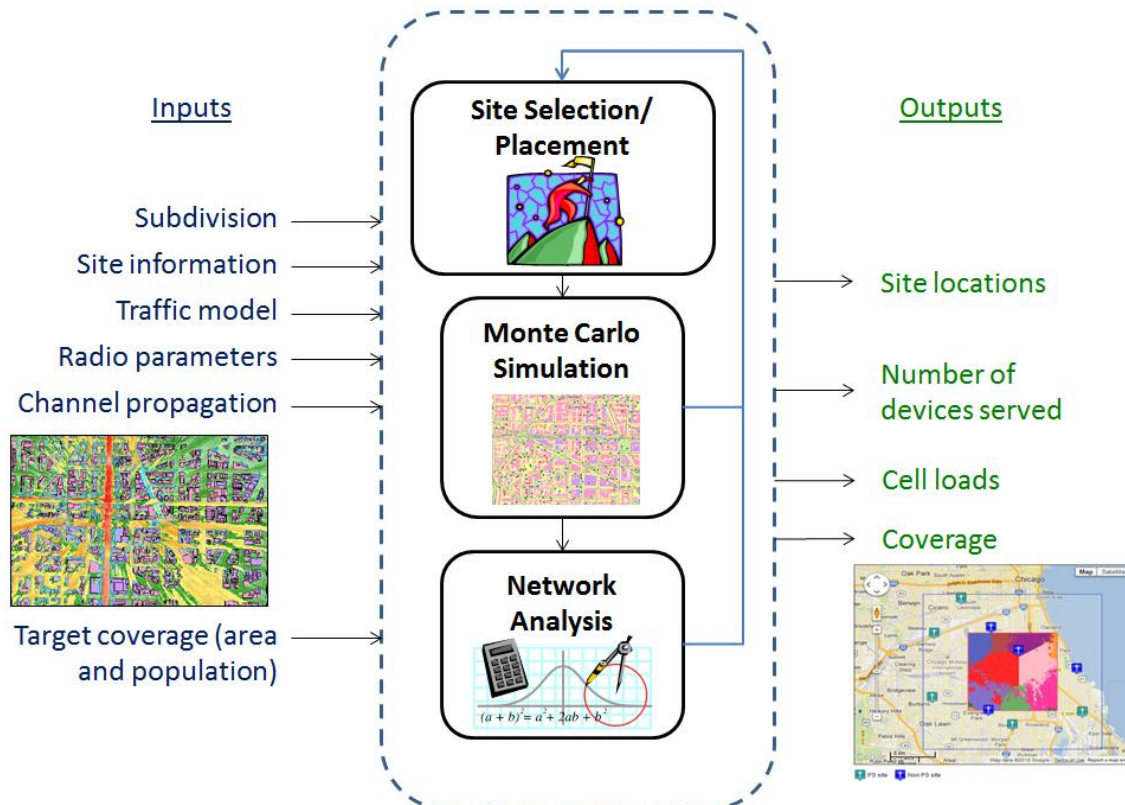


Figure 12: Performance analysis overview

2.2.1. Site Placement/Site Selection

The goal of the site placement algorithm is to select a subset of the available sites that meets a certain criterion. In the NPSBN modeling, the criterion used is the Reference

Signal Received Power (RSRP). A location is covered by a site if the predicted RSRP value is above a given threshold.

To reduce the computation time, a tiling algorithm [15] discretizes the user density map into demand points, each of which represents aggregated demand from a fraction of the user population. Because the tiling algorithm partitions the area so that equal populations are in each tile, the offered load from each demand point is the same. As an illustration of the tiling process, Figure 13 (a) shows user density map in Detroit, MI with a resolution of 30 m, and Figure 13 (b) shows the result of the tiling algorithm. Each demand point is located at the centroid of its tile. An iterative greedy algorithm with site swapping [16] is used to select a set of sites so that each demand point is covered by a site, while minimizing the total number of sites.

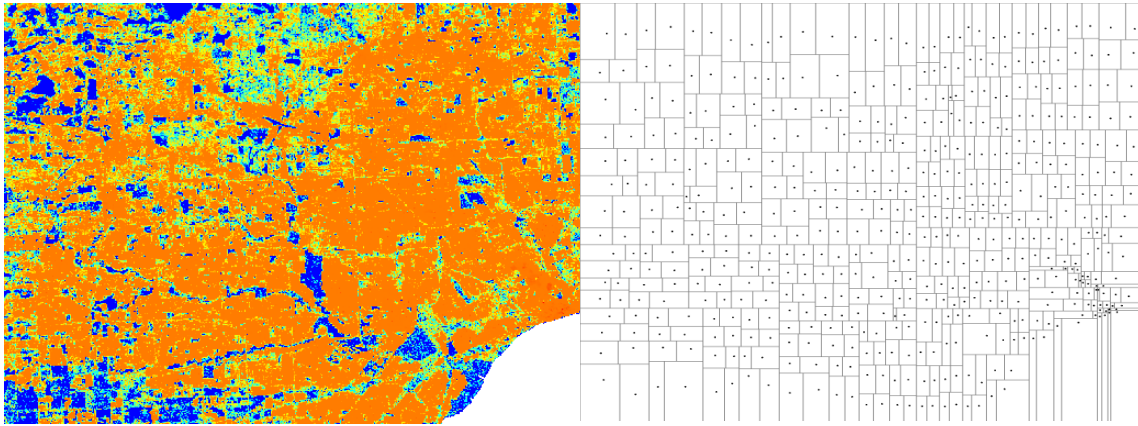


Figure 13 (a) User density map in Detroit, MI. (b) Discretized user density map with 1024 demand points

The site selection considers three types of sites: existing Public Safety sites, existing commercial sites, and greenfield sites. The list of existing sites is compiled using information from the Federal Communications Commission (FCC) [17] and other providers of wireless infrastructures (e.g., American Tower, TowerCo, AT&T). The sites in the FCC database with a service code value equal to 'GE', 'GF', 'GP', 'PA', 'PW', 'QM', 'SG', 'SL', 'SP', 'SY', 'YE', 'YF', 'YP' or 'YW' are used to identify the Public Safety sites. The remaining existing sites are deemed to be commercial. Candidate green field sites are added throughout the area of analysis and will be used to extend the coverage when needed. Each site type has an associated weight to prioritize the selection, with a higher weight meaning that the site is more likely to be chosen for the deployment. By default, a weight of 1 is used for Public Safety sites, 0.75 for commercial sites, and 0.5 for green field sites. The site selection algorithm uses the number of unique demand points a site can provide service for, multiplied by the appropriate weight, to sort the candidate sites. This means that a green field site will be preferred to an existing Public Safety site if it covers at least twice as many demand points; if the greenfield site covers fewer demand points, the Public Safety site will be preferred.

Since the criterion is based on a downlink metric that considers neither interference nor uplink conditions, the effective coverage can be determined only after running the Monte Carlo simulations and network analyses. Therefore, the algorithm is an iterative process that increases the required RSRP threshold, which effectively increases the number of

sites selected, until the coverage criterion is met. The initial value for the RSRP threshold is based on the target application data rate and coverage reliability.

2.2.2. Monte Carlo Simulations

Once the predictions have been generated and sites have been selected, Monte Carlo simulations estimate the uplink and downlink sector loads. To do so, users are first randomly distributed based on the user density maps. The coverage of each user and its impact on the sector loads and interference is computed based on the required data rate, coverage reliability, and resources available. This is an iterative process where the cell loads of an iteration is used to estimate the user coverage and loads on the next iteration until the results converge.

A key parameter is the coverage reliability. A higher reliability requires a larger Signal to Interference and Noise Ratio (SINR) margin to sustain a certain Modulation and Coding Scheme (MCS) and associated spectral efficiency. For 95 % coverage reliability, the SINR level needs to be 11.5 dB higher than that required by a given application, assuming a 7 dB shadowing standard deviation. At 50 % coverage reliability, the margin is 0 dB.

To account for possible interference and coverage provided by adjacent sectors, a margin area is added to the area of interest to take into account the neighboring users and sites. The optimum size of the margin is based on the type of region: in urban areas, sites usually have smaller coverage footprint and the margin area can be reduced to improve the performance; in rural areas with limited obstacles, however, the coverage can be 30 km or more, so larger margins are required. The users in the margin areas are served with 50 % reliability to prevent the site placement mechanism from targeting the coverage of those users, and only consider them in the estimation of their impact on the interference level.

Figure 14 (a) shows an example of an area to analyze with 12 sites selected during the site placement stage. The red line represents the boundary of the area of interest, where the target coverage must be met, while the blue line located at the figure boundary represents the outer edge of the margin area. Figure 14 (b) shows the users deployed along with their statuses for a single Monte Carlo iteration. The users that can be served are shown in green while other colors indicate failure: red indicates failure due to uplink power limitation, pink indicates failure due to lack of uplink resources, and blue indicates failure due to lack of downlink resources.

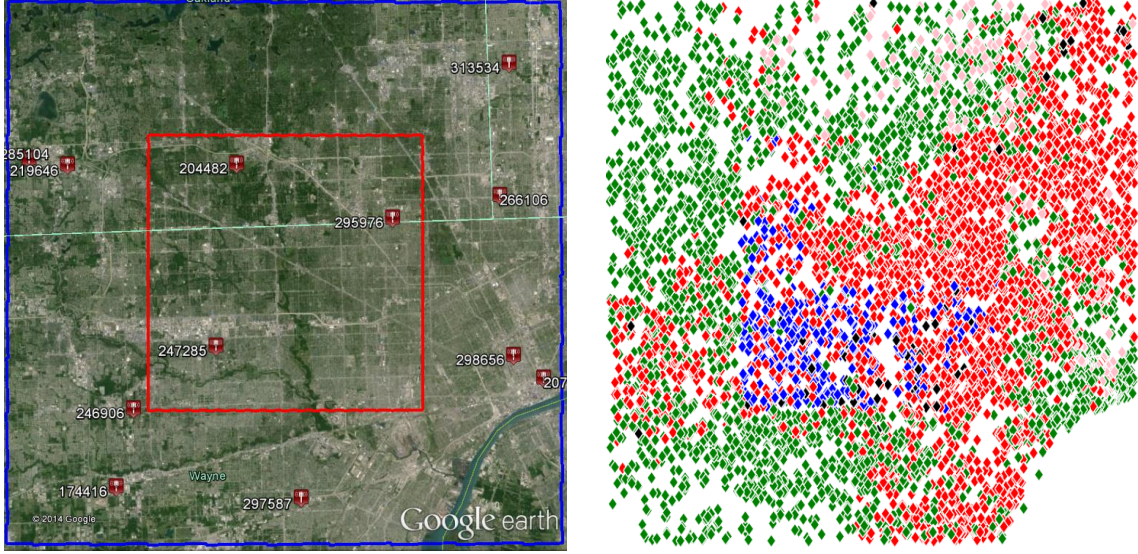


Figure 14 (a) Example of analysis area near Detroit, MI. (b) Deployed subscribers (green = served, other = not served)

The process uses two sets of Monte Carlo simulations: In the first set, the user coverage reliability is set to the target value (e.g., 95 %) and the percentage of users served is checked against the target user coverage. If the percentage is below the target threshold, it is assumed that more sites are needed and the performance analysis is repeated with a higher RSRP value threshold for the site placement. If the percentage exceeds the threshold, the second set of Monte Carlo simulations runs with the target user coverage reliability set to 50 % to obtain the estimated sector loads.

2.2.3. Network Analysis

With the sector loads estimated from the Monte Carlo simulations, a network analysis computes downlink and uplink coverage and capacity information. This step excludes the margin area used in the Monte Carlo simulations, which we showed in Figure 14. The results of the network analysis depend on assumptions being made regarding both the quantity of resources available to the user and the data rates. We define the cell edge data rate requirement based on the traffic model being used. Since the cell can contain a variety of applications with different usage, we compute the cell edge data rate by taking the maximum value of (a) the average user data rate and (b) the maximum data rate of any single application. For example, using the traffic model of Table 4, the average user data rate is 18.303 kbit/s in the downlink and 15.373 kbit/s in the uplink. The low values are mainly due to small activity factors. The maximum data rate is 300 kbit/s in the downlink and 256 kbit/s in the uplink. Those data rates correspond to data transfers and video transmissions, respectively. Since the maximum values are higher, they will be used as required cell edge data rate thresholds for deciding if the user can be covered. The output information available from the RF modeling tool does not directly give us the area and population coverages. Instead, the information indicates the maximum achievable data rate at any given location for both the uplink and the downlink. Examples of those outputs are shown on the leftmost side of Figure 15. Using the data rate thresholds, we can derive downlink and uplink coverage maps. Because the coverage

gaps in the downlink and uplink may occur at different locations, area coverage is derived by computing the areas where coverage is available in both directions, shown in the center of Figure 15. Finally, the area coverage is overlaid on the population distribution to compute the percentage of population covered.

If the area coverage or population coverage does not meet the target criteria, the performance analysis is executed with different thresholds in order to increase the number of sites.

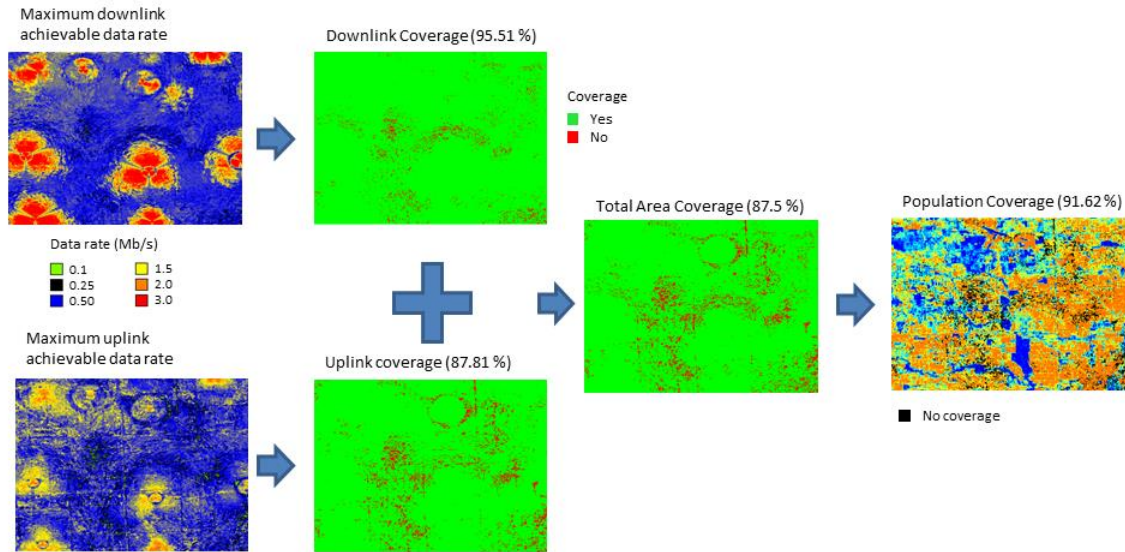


Figure 15: Process to determine area and population coverage

2.2.4. Sample Results

As mentioned previously, searching for a set of sites that meets a target coverage value is an iterative process. Figure 16 shows the results of the performance analysis in the sample area near Detroit, MI. The objectives were to obtain 95 % user coverage and 95 % population coverage. The x -axis shows the RSRP threshold used during the site placement. As the threshold increases, the number of sites selected also increases. The user coverage also increases with the number of sites selected, except when comparing the data points for the RSRP threshold of -130 dBm and -125 dBm. Even though both of these thresholds produced outcomes with the same number of sites, the selected sites were different and thus produced different user coverage values. We also observe that for RSRP thresholds above -110 dBm, the benefit of adding more sites diminished due to interference. Once the user coverage is above 95 %, the network analyses are also performed, which occurs when more than 30 sites are selected. For this particular example, the coverage objectives were met with 39 sites.

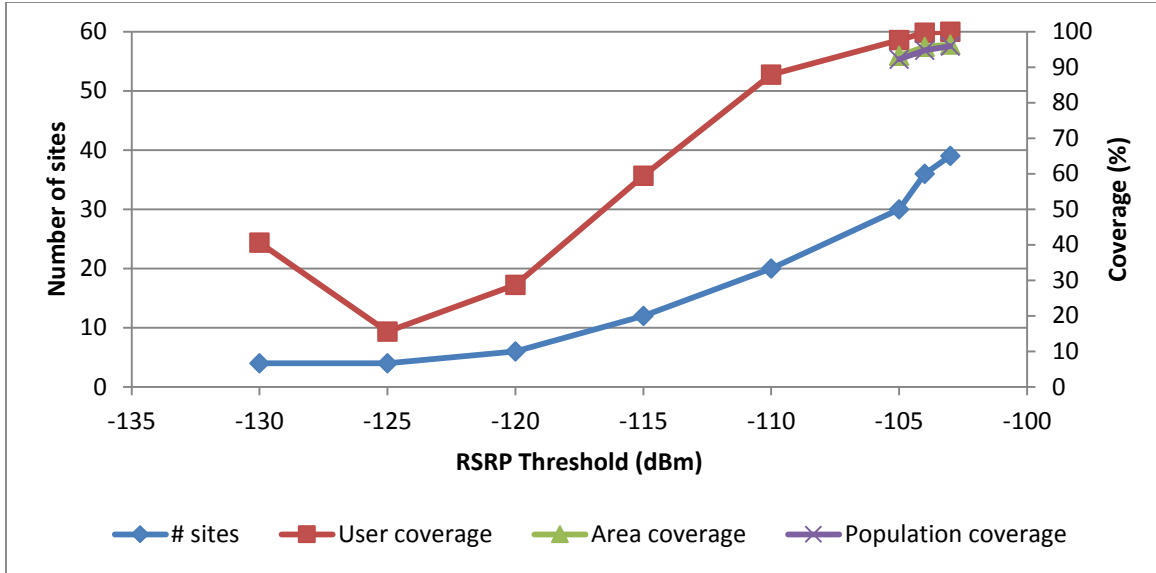


Figure 16: Example of the iterative process to find the set of sites that achieves the coverage requirements

3. Tackling Nationwide Scale Modeling

The computational time required to perform an RF analysis grows exponentially with the area to analyze. To handle the complexity associated with modeling a nationwide network, we analyzed a subset of areas throughout the nation and extrapolated the results following the process shown in Figure 17. This Section describes the mechanisms by which the nation is split into smaller areas called subdivisions, the areas to analyze are selected, and the results from the studied areas are extrapolated to obtain meaningful nationwide results.

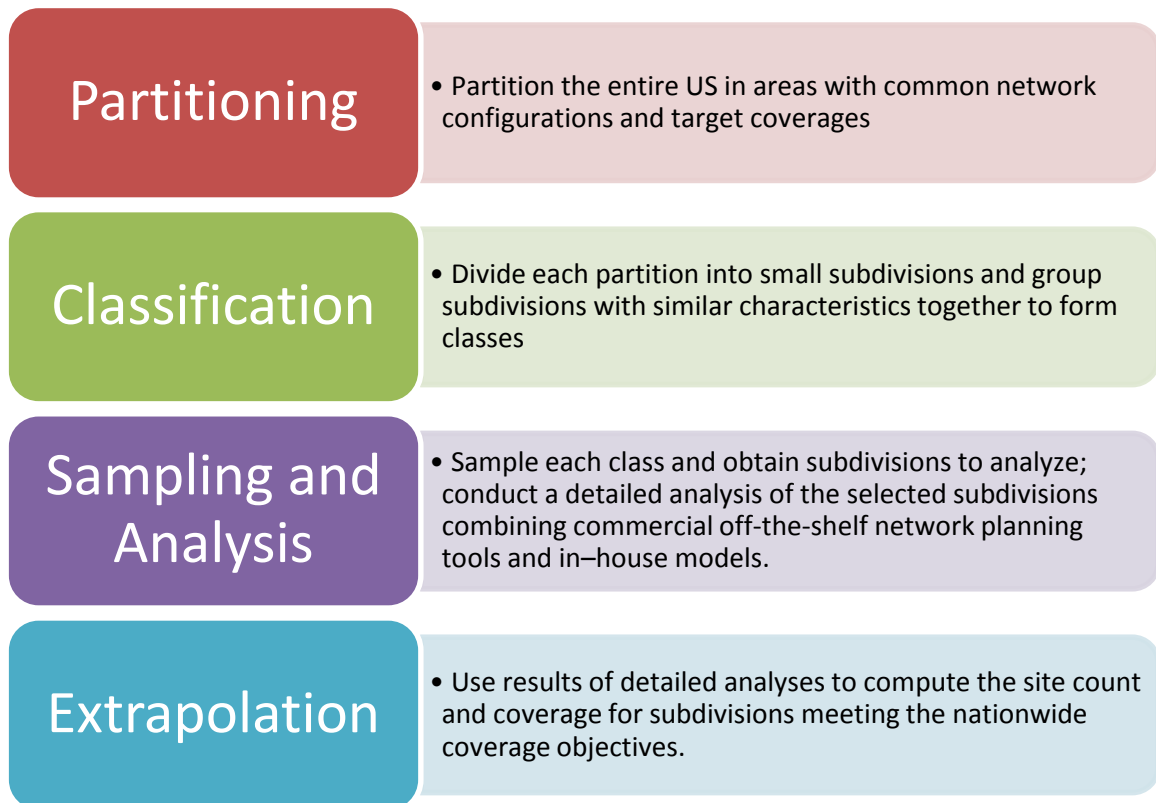


Figure 17: Nationwide modeling workflow

3.1. Partitioning the United States Area

When tackling the problem of analyzing the performance and behavior of a wireless network across such a great and diverse space as the whole United States area, it is soon apparent that the wide variety of user distributions and terrain characteristics make it ill-advised to uniformly define parameters and requirements for the whole analysis region. Areas with high population density and high-rise buildings, like major urban centers, will not present the same propagation characteristics as the farmlands of the Great Plains, which in turn exhibit completely different characteristics than the Rocky Mountains. As the equipment used in these areas is likely to differ, it does not make sense to analyze them uniformly. Similarly, and because of the different population and Public Safety user densities in these areas, the coverage requirements are likely to be different from those in the downtown area of a major city and both will be different from the coverage requirements in a heavily forested National Park.

In order to account for these idiosyncrasies, we characterized the whole area of the United States and defined the following criteria to distinguish the areas where different conditions, configurations and requirements may be defined for the analysis:

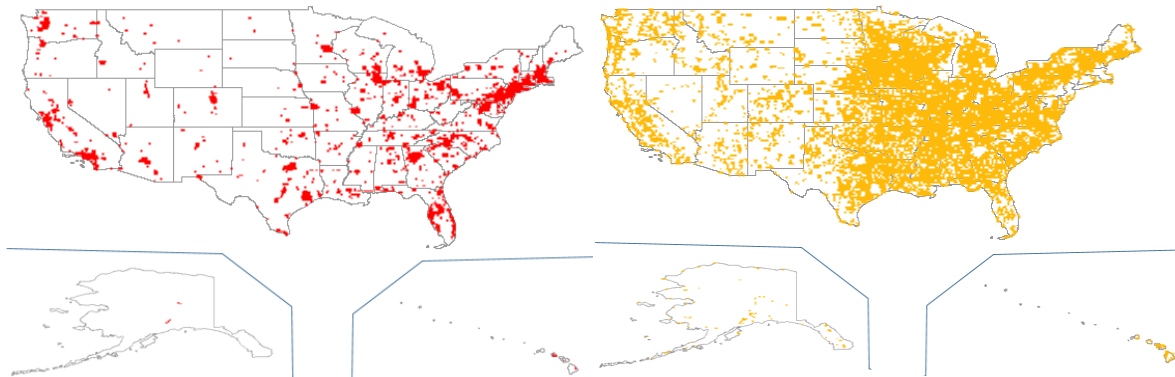
- Urban areas: Those defined by the Census 2010 to be of type ‘Urban’, ‘Urbanized Area’ or ‘Urban Cluster’ (“Urban”).

- Rural areas with an average population density over 5 people per square mile: Areas that, while not considered urban, still present a meaningful population (“Rural populated”).
- Rural areas with an average population density less than 5 people per square mile: The rest of the United States area (“Rural low population”).

These three partitions together fully enclose the whole area of the United States. However, a fourth category was identified as being a coverage target while having some particular configurations and requirements:

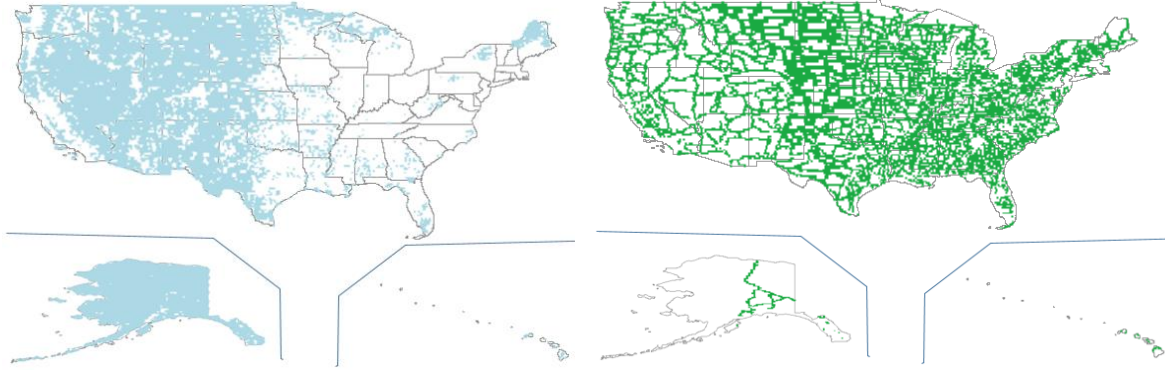
- Highway areas: Rural areas of the United States with one of the highways defined in the National Highway System (“Highways”) [18].

This fourth partition overlaps with the three previous ones, but as we will see in Section 3.4.2, only part of each of the different types of areas will be considered for coverage, depending on the target coverage for the nationwide analysis. Having this fourth type will ensure that we can target the National Highway System for coverage even if the areas the highways go through are not targeted themselves. The areas belonging to these four partitions are shown in Figure 18. In Figure 19 the map shows the three partitions that comprise the full area of the US, and Figure 20 shows the same map with the areas covering the highway partition overlaid.



Urban Areas

Rural Areas with More than 5 People per Square Mile



Rural Areas with Less than 5 People per Square Mile

Areas with part of the National Highway System

Figure 18: Maps of the different partitions defined

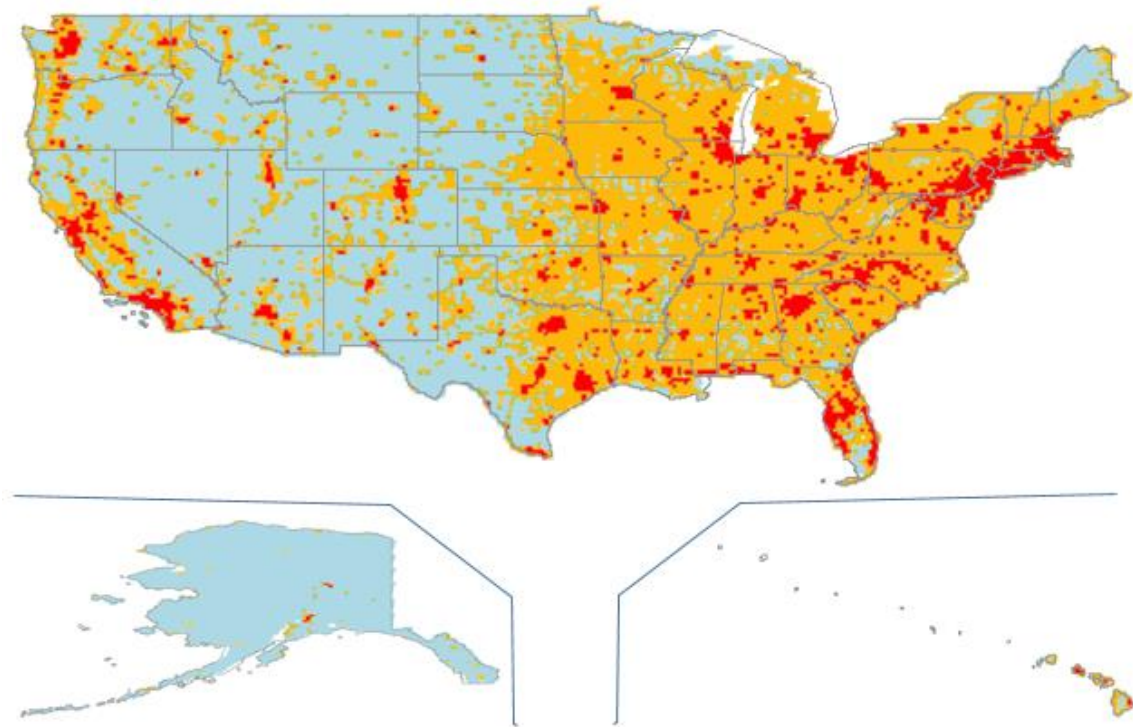


Figure 19: Map with the urban and rural partitions

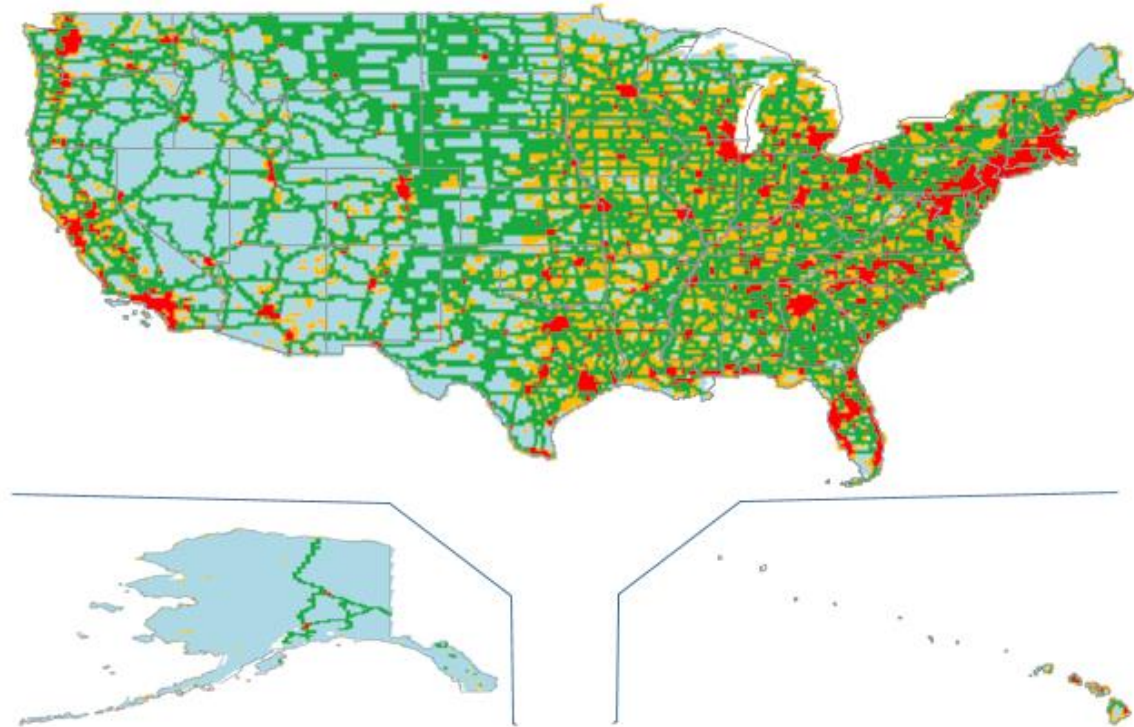


Figure 20: Map with the highway partition overlaid over the urban and rural partitions

3.2. Classification of the Analysis Areas

The complexity of the RF analysis described in Section 2 of this document is dependent on the number of users in the area and the number of sites. As both of these parameters increase with the area (either due to the increased population comprised in the area or to the additional number of sites required to provide coverage), analyzing large areas requires a significant amount of resources and time. In some cases, small areas with large user densities will also demand large amounts of CPU, memory and time to complete the analysis. Therefore, it is not feasible to perform an analysis of the whole United States area using a traditional approach.

To overcome this limitation, we developed an approach based on the *Divide and Conquer* principle, in which the total area of analysis (i.e., the whole area for one of the partitions defined in Section 3.1) is divided into smaller subdivisions. These subdivisions are then classified and grouped into clusters based on their characteristics, analyzed independently, and the individual results are later aggregated to provide the aggregated values for the original area. By combining these subdivisions with the sampling and extrapolation techniques described in Section 3.3 and Section 3.4, it is possible to significantly reduce the total number of RF analyses needed and allow for the parallelization of the study, while still obtaining results with the required confidence and precision.

The rest of this Section describes the parameters used for classifying the subdivisions, the process and parameters used for dividing the total analysis area in smaller subdivisions, and the classification process itself.

3.2.1. Input Characteristics

The classification of subdivisions in clusters or groups of subdivisions with similar characteristics is interesting for our approach because, if done properly, it will enable us to analyze a small number of these subdivisions in a given cluster, and then extrapolate the results to the rest of the subdivisions in the cluster, thus greatly reducing the number of analyses. For this assumption to be true, each subdivision has to be characterized with the parameters that may affect the analysis results, so subdivisions can be grouped based on the similarity of these aspects.

The goal of the RF analysis is to obtain information about the number of sites required in the study area and the associated coverage, so we will use attributes of a subdivision that may affect either one of these aspects. In particular, we identified two major features that will affect the analysis results:

- The topology of the terrain, as it will affect the propagation of the signal.
- The number of users in the network, as the load and interference increase with the number of users.

To characterize the terrain we considered in each subdivision the elevation and the clutter height using the following specific metrics:

- Average terrain elevation in an area.
- Minimum and maximum elevation, to provide information about the elevation gradient.
- Standard deviation of the elevation, to represent the variability of the terrain, (i.e., whether the terrain in the subdivision is mostly flat, hilly, or highly variable).
- Average clutter height, to complement the average terrain elevation.
- Standard deviation of the clutter height, to account for the variability of the clutter.

Similarly, the number of users of the network in an area is described using the following parameters:

- Average population density.
- Minimum and maximum population density.
- Standard deviation of the population density.
- Average Public Safety user density.
- Minimum and maximum Public Safety user density.
- Standard deviation of the Public Safety user density.

The reason for considering both the population and Public Safety user density is that, as we will describe in detail in Section 4.1.2, both the Public Safety user coverage and the population coverage are targets of the analysis and therefore the subdivisions must be characterized with both of these attributes to allow for adequate identification of similar subdivisions. Although Public Safety users are deployed for the analysis based on the population densities, the Public Safety user to population ratio varies, so we cannot assume that these features are redundant.

The information for each of these parameters in each subdivision is collected from the same sources as previously described in Section 2.1.

3.2.2. Subdivisions

In order to create a set of subdivisions that covers the whole area of the United States while minimizing the overlap between partitions, the process of creating the subdivisions is agnostic of the partitions, and considers only the total area to cover. Once this process is finished, the subdivisions are assigned to one of the partitions described previously according to the appropriate criteria.

The size of the subdivisions was chosen considering the propagation characteristics for LTE, so that they would be large enough to account for the maximum area a site can cover, while still being small enough that the cases in which a large number of users and sites are deployed in a single analysis are minimal. As a result of these conditions, we found the size of 20 km x 20 km to be the most appropriate across the nation. However, we noticed that in the Great Plains area it is possible for a single site to provide coverage for even larger areas if the network load is low, since the terrain presents minimal obstructions to the signal. Given that this is true only for the rural areas in the Great Plains (as urban areas require more sites due to the higher number of users deployed), the subdivision size is calculated as follows:

- If the subdivision is in the Great Plains and more than 90 % of its area is considered rural, the subdivision size is 40 km x 40 km.
- Otherwise, the subdivision size is 20 km x 20 km.

With this information, the process of creating the subdivisions begins by processing the whole US area, starting at the northwest corner and creating adjacent subdivisions first following the appropriate parallel (constant latitude). Once the process reaches the northeast corner it goes back to the west border of the area and creates another set of subdivisions right below the previous one. For each subdivision created, the input files are read to acquire the values for the elevation, clutter height, Public Safety user density, and population density with a resolution of one arc-second.

When all the values of a subdivision are read, we discard those subdivisions with less than 5 % of United States land, thus disposing of subdivisions over the oceans, or mostly across the Canada or Mexico borders. The final result of the subdivision creation process for the contiguous states can be seen in Figure 21.

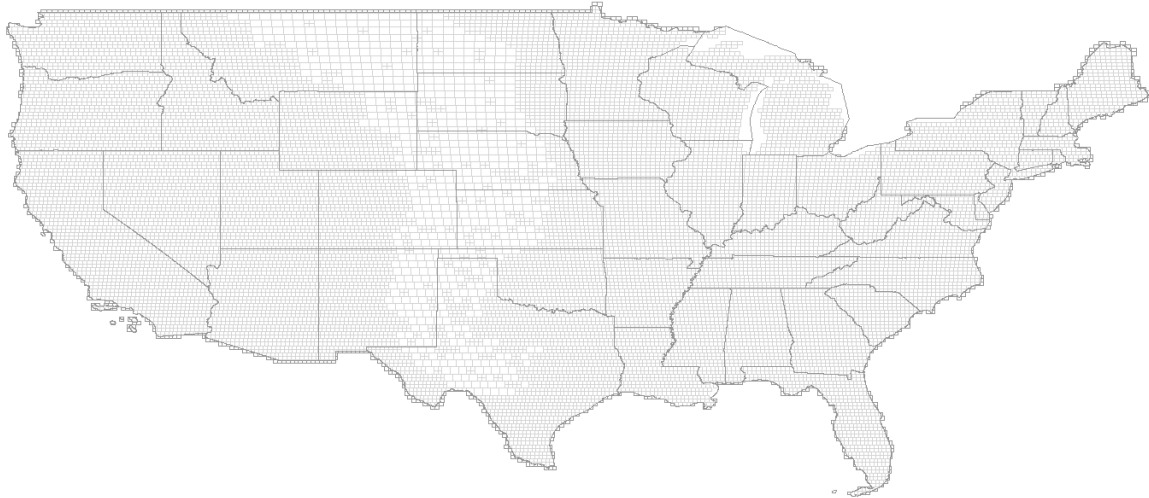


Figure 21: Subdivisions created over the contiguous states

Finally these subdivisions are assigned to one of the partitions described above according to the following process:

If at least 10 % of the subdivision is urban area, then the subdivision belongs to the “Urban” partition;

Else, the subdivision is rural:

If the average population density is greater or equal to 1.93 people per square km (i.e., 5 people per square mile), then the subdivision is allocated in the “Rural populated” partition;

Else, the subdivision is assigned to the “Rural low population” partition.

In any of these two cases, if the subdivision intersects with any of the highways in the National Highway System, the subdivision is also assigned to the “Highways” partition.

Once the subdivisions have been assigned to the appropriate partitions, we can proceed to classify the subdivisions in each of those partitions.

3.2.3. Classification

The classification of the subdivisions is performed by an unsupervised clustering algorithm: K-Means [19] using the K-Means++ initialization method [20]. This algorithm is capable of identifying items that have similar characteristics without previous training, through a series of iterations that associate each one of the items being classified to those that have the most similar set of characteristics.

One drawback of this classification method is that, as defined, it cannot identify on its own what is the best number of clusters to represent the diversity of subdivisions, as the number of clusters is an input to the algorithm. This problem can be solved by running series of classifications with an increasing number of clusters, until the gain (measured as the reduction of the total error in the classification over the total classification error) from adding additional clusters is negligible (in our case, we set the threshold to be 0.1 %,

which was found through experimentation to be the limit under which the error became asymptotic). Once we find the case where adding an additional cluster does not provide significant improvements in the classification ($C + 1$ clusters), we use the classification with one less cluster (C clusters), as that is the last classification in which all the clusters were meaningful. Therefore, a single classification process now consists of the following steps:

```

Initialize C, the number of clusters:  $C = 0$ .
Set the previous classification error to a large value.
Set the improvement to 1.
Do:
    Classify the subdivisions using  $C + 1$  clusters.
    Compute the total classification error.
    Compute the improvement.
    If the improvement is greater than 0.1 %:
        Increment C:  $C = C + 1$ .
    Else
        Discard the classification results.
    End If
While the improvement is greater than 0.1 %

```

As we have characterized each subdivision with a comprehensive set of parameters, we can improve the classification results by applying weights to these parameters, in order to emphasize some of those parameters while deemphasizing others. For example, while the minimum and maximum elevation provide a first indication of the variability of the terrain elevation in a subdivision, they are not as significant to the signal propagation as the combined average elevation and its standard deviation. Therefore, the weights applied to these parameters should be lower for the minimum and maximum elevation than for the average and standard deviation. After experimentation with different weights, the set chosen, which provided the best results regarding granularity of the results and differentiation of the different areas, were those shown in Table 5.

Table 5: Attribute weights used for the classification

Elevation Weights			
Average	Minimum	Maximum	Std. Deviation
1	0.5	0.5	0.8
Population Weights			
Average	Minimum	Maximum	Std. Deviation
0.5	0.1	0.1	1
User Weights			
Average	Minimum	Maximum	Std. Deviation
1	0.1	0.1	0.5

Clutter Weights	
Average	Std. Deviation
1	1

However, it is not possible to directly classify the subdivisions without considering the nature of the parameters being classified and the way K-Means works: In each iteration, K-Means computes the Euclidean distance between the items being classified, with each classification attribute being one dimension in an 18-dimensional space. This means that K-Means assumes that all the attributes have a similar scale, and therefore contribute equally to the distance. However, this is not the case for the subdivisions, as, for example, the elevation attribute ranges from -67 m to 6168 m, while the population density varies from 0 to 16 337 people per square kilometer. If we classified using these values directly, we would see how the different population densities are clearly identified by the classification algorithm, but the terrain features are lost. Adjusting the weights to compensate for the scale difference provides exactly the opposite result, as can be seen in Figure 22, where only the highest populated urban areas can be identified.

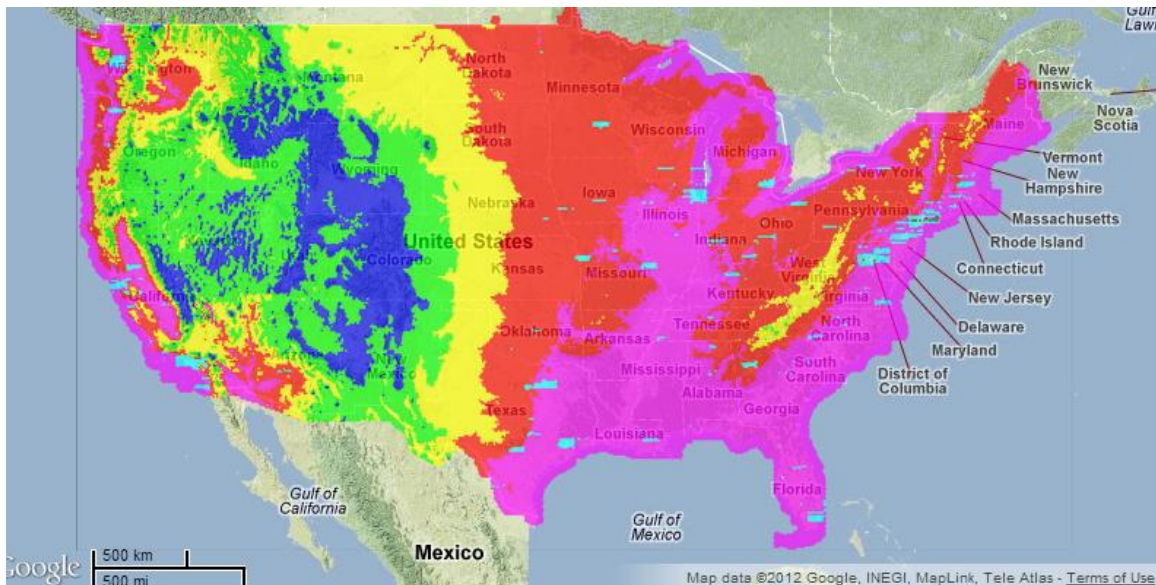


Figure 22: Classification result after using weights to adjust the scale differences between parameters

The explanation for this behavior can be found by looking at the distribution of the data for each of the attributes. In Figure 23 we can see the CDF of the four average attributes for the subdivisions (average elevation, average clutter height, average population density and average Public Safety user density). In this graph each series shows how many subdivisions we need to select (assuming they are sorted low to high value) to account for a given percentage of the maximum value for that series. For example, if we sort the subdivisions according to the average elevation, we will need to select 20 000 subdivisions to get an average elevation that is 50 % of the maximum value for that metric. Figure 23 shows that the data distribution varies greatly, with the population and Public Safety user density attributes having very low values in the vast majority of the

subdivisions; the steep increase in the CDF shows that there is a very small group of subdivisions that have high values of these attributes (i.e., the highest populated areas). On the other hand, the clutter height and the elevation attributes present more uniform distributions of values.

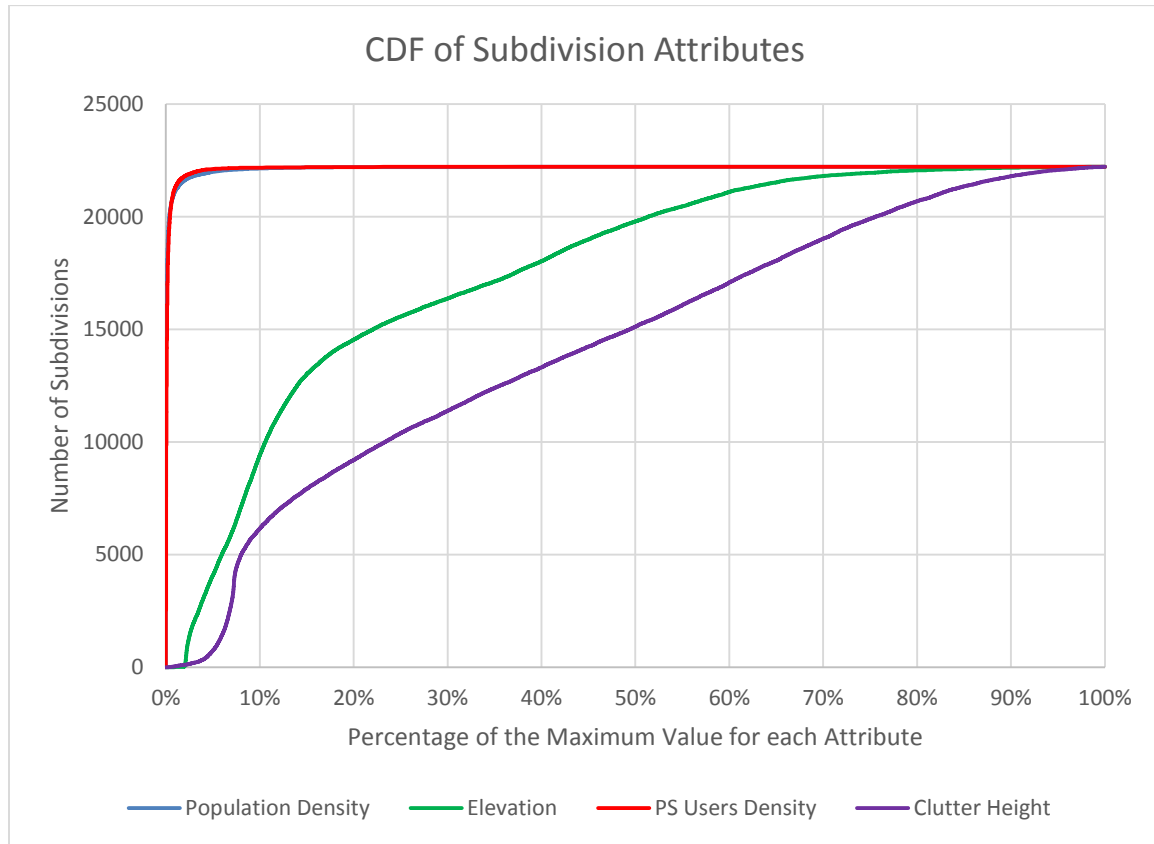


Figure 23 CDF of the average attributes of the subdivisions

In order to overcome the aforementioned problem, we treat each “set” of attributes (elevation, clutter height, population density, and Public Safety user density) independently, aggregating the results according to the following algorithm:

For each set of attributes:

If the population in the subdivision is 0, assign a special cluster ID;

Else classify the subdivisions based only on these attributes.

Store the resulting cluster ID for each subdivision.

End of for loop

Classify the subdivisions once more, this time using the cluster IDs of each of the previous classifications as the attributes, and redefining the distance function as follows:

If the cluster ID is the same, the distance is 0;

Else the distance is 1.

The subdivisions with population 0 are separated from the rest to avoid running RF analysis on them unless it is necessary to achieve the intended area coverage. Otherwise

they would be assigned to clusters with the rest of the subdivisions and would distort the results in the sampling and extrapolation stages.

With all these adjustments to the algorithm, conditions and processes, the results of the whole classification process can be seen in Table 6, Figure 24 (for the Urban partition); Table 7, Figure 25 (for the Rural Populated partition); Table 8, Figure 26 (for the Rural Low Population partition); and Table 9, Figure 27 (for the Highways partition).

Table 6: Cluster details for the Urban partition

Cluster ID	Number of subdivisions	Percent of total area	Mean Population Density (people per km ²)	Mean Elevation (m)	Mean Elevation Std Dev. (m)	Legend Color
1	332	19.774	744.45	151.32	41.12	Red
2	632	37.641	223.14	158.75	34.58	Green
3	254	15.128	226.91	257.88	64.06	Blue
4	248	14.771	684.32	148.2	47.84	Yellow
5	38	2.263	305.3	1392.42	230.71	Magenta
6	75	4.467	175.83	1128.95	151.74	Cyan
7	40	2.382	553.64	40.09	25.6	Grey
8	13	0.774	2166.79	60.65	34.46	Brown
9	47	2.799	581.29	1089.86	213.58	Dark Green

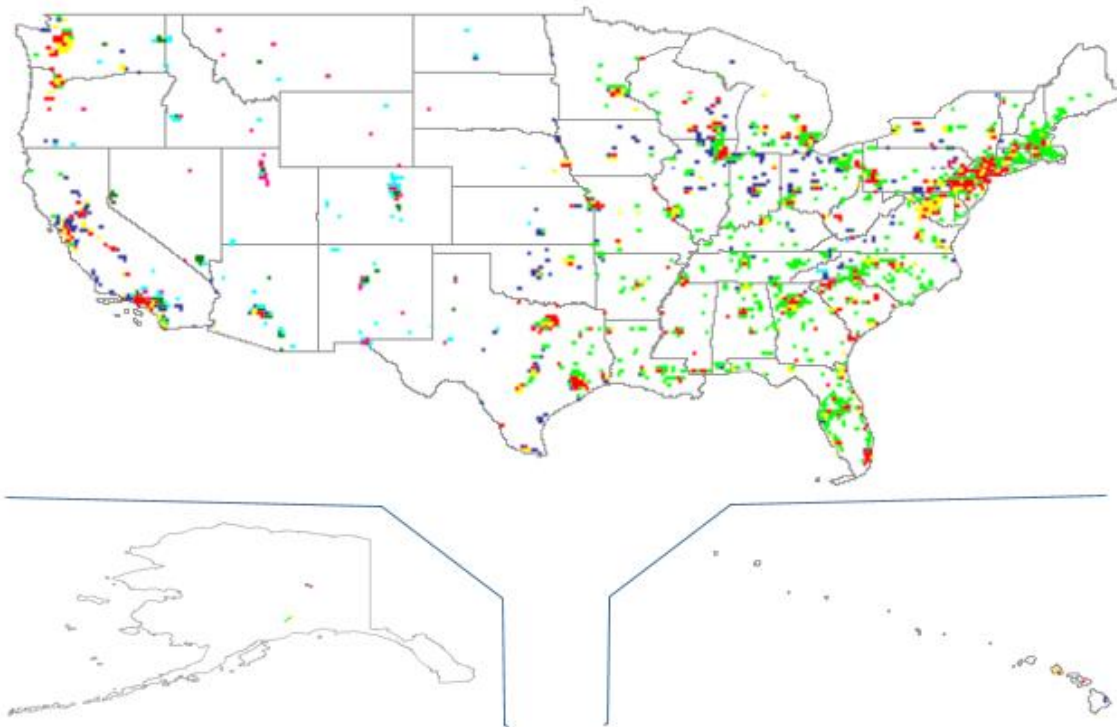


Figure 24: Classification results for the Urban partition

Table 7: Cluster details for the Rural Populated partition

Cluster ID	Number of subdivisions	Percent of total area	Mean Population Density (people per km ²)	Mean Elevation (m)	Mean Elevation Std Dev. (m)	Legend Color
1	1663	18.855	30.35	324.14	51.73	Red
2	3135	35.544	22.65	309.98	73.24	Green
3	1563	17.721	9.11	431.72	96.32	Blue
4	1189	13.481	17.32	494.49	55.65	Yellow
5	1270	14.399	29.41	620.61	59.85	Magenta

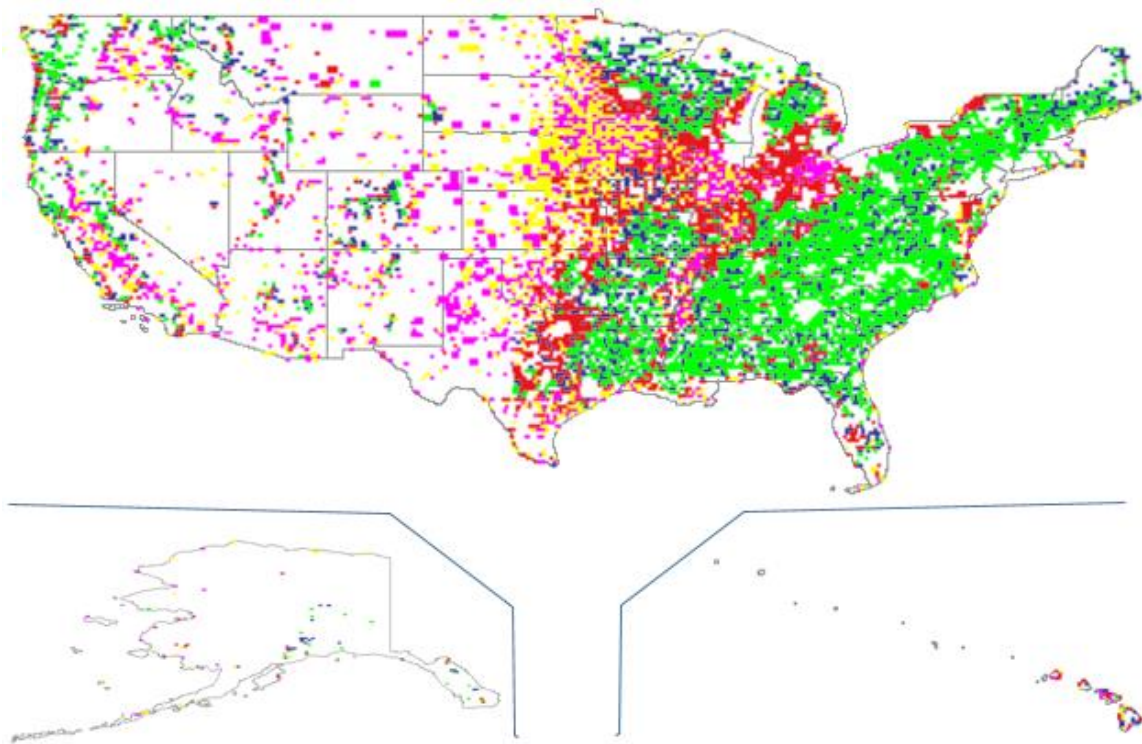


Figure 25: Classification results for the Rural Populated partition

Table 8: Cluster details for the Rural Low Population partition

Cluster ID	Number of subdivisions	Percent of total area	Mean Population Density (people per km ²)	Mean Elevation (m)	Mean Elevation Std Dev. (m)	Legend Color
1	1262	10.773	0.18	347.62	92.33	Red
2	2168	18.508	0.17	1581.96	189.19	Green
3	1248	10.654	0.25	631.5	133.68	Blue
4	1823	15.563	0.16	579.59	233.12	Yellow

5	1399	11.943	0.24	1782.75	324.98	
6	1638	13.983	0.24	651.98	117.62	
7	1328	11.337	0.85	956.52	125.31	
8	848	7.239	1.04	822.18	145.54	

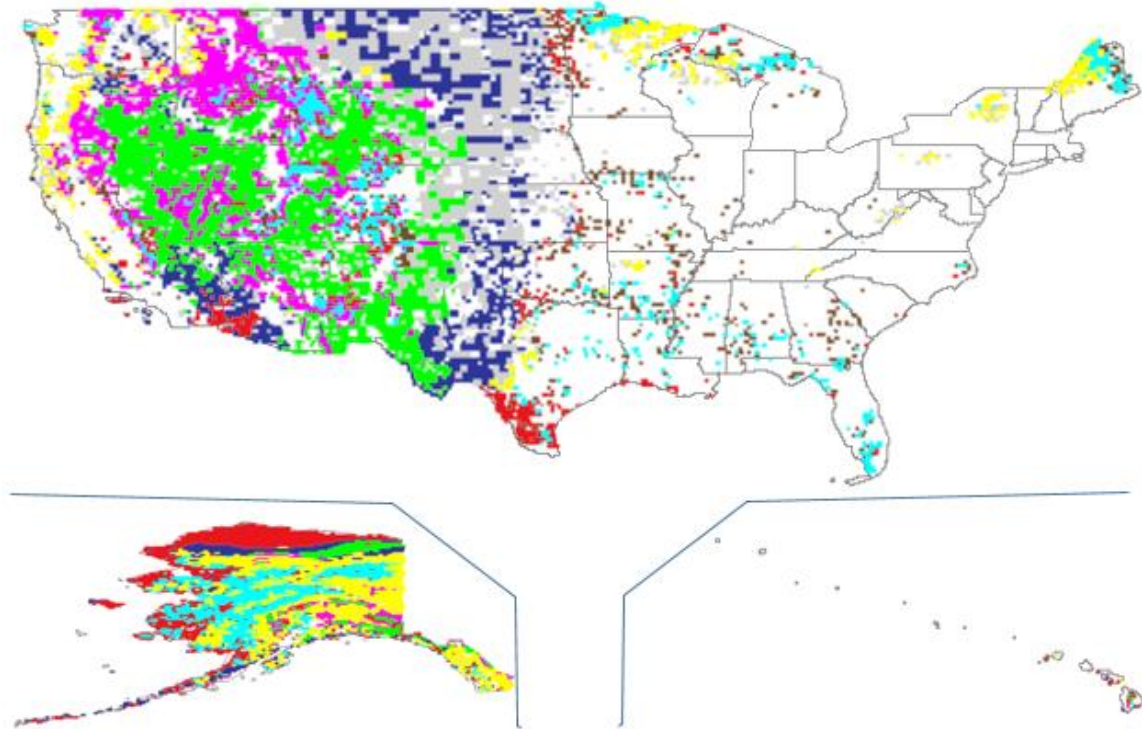


Figure 26: Classification results for the Rural Low Population partition

Table 9: Cluster details for the Highways partition

Cluster ID	Number of subdivisions	Percent of total area	Mean Population Density (people per km ²)	Mean Elevation (m)	Mean Elevation Std Dev. (m)	Legend Color
1	1483	17.71	6.57	566.09	130.55	
2	1434	17.124	39.57	475.12	54.74	
3	2315	27.645	29.13	291.46	66.89	
4	1171	13.984	9.51	743.99	126.6	
5	1971	23.537	5.83	974.93	91.63	

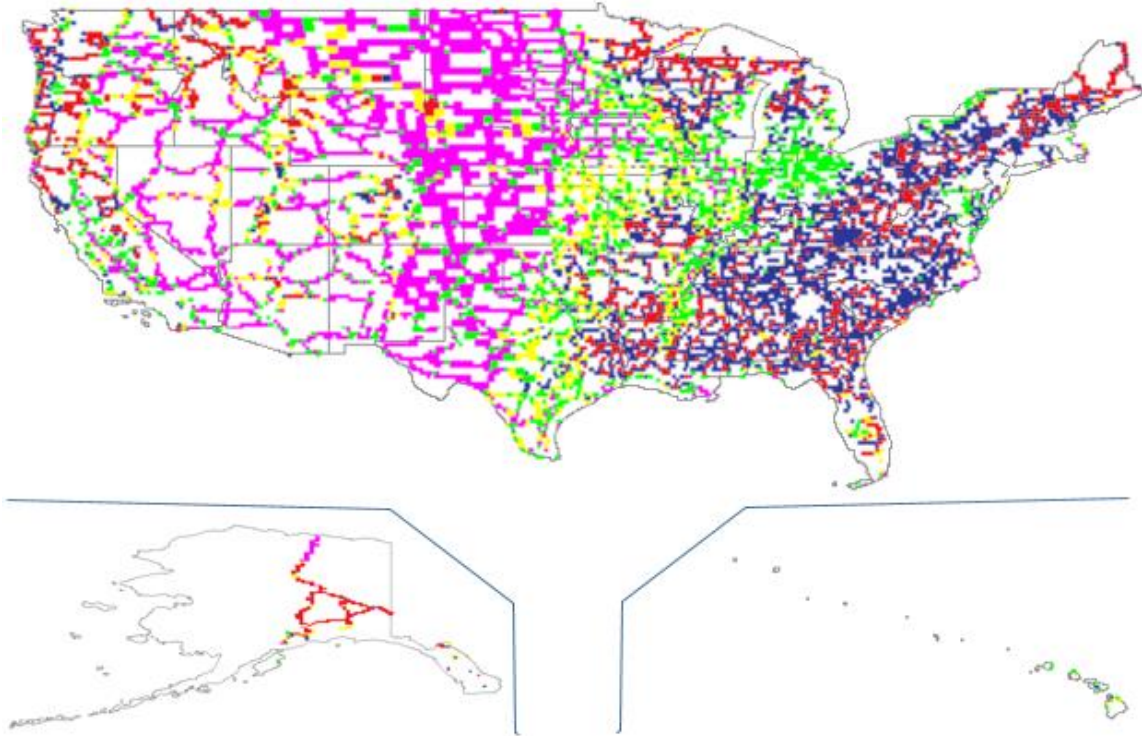


Figure 27: Classification results for the Highways partition

3.3. Sampling and Analysis

We use random sampling to develop an estimate of the number of sites required at the national level. Because a nationwide random sample may miss areas with high population counts, we rely on a stratified sampling approach that uses the subdivision classification technique described in Section 3.2. As we do not know the number of samples to take from each class *a priori*, we developed an iterative algorithm that operates on each class, which we show in Figure 28. In this Section, we describe the algorithm in detail.

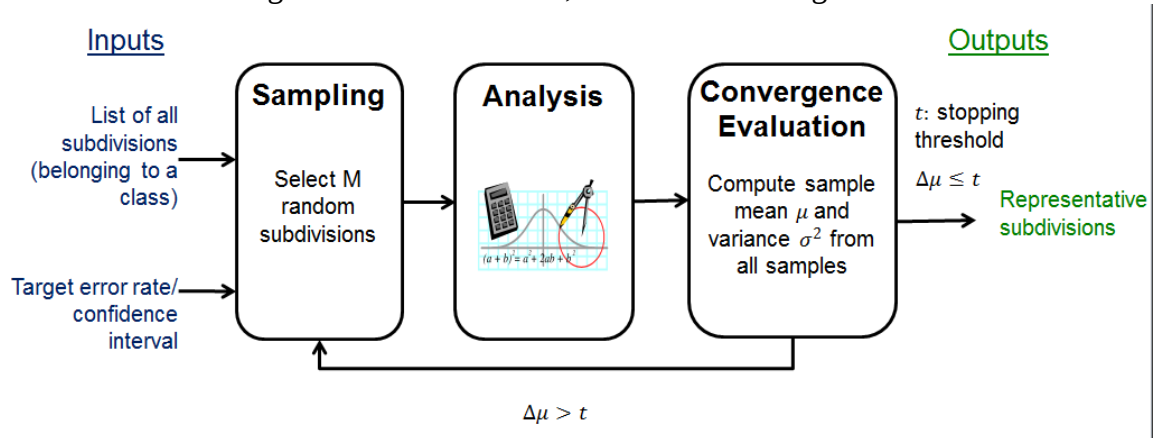


Figure 28: Sampling and analysis algorithm.

When we perform a random sample from a population of N objects, we choose n of the objects at random. In this case, the objects that compose the population are the

subdivisions. We measure some quantity of interest y (in our case, the number of cell sites in a subdivision) in each sampled object, and we define y_i to be the observed value of y in the i th subdivision in the sample. The sample average is $\hat{\mu} = \sum_{i=1}^n y_i / n$. The sample mean is an unbiased estimator, meaning that its expected value, $E\{\hat{\mu}\}$, is equal to μ , the true mean of y .

The primary performance metric for the estimate $\hat{\mu}$ is the mean squared error (MSE), which is also the variance of $\hat{\mu}$: $\text{MSE} = E\{(\hat{\mu} - \mu)^2\} = E\{\hat{\mu}^2\} - \mu^2$. If the observed values of y in different samples are independent, then $E\{y_i y_j\} = \mu^2$ when $i \neq j$; when $i = j$, we get $E\{y_i^2\} = \sigma^2 + \mu^2$, where σ^2 is the variance of y . Since $\hat{\mu}^2 = \sum_{i=1}^n \sum_{j=1}^n y_i y_j / n^2$, we get

$$\text{MSE} = \frac{(n^2 - n)\mu^2 + n(\sigma^2 + \mu^2)}{n^2} - \mu^2 = \frac{\sigma^2}{n}.$$

The MSE shrinks as the number of samples increases. Also, if n is sufficiently large, the Central Limit Theorem tells us that $\hat{\mu}$'s distribution converges to a normal distribution with mean μ and whose variance is equal to the MSE. Thus, the p -confidence interval centered on $\hat{\mu}$ is $[\hat{\mu} - t\sigma, \hat{\mu} + t\sigma]$; the probability that the true mean μ lies in the interval is $p = \Pr\{\hat{\mu} - t\sqrt{\text{MSE}} \leq \mu \leq \hat{\mu} + t\sqrt{\text{MSE}}\}$, since the MSE is the variance of $\hat{\mu}$. If we know p , t is the quantity that solves

$$p = 1 - \frac{2}{\sqrt{2\pi}} \int_t^\infty \exp(-w^2/2) dw.$$

For a 95 % confidence interval, $t = 1.96$.

Once we have our estimate of the average value of the quantity of interest, y , we can estimate Y , the total amount of x in the entire population, by computing $\hat{Y} = N\hat{\mu}$. This is also an unbiased estimate, since its mean is $E\{\hat{Y}\} = N\mu = Y$. The MSE of the estimate of the total is $N^2\text{MSE}$, which we can use to generate a confidence interval as described above.

To do a stratified sample, we break the population of N objects into L classes and sample from each one. The number of objects in the h th class is N_h , and we define the ratio $W_h = N_h/N$ to be the weight of the h th class. Then the stratified estimate is $\hat{\mu}_{st} = \sum_{h=1}^L W_h \hat{\mu}_h$, where $\hat{\mu}_h = \sum_{i=1}^{n_h} y_{h,i} / n_h$. In this estimate, n_h is the number of samples taken from the h th class, and $y_{h,i}$ is the i th sample from the h th class. The resulting sampling fraction for the h th class is $f_h = n_h/N_h$. This estimate is also unbiased, since $E\{\hat{\mu}_h\} = E\{y|h\}$ and $\mu = \sum_{h=1}^L E\{y|h\}\Pr\{h\}$ by Bayes' Theorem, and $\Pr\{h\} = N_h/N$. The mean squared error of the estimate is

$$\sigma_{\hat{\mu}_{st}}^2 = \sum_{h=1}^L W_h^2 E\{(\hat{\mu}_h - \bar{y}_h)^2\} = \sum_{h=1}^L W_h^2 \frac{S_h^2}{n_h} (1 - f_h),$$

Unfortunately, this expression for the mean squared error relies on S_h^2 , which can't be known without examining the entire population of the h th class. To compute the confidence intervals, we must use the following estimator of $\sigma_{\hat{\mu}_{st}}^2$:

$$\varsigma_{\hat{\mu}_{st}}^2 = \sum_{h=1}^L W_h^2 \frac{s_h^2}{n_h} (1 - f_h),$$

Where s_h is the sample standard deviation of y within the h th class: $s_h^2 = \sum_{i=1}^{n_h} (y_{h,i} - \hat{\mu}_h)^2 / (n_h - 1)$. The p -confidence interval is $\hat{\mu}_{st} \pm t(p) \varsigma_{\hat{\mu}_{st}}$. Once we have the stratified estimate, the estimate of Y is $\hat{Y} = N \hat{\mu}_{st}$, with p -confidence interval $N \hat{\mu}_{st} \pm N t(p) \varsigma_{\hat{\mu}_{st}}$. If the statistics of the population being sampled are known, they can be used to compute what proportion of samples should be taken from each class; unfortunately, this is not the case here. Thus, we use the following iterative procedure, which is depicted in Figure 28. After taking a small pilot sample and generating site placements in each of the chosen subdivisions, we compute the sample mean and sample variance of the site counts from the pilot sample. We also use the sample standard deviation to generate the confidence interval for the estimated site count. Next, we choose additional subdivisions at random from within the class and update the sample statistics. Once the confidence interval is below a predefined target, we stop the sampling and return the mean site for the class. We multiply this quantity by the number of subdivisions in the class to get the total site count for the class.

3.4. Extrapolation

The results of detailed analyses are used to compute the site count and coverage for subdivisions not analyzed, meeting the nationwide coverage objectives. This process involves two steps, namely, class extrapolation and nationwide aggregation.

3.4.1. Class Extrapolation

The sampling and analyses performed in each class provides an estimation of the mean site density S_c with a given confidence interval and coverage reliability. The total number of sites needed to cover all the subdivisions in a class can then be computed as: $T_c = \sum_{n=1}^N (S_c \cdot A_n)$, where N is the number total number of subdivisions in the class and A_n is the land area of subdivision n . Using the land area allows to adjust the site count in regions partially covered with water or near the borders. To some extent, the same formula can be applied to estimate the number of sites needed to cover a subset of the subdivisions in a class for a large subset. Using the average on a small subset will decrease the confidence of the estimated site count.

3.4.2. Nationwide Aggregation

In order to obtain an estimation of the number of sites for a nationwide coverage scenario, the results of different partitions are aggregated together. The first step is to select the configuration for each of the partition (Urban, Rural Populated, Rural Low Population, and highways). The second step is to define a nationwide coverage scenario that specifies the areas to cover, which can be based on population, Public Safety users, or area covered, and may also include minimum coverage per state or county. The selection method used is significant, as it will determine which areas are selected first, which areas require a minimum coverage, etc...

The results presented in this document use the following selection method:

1. Sort the subdivisions by population, high to low.
2. Select the subdivisions that meet the target population to cover, starting with the highest population.
3. Add the subdivisions from the national highway systems that have not been selected in step 2.

Figure 29 shows the areas selected to cover 99.9 % of the population and the national highway systems. With this target coverage, all the subdivisions in urban areas and rural areas with a population density of at least 5 people per square mile have been selected. The set of selected areas also includes a large portion of the Great Plains. We can also observe that the highways are covered in the rural areas with low population densities. The land area in this selection represents 64.6 % of the total land area.

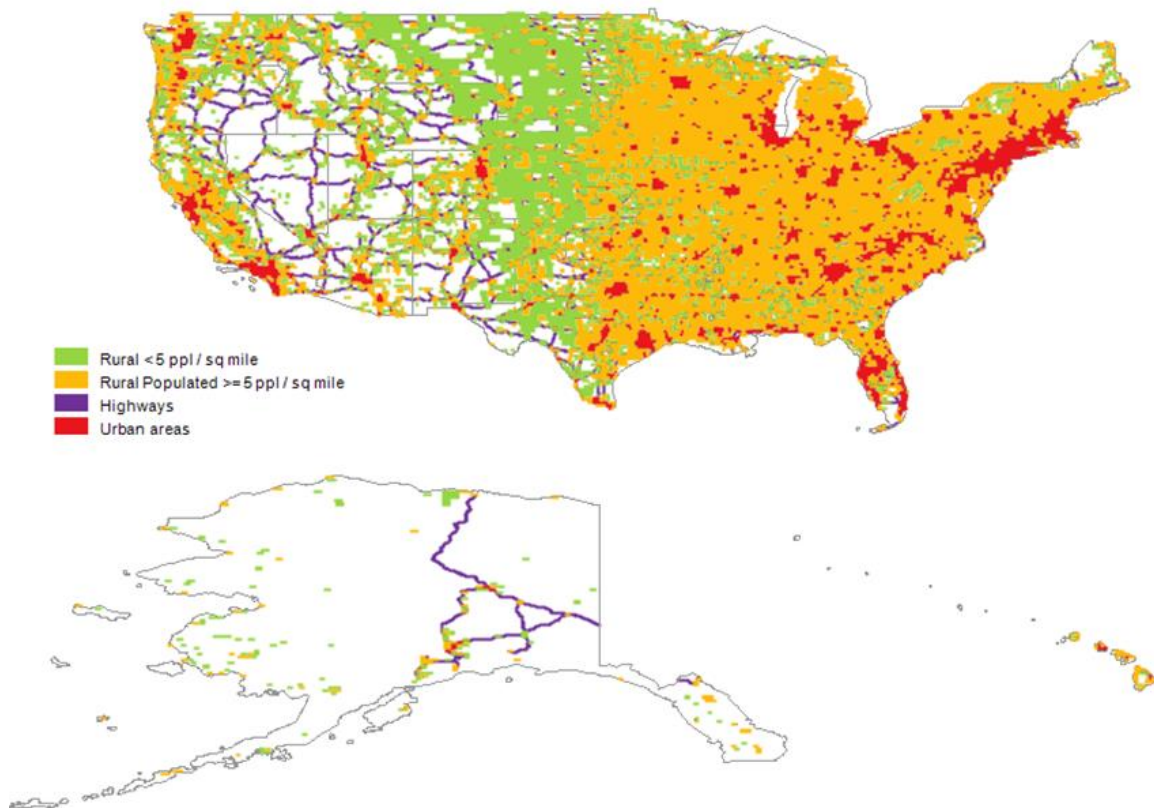


Figure 29: Coverage map for 99.9 % population coverage

4. Sensitivity Analyses

Several analyses were conducted to evaluate the impact of specific parameters on the performance of an LTE network for Public Safety, and more specifically, their impact on the number of sites required to achieve the target coverage criteria.

4.1. Nationwide Coverage

To perform the nationwide sensitivity analyses, several sets of configuration parameters were considered for each partition, and then aggregated into possible nationwide deployment scenarios. The details of these sets are described below.

4.1.1. Assumptions

The results of the analyses presented are directly based on the assumptions described in this document and those assumptions should be kept in mind when interpreting the results. Changing assumptions will lead to different results; however, the overall trends highlighted in this document (e.g., effect of UE power on site count) will not change.

4.1.1.1. Clutter configuration

The results presented in this document use a 30 m resolution with no building information. As such, a single clutter height value was assigned to each class. In addition, a penetration loss was added to the clutter classes that are likely to represent buildings and are used to simulate the loss as the signal penetrates through the wall. This loss was added to the clutter losses used by the propagation model to compute the predictions. The sets of values used (both height and loss) are shown in Table 10.

Table 10: Clutter configuration

Class	Height (m)	Indoor Penetration Loss (dB)
Airport	0	0
Commercial - Industrial	10	10
Forested - Dense Vegetation	7	0
Grass - Agriculture	0.7	0
High Density Urban	20	20
Marsh - Wetland	1	0
Open	1	0
Residential with Few Trees	3	6
Residential with Trees	4	6
Transportation	1.4	0
Water	0.5	0
Shrubland - Woodland	1	0

4.1.1.2. Network Configuration

The assumptions made about the network configuration have a large impact on the performance of the network, such as the type of antenna used or interference coordination scheme. Table 11 shows the list of parameters common to all partitions while Table 12 shows the list of configuration specific parameters applied to each partition.

Table 11: Common network parameters

Parameter	Value
Propagation model	CRC Predict
Frequency bandwidth (MHz)	2x10
Slow fade standard deviation (dB)	7
eNodeB	
Transmit power per antenna (dBm)	Configuration-based
Uplink power control	Fractional (P0=-32.6 dBm, alpha=0.4)
Antenna model	LNx-6515DS-VTM_0725
Max antenna gain (dB)	16.7
MIMO	2x2
ICIC scheme	Dynamic, RSRQ Threshold = 0 dB, 33 % outer cell resources
Sector azimuths (degree)	0, 120, 240
Sector downtilt (degree)	0
Noise figure (dB)	2.5
Subscriber	
Transmit power (dBm)	Configuration-based
Antenna gain (dBi)	Configuration-based
MIMO	1x2
Noise figure (dB)	9
Body loss (dB)	0
User Distribution	See Section 2.1.3
Traffic Model	See Section 2.1.4
Number of RBs per user	Configuration-based
Site placement	
Weight PS sites	1
Weight Commercial sites	0.75
Weight green field sites	0.5
Antenna height for green field sites (m)	75 m in Great Plains 30 m in urban areas and mountains 50 m in rural areas not in the mountains or Great Plains
Green field site spacing (km)	4
Analysis	
Coverage reliability (%)	Configuration-based
Population coverage (%)	95
Geodata resolution (m)	30 m in urban, 60 m in rural

Table 12: Configuration specific network parameters

Partition	Configuration	eNodeB Tx Power (dBm)	UE Tx Power (dBm)	UE Gain (dBi)	Coverage Reliability (%)	Uplink RBs	Down-link RBs
Urban	95 % reliability, indoor	46	23	-4	95	1 to 5	5
	95 % reliability, outdoor	46	23	-4	95	1 to 5	5
Rural ≥ 5 people per mile²	95 % reliability, outdoor	46	23	0	95	1 to 5	5
	85 % reliability, outdoor	47	31	0	85	1 to 5	7
Rural < 5 people per mile²	85 % reliability, outdoor	47	31	0	85	1 to 5	7
Highways	95 % reliability, outdoor	46	23	0	95	1 to 5	5
	85 % reliability, outdoor	47	31	0	85	1 to 5	7

4.1.2. Results

Various configurations have been studied to investigate potential nationwide deployment scenarios, and Table 13 shows the values used in each of those configurations. The first scenario provides an indoor environment for the urban areas (using the penetration losses described in Section 4.1.1.1), with 95 % reliability except in rural areas with low population, and is considered the baseline. In the second scenario, the urban coverage is changed to be outdoor only. The third scenario mainly considers a lower reliability in all the rural areas, while the last scenario combines an outdoor coverage for the urban areas and a lower reliability in the rural areas.

Table 13: Nationwide scenarios

Scenario	Urban areas	Rural areas with pop. density ≥ 5 people per mile ²	Rural areas with pop. density < 5 people per mile ²	National Highway System
Urban indoor and 95 % rural reliability (except very rural)	95 % reliability Indoor environment	95 % reliability Outdoor environment	85 % reliability Outdoor environment	95 % reliability Outdoor environment

Urban outdoor and 95 % rural reliability (except very rural)	95 % reliability Outdoor environment	95 % reliability Outdoor environment	85 % reliability Outdoor environment	95 % reliability Outdoor environment
Urban indoor and 85 % rural reliability	95 % reliability Indoor environment	85 % reliability Outdoor environment	85 % reliability Outdoor environment	85 % reliability Outdoor environment
Urban outdoor and 85 % rural reliability	95 % reliability Outdoor environment	85 % reliability Outdoor environment	85 % reliability Outdoor environment	85 % reliability Outdoor environment

The population and area coverage that can be achieved for each scenario based on the number of sites available, assuming that the most populated areas are selected first, is shown in Figure 30 and Figure 31 respectively. These figures also compare the number of sites needed to achieve a particular population or area coverage with the different configurations. The sampling was configured to estimate the site count with a margin of error of $\pm 10\%$ at an 85 % level of confidence. We observe that the baseline scenario requires the highest number of sites for any population target with 39 000 sites to cover 95 % of the population and 50 % of the US land area. This is to be expected, since it has the most stringent coverage criteria (indoor and high reliability). By providing outdoor coverage in urban areas or reducing the coverage reliability the number of sites needed is lower, estimated at 33 600 and 27 800 sites respectively. Reducing the coverage requirements in both urban and rural areas further decreases the number of sites, down to 22 000 for 95 % population coverage. In all cases, we notice that the number of sites needed increases exponentially when targeting more than 99 % population coverage. Figure 31 shows that the last 1 % of the population is spread through 38 % of the US land.

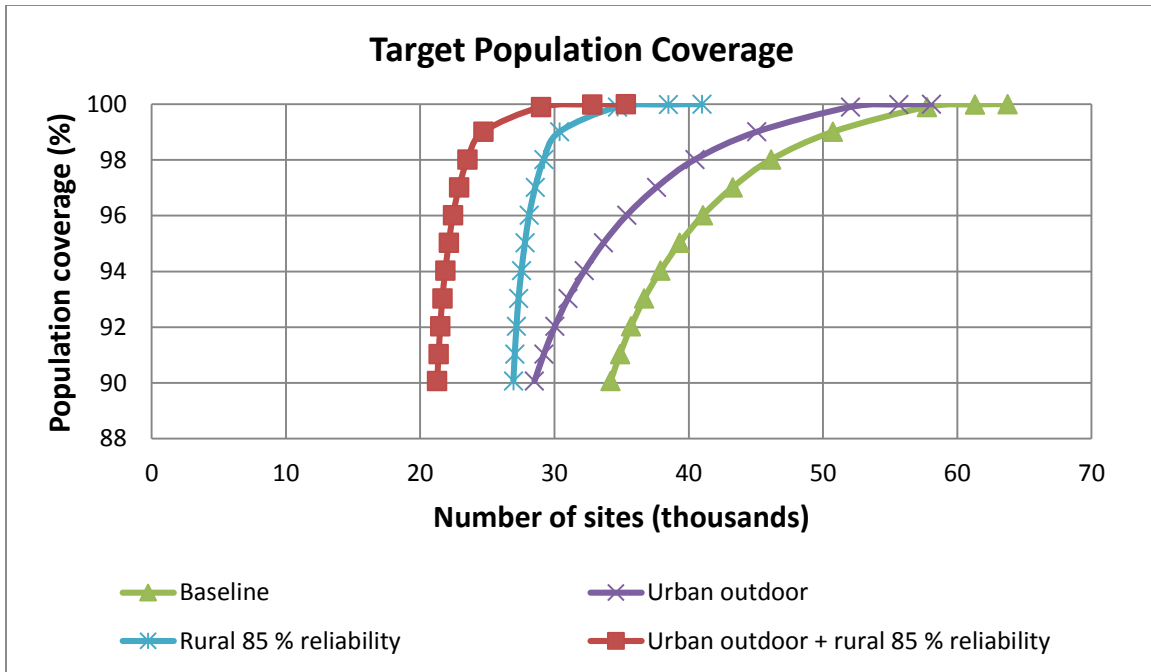


Figure 30: Nationwide population coverage

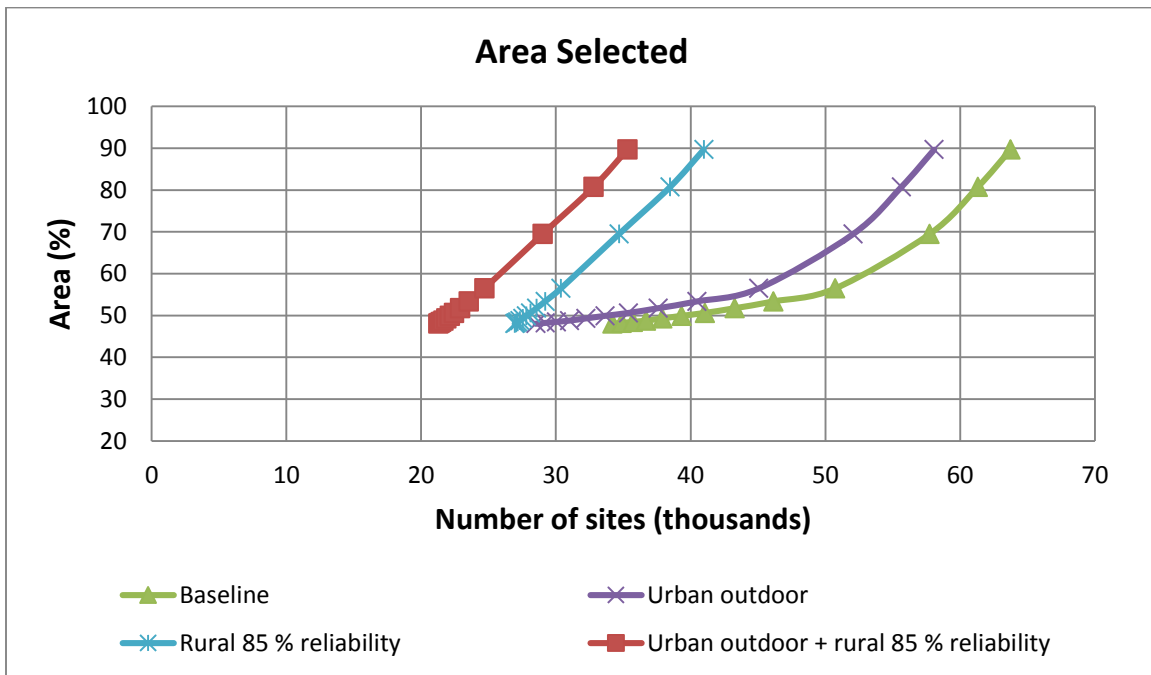


Figure 31: Nationwide area coverage

4.2. Impact of High Power devices

The use of high power devices by Public Safety in Band 14 has been proposed in order to extend the coverage [22], with the assumption that the communication is more likely to be limited by the uplink transmissions. However, increasing the transmit power of the user devices also creates challenges because it potentially increases the interference to the

neighboring sectors, thus reducing the SINR values for the users on those sectors, which in turns limit the MCS that can be used by the UEs, resulting in an increase of the sector loads to maintain the user data rates. Since the level of interference generated is based on the number of users transmitting in a sector, an analysis was performed to characterize the limitations of using a higher transmit power. In this analysis, a site plan was generated for various regions with increasing user densities. The site plans were generated using two UE power configurations, 23 dBm and 31 dBm. To take full advantage of the additional power, the fractional uplink power control settings were also adjusted, changing the target received power P_0 and keeping the compensation factor Alpha the same, as shown in Table 14.

Table 14: Fractional power control settings

UE Power (dBm)	UE Power (W)	P0	Alpha
23	0.2	-32.6	0.4
31	1.2	-27.8	0.4

Figure 32 plots the differences in the number of sites required when using a UE transmit power of 0.2 W compared to 1.2 W as a function of the user density of that area, i.e., $\text{Difference} = N_{0.2\text{ W}} - N_{1.2\text{ W}}$. Each area analyzed has unique features, leading to different number of sites even for similar user densities. For the same reason, one configuration is not always going to be better than the other one and we need to look at the trend (dashed red line on the graph). The fit curve shows that for areas with low user densities, the number of sites with 0.2 W UEs is higher than the one with 1.2 W UEs. However the benefit decreases as the user density increases and beyond a user density of 1.35 users per km^2 , on average, more sites are required when using the higher power UEs.

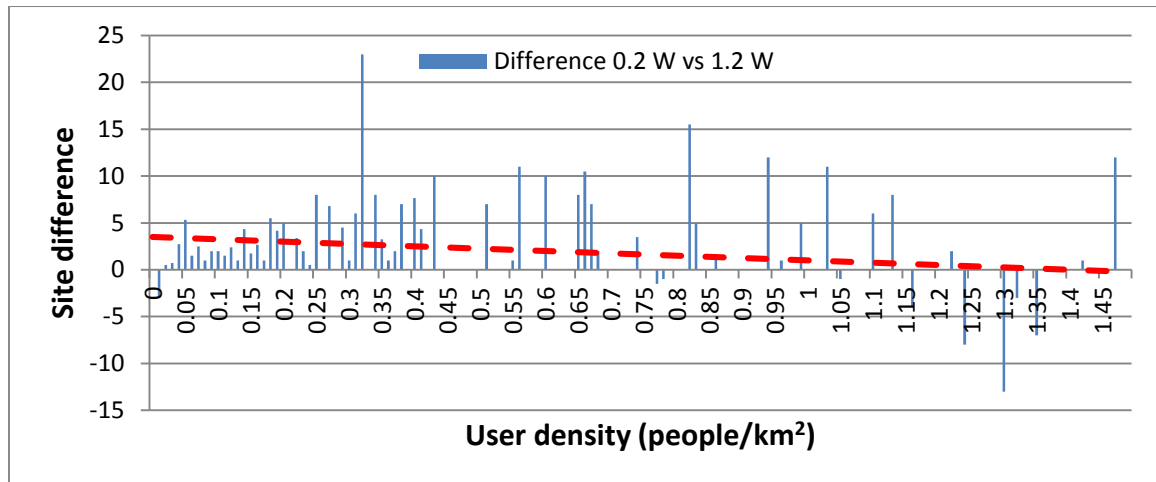


Figure 32 Differences in the number of sites required between 0.2 W and 1.2 W

Figure 33 shows the Cumulative Distribution Functions (CDF) of the user density for all the subdivisions in the US and in various nationwide coverage criteria. We observe that about 92 % of the US territory has a user density less than 1.35 users per km^2 . However,

not all of that area will be covered, and it is necessary to look at the actual selection method (as described in Section 3.4.2) to determine how much of the area selected will benefit from the high power UEs. For example, if we consider target population coverages of 99.9 % and 99 %, respectively 88 % and 85 % of the coverage area can benefit from the high power UEs.

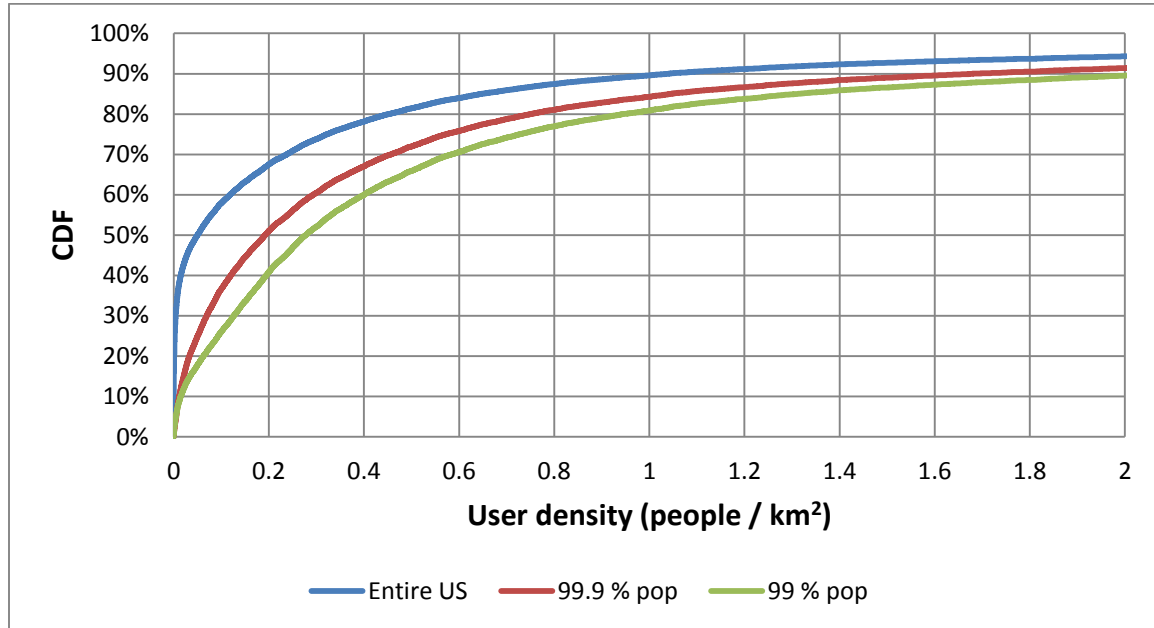


Figure 33: CDF of user density

5. Conclusion

This document described a method for modeling site deployments for the future NPSBN. We discussed the types of inputs that must be provided for accurate RF modeling, such as elevation and clutter data, as well as assumptions regarding the network configurations and the traffic models. We also described the steps involved in performing an area analysis, starting with the site selection and continuing up to the validation of the network performance. To resolve the scalability issues caused by modeling the entire United States, our approach performs detailed analysis of sample areas and then extrapolates the results nationwide. The statistical selection of those samples ensures that the extrapolated results are correct and precise within known limits, and quantifies the uncertainty of the results.

The nationwide results presented demonstrate the impact of a few key parameters, namely indoor coverage and reliability. Intuitively, it is clear that trying to provide indoor coverage or increasing the coverage reliability will increase the number of sites needed. The benefit of the model is to be able to quantify the impact associated with varying the parameters in order to perform cost-benefit analyses. We also looked at the use of high power UEs to increase the cell coverage, thus reducing the number of sites needed. However, we observed that there are some limitations and the improvements are noticeable only in rural areas.

As LTE networks are being deployed, data usage will be collected and used to refine the traffic models. Additionally, the LTE standard is still evolving and will affect the operations of the network by providing new capabilities (e.g., higher modulation, advanced interference coordination, carrier aggregation). For all these reasons, the actual figures obtained may change in the future, as further analyses are run to ascertain the impact of these new or updated inputs to the process. The method presented, on the other hand, is resilient to these changes and will remain unmodified.

6. References

- [1] United States Government, "MIDDLE CLASS TAX RELIEF AND JOB CREATION ACT OF 2012," 2012.
- [2] R. Rouil, A. Izquierdo, M. Souryal, C. Gentile, D. Griffith and N. Golmie, "Nationwide Safety: Nationwide Modeling for Broadband Network Services," *IEEE Vehicular Technology Magazine*, vol. 8, no. 2, pp. 83-91, 2013.
- [3] I. M. Chakravarti, R. G. Laha and J. Roy, *Handbook of Methods of Applied Statistics, Volume I*, John Wiley & Sons, 1967.
- [4] Bureau of Justice Statistics, "Federal Law Enforcement Officers," 2008. [Online]. Available: <http://www.bjs.gov/index.cfm?ty=pbdetail&iid=4372>.
- [5] Bureau of Justice Statistics, "Census of State and Local Law Enforcement Agencies," 2008. [Online]. Available: <http://www.bjs.gov/index.cfm?ty=pbdetail&iid=2216>.
- [6] United States Office of Personnel Management, "Federal Employment," 2012.
- [7] United States Fire Administration, "National Fire Department Census Database," [Online]. Available: <http://apps.usfa.fema.gov/census>.
- [8] United States Department of Labor, "Occupational Outlook Handbook," Bureau of Labor Statistics, 2010. [Online]. Available: <http://www.bls.gov/ooh/>.
- [9] United States Department of Labor, "Occupational Employment and Wages," Bureau of Labor Statistics, May 2013. [Online]. Available: <http://www.bls.gov/oes/current/oes292041.htm>.
- [10] Oak Ridge National Laboratory, "LandScan Home," Computational Sciences & Engineering Division, [Online]. Available: <http://web.ornl.gov/sci/landscan/index.shtml>.
- [11] N. Golmie, C. Gentile, D. Griffith, A. Izquierdo, R. Rouil, M. Souryal and W.-B. Yang, "Network modeling and analysis of a public safety broadband network," in *Public Safety Broadband Stakeholder Conference*, Westminster, CO, USA, 2013.
- [12] National Public Safety Telecommunications Council, "Public Safety Communications

Assessment 2012-2022," Littleton, CO, USA, 2012.

- [13] First Responder Network Authority, "Use cases for interfaces, applications, and capabilities for the nationwide public safety broadband network," Public Safety Advisory Committee , 2014.
- [14] TeleVate, "Public Safety Wireless Data Network Requirements Project - Wireless Data Network Implementation Model," Minnesota Department of Public Safety, Falls Church, VA, USA, 2012.
- [15] K. Tutschku and P. Tran-Gia, "Spatial traffic estimation and characterization for mobile communication network design," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 804-811, 1998.
- [16] R. Church and C. R. Velle, "The maximal covering location problem," *Papers in regional science*, vol. 32, no. 1, pp. 101-118, 1974.
- [17] Federal Communications Commission, "The Universal Licensing System (ULS) database," [Online]. Available: <http://wireless.fcc.gov/uls/index.htm?job=transaction&page=weekly>. [Accessed 24 10 2012].
- [18] United States Department of Transportation, Federal Highway Administration, "The National Highway Planning Network," 2011. [Online]. Available: <http://www.fhwa.dot.gov/planning/processes/tools/nhpn/>. [Accessed 29 09 2014].
- [19] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1*, pp. 281-297, 1967.
- [20] D. Arthur and S. Vassilvitskii, "K-Means++: The Advantages of Careful Seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.*, Philadelphia, USA, 2007.
- [21] W. G. Cochran, Sampling Techniques, New York: John Wiley & Sons, Inc, 1963.
- [22] 3rd Generation Partnership Project, "TR 36.837 - Public safety broadband high power User Equipment (UE) for band 14 (Release 11)," 2012.