Chapter 7

Drive-test and Crowd-Sourced Coverage Mapping

The previous chapter proposed a method for wireless network coverage mapping using geostatistical techniques and carefully selected measurements. This thesis argues that the most interesting and opportune area for coverage mapping is where the rubber meets the road: empicial, measurement-based techniques. However, it is not always possible or affordable to collect measurements on a uniform lattice. This chapter investigates two methods of more "convenient" sampling: drive-test measurements and crowd-sourced measurements. Drive-test measurements are the defacto method of measurement used by cell providers and many network planning consultants. These measurements are made by vehicles and are confined to city streets. While previous chapters have argued that this sort of measurement has clear sources of systematic bias, particularly from the wave-guiding effects of street-canyons, whether these measurements may still be useful for some general prediction is an open question this chapter will seek to address.

The second type of "convenience" sampling, is crowd-sourced measurements where a possibly large number of volunteers may choose to use their mobile devices to collect and contribute mesurements to a coverage mapping campaign. This approach has the pleasing attribute that crowd-based sampling will oversample the regions of highest interest (i.e., where people go) and undersample the areas of lowest interest (where they do not). Although this is still an emerging idea with a great deal of work needed to understand how well this may work, this chapter offers an initial step. First, a mobility model is used to simulate walking patterns and understand where measurements might be made in a system like this, and how often. These results show that crowd-sourced coverage mapping is likely feasible in terms of practical participation levels and the data generated as a result are likely sufficiently complete. Then, a case study

is performed using data collected by the Open Signal Maps (OSM) application on real users cell phones in Colorado, so that the limitations of real data collected this way can be understood. Although the data is sparse, geostatistical approaches still perform admirably, suggesting that they may be successfully applied in this domain.

7.1 Drive Test Coverage Mapping

As a first example, consider the data collected from the municipal wireless network in Portland, Oregon. The measurements collected for this network are described in sections 3.1 and 4.2.1. Figure 4.1, shows the collected measurements overlayed on orthoimagery. This dataset is a typical exhaustive drive-test, of the sort commonly performed by network engineers and cell providers in which a vehicle with RF measurement hardware is driven on every publicly accessible street within a 2km by 2km region and RSS values are recorded. This method results in a large number of measurements, contained entirely along public streets. This data is interesting because it is both typical and convenient to obtain, and because it has an obvious source of bias.

7.1.1 Method

Three drive-test data sets are used in this experiment:

- tfa: measurements taken in a suburban environment in Houston, Texas, described in section 3.1.3.2
- pdx: measurements taken in an urban environment in downtown Portland, Oregon, described in section 3.1.3.2
- google: measurements taken in an urban/suburban environment in Mountainview California, described in section 3.1.3.2

These data sets were chosen to provide a large number of representative drive-test measurements with easy comparability to prior work. All data sets were collected passively at 2.4 GHz, using commodity hardware and packet-based measurement strategies. All three data sets have been made publicly available for research purposes.

As a first processing step, the GPS receiver used for measurement reports 6 digits of precision, in order to reduce the data and perform spatial averaging per the method of Lee *et al.* [132], a precision if 5 digits, which corresponds to a resolution of 1.11 meters, or an averaging radius of 8.9 wavelengths at 2.4 GHz is used. For collocated measurements, the median value is used. The next step involves resampling along a uniform equilateral triangular lattice, as described in section 5.4.2. At each point on the sample lattice, the nearest neighbor point is used. Here, two different approaches to sampling are used: careful and aggressive sampling. In careful resampling, if there is a point within 40 wavelengths of the sample point, it is used but if there is not, no measurement is recorded. In aggressive sampling, the nearest point which is no further away than the sampling lag distance is taken. One of the research questions this section seeks to address is how harmful aggressive resampling may be to the fidelity of the ultimate model. These resampled measurements are used to automatically fit a variogram, and Krige a map, using an identical process to that described in chapter 5.

7.1.2 Experiment

The experimental design used here involves resampling, fitting, and mapping measurements from each BS in each dataset, at each of a number of lag distances (h). Lag distances from 10 meters to 500m are used, both with and without careful resampling. Due to the fact that even in careful resampling, measurements may be relocated up to 4 m, 10 m is the smallest reasonable sampling lag. An attempt is made to produce fits for $h \in \{25, 50, 100, 250, 500\}$, although some combinations produce too little data to fit (e.g., $h = 500m$ with careful resampling produces very few measurements in all scenarios), and some produce an intractible number (e.g., $h = 10m$ with will easily exhaust 20 GB of memory). In these scenarios, the experiment proceeds without a fit, and results are derived from those configurations that produce sufficient data and are tractable.

In order to analyze the performance of the fitted models, the method of Robinson *et al.* is used [200]. The original points are used to create an oracle map with a resolution of 0.2 pixels per meter (5 meters per pixel). To determine the fidelity of the fit, each cell on the oracle map (that contains data) is compared with the corresponding cell on the fitted map. Figure 7.1 shows this process schematically. Both the median

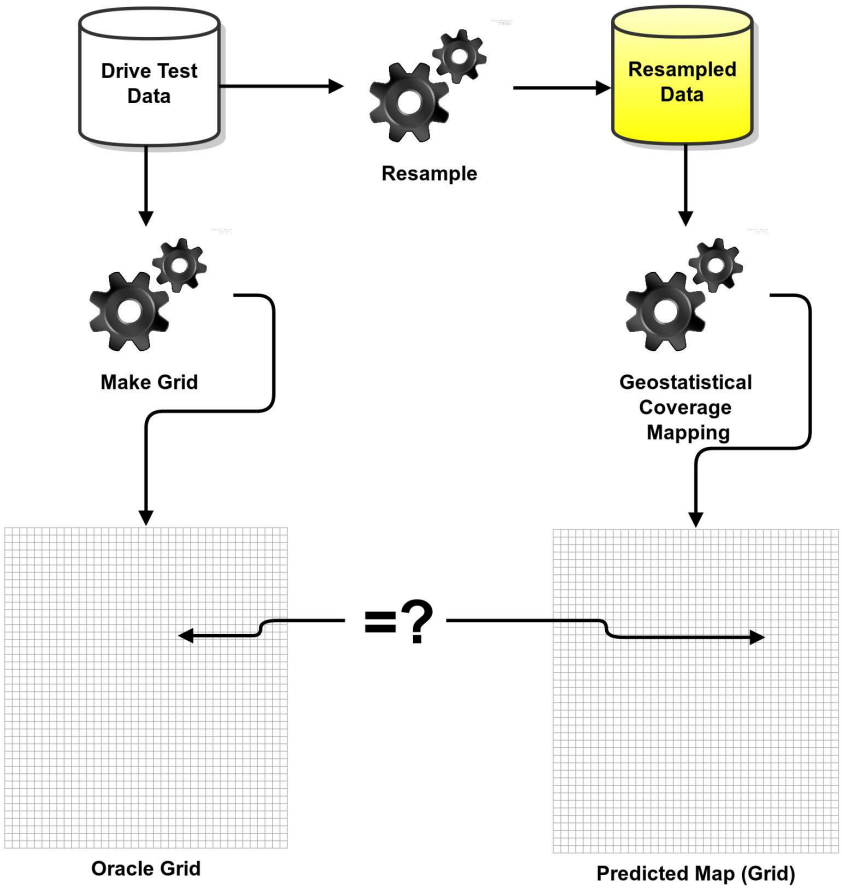


Figure 7.1: Schematic describing validation process for drive-test coverage mapping.

and maximum value are compared, and prediction accuracy for finding holes as well as aggregate RMSE are computed. The same definition of a coverage hole is used here as in [200], measured areas where the SNR dips below 20 dB is are considered to be coverage holes. These performance statistics are calculated in addition to the 10-fold cross validation statistics which measure goodness of fit during the variogram fitting and Kriging process described in chapter 5.

All told, this requires a substantial amount of computation: the complete fitting and kriging process is performed for each combination of BS and lag distance, the corresponding oracle map is extracted from the database of measurements, and the predicted and measured maps are compared. To make this computation feasible in a reasonable amount of time, each combination of BS and lag can be computed in parallel on the Janus supercomputer [156]. Without the level of parallelism offered by this system, performing such an analysis would require many hundreds of days.

7.1.3 Results

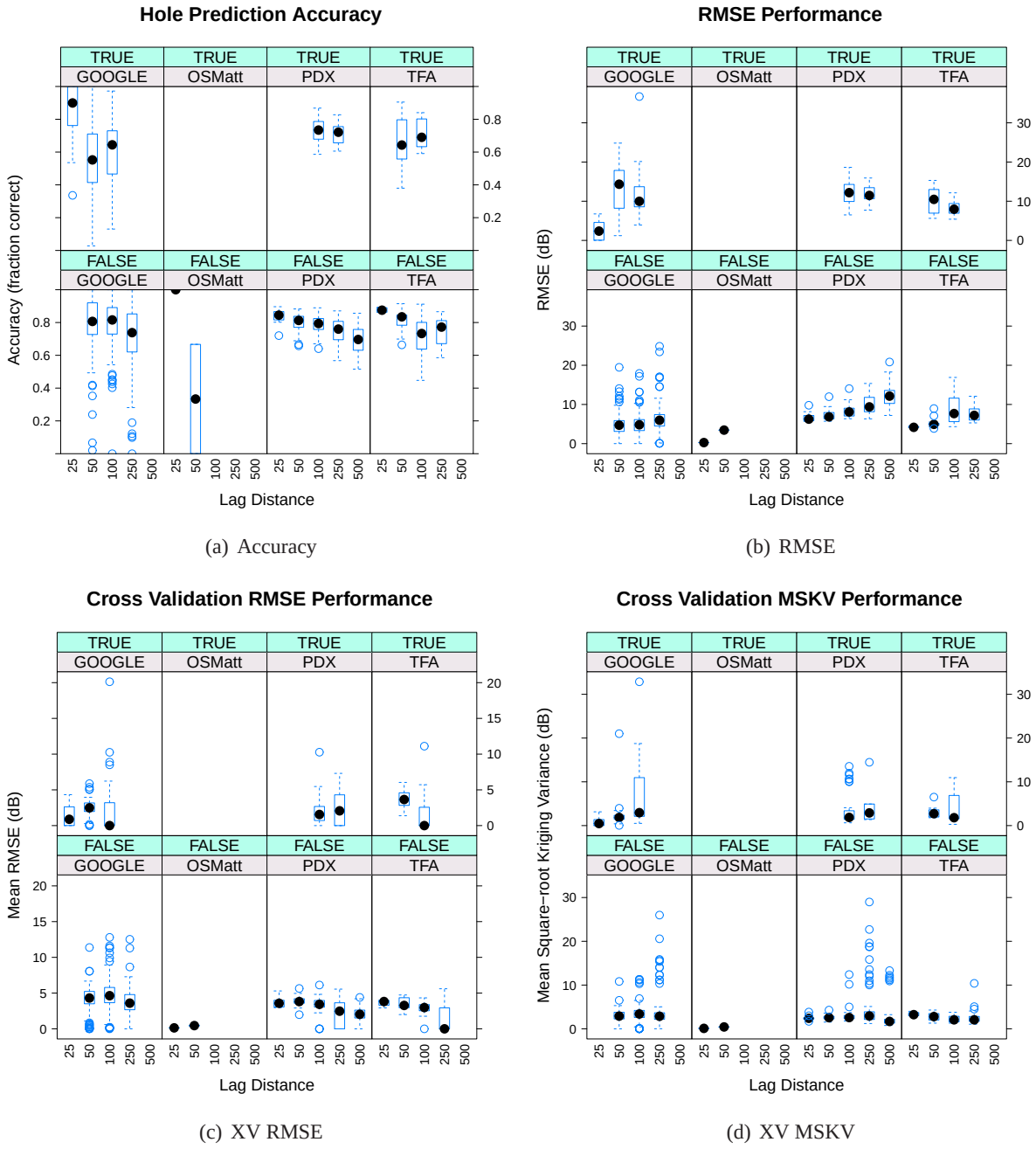
Figure 7.2 shows the overall performance of each resampling for each dataset both with and careful and aggressive resampling. Four metrics are provided, (a) Accuracy, the percentage of holes predicted correctly (areas where the SNR drops below 20 dB), (b) RMSE of the oracle measurements as compared to the predictions, (c) 10-fold cross-validation RMSE from the Kriging process, and (d) 10-fold cross-validation MSKV, also from the Kriging process. The first two metrics describe the resampled map's ability to predict the original data and the latter two metrics describe the residual error in the fitting process.

The first question is whether careful or aggressive resampling produces a more harmonious map. As can be seen in figures 7.2a and 7.2b, in almost every scenario aggressive sampling (bottom row) outperforms careful resampling (top row). The difference is both apparent and has been shown to be statistically significant using a Welch two-sample t-test. This is an unexpected and pleasing result, because it means that it is permissable to "move" measrements by some small amount when resampling without negatively impacting the fidelity of the ultimate map. In fact, the additional data available when doing aggressive sampling appears to produce better predictions.

The next question this section seeks to answer is what is the right resampling density. Clearly this

implies a tradeoff between precision and effort, and the plots support this. For the Google data set, an accuracy of 80% can be acheived with 100m or 50m lags and approximately 75% with 250 meter lags. For The PDX and TFA data sets, close to 90% accuracy can be acheived when resampling at a lag of 25m, and approximately 60% with a lag of 500m. The plots of RMSE tell a similar story, as the sampling lag is increased, the RMSE increases linaerly. Each data set is able to obtain nearly 5 dB RMSE in the best case, and rises to more than 10 in the worst. The cross validation metrics appear to be less sensitive to the sampling lag and regime, however there does appear to be a clear reduction in harmful outliers as the density is increased. A different view of this data is given in figure 7.3, which provides a plot of the Receiver Operating Characteristic (ROC) and Detection Error Tradeoff (DET) space for each sampling lag. Clearly, the choice of lag is application dependent, as one should choose the lag that requires the least amount of work while meeting the accuracy requirements of the application. However, based upon the results here, $h = 100m$ is easy to advocate as a good middle ground between precision and ease of collection for frequencies around 2.4 GHz, supporting the choice of $h = 100m$ in previous chapters.

A final question is how this strategy for selecting samples compares to the state-of-the-art method of iterative heuristic refinement proposed by Robinson [200]. Figure 7.4 addresses this, by explicitly showing the relationship between sampling density and hole prediction accuracy. As compared to the method of Robinson almost exactly the same performance is acheived with a similar number of samples. For the Google dataset, Robinson's framework is able to acheive approximately 75% accuracy with slightly more than 10 measurements per km^2 and nearly 80% accuracy with 70 measurements per km^2. The same result is acheived here, and indeed the performance appears to flatten around 0.8 with increasing samples providing no improvement in performance (incidentally, the 65 measurements per km^2 result corresponds to the $h = 100m$ resampling). With the TFA dataset, Robinson's framework acheives slightly better than 70% accuracy with 15 measurements per km^2 and is able to ultimately acheive slightly better than 80% with approximately 60 measurements. The method proposed here is acheives a higher starting accuracy, but appears to acheive the 80% mark more slowly. Although the PDX dataset was not studied by Robinson, the results are similar, and perhaps more consistent than the others; an average accuracy of 80% is acheived with approximately 70 measurements and tops out near 85% with more then 600 measurements. Overall these results, and their



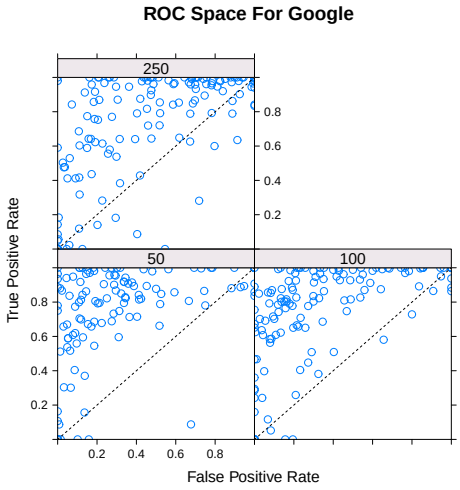
(a) Accuracy

(b) RMSE

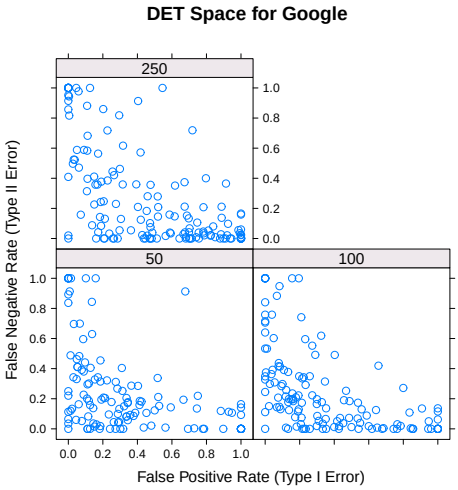
(c) XV RMSE

(d) XV MSKV

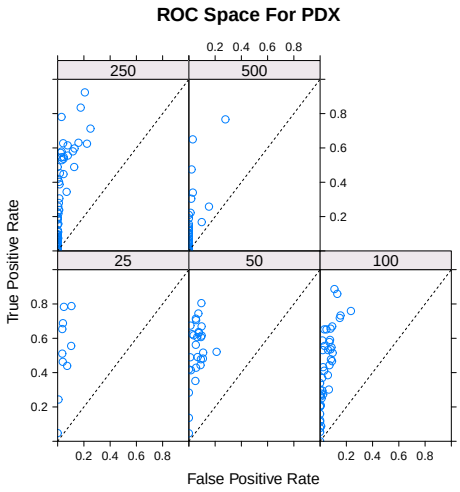
Figure 7.2: Performance results for different resampling lags and careful and aggressive resampling using four metrics of interest.



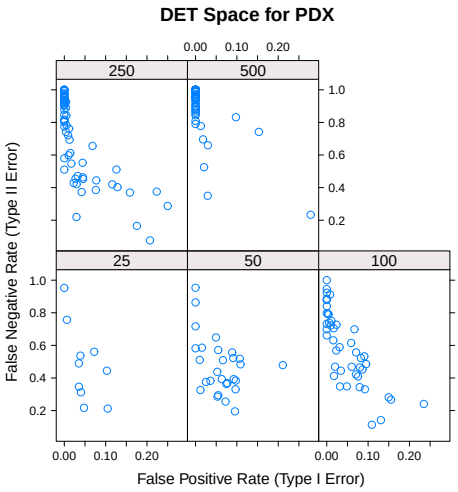
(a) Google, ROC



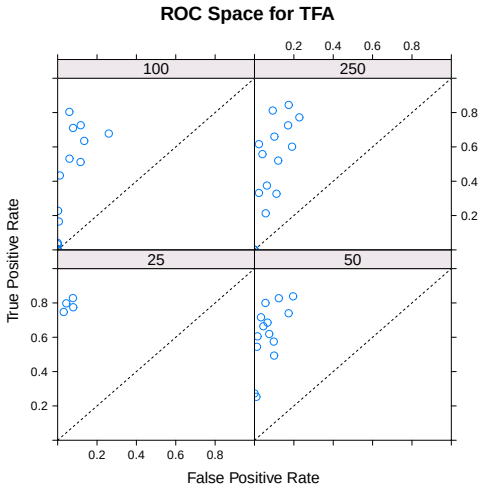
(b) Google, DET



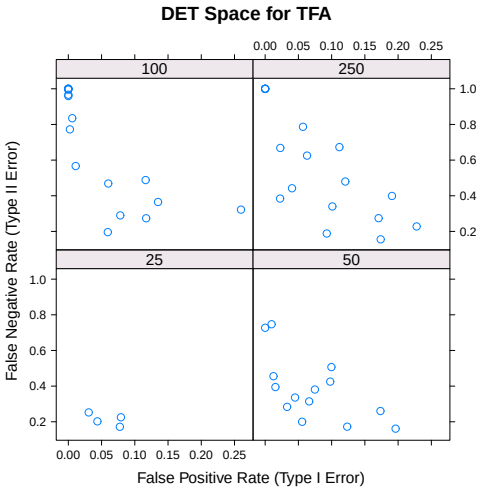
(c) PDX, ROC



(d) PDX, DET

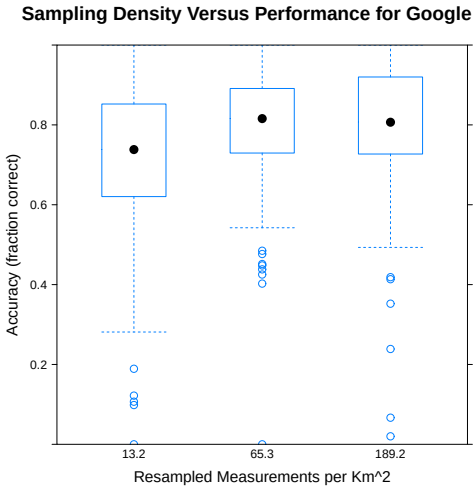


(e) PDX, ROC

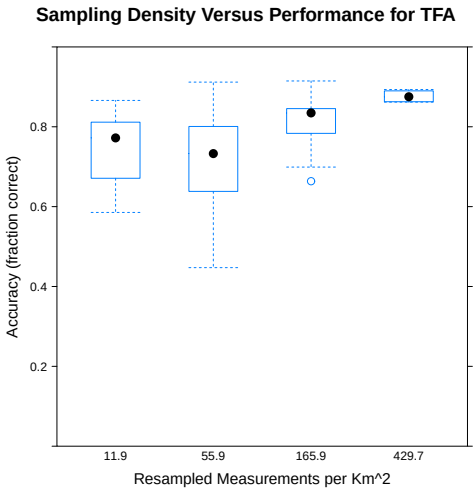


(f) PDX, DET

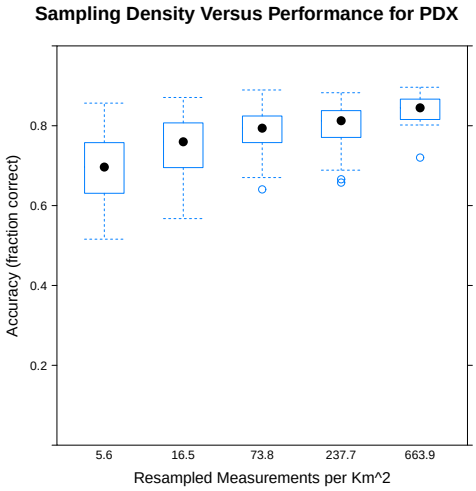
Figure 7.3: ROC and DET curves for each dataset using “aggressive” resampling.



(a) Google



(b) TFA



(c) PDX

Figure 7.4: Hole prediction accuracy as a function of sampling density.

strong agreement with the performance of Robinson's method, hints at an upper-bound in terms of accuracy using this metric. However, the most important conclusion these results highlight is that approximately the same performance can be acheived with Kriging and iterative heuristic refinement. Since Kriging provides a substantially richer perspective on propagation (providing a value distribution at each point, and not simply a boolean covered/not-covered value), it seems easy to advocate for this type of application in almost every circumstance.

7.1.4 Discussion

This section described and analyzed the application of the geostatistical methods proposed in chapter 5 to the problem of mapping coverage using drive-test data as input. This is an important extension becuse drive test measurements are generally easy to obtain and are widely used by cell providers for site-surveying. In order to cope with the intrinsic sampling bias of making measurements in city streets, this section proposed a resampling methodology where the closest measurement is taken at each vertex on a uniform equalateral triangular sample. Ultimately it was found that there is a linear relationship between the resampling density used and the predictive ability of the resulting map. The mean error observed here is typically larger than was observed in the case studies described in section 5, however those studies involved the careful placement of samples and measurement with calibrated equipment. One higher-order result of this section may be that while *yout get what you pay for* in terms of measurements, even rough measurements can produce a reasonably accurate map when an appropriate resampling scheme and statistical modeling process is used. Indeed, the error is substantially smaller than would be acheived with an *a priori* predictive method or simple data fitting approach (see chapter 3), while producing nearly identical results to the iterative heuristic refinement method proposed in [200]. Figures 7.5, 7.6, and 7.7 show the map-combined maps for these datasets, which bring to light the rich detail that can be derived of the coverage using these measurements, especially as compared to the simplistic maps produced by *a priori* models (e.g., see figure 3.28) and iterative heuristic refinement (e.g., [200]). The next section will take this idea a step further by introducing and investigating the idea of crowd-sourced coverage mapping, where many individuals contribute to the data collection process voluntarily.

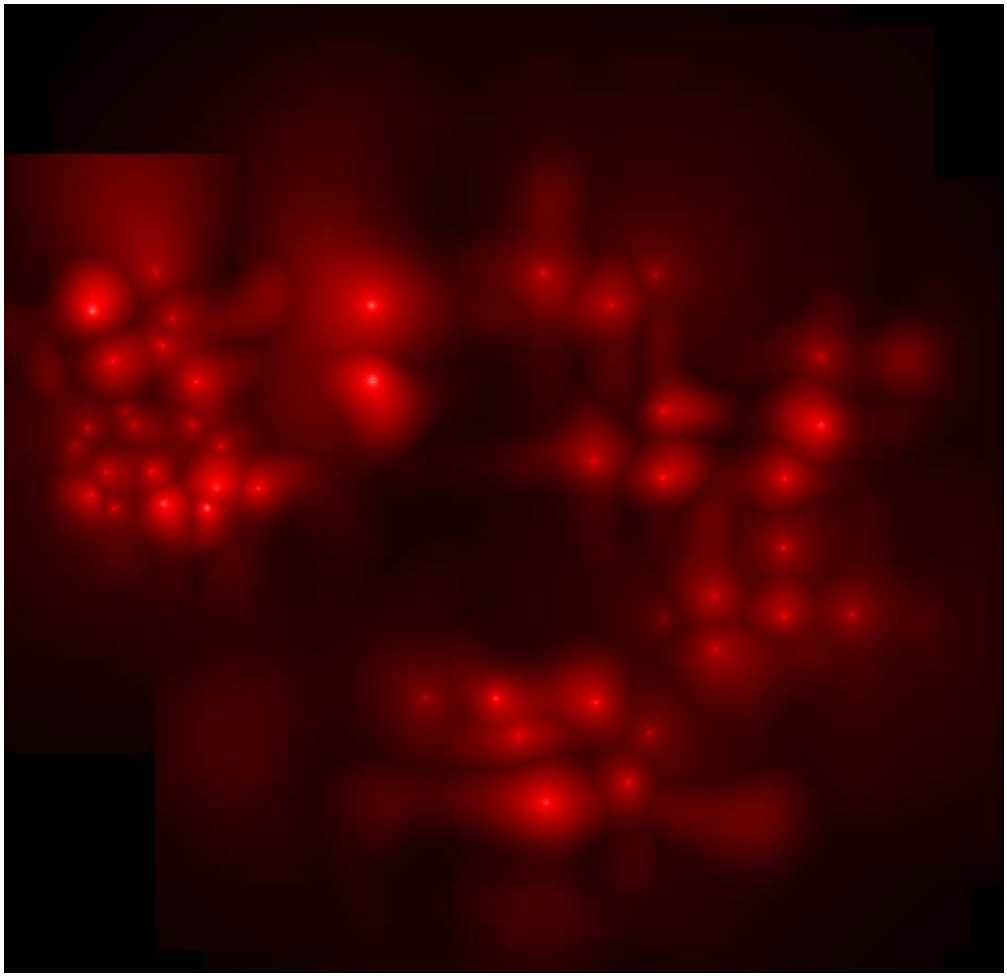


Figure 7.5: Map-combined (maximum combining) Kriged coverage map for PDX Data using best variogram and h=100m.

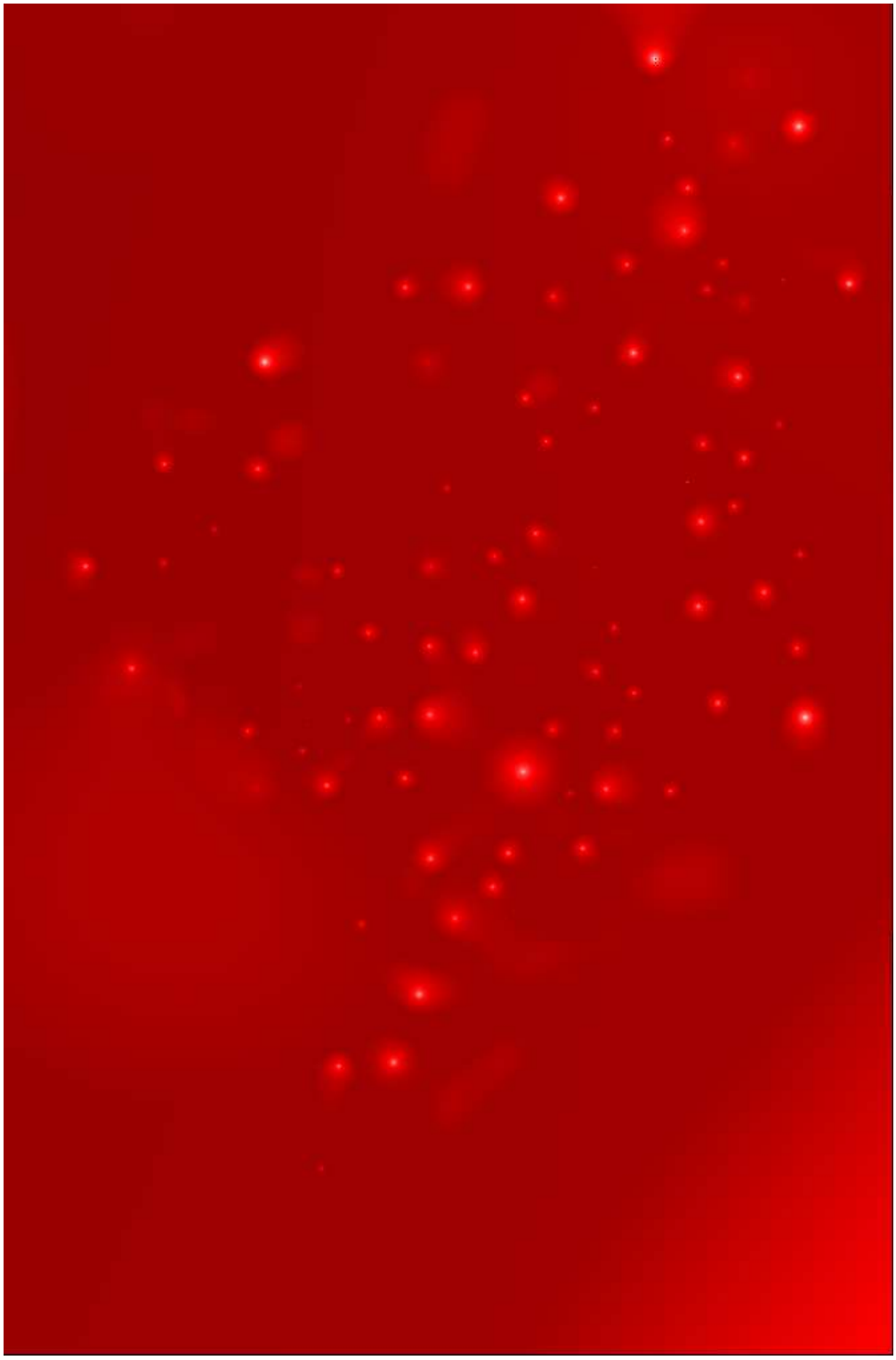


Figure 7.6: Map-combined (maximum combining) Kriged coverage map for Google Data using best variogram and h=100m.

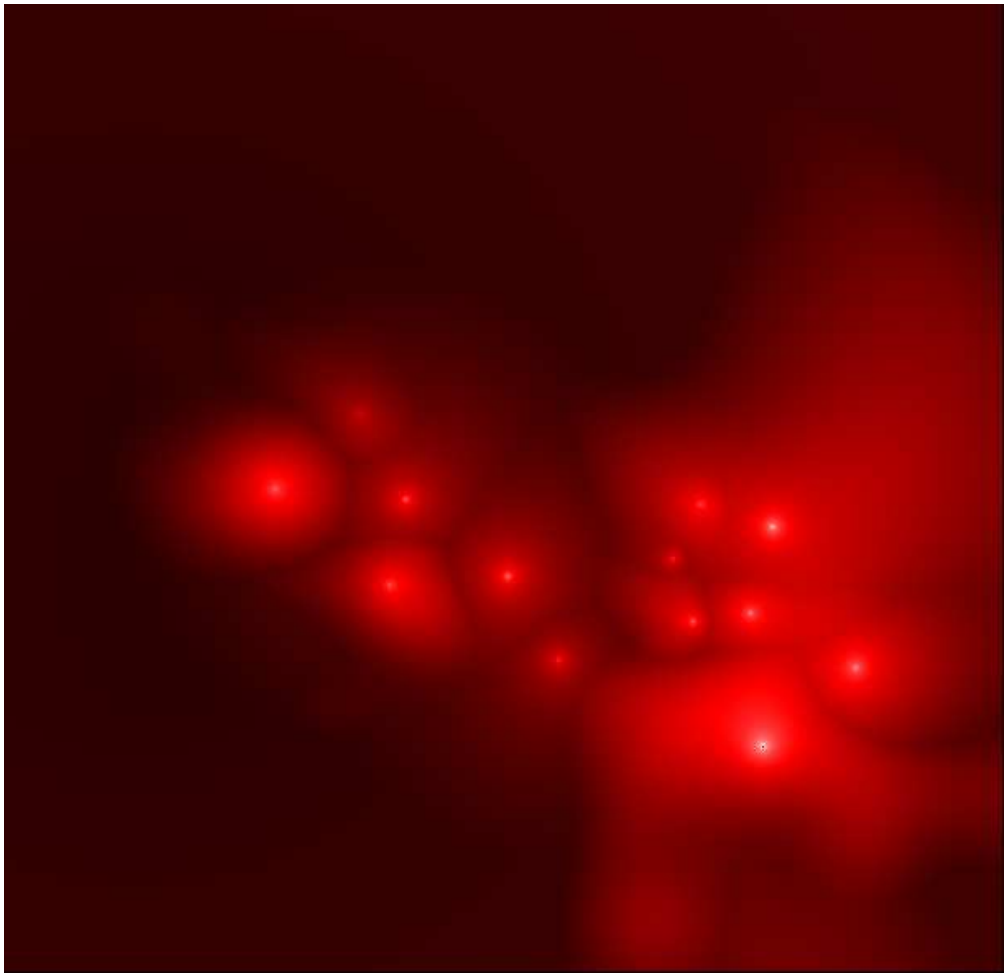


Figure 7.7: Map-combined (maximum combining) Kriged coverage map for TFA Data using best variogram and h=100m.

7.2 Crowd-Sourced Coverage Mapping

This section introduces the topic of crowd-sourced coverage mapping, where mobile devices and their users are solicited (wittingly or unwittingly) to collect measurements of a wireless netowork. The idea of using mobile devices for sensing applications is not new, and in fact has been applied to a large number of interesting applications. For instance, in [70] the authors use sensing mobile devices to generate better cycling maps, and in [137] the authors describe a system for mapping sources of noise in an urban environment to enable the discovery of quiet places. [126] provides a survey of various mobile sensing applications. The prospect of using these mobile devices to map the radio environment is a fairly recent proposal which has seen some modest interest (e.g., [244, 47]). This section will try to understand the feasibility of crowd-sourced coverage mapping in the context of the geostatistical approach taken in this thesis. In the next section, the key question of how many participants would be needed to map an area of typical size and complexity is analyzed via simulation using a human mobility model. Then in section 7.2.2, real crowd-sourced traces collected by the OSM application will be analyzed to understand the practical difficulties of fitting sparse and incomplete data using geostatistical methods.

7.2.1 Feasibility Study using Simulated Mobility

In order to understand how a crowd-sourced sampling scheme might scale to a large region with a large number of users, this section looks to human mobility models and verification via simulation. For this, the Self-similar Least Action Walk (SLAW) model proposed by Lee *et al.* is used [130]. This model is based on an extensive modeling and data-fitting campaign using GPS measurements from theme parks, college campuses and cities. SLAW assumes that there are some fixed number of way-points, which the individuals being simulated visit in a stochastic manner. For trace generation, at each time step, some number of users will move from one waypoint to another along a straight line path and reside there for an amount of time also determined by the model.

To simulate the dynamics of the CU campus, 6,205 unique users were modeled. This constitutes 20% of the students and faculty at the University, and is meant to approximate a rough lower bound on the

number of individuals present on campus on a given day. The main CU campus has 145 buildings and other facilities, which are used as way-points. The campus is roughly a right triangle, with a long side at 1.5km and s short side of 850m. Hence, for this simulation, a square space 1.5 km on a side is simulated, which is approximately twice the size of the existing campus. This further skews results towards a conservative estimate. It is assumed that a building or facility has a radius of 20m (i.e., any movement within 20m of a waypoint is located at that waypoint), and that users pause at waypoints for some amount of time between 1 minute and 2 hours (a typical class length). The SLAW model also requires some additional model-specific parameters that control the stochastic self-similar nature of the generated walks. For these, values that are similar to those observed in data collected at the University of California San Diego (UCSD) and Korea Advanced Institute of Science and Technology (KAIST) are used: $\alpha = 1.6$ and $\beta = 0.8$. To gain enough data, MATLAB code provided by [130] is used to generate a 170 hour trace. The first 50 hours of the trace are excluded to avoid transient behavior and the remaining 120 hours (5 days) are used below.

In order to understand how the number of users and the time elapsed effects the sampling coverage, the number of users participating and the duration of the study are varied. A random subsample of users from 0.01% to 100% is used to generate a trace with a max time from one hour to 150 hours. For each trace the study area is subdivided into a 5 m grid (approximately 40 wavelengths at 2.5 GHz). How many times the path of the particular user enters a given grid square is counted. Perfect coverage for a given participation rate and time would have every square entered at least once. In an imperfect sample, those areas not visited by a user form measurement “holes” and limit the final resolution (and prediction accuracy) of a map generated from the sample coordinates.

Figure 7.8 gives an example of the aggregate coverage with several different participation percentages and durations. The left hand plots are colored so that grid cells which see the largest number of visits are darker (number of visits is on a log scale). The right hand plots give a boolean version of this map, where cells that have seen any visits are colored purple and cells that have not seen any are cyan. With just one user and one hour, hardly any of the map is covered. However, with a modest increase in the percentage participating to 6% (372 individuals) and a single work day at 7 hours, much of the map is covered. After 81 hours, with 96% of people participating (5957 individuals), the sample is completely covered except near

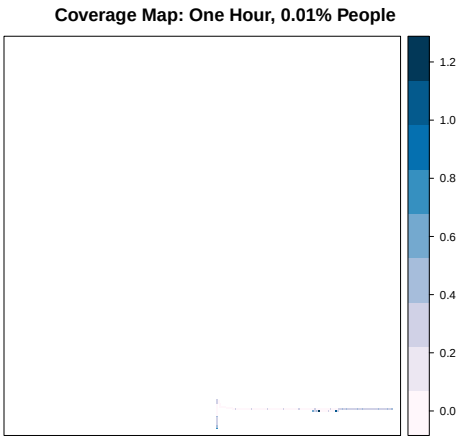
the edges.

Figure 7.9a shows the number of holes remaining as a function of the fraction of participating users and time elapsed. This plot is on a logscale, since there are a very large number of holes when there are few participants or a small amount of time. To find holes,the sample grid is searched for contiguous regions that have zero measurements. By increasing the percentage of participating users, or the duration of the study, the number of holes decreases exponentially. Figure 7.9b plots a different metric, which is the maximum "effective hole diameter" as a function of participation and time. To determine the effective diameter of a hole, we calculate the width of the hole and the height of the hole. The effective diameter is the diagonal line scaled by the fraction of cells that are empty in the square bounded by the height and the width:

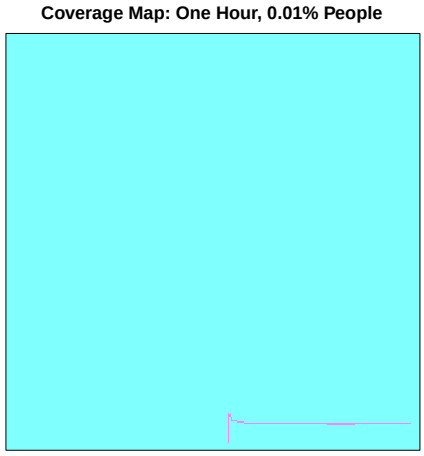
$$d = (a/n) * \sqrt{h^2 + w^2} \tag{7.1}$$

where a is the number of cells that are empty in the square region contained in a box centered on this hole of height h and width w and n is the number of cells total in that square (i.e., $n = h * w$). Although this is not a perfect description of the shape of a given hole, it acts as a conservative single value description of the limitations of a given sample. Compared to figure 7.9a, we can see that many levels of participation and durations that produce a large *number* of holes, do not produce holes of a very large size. In fact, most experiments except the shortest or those with the least participation, have a maximum effective hole diameter of 50*5 = 250 meters. Finally, figure 7.9b plots the number of holes greater in diameter than 50 squares as a function of participation and duration.The majority of cases have one or two "large" holes, which are presumable at the edges as in figure 7.8e and 7.8f.

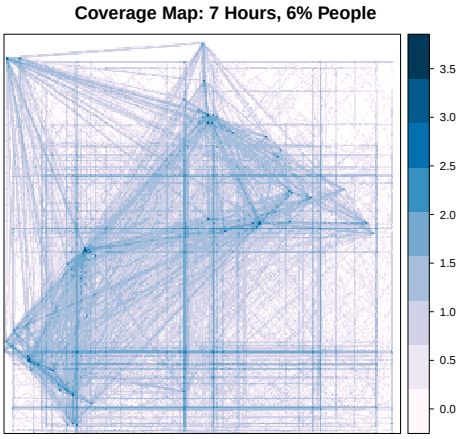
From these experiments it can be concluded that a reasonably small fraction of the CU students and faculty would be required to participate in a crowd-sourced mapping campaign in order to collect sufficient samples required to produce a high resolution coverage map. For instance, with a total of 14.2% of the campus participating (4506 individuals), the complete campus could be mapped within three hours. And, with 4.2% total participation (1304 individuals), a similar level of coverage would be acheived after 25 hours. Although promising, there are few criticisms that can be made of these results which may affect their



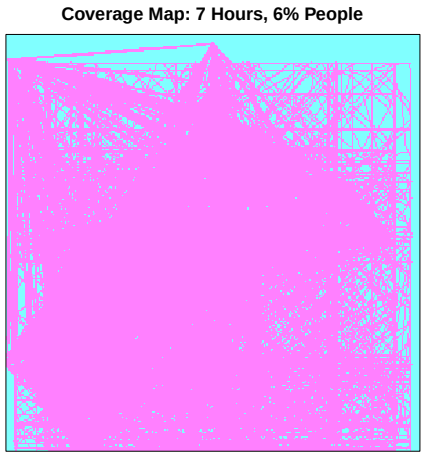
(a) 0.01 Percent, 1 Hour



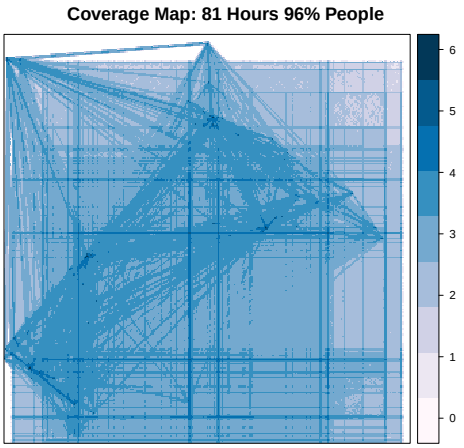
(b) 0.01 Percent, 1 Hour



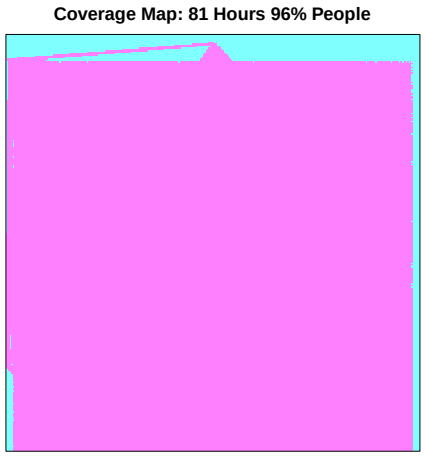
(c) 6 Percent, 7 Hours



(d) 6 Percent, 7 Hours

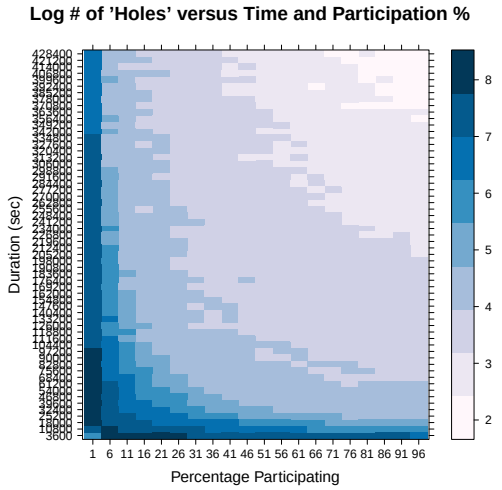


(e) 96 Percent, 81 Hours

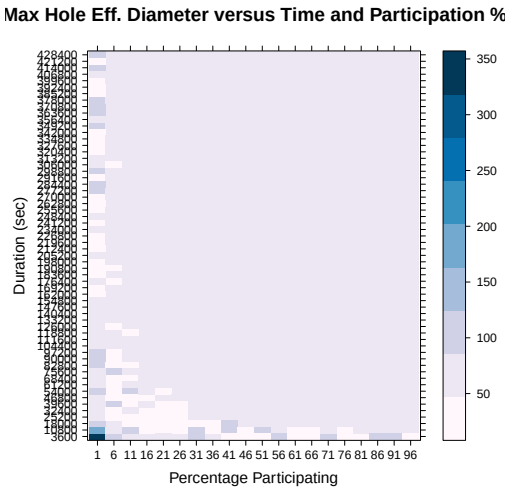


(f) 96 Percent, 81 Hours

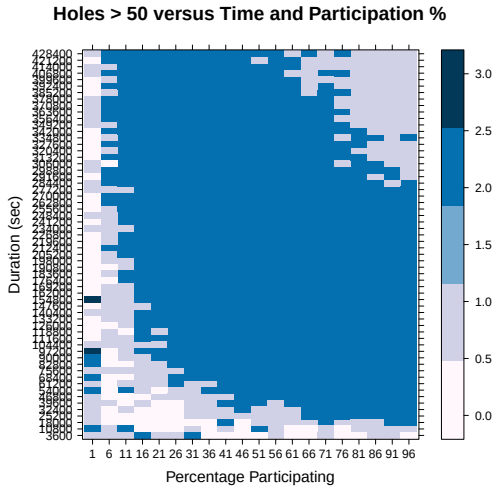
Figure 7.8: Sample coverage as a function of percentage of participating users and length of experiment. The left-hand plots show the number of visits to each 5 meter cell on a log scale. The right-hand plots show a boolean map of cell coverage.



(a) Number of Holes



(b) Max Hole Size



(c) Number of Big Holes

Figure 7.9: Sample coverage as a function of percentage of participating users and length of experiment.

generalizeability:

- Obstructions are not modeled. In this simulation it is assumed that users can walk directly between waypoints on a straight line without routing around obstructions, people, and other waypoints. In realitiy, users likely constrian themselves to predefined paths and do not visit some (i.e., fenced and walled) areas. This would cause more holes to exist, but centered at the points where information is likely least valued (i.e., users do not go there, so a high resolution map of that particular spot may not be valuable).

- The CU campus is very user-dense, with more then 30,000 individuals sharing the same 2.25 square kilometers. Areas with fewer users and sparser waypoints would be sampled less readily. However, the crowd-sourcing model is inherently democratic: the areas with the most users will recieve the most samples and therefore have the most accurate measurements. In scenarios where measurements are being used to plan future deployments, a crowd-sourced data collection may need to be augmented with directed measurement in the area of interest.

- It is assumed that every visit to a given grid cell produces a usable measurement, when in practice mobile devices may produce noisy measurements, or fail to make useful measurements in some location (e.g., a failure to obtain good location information through GPS constitutes a general failure of measurement. However, we argue that there is a steady trend of improvement in the measurement and locationing abilities of mobile devices and any limitations along these lines are likely to be mitigated in the near future by the advent of better mobile systems.

Overall, the results in this section make a compelling argument in favor of the abilities of crowd-sourced coverage mapping systems. Given a reasonably small fraction of participating users (4 to 15%) a fairly large and diverse area can be extensively measured in a reasonable amount of time (three to 25 hours). One can imagine such a system producing real-time wireless coverage maps (perhaps using a geostatistical interpolation method similar to that described in chapter 5 that regenerate periodically in response to measurements provided by users in real time. The next section looks at maps generated from real measurements collected by a prototype crowd-sourced sampling system.

7.2.2 Case Study: Open Signal Maps

In order to understand how well the crowd-sourced coverage mapping method may work in practice, this section describes experiments using real data collected by the OSM application [101, 47]. The OSM application is a project of the United Kingdom (UK)-based web development firm, Staircase3 [212]. Their application runs on Google Android-based phones. Users voluntarily download and install the application which collects measurements of signal strength and location in the background and sends them to the company via a web-service. Figure 7.10 gives an example of a map generated from crowd-sourced measurements on the OSM website. The mapping method used is basic heatmapping without interpolation. The heatmapping regime assumes that an absense of measurements indicates weaker signal and places red areas (strong signal) over regions with many strong meausurements.

In October 2011, a meeting was held with developers at Staircase3 to discuss the possibility of a research data collaboration and in particular the prospect of using geostatistical techniques to map the data. The developers agreed to share all the data they had collected contained within the state of Colorado for this analysis. The data contains measurements of multiple networks by a number of distinct users. For the purpose of this case study, results for the AT&T network are presented, as it was the most prevalent network in the measurements, and a subset of measurements confined to the city of Boulder, within the bounding polygon defined by the points $(39.995057, -105.249195, 40.011658, -105.277476)$ were used. Because the data is sparse, and often confined to straight lines (similar to measurements made in drive-test studies), "aggressive" resampling was used.

The complete dataset contains measurements from 581 distinct BSs with as few as one and as many as 1,257 measurements each. These measurements were collected by 190 distinct devices, and 74 total devices making measurements of the AT&T network. Within the bounding polygon defined above, 50 unique AT&T BSs are observed to varying degrees by 13 unique devices. Figure 7.11 shows the resampled measurements at different lags for a representative BS. Clearly these measurements are data-sparse, especially considering that this BS has more measurements present in the dataset than most. Figures 7.12 and 7.13 show the excess map and coverage map for this BS, mapped using the same method described in section 5. Figure 7.19 shows

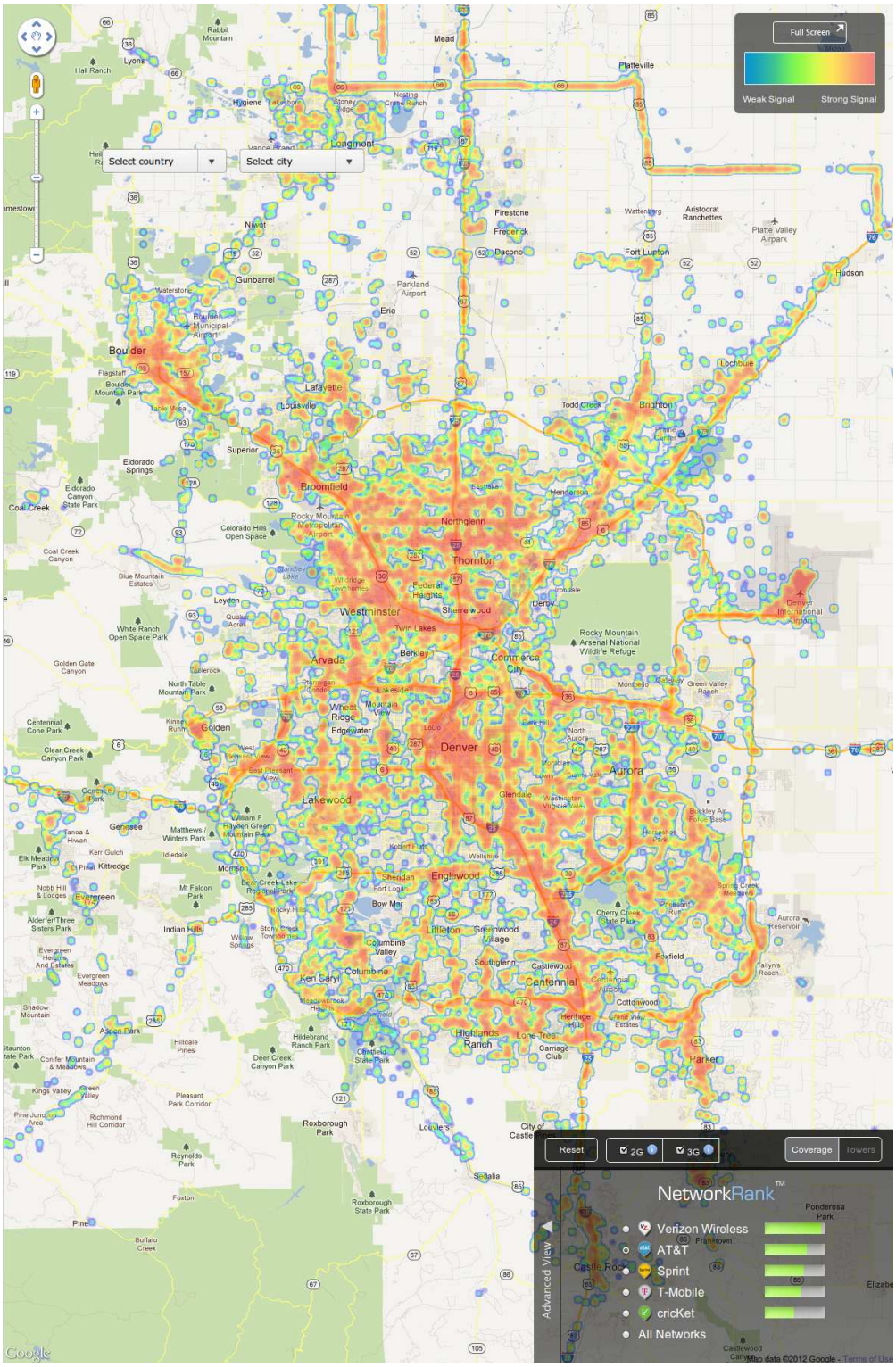


Figure 7.10: Open Signal Maps measurements for AT&T network overlayed on Google orthoimagry.

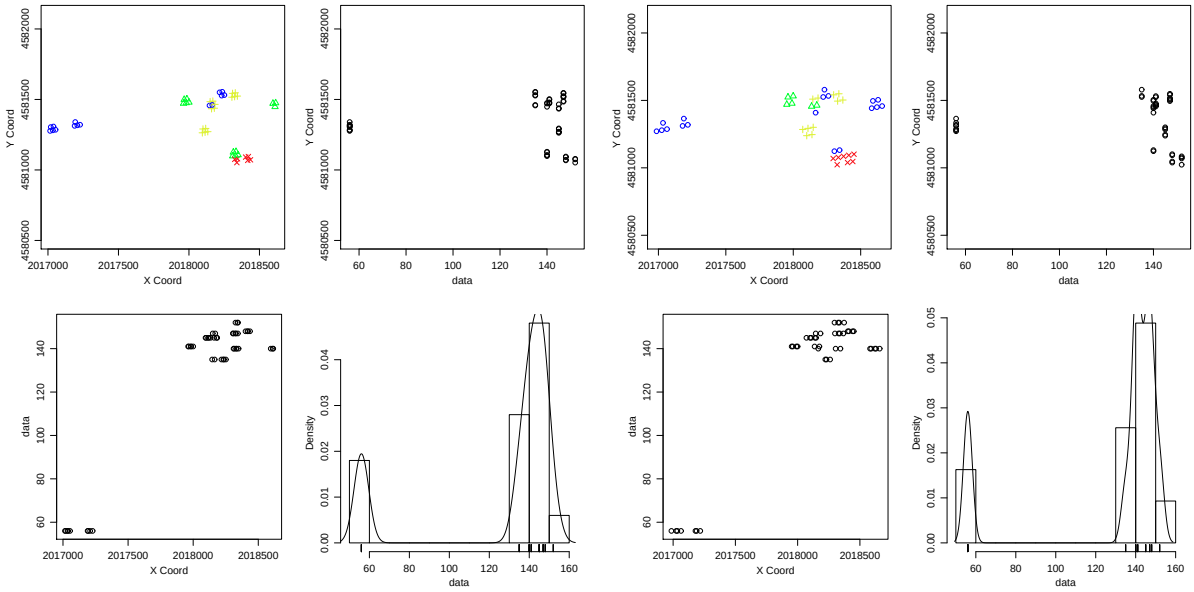
Lag (m)	Model	ϕ	τ^2	σ^2	N	Trunc/Neg	Mean K-Var	Mean RMSE	Gain
25	cubic	3290.48	25.00	6259.02	417	FALSE/FALSE	5.54	4.02	8.79
50	cubic	2182.09	75.54	190.31	361	FALSE/TRUE	8.93	9.24	4.78
100	cubic	29648.22	41.28	49124.86	219	TRUE/TRUE	6.59	10.74	8.22
250	gaussian	792.46	237.22	606.03	83	TRUE/TRUE	17.09	14.74	5.95

Table 7.1: Best fit statistics for variogram fitting of resampled OSM AT&T data.

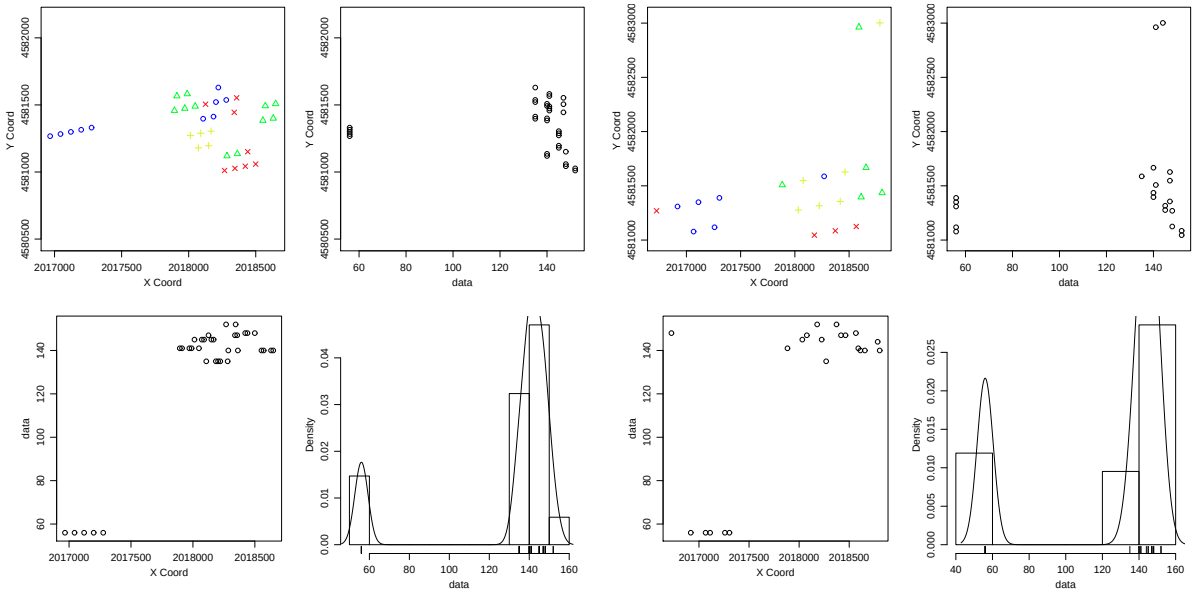
the threshold-based map using an SNR of 20 dB as the coverage criterion. These maps highlight the fact that each resampling density produces a considerably different map when the data is sparse. Table 7.1 proves cross-validation performance results, which show that the smaller lag distances (i.e., denser resampling) generally produce less residual error, with a RMSE of 4.02 dB acheived with $h = 25$, and gain of 8.79 dB over simple log/log data fitting.

The other 49 BSs have similar fit statistics, although those with very little data (i.e., only a handful of actual observations) fit quite poorly as would be expected. Figures 7.15 and 7.16 show the map-combined maps for all the BSs combined. These maps make clear the fact that most measurements present are from a small fraction of the 50 BSs, presumbaly because those users participating in the measurement spend most of their time in the coverage of those cells. In this way, this dataset is similar to the LTE case study in chapter 6, where the combined maps provide greater insight into the coverage of the network since individual cells only provide a partial picture (and many cells have been insufficiently measured). However, due the limited number of measurements for most BSs in this map, the picture of coverage is likely far from complete—more participation is needed in order to obtain a sampling coverage sufficient to produce a consistent and meaningful map. Simply put, the 13 participating users are not quite 1% of the 1304 participants shown in prior section to be the number of individual participants required to sufficiently map a slightly smaller area within 25 hours.

Figure 7.17 provies an aggregrate view of performance for all 49 BSs, which is similar to the analysis done in the previous section. This plot shows that in spite of the data sparsity, the predictive ability of the maps is still fairly strong, at least with respect to the measurements that are available. The mean coverage hole prediction accuracy is approximately 90%, and mean RMSE is around 7 dB. The residual error in the fits is small, generally less than 5 dB. The ROC and DET plots are given in figure 7.18, which show similarly

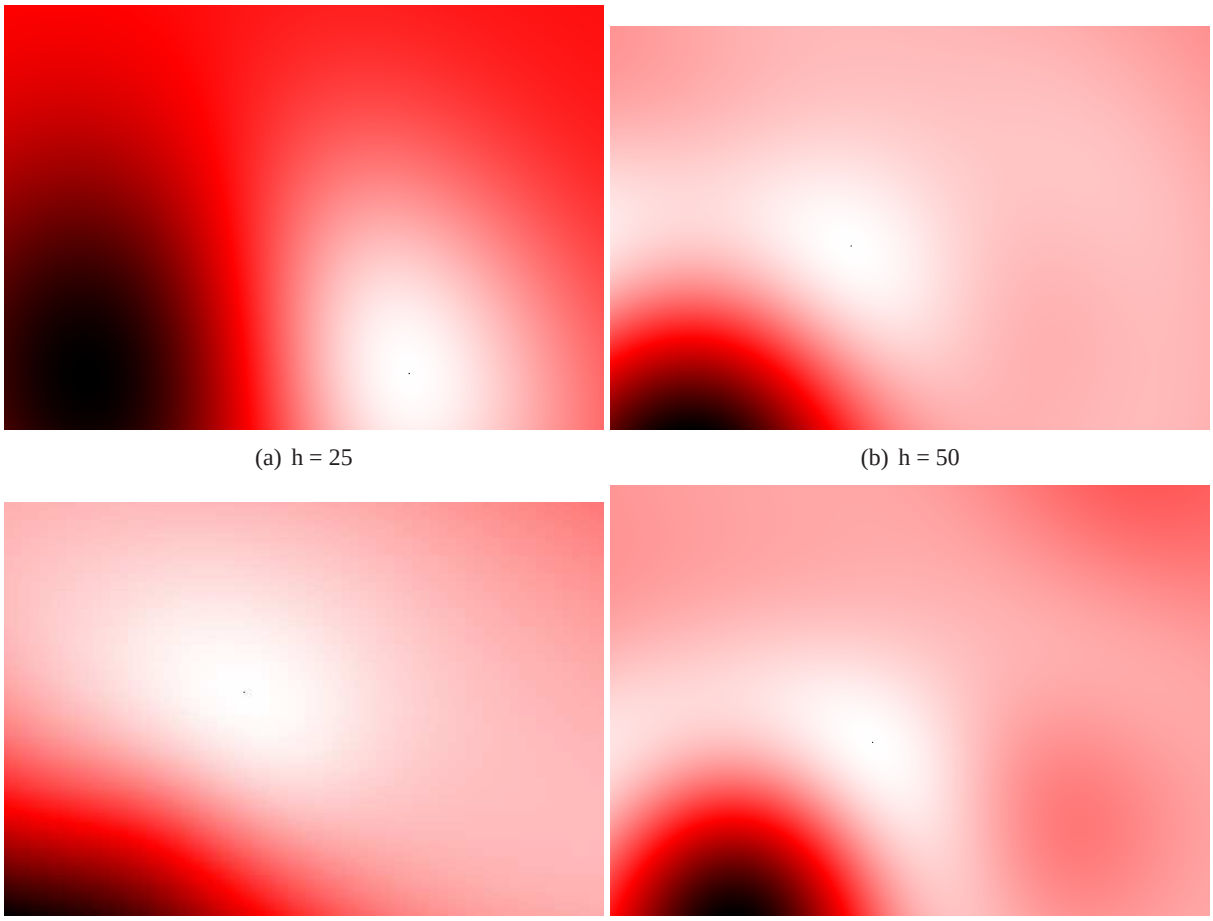


(a) h = 25 (b) h = 50



(c) h = 100 (d) h = 250

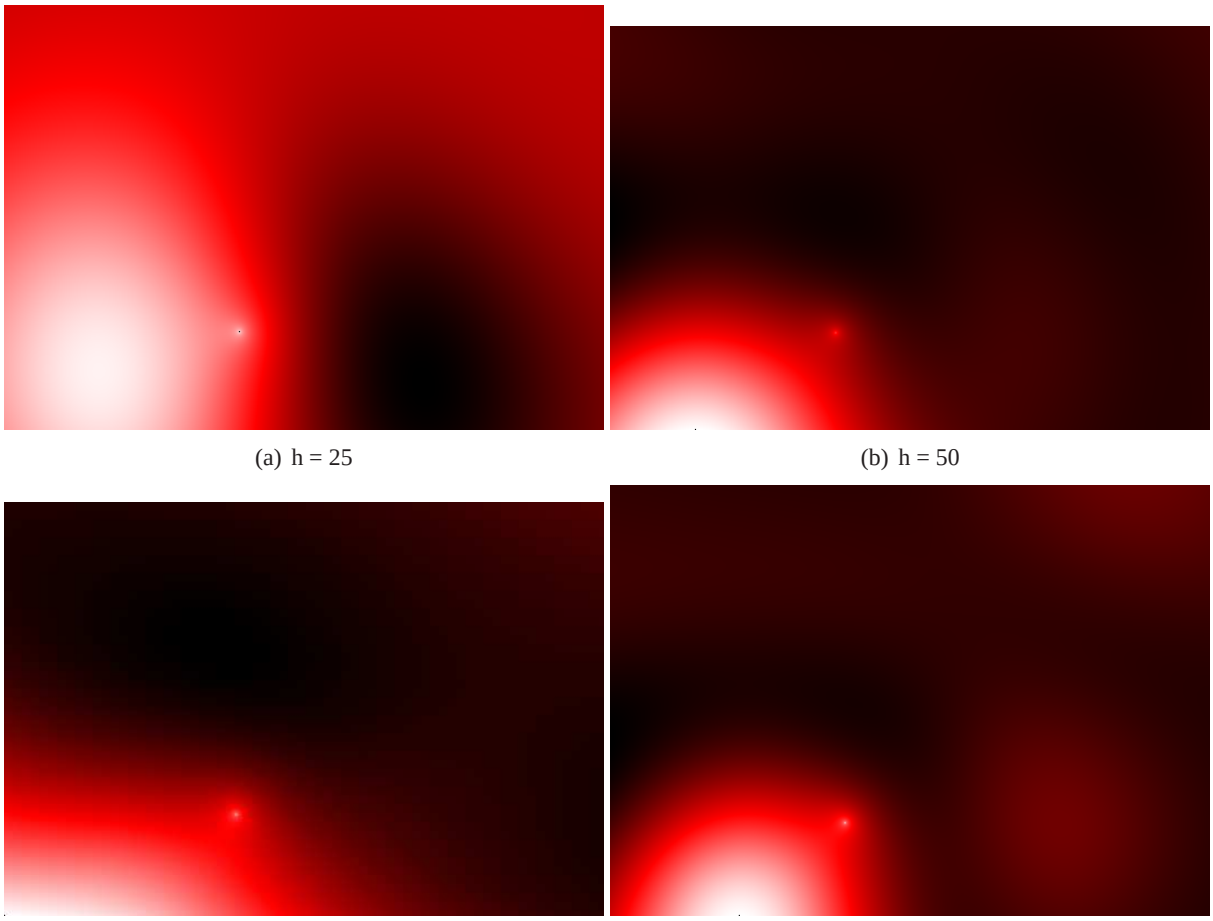
Figure 7.11: Performance results for OSM data at different resampling lags using four metrics of interest.



(a) h = 25 (b) h = 50

(c) h = 100 (d) h = 250

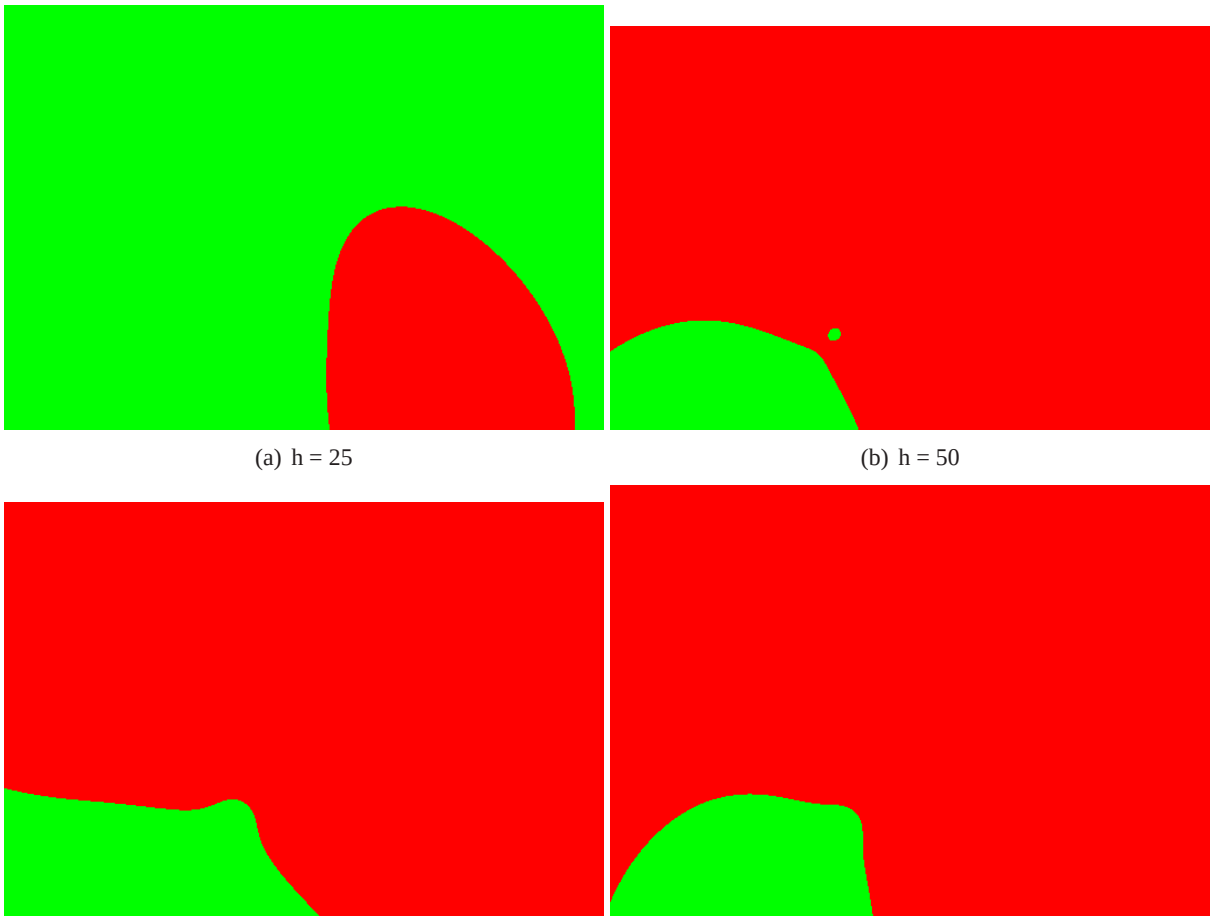
Figure 7.12: Performance results for OSM data at different resampling lags using four metrics of interest.



(a) h = 25 (b) h = 50

(c) h = 100 (d) h = 250

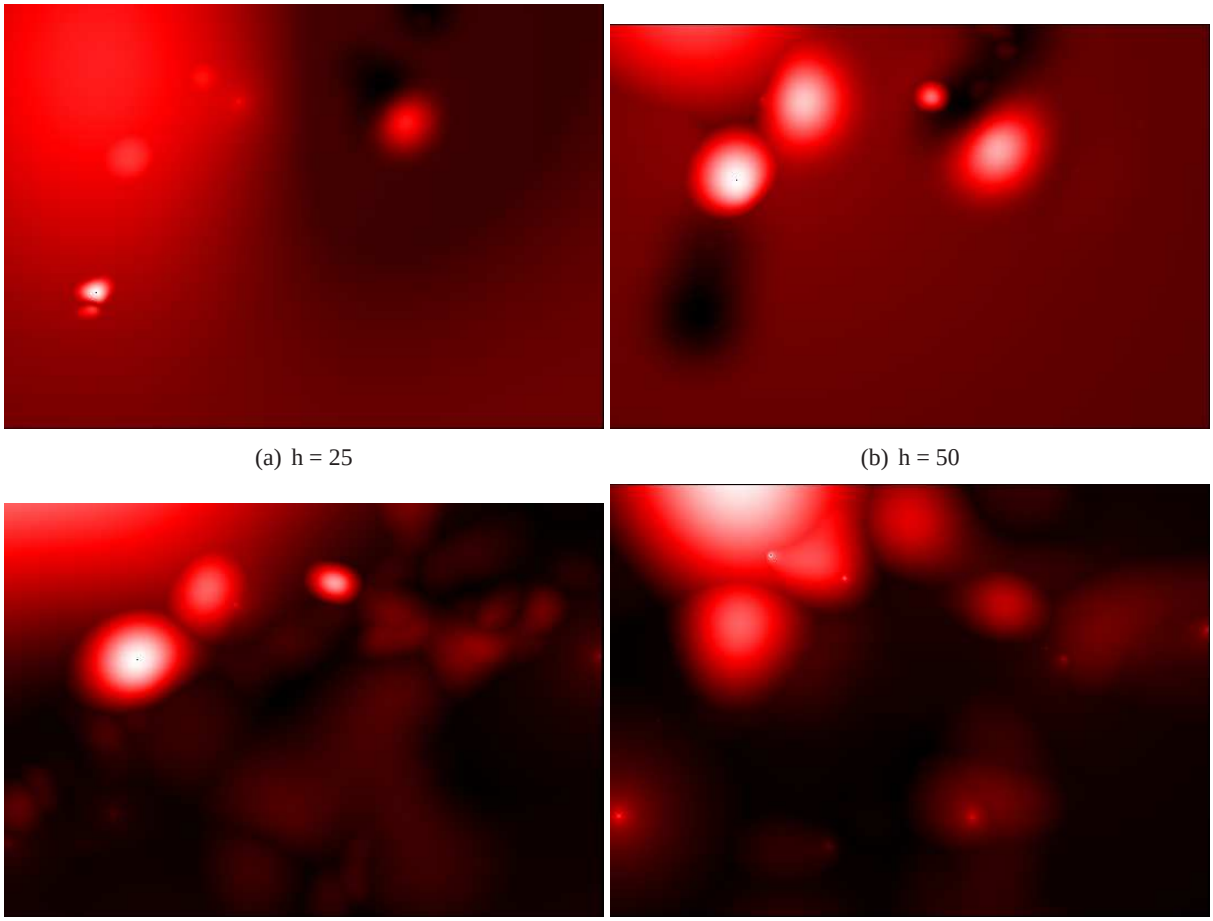
Figure 7.13: Performance results for OSM data at different resampling lags using four metrics of interest.



(a) h = 25 (b) h = 50

(c) h = 100 (d) h = 250

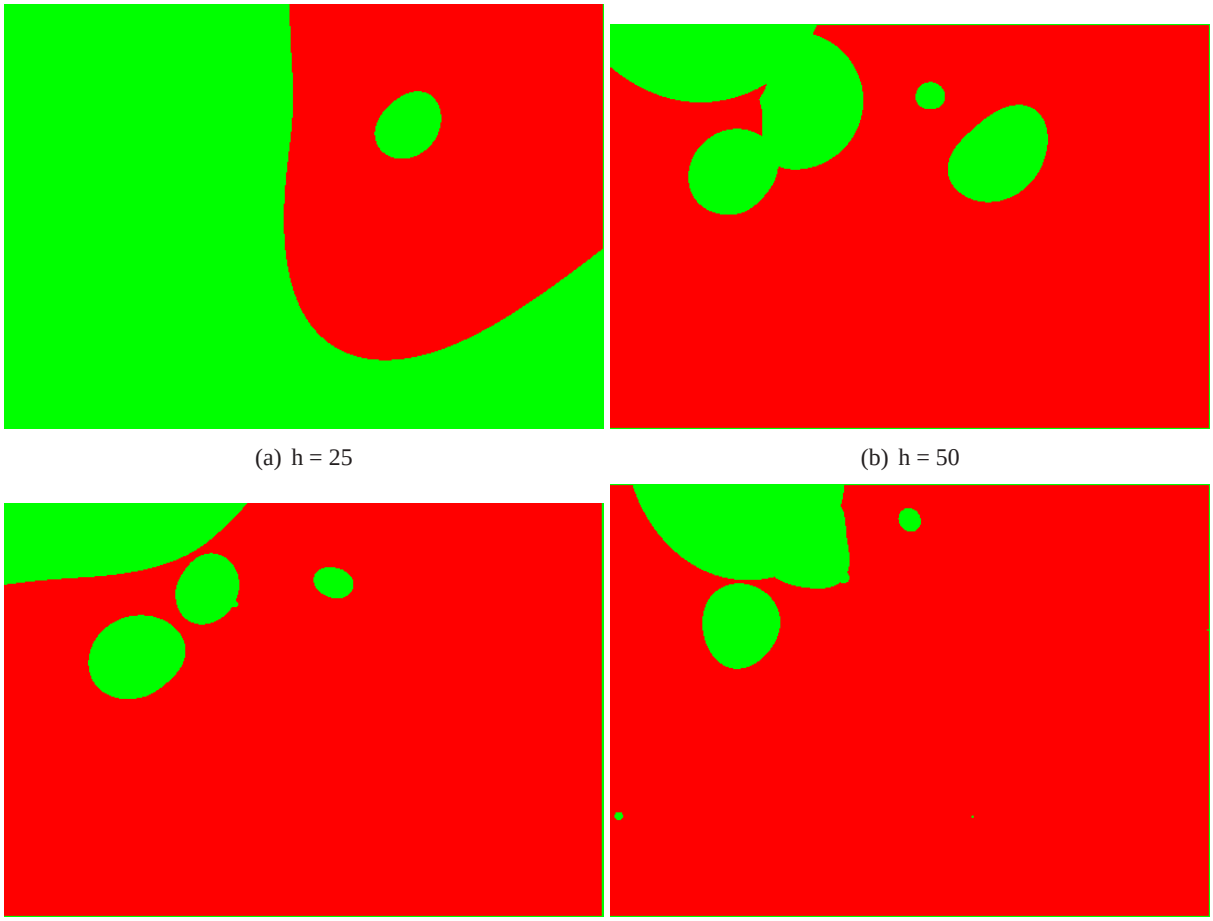
Figure 7.14: Performance results for OSM data at different resampling lags using four metrics of interest.



(a) h = 25 (b) h = 50

(c) h = 100 (d) h = 250

Figure 7.15: Map-combined maps using maximum-based combining for OSM data.



(a) h = 25 (b) h = 50

(c) h = 100 (d) h = 250

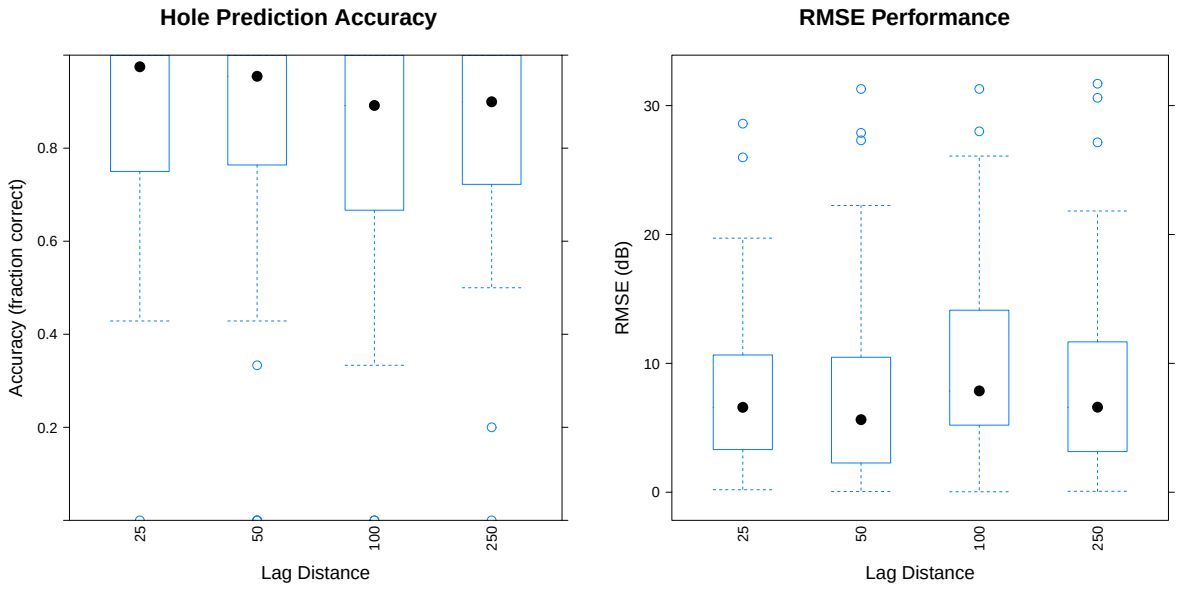
Figure 7.16: Map-combined threshold maps using maximum-based combining for OSM data.

strong results in terms of hole prediction, with very few false negatives and a reasonably small number of false positives. Finally, figure 7.19 shows the relationship between resampling density (samples per km^2) and hole prediction accuracy. Here the number of samples to acheive an accuracy of nearly 90% is less than 60 per km^2, which is quite good by the standard set for municipal WiFi networks in [200].

Overall, these results bring to light two conclusions. First and foremost, geostatistical coverage mapping appears to be a reasonable way to produce maps of crowd sourced measurements. However, real crowd-sourced data is extremely sparse as compared to data collected for the express purpose of coverage mapping. A larger fraction of users need to participate in order to acheive the sampling density required to generate a coverage map that is as detailed as would be necessary for it to be useful in practice. Although the previous section showed that only a small fraction of all people need to participate to create a complete map, that fraction does not appear to be met in this dataset. With the current level of measurement, only broad conclusions can be drawn (e.g., which BSs are most prevalent in a given area and their general propagation pattern). For the cell providers themselves, extensive measurements could be collected simply by requiring users to provide periodic measurements of the channel to a central source. However, such an application may have privacy implications, exposing the location and movement behavior of end-users to the providers. In addition to increased participation, further work is needed to understand the practical fidelity of mobile phones as measurement devices. However, once the data is available, and at sufficient density, crowd-sourced coverage mapping using geostatistical approaches appears to be a feasible design.

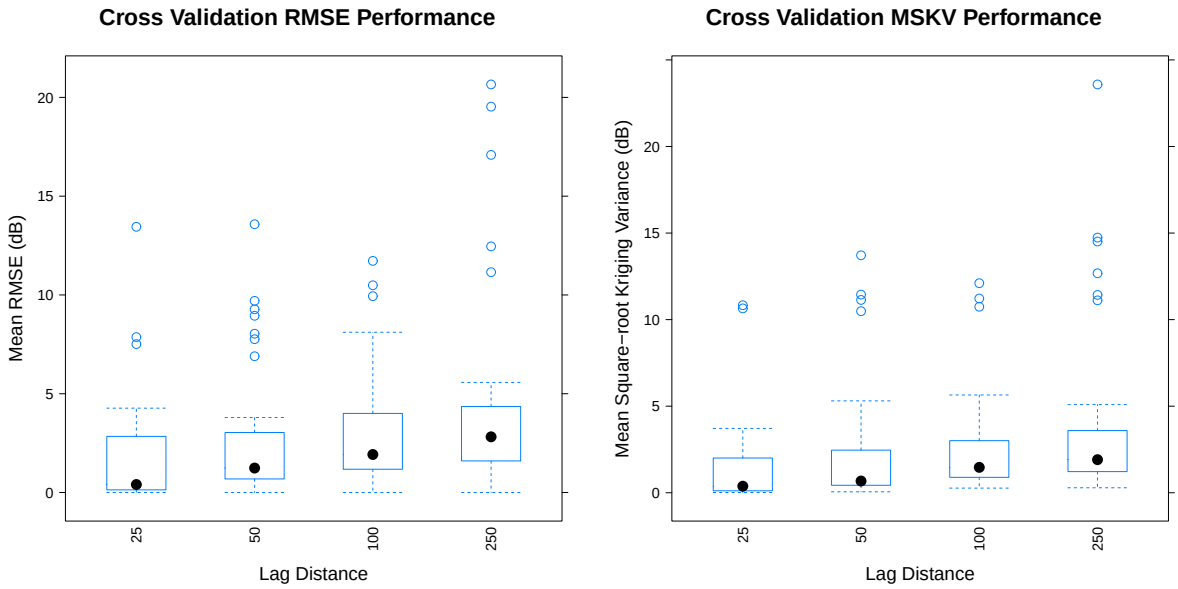
7.3 Summary and Conclusion

This chapter took a practical look at geostatistical coverage mapping by considering two methods of measurement that prioritize ease of collection and convenience. Although a careful and principled sample is clearly the best strategy for producing an accurate coverage map, the practical adoption of the methods proposed in this thesis requires that they work with existing measurement methodologies as well. First, drive test measurements were studied as a basis for mapping. It was found that resampling measurements on an equilateral triangular lattice is a reasonable way to cope with the inherent sampling bias in these datasets. The resulting Kriged maps provide a rich picture of the RF landscape, with a small error as compared to



(a) Accuracy

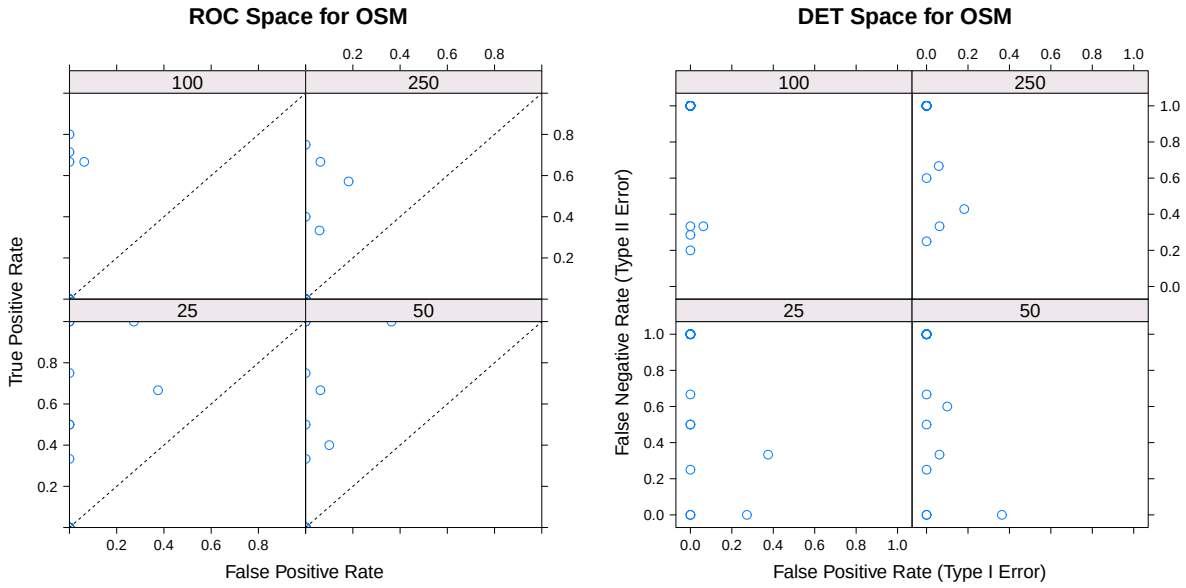
(b) RMSE



(c) XV RMSE

(d) XV MSKV

Figure 7.17: Performance results for OSM data at different resampling lags using four metrics of interest.



(a) ROC (b) DET

Figure 7.18: ROC and DET curves for OSM measurements using “aggressive” resampling.

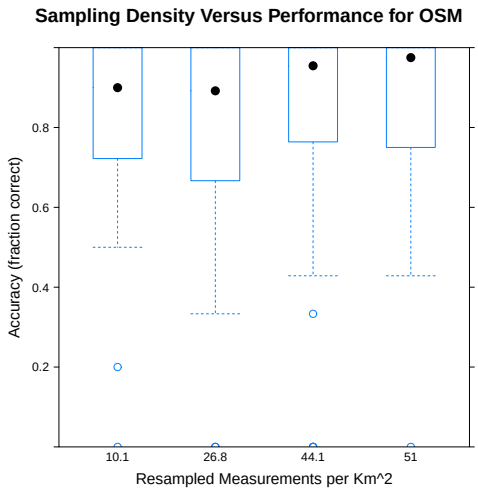


Figure 7.19: Hole prediction accuracy as a function of sampling density for OSM data.

state of the art *a priori* predictive methods or simple data fitting approaches. It also acheived nearly identical performance at predicting coverage holes to the sophisticated method of Robinson et al. [200], using the same number of measurements.

Next, crowd-sourced coverage mapping was investigated to understand the promise and practical limitations of a measurement methodology which requires no work from the network operators themselves. Although a simulation-based analysis of measurement coverage showed that only a small fraction of the population needs to participate to generate sufficient data, a case study using real measurements from the OSM project showed that in practice participation is too small to generate enough measurements for general purpose mapping. However, even with sparse measurements, the geostatistical mapping approach proposed in this thesis performs admirably. Although promising, a great deal more work is needed to understand the abilities and limitations of crowd-sourced coverage mapping in practice. Thankfully, from the perspective of further research, the data will just keep piling up on its own.