On Sampling Procedures for Detection of *Heterodera glycines*, the Soybean Cyst Nematode, and Other Soil Dwelling Organisms

by

Alexander McLellan

B.S., Louisiana State University, 2013

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Perla Reyes

Copyright

Abstract

Heterodera glycines, or the soybean cyst nematode (SCN), is a parasite that targets and damages the roots of soybean plants. It is the most yield-limiting pathogen of soybean in the U.S. and the reliable detection and accurate estimation of population densities is crucial to research and management of this pathogen. A study was performed to understand the effects of crop rotation on the prevalence of SCN. Standard sampling procedures in the plant pathology community dictate taking soil samples from potentially infected fields, processing them and counting the number of eggs in one 1 mL subsample via microscope. Suspecting the traditional procedure may lead to invalid results, false negatives in particular, the researcher created and implemented a sampling procedure based on his knowledge of sampling methods and constraints of sampling in the field. Using the data collected, we will discuss the strengths and limitations of the procedure in estimating the population density of SCN in the field. In addition, a simulation study informed by the data will be conducted to determine a sampling strategy that will yield accurate results while still considering the conditions in the field. Knowledge on how the different stages of the sampling procedure for SCN affect the accurate detection of the pathogen would extend to experimental designs and sampling methodologies for other soil dwelling organisms.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This report summarizes the analysis of data provided by Dr. Oscar Perez-Hernandez, a plant pathologist from the University of Central Missouri who is currently focusing on *Heterodera glycines*, or the Soybean Cyst Nematode (SCN). Dr. Perez-Hernandez is seeking to model the population distribution of the parasite in order to better understand the effect of crop rotation on the severity of infections in fields, which is the ultimate goal of this study. SCN is a plant parasite that damages and kills soybean plants by targeting their roots. Symptoms include stunted root growth, yellow leaves, and reduced crop yield, but distinctive symptoms may not be readily visible above ground or may be misdiagnosed as another illness since they are not unique. Its presence in a field is tested by processing soil core samples and then counting the number of SCN eggs in a 1 mL subsample from the processed sample. Once a field is determined to be infected, it is believed to be counteracted by planting SCN resistant varieties of soybean or with crop rotation since the prevalence of the parasite tends to drop when farmers change from soybean to other crops, such as corn.[1] Perez-Hernandez, and our own analysis of the data, showed some evidence of this reduction on estimated population counts. However, limitations on the data and statistical modeling tools available forced us to treat these results with caution.

Rotation	Field	Sample	Longitude	Latitude	count1	count2	count3
1	1	1	-96.4876	41.95298	25	15	20
1	1	2	-96.4899	41.95294	0	0	0
1	2	1	-96.3479	41.90825	40	35	37
1	2	2	-96.3492	41.90727	265	260	300
1	46	9	-96.6552	41.69038	10	12	5
1	46	10	-96.656	41.69101	0	0	0
2	1	1	-96.4876	41.95298	10	15	9
2	1	2	-96.4899	41.95294	0	0	0
2	2	1	-96.3479	41.90825	2	1	4
2	2	2	-96.3492	41.90727	125	120	130
2	46	9	-96.6552	41.69038	0	0	0
2	46	10	-96.656	41.69101	0	0	0

Figure 1.1: *An example of the data provided by Dr. Perez-Hernandez. Values for the counts have been changed for proprietary reasons.*

1.1 Data Collection

A selection of 25 fields, known to be infected with SCN, were tested to determine the effect of crop rotation on the parasite. For each field, the researcher selected 10 3-by-3 meter sampling grids. These sampling locations were selected by using a zig-zag random walk pattern while avoiding the edges of the field. The edges were avoided since the researcher expected concentrations to be higher close to roads and borders with other fields, due to outside contamination, and thus desired to avoid bias. The researcher began at a convenient spot in the field, set up a sampling grid, walked a certain distance, set up another sampling grid, turned approximately 45°, walked the same distance, and repeated to create the zig-zag pattern. The researcher stayed to one side of the field, until he reached the end, and then repeated the process for the other side. Within each grid, 20 randomly selected 2.5 cm diameter solid cores (15-20 cm deep) were collected and mixed into a composite sample for that grid.

In the lab, a 100 cm^3 subsample was taken from each composite sample and was processed according to standard procedure for SCN. This involved sifting out the eggs and suspending in water. Once processed, three 1 mL sub-subsamples were taken and the number of parasite eggs were counted manually under a microscope. This resulted in three counts for each sampling grid. This entire procedure was done twice for each field and sampling grid, before

and after rotation of crops. The location of the grids were recorded to facilitate sampling from the same points after crop rotation. An example of the data is shown in Figure 1.1.

The only variables available to us were rotation, field, sample, and sub-subsample. Since our focus was on estimating population density rather than explaining it, we did not include the additional variables recorded by the researcher in the analysis; such as soil pH, soil organic matter, accumulated rainfall, number of days with soil temperature below freezing, et al.[1] As a consequence of the sampling procedures, the true samples were the soil cores that were mixed into the composite sample. The composite sample made it impossible to account for the variance at this level, which left us with only what is really the sub-subsamples taken from the $100cm^3$ subsample that was taken from all of these samples mixed together. Potential solutions for this include either abandoning the composite sample idea completely, and only take single soil cores for each sample, or taking multiple subsamples from each composite sample in order to estimate and account for some of the variation within them.

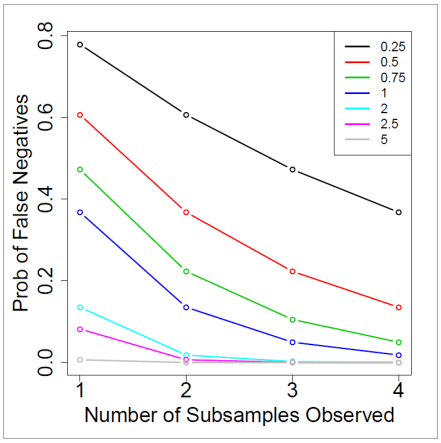
Collecting a composite sample was typically recommended to farmers seeking to detect infestation of the parasite in their fields. It seemed that this method may have been developed with a focus on determining *if* a field was infected while limiting the number of samples sent to the testing facility. Processing and counting eggs under the microscope is a time consuming and costly procedure, so it is advantageous to keep sample size low while simultaneously sampling from large areas of the field. Unfortunately, the sampling method performed is generally not ideal for determining a model for population densities. A number of issues arose in our attempts, which will be discussed in the next chapter.

The zig-zag random walking pattern potentially introduced correlation between the samples. There was also some concern about the 'human element' introduced in this method as people are not typically able to make completely random selections and often introduce unquantifiable bias. This pattern can also be prone to missing large areas of a field and

some regions may not be sampled. Thus a random walk pattern may not be ideal for population density estimation. Faced with these issues and concerns, the focus of this research switched to determining what information could be gathered from the available data and what sampling method could be recommended to Dr. Perez-Hernandez for a study that would better estimate SCN population densities in infected fields.

1.2 Analysis

Before the beginning of this project, Dr. Reyes had already worked with Dr. Perez-Hernandez on this data for some time searching for ways to avoid false negatives in the lab. Typically, only a single 1 mL sub-subsample is taken from the 20 mL processed subsample. They wanted to determine if taking additional sub-subsamples would prevent counts of zero when there were actually eggs in the beaker. The following Bayesian model was defined:



$$U_i \sim Bernoulli(\pi)$$

$$X_{i,j}|U_i = 1 \sim iidPoisson(\lambda_1)$$

$$X_{i,j}|U_i = 0 \sim iidPoisson(\lambda_0)$$

$$\pi \sim Beta(a, b)$$

$$\lambda_1 \sim Gamma(\alpha_1, \beta_1)$$

$$\lambda_0 \sim Gamma(\alpha_0, \beta_0)$$

Figure 1.2: *Theoretical false negative rates based on the data, by expected number of eggs in 1 mL subsamples (referred to as sub-subsamples in this report).*

where U_i is a binary random variable representing whether or not the i^{th} beaker contained eggs, with a success probability π modeled with a Beta distribution, and $X_{i,j}$ is a discrete random variable indicating the number of eggs counted in the j^{th} 1 mL sub-subsample from the i^{th}

beaker. The counts followed a Poisson distribution with one of two means depending on the value of U_i, each of these means were modeled with Gamma distributions.

The model showed that the second count greatly reduced the incidence of false negatives for low density SCN populations but did not add much additional information when the first count was not a zero. A third count would further reduce the false negative rate, as would be expected. However, the reduction would be marginal for most cases, except for lower expected egg densities. Considering the additional work that each new count required, it was determined that the best approach was to take a single count in general but to take a second if that count yielded a zero. This would allow researchers to reduce the incidence of false negatives while not requiring multiple counts for every subsample. The results from this work are shown in Figure 1.2.

The prospect of spatial correlation was investigated for obvious reasons. Variograms, measurement sequence graphs, and bubble plots were created. However no spatial correlation was apparent in the data, likely due to the small sample size, the nature of the composite samples, and/or the distant spacing of the samples. This was thus excluded from the models that were be considered.

A Poisson Mixed Effects model was fitted and tested for goodness of fit. Issues with overdispersion prompted switching to a Negative Binomial Mixed Effects model. However, in both of these models, there was severe zero inflation. The Zero-Inflated Poisson and Negative Binomial Mixed Effects models are currently not able to be fitted in SAS. Alternative modeling approaches, such as building and fitting a Hierarchical Bayesian Model, were considered. However they were beyond the time constraints and scope of this project. Thus we focused our efforts on a simulation study in order to determine a better sampling method for future studies.

1.3 Simulation Study

Custom R code was written to generate Gaussian random fields, take samples from those fields, fit a basic Poisson model to each set of samples, estimate values at a set of locations, and calculate three criteria utilizing the true intensities to judge the performance of each sampling method. At first, three sampling methods were evaluated: a completely random sample, a stratified grid sample, and the 'zig-zag random walk' method that Dr. Perez-Hernandez originally performed. Poor results for the criteria and problems with estimablility when the models were calculated lead to the conclusion that using only a sample size of 10 was not enough to accurately estimate the population distribution.

Further research and thought lead to some changes to the code which accommodated a better model that incorporated spacial correlation and larger sample sizes while it included two additional sampling methods. The 'random walk' pattern was dropped while fixed grid and fixed grid with random sub-grids methods were added. Several different methods for generating the random fields were explored and increasing sample sizes were tested. Finally, higher resolution fields were also tested.

1.4 Results

Ultimately, it was determined that the sample size that was used by Dr. Perez-Hernandez is much too small to accurately estimate the population density of a field. Furthermore, it was determined that the fixed grid method performed the best according to the criteria proposed in all cases. However, the stratified grid method also performed very well. Thus, it was recommended that the researcher use an 'almost' fixed grid method with some tolerances for situations were a strictly fixed grid may not be viable due to limitations in the field. As in most situations, the larger sample sizes produce more accuracy in the results. Although sample sizes as small as 25 can provide reasonable results, a larger sample would

be preferable. It is also determined that, due to the nature of the Soybean Cyst Nematode, there will likely be severe zero-inflation. While it is possible that the larger sample size will result in the Poisson model being sufficient to study the data, as opposed to the Negative Binomial currently required for the data available, a Zero-Inflated Mixed Effects model would be required to fit the new data.

Chapter 2

Data Analysis

2.1 Exploratory Analysis

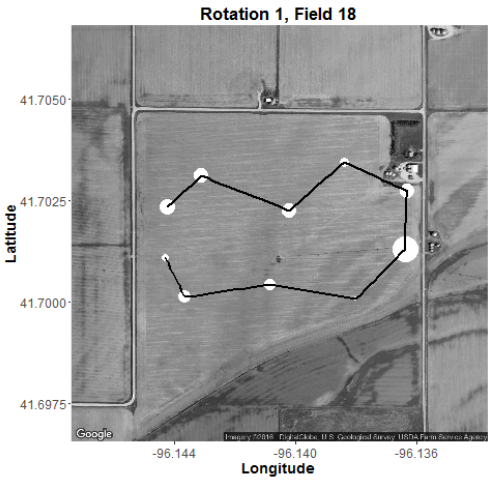


Figure 2.1: *The bubble plot for field 18, before crop rotation, overlaid on a satellite image of that field. More of these plots can be seen in Appendix A.*

In Figure 2.1, we can see a satellite map from *Google Maps*^{TM} of one of the sampled fields showing the path that the researcher took in collecting the data, the rest of the maps are provided in Appendix A. The size of the bubbles indicate the number of eggs averaged across the three counts, on a square root scale due to massive differences between the average counts. Henceforward, "Rotation 1" will refer to before crop rotation while "Rotation 2" will refer to after crop rotation. No particular patterns can be observed beyond the large variability from sampling to sampling locations. The histograms and box-plots in Figure 2.2 show the distribution of counts across all fields comparing before and after crop rotation.

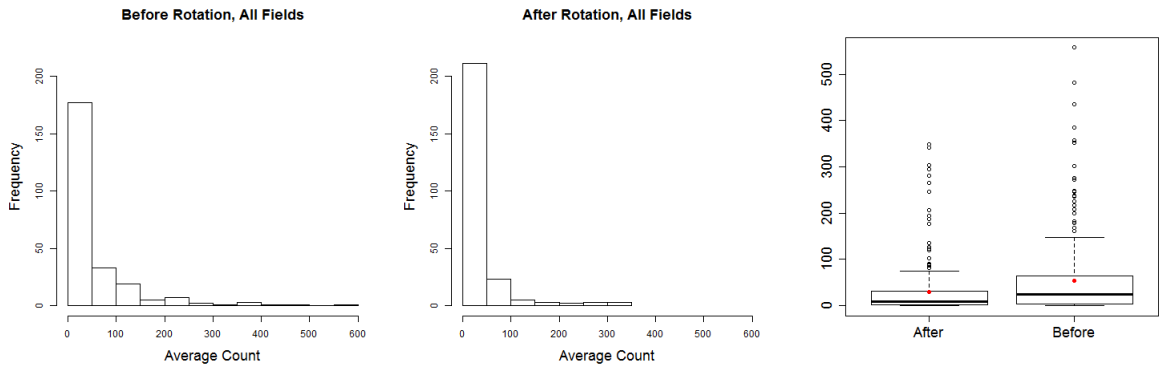


Figure 2.2: *Histograms and box-plots (with mean indicated by a red dot) of the averages of the three counts from each composite sample show a large number of samples with counts of zero.*

Before rotation, 14.8% of the composite samples had zeros for all three counts and 17.5% of the individual counts were zero. After rotation, there were 18.8% of samples with all zeros and 22.4% of the individual counts were zero. This was out of 250 composite samples and 750 individual counts for each rotation. The average counts were 53.18 and 30.15 for before and after rotation, respectively. However, as can be seen in the histograms, both of these distributions are heavily right skewed. These results, and the descriptive plots, seemed to imply that there was a reduction in the severity of SCN infection in the fields after crop rotation, but we could not make the claim that this reduction was significant.

2.2 Spatial Analysis

Due to the fact that the data was collected in enclosed fields, the possibility of spatial correlation between the samples within a field was investigated. The different fields were not near to each other so correlation between fields was not a concern. Sample variograms were calculated using the average of the counts for each sampling grid instead of each of the three counts. The repeated counts were not providing additional information about the spatial associations, but they would introduce false variability. We opted for average over

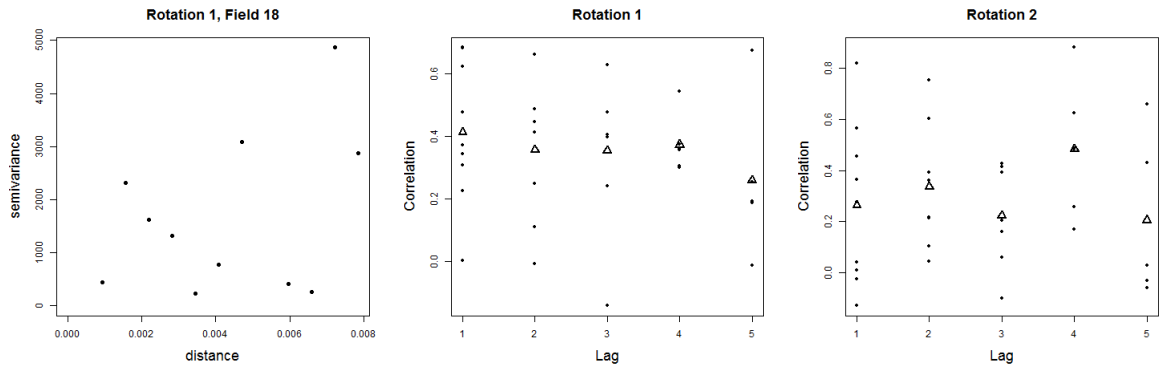


Figure 2.3: *This variogram (left) of field 18, before crop rotation, as well as the measurement sequence correlation graphs from before crop rotation (middle) and after (right) show that there did not appear to be spacial or sequential correlation in the data collected. More variograms can be seen in Appendix B.*

total to keep quantities in a smaller scale. A variogram for one location, from before crop rotation, is presented in Figure 2.3, all of the other variograms show similar patterns (see Appendix B). All variograms produced patterns inconsistent with spatial correlation and, in fact, did not show any recognizable patterns. The bubble plots further reinforced this. This analysis was done using R software.[2]

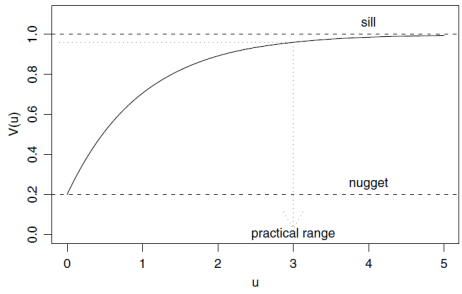


Figure 2.4: *The theoretical variogram indicating spacial correlation.*

A variogram is a method of analyzing spatial correlation by calculating the semivariance of pairs of data points with respect to the distance between them. The semivariance is half of the variance of the difference between the values for all pairs of locations at a particular distance apart.

$$V(u) = \frac{1}{2} Var[S(x) - S(x-u)]$$

with $S(x)$ as the measurement of the attribute at a random location, x, and $S(x-u)$ as the measurement of the attribute at locations that are at u distance from location x.

If a field has significant spatial correlation, the semivariance will start low with small distances and increase rapidly until reaching a maximum value. This would mean that the attributes tend to be fairly similar for locations close to each other. As the distance between locations increases, the variance of the difference between the attributes of pairs of points at that distance apart would also increase indicating that locations separated at that distance are less related to each other. The semivariance then reaches a maximum, called a sill. Pairs of points very far apart will have about the same semivariance, equal to the overall variance of the field, due to lack of correlation at large distances. At a particular distance, u, the semivariance is estimated with half of the mean of the squared differences between the attributes within all pairs of points that are that distance apart, within a certain tolerance.[3]

$$\hat{V}(u) = \frac{1}{2N} \sum [Y(x_i) - Y(x_j)]^2$$

For estimation, there is typically a tolerance for each level of distance as it is difficult to find a large number of pairs of locations that are, for example, exactly 100m apart. In the data provided, there was no smooth curve as what would be expected with spatial correlation. The estimated semivariance was very erratic for all locations, as can be seen with the example in Figure 2.3 and the rest of the variograms in Appendix B.

Due to these results we had reason to believe that, either, there was no spacial correlation between the composite samples within each field, or, any correlation was on a smaller scale than what the observed data measured. The latter possibility is mainly derived from the erratic nature of the variograms. It may be that the data points were already too far away from each other and the sill had already been reached. This also may have been due to the composite sample being a combination of samples collected across a small area in a sampling grid or a result of the small sample size not being suitable to accurately measure the semivariance. Whatever the reason, there was not enough information to determine the

nature of correlations between each of the individual samples within the fields. Based on this, we did not specify a spatial correlation structure in the modeling of this data.

To identify any correlation effect by the sequence in which samples in a field were collected, correlations between samples and the next few samples along the path in the random walks were also investigated. The average counts for each index, i.e. first sampling grid, in all of the fields were collected into a vector and correlations between these index vectors were calculated. The measurement sequence correlation graphs are also shown in Figure 2.3. The 'lag' is the difference in index numbers of the two vectors for each correlation. The points plotted represent the correlations between each pair of indices for each value of lag, incorporating all 25 fields. The average sequential correlations, marked with triangles, in Figure 2.3 appeared to be constant around 0.35. However, considering the large variability and small sample sizes, the observed sequential correlations seemed mostly low for both before and after crop rotation. Given the lack of evidence for sequential correlation, we did not introduce it into the model.

2.3 Generalized Linear Mixed Modeling

Since count data was being used, a Poisson regression model was the logical starting point. The available variables were: before and after rotating crops, each field that was sampled, the sampling grids within each field, and the three counts from each sample. The counts came from the single subsample taken from each composite sample and each sampling grid produced a single composite sample, so it would have been redundant to include separate effects for these levels and they could be bundled together into a 'sample' effect. This sample effect was considered a random effect as it was randomly selected from all potential sampling grids across the field. The rotation and field effects were considered

fixed effects. The model used was

$$\eta = log\left(E\left[y|\lambda\right]\right) = \beta_{rotation} + \beta_{field} + \beta_{rotation*field} + u_{sample(field)}$$

In fitting this model, the estimated dispersion factor was calculated to be over 5, that is the generalized chi-square test statistic divided by the degrees of freedom. This indicated over-dispersion and prompted the use of a Negative Binomial model instead. For a Poisson random variable, the variance is meant to be equal to the mean. Over-dispersion occurs when this is not the case and the variance is larger than the mean. This indicates that the Poisson model is not appropriate, which is usually resolved by applying the Negative Binomial model. The Negative Binomial model is the result of applying a Poisson model while considering the rate parameter as a random variable with a Gamma distribution.

$$Y|\lambda u \sim Poisson(\lambda u),\ u \sim Gamma(\frac{1}{\phi}, \phi) \Rightarrow Y \sim NegativeBinomial$$

with mean λ and variance $\lambda(1 - \lambda\phi)$. This allows for random count variables to have variances different from their respective means.[4]

Another problem that was encountered involved a large numbers of zeros in the observed data, with more zeros after crop rotation than before. This was further evidenced in the histograms of the data illustrated in Figure 2.2 and most apparent in the residual and Q-Q plots produced when attempting to fit the Poisson and Negative Binomial models to the data. These plots showed a flat line, representing the residuals associated with the zero values, containing more values than the model could accommodate. These values influenced the estimated parameters, resulting in a poor fit. Observing a large number of zeros can occur when there are two different distributions or processes generating the data: one distribution when the area is infected and another when it is not infected. This is

handled by utilizing a zero-inflated model described as:

$$Pr(Y = y) = \begin{cases} \pi + (1-\pi)f(0) & \text{for } y = 0 \\ (1-\pi)f(y) & \text{for } y = 1, 2, \ldots \end{cases}$$

with Y as the random variable of interest, π as the probability that the area is not infected, and $f(y)$ as the discrete probability function. The probability function was a Negative Binomial for this study, but could also be a Poisson. This type of model considers the two cases, utilizing a binomial process to determine the probability that the area is infected similar to the Bayesian model described in Section 1.2. It still recognizes the possibility of a zero count in an infected area. If the area is infected, then it will follow the specified model. If not, the expected count will simply be zero.[4] It was thought that this may have been the source of over-dispersion in the previous fit but, when a Zero-Inflated Poisson Model without mixed effects was fitted to the data, there was still over-dispersion. Due to this, a Zero-Inflated Negative Binomial model was required. Both fitted models showed a significant effect for crop rotation, however we could not be sure of the validity of these results given that neither model fitted the data properly.

Unfortunately, the GLIMMIX procedure was not able to apply this model and the GENMOD procedure was not able to incorporate random effects.[4] Further research showed that fitting a Zero-Inflated Negative Binomial Mixed Effects model is an open question that still has not been fully explored. There was some promising work involving the NLMIXED procedure towards this end,[4] however in the interest of keeping the project within the time constraints of a master's project, fitting and validating a model using NLMIXED was left for a follow up analysis. Future work may also involve writing a custom program to fit this type of model or taking a Bayesian approach, which is beyond the scope of a Master's project. Model fitting was completed using SAS® software, version 9.4 of the SAS system

for Windows.[a]

Unable to analyze the data provided while faced with the concerns over using composite samples and the 'random walk' method performed, it was decided to instead focus on recommending a sampling procedure for Dr. Perez-Hernandez to utilize that will produce accurate results. With a recommendation from a statistician, he may apply for a grant to perform a new study with a sampling procedure capable of gathering more information about the pathogen's population distribution in an infected field and how it is affected by crop rotation.

Chapter 3

Simulation Study

3.1 Small Sample Study

A series of Gaussian Random Fields with a Matérn covariance model was generated in order to test the various sampling procedures. This model has a 'smoothness' parameter that can be adjusted so that different sets of simulations can be run with relatively 'smooth' and 'patchy' fields, such as the ones in Figure 3.1. For each field, a set of center-points were aligned in a 16-by-9 grid in the x-y plane. Without loss of generality, the fields could have been a unit square. We chose a rectangular shape to mimic more traditional fields for the researcher. Each center-point was randomly assigned, according to the model, a mean intensity value which was associated with a fixed area around the center-point, this fixed area will be called a "cell." The cells were evenly sized so as to completely cover the field and not overlap.

The fields were generated using a Gaussian spatial process where for any set of coordinates $x_1, x_2, \ldots, x_n$ with $x_i \in \Re^2$, the joint distribution of $S(x_1), S(x_2), \ldots, S(x_n)$ was Multivariate Gaussian with a mean of 0 and a Matérn covariance function (Eq. 3.1). The values generated from the process were then converted into Poisson intensities, with an arbi-

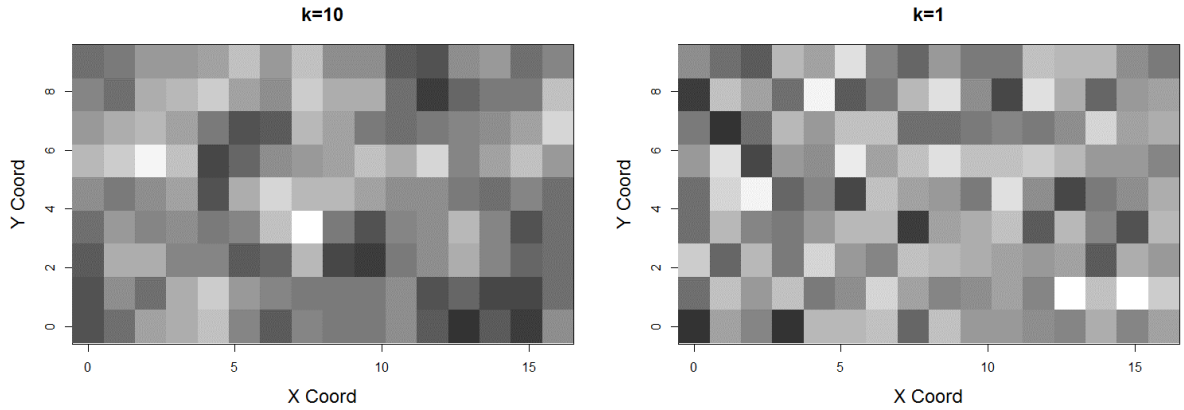


Figure 3.1: *Examples of the random fields generated in the simulation study, including 'smooth' (left) and 'patchy' (right) variations.*

trarily chosen mean of 0.5, by applying the exponential function $\lambda = e^{0.5+data}$. The Matérn family of covariance functions relies on two parameters and follows the form:

$$\rho(u) = \left[2^{k-1}\Gamma(k)\right]^{-1} \left(\frac{u}{\phi}\right)^{k} K_k(u/\phi) \tag{3.1}$$

with $\phi > 0$ as a scale parameter and $k > 0$ as a smoothness parameter. K_k is a kernel function, based on the modified Bessel function, of order k.[3]

At first, three sampling procedures were tested: the random walk pattern used by Dr. Perez-Hernandez, a completely random sampling, and a stratified grid method. Examples of each of these are shown in Figure 3.2. Since the researcher used ten samples per field, the same number was used in order to mimic what was actually performed. Due to lack of information about the composite sample, it was not incorporated into the simulation study. Moreover, previous research, completed by Dr. Reyes, showed that the additional counts taken from each sample added little information unless the initial count is zero.[1] While it could greatly reduce the false discovery rate in this instance, that was not the focus of this study and thus the additional counts were not included in the simulations.

Custom code was written with R software,[2] making use of the geoR package,[5] to run the

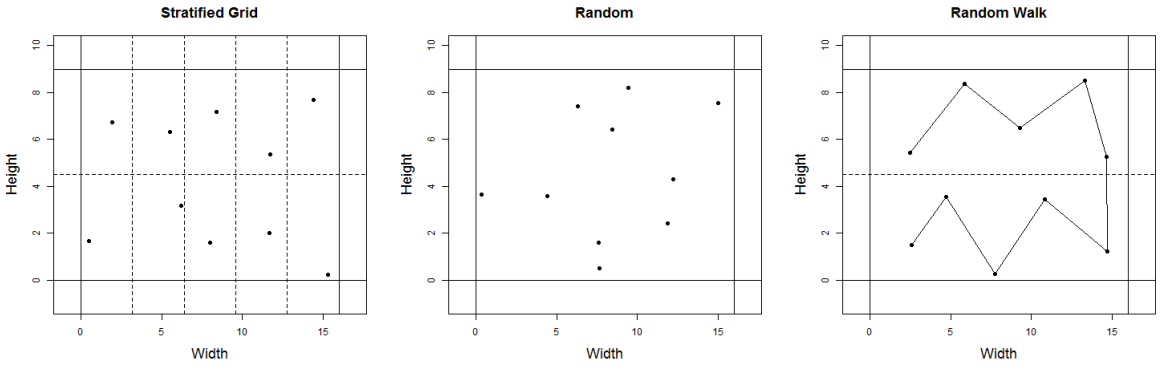


Figure 3.2: *Examples of the sampling procedures generated in the simulation study. Solid lines indicate hypothetical field borders*

simulations. 100 iterations were run to test the three procedures. Within each iteration, a single field was randomly generated and 500 different groups of sampling locations with size 10 were randomly selected for each procedure. Each of these groups of 10 sampling locations will be called a "sample" from this point forward. At each sampling location within each sample for each procedure, a single random Poisson variable was generated using the true mean associated with the cell containing the location. Once all of these counts were taken, a standard Poisson model was fitted for each sample using the x and y coordinates as predictors.

$$\eta = log\left(E\left[y|\lambda\right]\right) = \beta_1 x + \beta_2 y \tag{3.2}$$

The fitted model was as simple as possible for ease of computation and, regardless of its appropriateness, the focus was to compare sampling methods with all other things being equal. This model was then used to estimate the means for each cell in the entire field, specifically by predicting the mean at each center-point. With these estimates, three different goodness of fit criteria were calculated for each iteration. They were as follows:

$$\underset{i}{maximum}\ \widehat{MSE}(x_i) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\lambda}_b(x_i) - \lambda(x_i)\right]^2 \tag{3.3}$$

Smooth					Patchy				
Method		$\widehat{MSE}(x_i)$		$\widehat{MSE}(\mu_\lambda)$	Method		$\widehat{MSE}(x_i)$		$\widehat{MSE}(\mu_\lambda)$
		Mean	Max				Mean	Max	
Grid	Min	15	244	1	Grid	Min	41	447	3
	Lower	133	3,250	10		Lower	163	5,160	10
	Mean	1.31E+77	1.88E+79	9.06E+74		Mean	1.73E+126	2.49E+128	1.20E+124
	Upper	27,100	2,330,000	1,150		Upper	27,800	2,240,000	1,080
	Max	1.27E+79	1.82E+81	8.79E+76		Max	1.73E+128	2.49E+130	1.20E+126
Random	Min	244	3 310	27	Random	Min	156	6 200	12
	Lower	50,800	5,270,000	1,560		Lower	28,300	2,070,000	1,370
	Mean	1.11E+213	1.59E+215	7.67E+210		Mean	4.76E+296	6.86E+298	3.31E+294
	Upper	2.71E+12	3.54E+14	4.22E+10		Upper	2.25E+11	2.96E+13	3.43E+09
	Max	1.07E+215	1.54E+217	7.44E+212		Max	4.71E+298	6.79E+300	3.27E+296
Walk	Min	11	103	1	Walk	Min	18	238	2
	Lower	77	1,960	6		Lower	100	2,470	8
	Mean	2.48E+109	3.57E+111	1.74E+107		Mean	2.84E+134	4.08E+136	1.97E+132
	Upper	1,450	61,200	106		Upper	1 640	110,000	113
	Max	2.48E+111	3.57E+113	1.74E+109		Max	2.84E+136	4.08E+138	1.97E+134

Table 3.1: *Results from 100 fields with 500 samples, of size 10, taken from each.*

$$\underset{i}{mean}\ \widehat{MSE}(x_i) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\lambda}_b(x_i) - \lambda(x_i)\right]^2 \tag{3.4}$$

$$\widehat{MSE}(\mu_\lambda) = \left(\tfrac{1}{B}\sum_{b=1}^{B}\bar{\hat{\lambda}}_b - \bar{\lambda}\right)^2 + \frac{1}{B}\sum_{b=1}^{B}\left[\bar{\hat{\lambda}}_b - \tfrac{1}{B}\sum_{b=1}^{B}\bar{\hat{\lambda}}_b\right]^2 \tag{3.5}$$

with $\hat{\lambda}_b(x_i)$ as the estimated value for the b^{th} sample at the i^{th} cell, $\lambda(x_i)$ as the true mean for the i^{th} cell, B as the number of samples (500) within each iteration, $\bar{\hat{\lambda}}_b$ as the average of the estimated values for the b^{th} sample across all cells, and $\bar{\lambda}$ as the average of the true means across all cells. For these criteria, lower values are preferable. The first two criteria take the mean squared difference between the predicted and true mean at each cell within each iteration, averaged across all of the B samples taken for each simulated field. Equation 3.3 took the maximum mean squared difference for each iteration while Equation 3.4 took the average across all of the cells. Equation 3.5 instead determined the squared difference in the average true mean for each field and the average predicted mean, averaged across all of the cells and again across all simulations within an iteration, plus the variance of the average predicted mean.

The results of these simulations, presented in Table 3.1, were less than encouraging. The exponential model would sometimes not be estimable and the criteria would have a great

Smooth					Patchy				
Method		$\widehat{MSE}(x_i)$		$\widehat{MSE}(\mu_\lambda)$	Method		$\widehat{MSE}(x_i)$		$\widehat{MSE}(\mu_\lambda)$
		Mean	Max				Mean	Max	
Grid	Min	271.5	10.51	0.06891	Grid	Min	277.1	13.28	0.1207
	Lower	1,190	36.46	0.2315		Lower	1,788	45.36	0.3169
	Mean	6,802	100.8	0.7258		Mean	8,919	113.5	0.9189
	Upper	6,470	100.9	0.759		Upper	7,009	120.3	0.8995
	Max	150,300	1,522	11.72		Max	136,400	994.4	8.307
Random	Min	268.4	11.01	0.1379	Random	Min	281.9	13.56	0.1797
	Lower	1,190	37.74	0.4985		Lower	1,781	46.21	0.56
	Mean	6,735	106.8	1.85		Mean	8,854	118	1.803
	Upper	6,364	113.4	1.989		Upper	7,058	123.1	1.682
	Max	148,800	1,578	33.94		Max	136,200	1,011	19.82

Table 3.2: *Results from 100 fields with 500 samples, of size 100, taken from each.*

many large values in the order of $\times 10^{200}$. Also, the average value of the criteria would consistently be much higher than the upper quartiles. There would also be situations where the estimated means would be infinite or the criteria would have values larger than what the program could store. These problems occurred for all of the sampling procedures tested resulting in the appearance that they all performed equally poorly. However, among the models that would actually estimate the whole field, the stratified grid seemed to perform best. The possibility of coding error was suspected so a run with a sample size of 100 was performed, leaving out the random walk method as that would have been very difficult to code for such a large sample size and the sample produced would have been similar to the stratified grid method. The results for a sample size of 100 are shown in Table 3.2. This run produced much more favorable results, without encountering problems with estimability, hinting that the issues encountered were a result of a small sample size. It was concluded that a sample size of 10 is too small to accurately estimate the population distribution regardless of the sampling procedure performed.

3.2 Increasing Sample Size

Further research revealed that similar studies have been done before. The book "Model Based Geostatistics" by Peter J. Diggle and Paulo J. Ribeiro Jr. showed that one of the

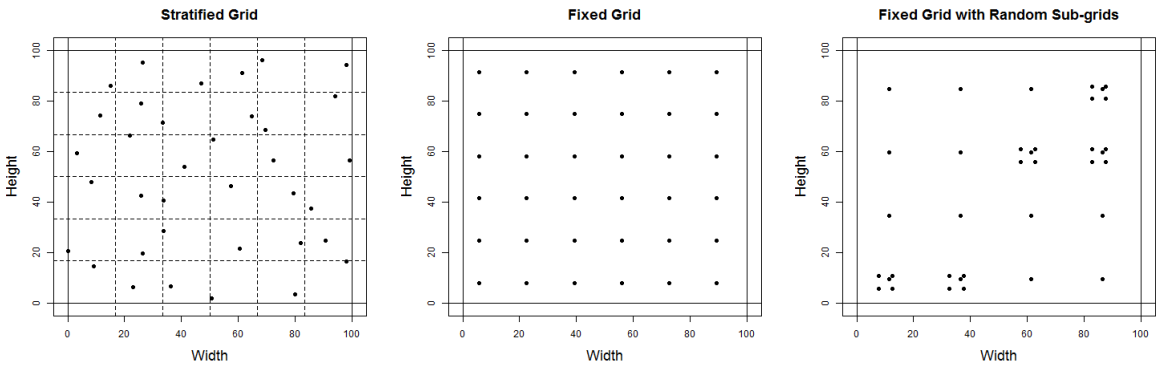


Figure 3.3: *Examples of the sampling procedures of size 36 generated in the simulation study. Solid lines indicate hypothetical field borders*

better sampling methods was a fixed grid pattern with random smaller grids inlaid within. Diggle and Ribeiro argued that inlaid grids allowed studying the entire field while also being able to determine the small scale spatial correlation structure.[3] Their results were not directly applicable for our problem since they assumed very large sample sizes in the hundreds, much larger than what would be feasible in the field, and they observed continuous data while we had counts. In light of this, fixed grid and fixed grid with sub-grids sampling methods were included in the simulations and work began to find a sample size that balanced accuracy in estimates and real-world feasibility while simultaneously determining which methods perform best. It was also determined that, for larger sample sizes, the random walk method was indistinguishable from the stratified grid procedure while being much harder to code and thus it was dropped.

With these four sampling methods; completely random, stratified grid, fixed grid, and fixed grid with sub-grids; larger sample sizes were tested. Due to code limitations we made the following two modifications to the simulation study to generate the fixed grid and sub-grid methods: 1) sample sizes were limited to square values; sizes of 25, 36, 49, 64, and 81 were tested; and 2) simulated fields were switched to squares, going to a 10-by-10 grid from a 16-by-9. These modifications were only for simplicity in the code and can be performed

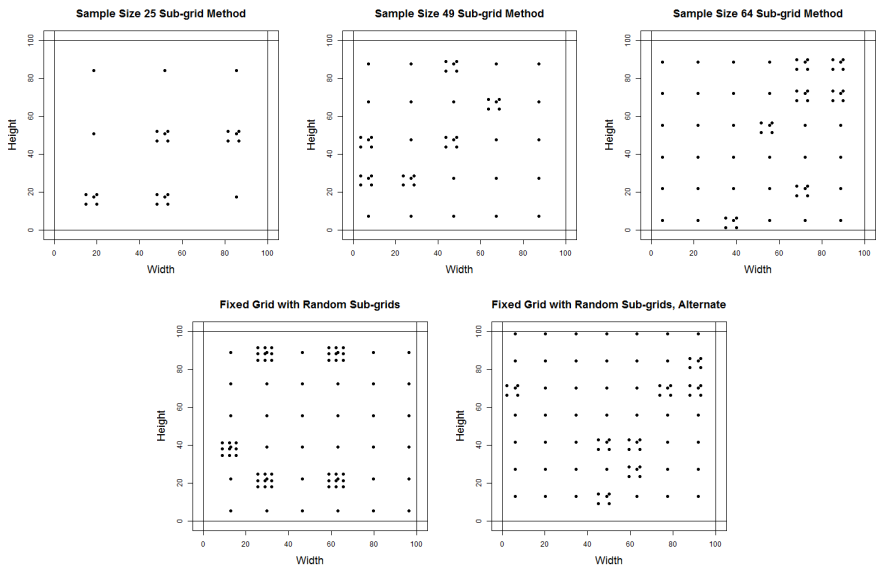


Figure 3.4: *Examples of the sub-grid sampling procedures generated in the simulation study for each of the sample sizes tested. The bottom row shows the two different types for the sample size of 81. Solid lines indicate hypothetical field borders*

without loss of generality; while in real world studies these restrictions need not be applied and these methods can be adapted to any field's shape or sample size without affecting their performance. Examples of the sampling methods are provided in Figures 3.3 and 3.4. Two similar sampling methods were used for the sub-grid method with sample size of 81, with the alternate method reducing the number of samples in the sub-grids while increasing the number of sub-grids, as illustrated in the bottom row of Figure 3.4.

Furthermore, the covariance model used in generating the fields was changed from a Matérn covariance model (Eq. 3.1) to a Gaussian one. This was to match our results to those from Diggle and Ribeiro, who used the Gaussian model for their study.[3] The Gaussian covariance model follows the function

$$\rho(u) = c_0 + c_1 \left[1 - e^{-(u/r)^2}\right] \tag{3.6}$$

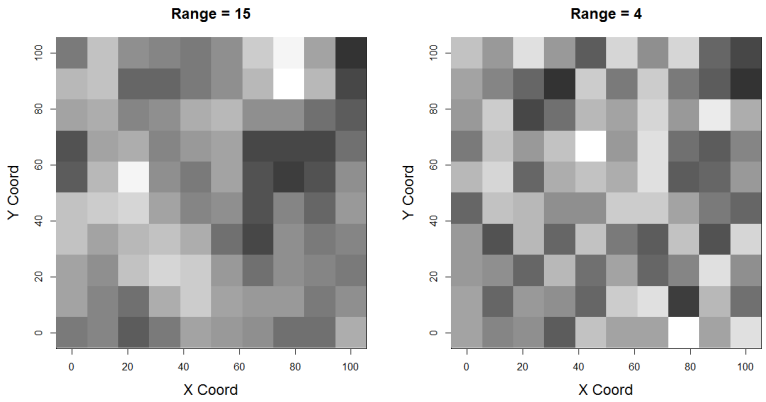


Figure 3.5: *Examples of the random fields generated with the Gaussian covariance model.*

which describes the semivariance of the field, depicted in Figure 2.4, with c_0 as the nugget, c_1 as the sill, and r as the range.[3] The 'smooth' and 'patchy' versions of the fields, were accomplished by manipulating the range parameter as this covariance model did not have a smoothness parameter. Reducing the range resulted in a more rugged terrain as shown in Figure 3.21. Additionally, the simple exponential model (Eq. 3.2) used previously was not a particularly 'good' model to fit to the data. The model generated a plane to represent the expected intensity across the field when the parameter of interest was distributed more like a mountainous or hill-covered region, as shown in Figures 3.1 and 3.21. Given this and the increased sample size, Kriging was applied to the simulated samples and a spatial prediction model was applied by way of the geoRglm package[6] in R (version 3.2.5).[2] Using the predictions obtained from the pois.krige command, the three criteria were once again calculated using this model. Kernel density plots and box-plots were produced for all of the results and provided in Appendix C. For these graphs, Eq. 3.3 is called "Criteria 1," Eq. 3.4 is called "Criteria 2," and Eq. 3.5 is called "Criteria 3" while there are also zoomed in graphs as there are many incidents of very large values.

Kernel density graphs show how susceptible to extreme cases these methods are. Smoother fields tend to concentrate mostly toward smaller values, but have a few very bad results that

produced long right tails for the criteria. In contrast, box plots present a clearer picture of the criteria's overall distribution of values. For all cases, there are a few extreme values which necessitate providing zoomed in graphs to see the details in addition to the full sized ones. Recall that, for the three criteria, smaller values are better. Thus kernel density estimators with most of the points, and their peaks, lower along the horizontal axis are considered as performing better and box plots with most of the 'box' being lower on the number line are preferable.

The following is a selection of the box-plots for the three criteria, or Equations 3.3, 3.4 and 3.5, organized by sample size. The 'high resolution' plots will be elaborated on in the next section. Sample sizes of 25, 36, 64, and 81 are included and the scale has been adjusted so that all graphs for the same criterion and resolution have the same scale. In each page, the plots in the left two columns have the same scale and so do the plots in the right two columns. Right tails and extreme values are not shown in these adjusted scales. Unadjusted box plots and Kernel density plots of the results are included in Appendices C and D.

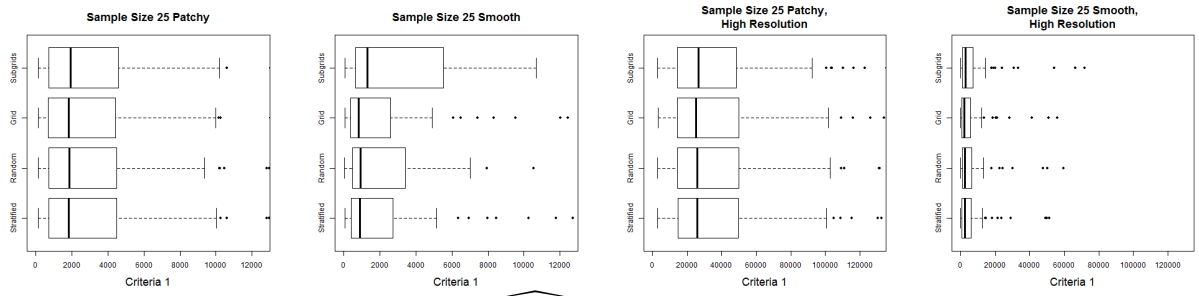


Figure 3.6: *n=25 criterion 1 (Max* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

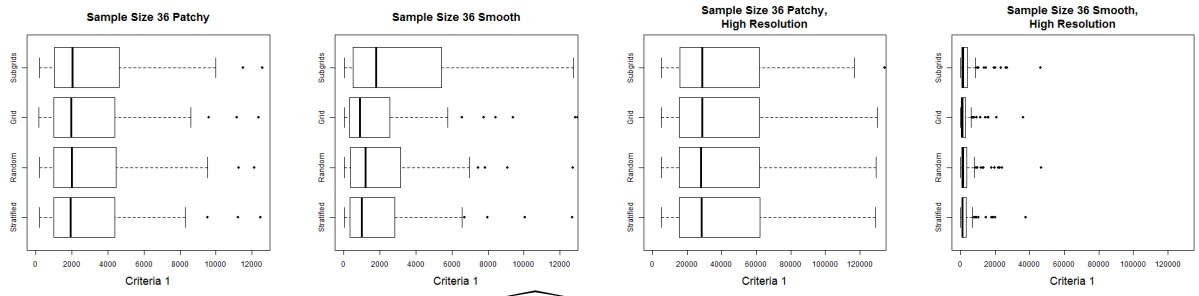


Figure 3.7: *n=36 criterion 1 (Max* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

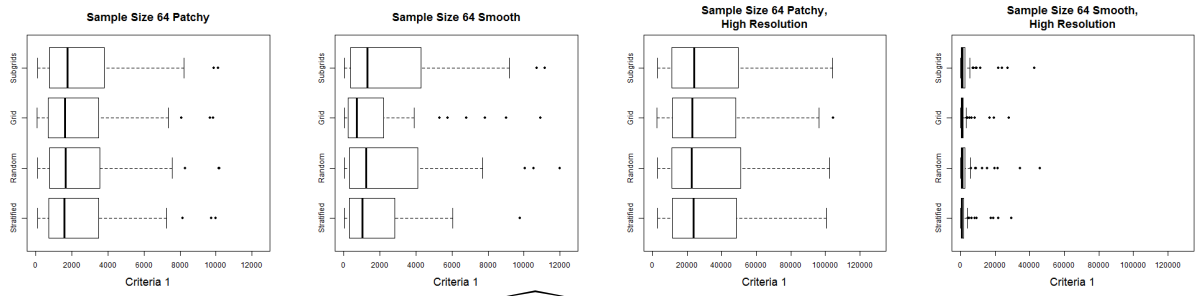


Figure 3.8: *n=64 criterion 1 (Max* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

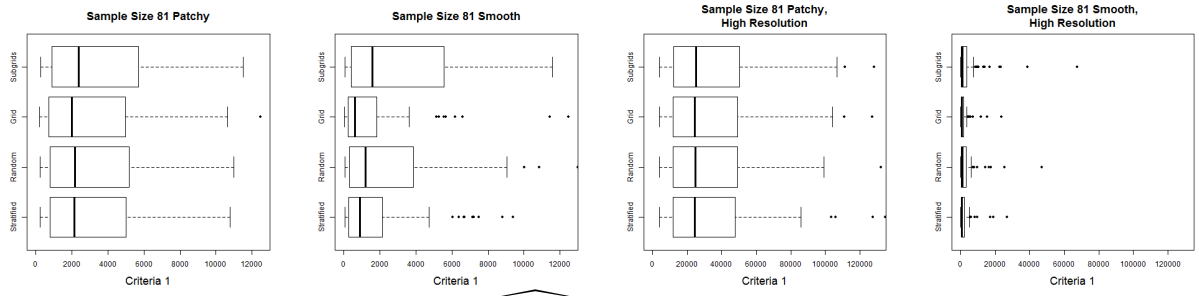


Figure 3.9: *n=81 criterion 1 (Max* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

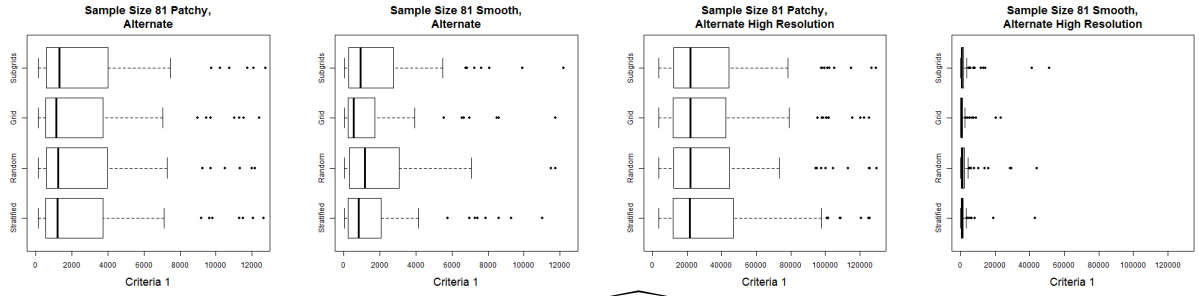


Figure 3.10: *n=81 (alt) criterion 1 (Max* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

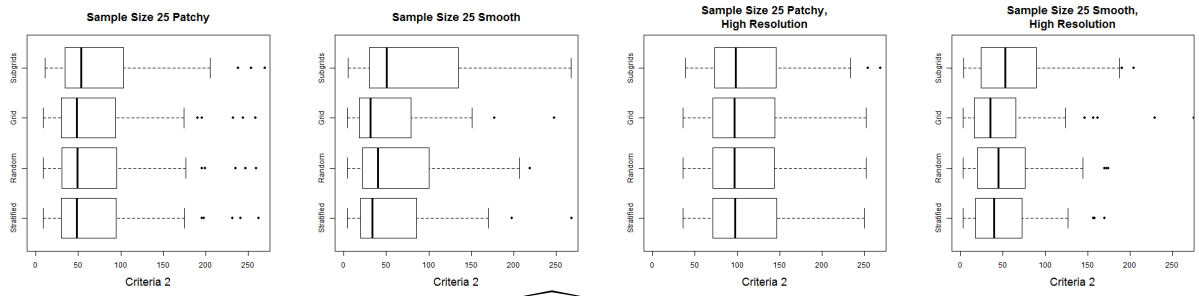


Figure 3.11: *n=25 criterion 2 (Mean* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

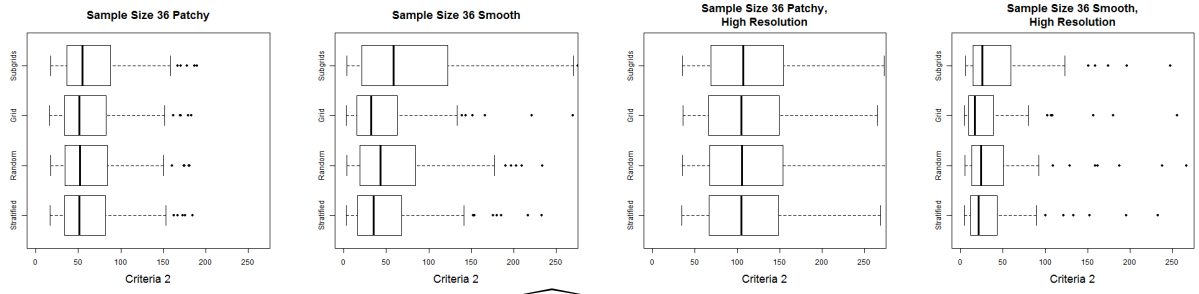


Figure 3.12: *n=36 criterion 2 (Mean* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

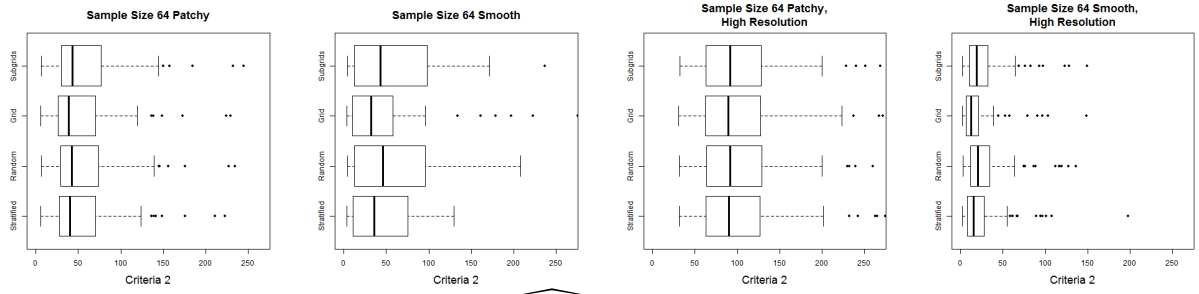


Figure 3.13: *n=64 criterion 2 (Mean* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

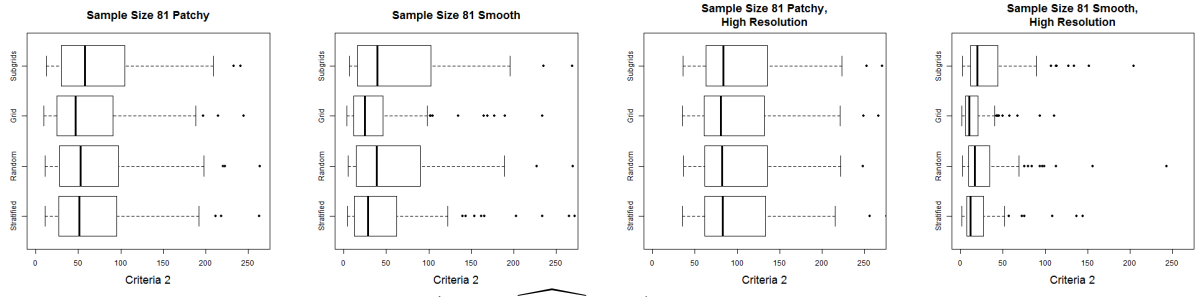


Figure 3.14: *n=81 criterion 2 (Mean* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

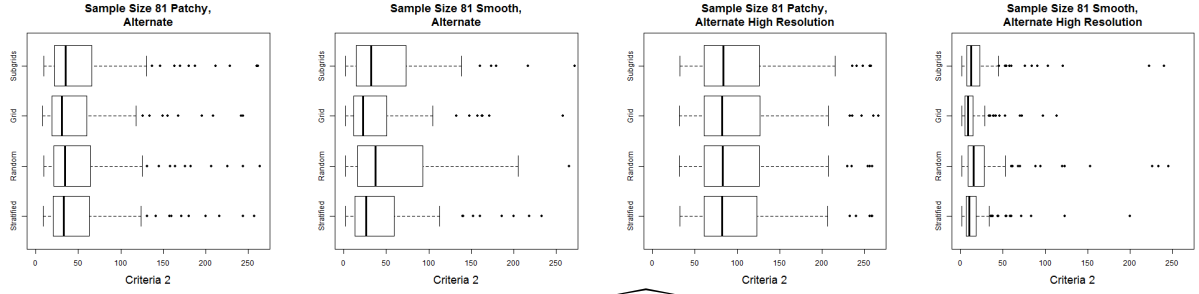


Figure 3.15: *n=81 (alt) criterion 2 (Mean* $\widehat{MSE}(x_i)$*). Same x-axis across plots in same column.*

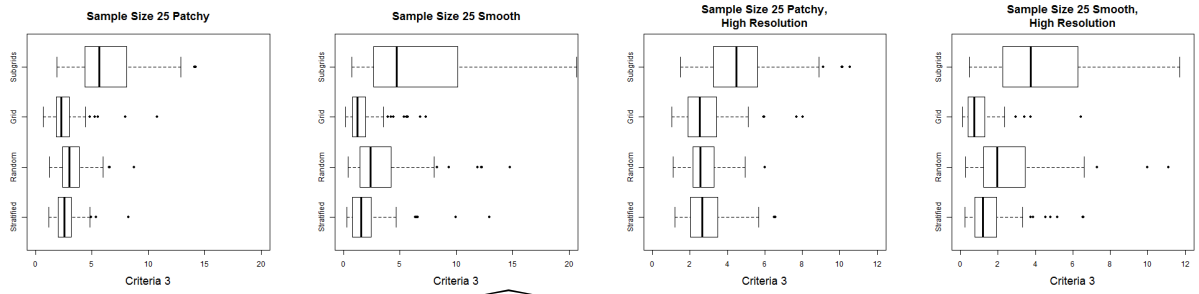


Figure 3.16: *n=25 criterion 3 ($\widehat{MSE}(\mu_\lambda)$). Same x-axis across plots in same column.*

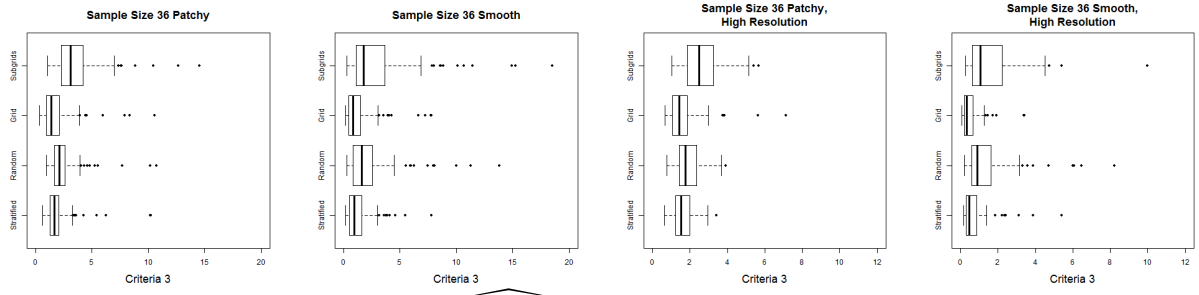


Figure 3.17: *n=36 criterion 3 ($\widehat{MSE}(\mu_\lambda)$). Same x-axis across plots in same column.*

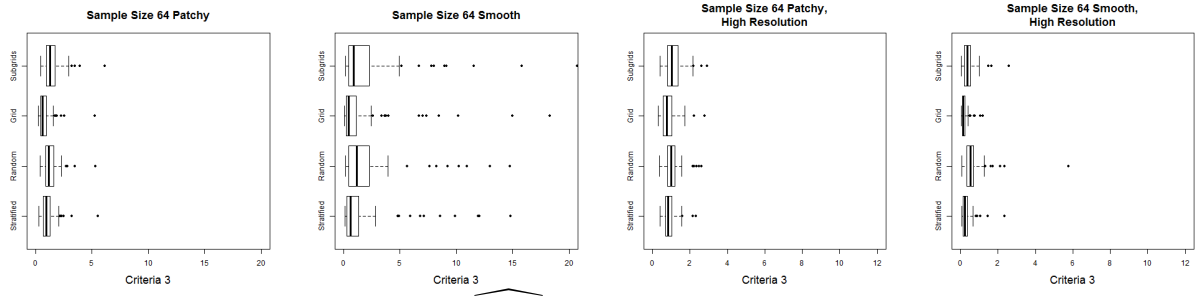


Figure 3.18: *n=64 criterion 3 ($\widehat{MSE}(\mu_\lambda)$). Same x-axis across plots in same column.*

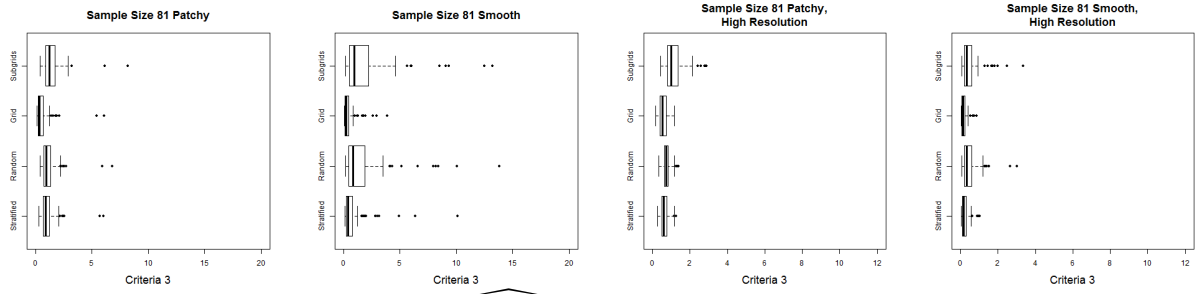


Figure 3.19: *n=81 criterion 3 ($\widehat{MSE}(\mu_\lambda)$). Same x-axis across plots in same column.*

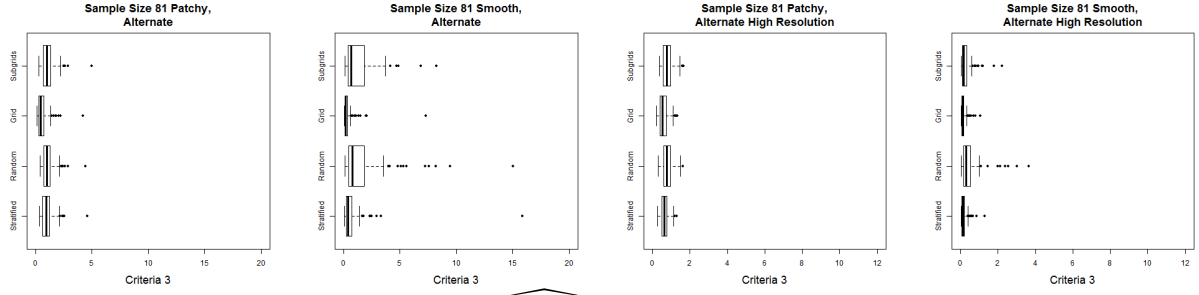


Figure 3.20: *n=81 (alt) criterion 3 ($\widehat{MSE}(\mu_\lambda)$). Same x-axis across plots in same column.*

The results stay mostly the same across the different sample sizes and each of the types of fields. For the patchy fields, there are no appreciable differences between the sampling methods for either of the first two criteria. Criterion 1, the maximum MSE, seems to show a slight improvement as sample size increases but they are fairly close together, so it is difficult to choose one way or the other. For criterion 2, the average MSE, this benefit from increased sample size seems to be more pronounced, as would be expected, however there are still no major differences between the methods. For the smooth fields, the fixed grid and stratified grid methods tend to perform better than the others in all cases. The sub-grids method would perform worse than the completely random method for the smaller sample sizes, but its relative performance seems to increase along with the sample size. Again, differences between the sample sizes are small for the first two criteria.

More dramatic differences appear in the last criterion, shown in Equation 3.5 as the MSE for the overall mean intensity for each field. Here, the fixed grid method tends to perform best, followed by the stratified grid method, the completely random method, and, finally, the sub-grid method seems to be performing the worst in general. This pattern is maintained for all sample sizes and field types. This may be due to the sub-grid method producing a reduced density of samples in order to accommodate the sub-grids, as is apparent in Figure 3.3. That being said, the sub-grid method rapidly increases in performance, relative to the other methods, as the sample size increases. It eventually catches up to the completely random method, for the larger sample sizes, but it never does as well as the fixed or stratified grid methods. It can also be noted, that all of the methods clearly perform better as the sample size increases for the third criterion.

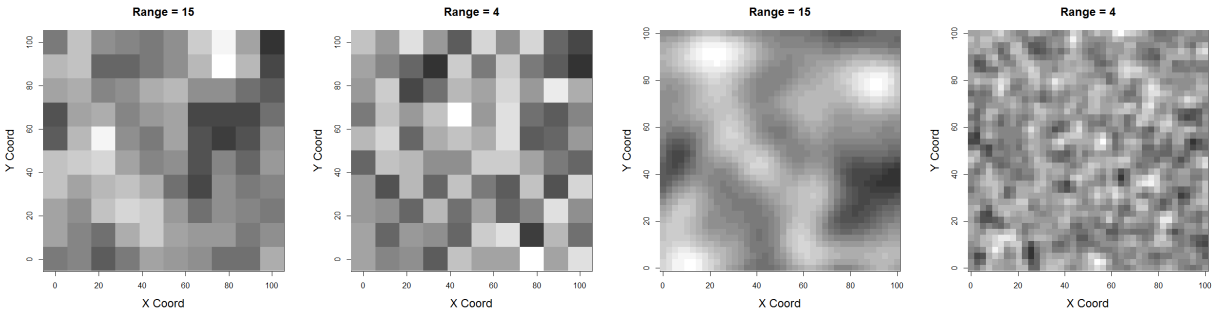


Figure 3.21: *Examples of the Gaussian Random Fields with Gaussian Covariance Models. Left to right: Smooth 10x10 grid, patchy 10x10 grid, smooth 40x40 grid, and patchy 40x40 grid.*

3.3 High Resolution Study

In the interest of thoroughness, the simulations were run again but increasing the resolution with which the fields were being generated to a 40-by-40 grid up from 10-by-10. The reasoning behind this was to determine if the advantages of the sub-grid method would be more apparent if the simulated fields had more variance on the smaller scale. Also, there was some concern that the cells were too large, resulting in multiple samples within the sub-grids sampling from the same cell. This would negate any advantages that the sub-grid method might provide. It could have also been argued that sampling from larger cells mimicked the composite samples utilized by Dr. Perez-Hernandez, assumed constant mean intensity for large areas of the field and we would prefer to avoid. With the new 40-by-40 grids the kriging model needs to estimate more points to compare to true values, causing the simulations to take considerably longer. In order to make them run within a reasonable time frame, the number of samples taken per field was reduced from 500 to 50 while leaving the number of generated fields at 100. The complete results for these simulations are provided in Appendix D, the term "High Resolution" in the plot titles aids in distinguishing them from the others.

The higher resolution fields yield similar results to before. For the patchy fields, once

again, the first two criteria have all of the methods performing similarly and most of the differences are in the third criterion. For the smooth fields, the fixed and stratified grid methods consistently perform better than the other methods, with the fixed grid surpassing the stratified grid for larger sample sizes. It can be seen, however, that the sub-grid method performs just as well as the completely random, even for small sample sizes, and began to perform better for larger sample sizes. Still, it never catches up to the fixed and stratified grid methods. It can also be noted that the smooth fields have all of the methods perform much better than in patchy fields, for all sample sizes.

The third criteria once again maintains the same order as before, but the sub-grid method surpasses the completely random method much faster than in the low resolution plots. It still does not come up to par with the fixed and stratified grid methods but, for the larger sample sizes, it starts to get close. This could imply that, if the sample size had continued to increase, the sub-grid method may have performed as well as the other grid methods, which would be consistent with the results by Diggle and Ribeiro. However, for the smaller sample sizes, it is overshadowed by the fixed grid method. Similar to the low resolution fields, the performance of all of the methods once again improve as the sample size increased.

In addition, there are no cases where the kernel density estimators produce extreme values, as was encountered previously. Ultimately, in all cases, the fixed grid method always has better or similar results to the other methods, with the stratified grid method close behind. Thus, we must conclude that the fixed grid method, shown in the center of Figure 3.3, performs the best among the methods tested at all sample sizes between 25 and 81 for both smooth and patchy fields, although the difference is minuscule for patchy fields.

Chapter 4

Conclusions

In order to quantify the observed effect of crop rotation on the reduction in Soybean Cyst Nematode populations, accurate population distribution estimation is required. For this end, data was collected from various fields by Dr. Perez-Hernandez, who has been studying the plant parasite, and analyzed. To be applicable in the field, any sampling procedure will need to be easily implementable within the constraints of fieldwork and have a reasonably small sample size, since the process of collecting and measuring each sample is very time consuming.

It was realized that, by the nature of the SCN, there were regions of high intensity and areas of low intensity, with the latter having an increased occurrence after crop rotation. Moreover, the data contained many zeros, some coming from low intensity areas and others from the result of pathogen free areas. Therefore, a zero-inflated model will be necessary to properly analyze the data. In addition, the data provided presented over-dispersion which requires the use of a Negative Binomial model. It is still possible that future data sets with larger sample sizes or different sampling methods will allow the Poisson model to be used, but it seems that the Negative Binomial model will likely be required from the very high dispersion parameter observed in the data. Regardless of the type of model

used, zero-inflation will still need to be considered in combination with a Mixed Effects model to account for the sampling locations. Possible solutions to this include a Bayesian model, which could incorporate the false negatives model defined by Dr. Reyes into its larger structure, or using the NLMIXED procedure available in SAS software.[4] However, models to estimate the population distribution of the parasite in infected fields could not be completed within time constraints.

Then, after an initial simulation study, it was determined that a sample size of 10 was simply not sufficient to accurately estimate the population distribution of a large field. Larger sample sizes would be needed, and this may be accomplished with similar cost to the previous study completed by Dr. Perez-Hernandez due to the lack of need to take three counts for every sample. In total, he performed 30 counts: three per each of 10 samples. The simulation study showed reasonable results with 36 samples as well as 25. But, of course, more samples is always better. Thus, it was recommended to utilize a sample size as large as possible.

It was also determined, by the simulation study, that the fixed grid method performs the best across all sample sizes and all types of fields. Throughout the study, it consistently produced better or similar results for all three criteria compared to the other methods that were tested. There was not a single scenario where another method outperformed the fixed grid method in the study. Unfortunately, a strictly fixed grid method may not always be feasible due to the limitations of fieldwork. However, since the stratified grid method also performed very well, compared to the other methods tested, an 'almost' fixed grid method was considered sufficient. This method would attempt to maintain a fixed grid but would allow tolerances when the exact location of the samples is not available for testing.

Due to the lack of knowledge regarding the small scale spatial correlation of SCN in the field, it may be prudent to perform smaller real-world studies, perhaps utilizing fixed grid methods on small sections of fields, to thoroughly investigate these properties of the

parasite. This is especially important for the simulation study that was performed since the spatial correlation model used for estimation took advantage of a known covariance model and known parameters. Also, all of the methods performed similarly for the patchy fields. In personal correspondence with Dr. Perez-Hernandez, he implied that it was believed that SCN follows a distribution similar to a patchy field. If this is the case, then the sub-grid method would be viable and also serve to provide additional information about the small scale spatial correlation of the parasite.

Therefore the final recommendation, that will be given to Dr. Perez-Hernandez for another study, will be to use an 'almost' fixed grid method that encompasses the entire field and to use as large a sample size as is feasible, greater than 25 samples, while only taking one count per sample unless the count is zero, in which case a second count should be taken. Furthermore, more studies will need to be conducted to better understand the nature of SCN in real fields at the local level, so that small scale spatial correlation models can be determined with confidence. It is also recommended to either abandon the composite sample method or take multiple sub-samples from them in order to measure the variation at this level.

Bibliography

[1] O. Perez-Hernandez and L. J. Giesler. *Multifactorial Analysis of Mortality of Soybean Cyst Nematode, Heterodera glycines Ichinohe, Populations in Soybean and in Soybean Fields Annually Rotated to Corn in Nebraska*. PhD dissertation, University of Nebraska-Lincoln, Lincoln, NE, 2013.

[2] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL https://www.R-project.org/.

[3] P. J. Diggle and P. J. Ribeiro Jr. *Model Based Geostatistics*. New York: Springer, 2007.

[4] W. W. Stroup. *Generalized Linear Mixed Models: Modern Concepts, Methods, and Applications*. Boca Raton; London; New York: Taylor & Francis Group, 2013.

[5] Paulo J. Ribeiro Jr and Peter J. Diggle. *geoR: Analysis of Geostatistical Data*, 2015. URL https://CRAN.R-project.org/package=geoR. R package version 1.7-5.1.

[6] O.F. Christensen and P.J. Ribeiro Jr. georglm - a package for generalised linear spatial models. *R-NEWS*, 2(2):26–28, 2002. URL http://cran.R-project.org/doc/Rnews. ISSN 1609-3631.

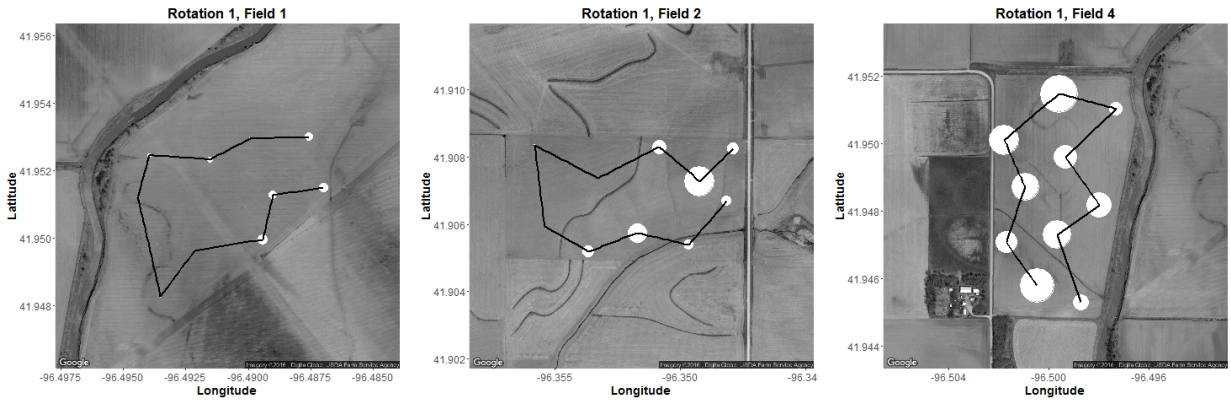
[7] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. ISBN 978-0-387-98140-6. URL http://ggplot2.org.

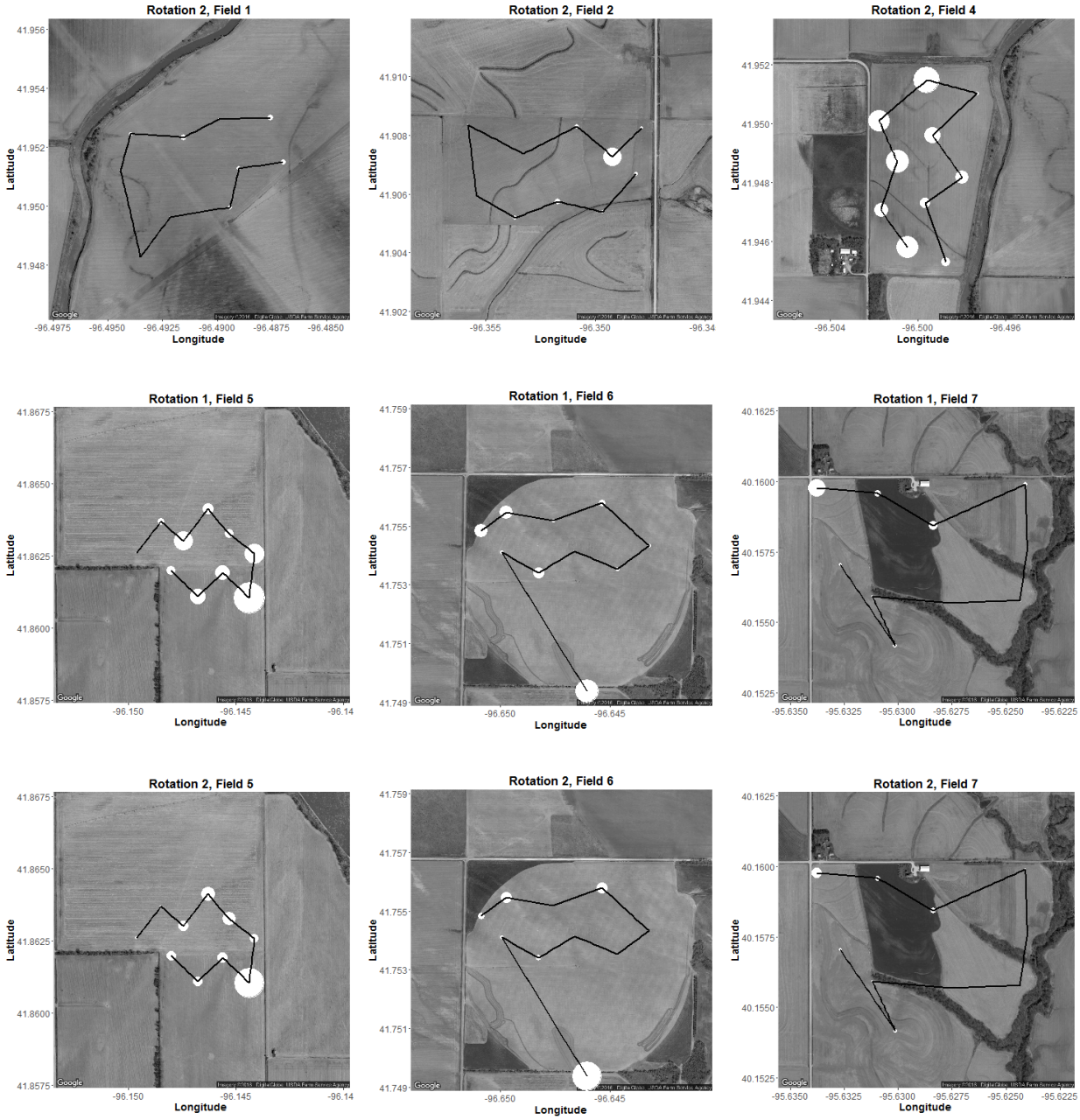
[8] David Kahle and Hadley Wickham. ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161, 2013. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

Appendix A

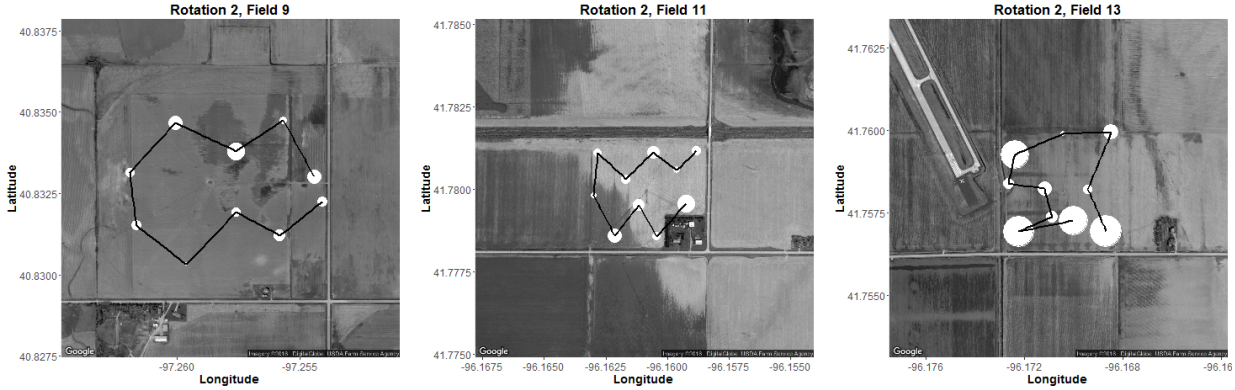
Data and Sampling Location Maps

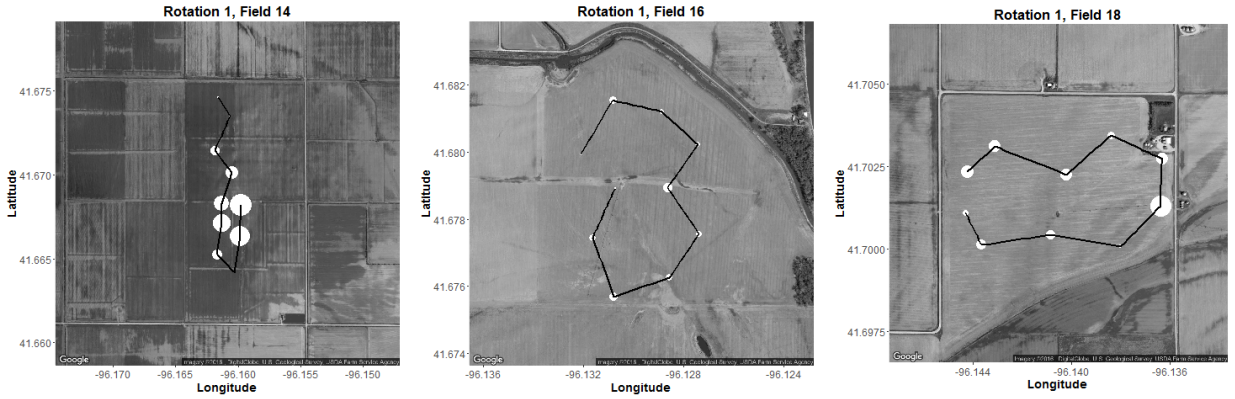
Sattelite Maps, from *Google Maps*^{TM}, of the fields that were sampled. The lines show the path that was taken through the field and the size of the circles represent the average of the three counts taken at that location, on a square root scale. 'Rotation 1' refers to before crop rotation and 'Rotation 2' refers to after crop rotation. The plots were made using the ggplot2[7] and ggmap[8] packages with R software.[2]

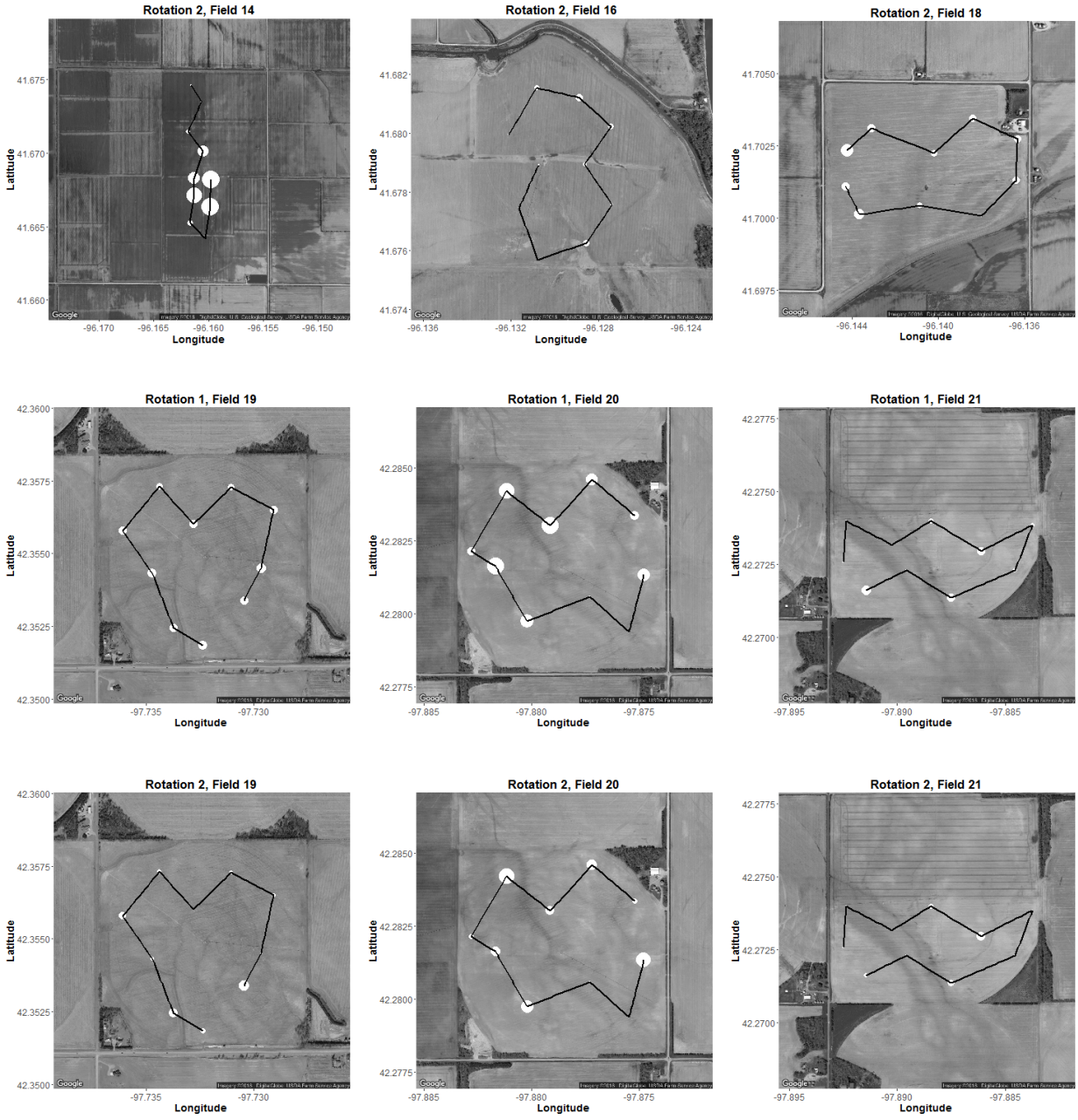


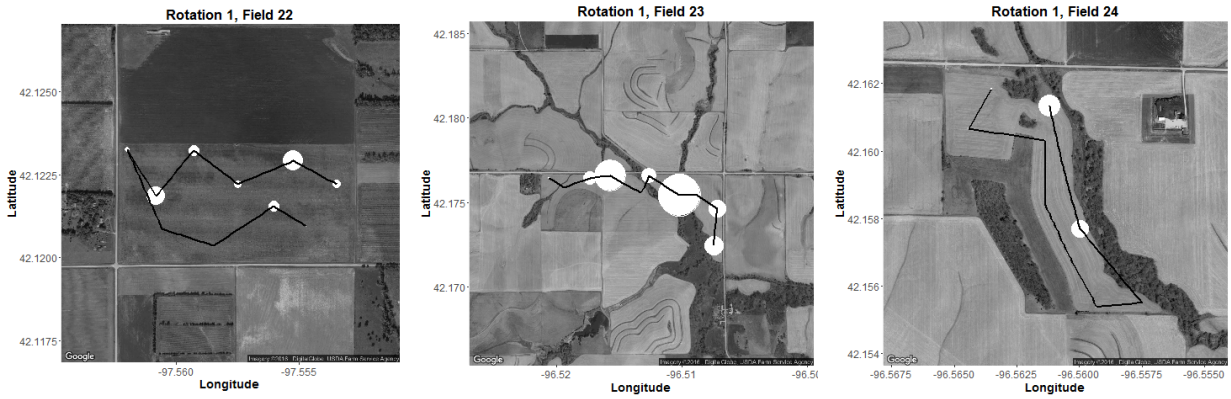

Rotation 2, Field 1
Latitude
Longitude
41.956
41.954
41.952
41.950
41.948
-96.4975
-96.4950
-96.4925
-96.4900
-96.4875
-96.4850
Rotation 2, Field 2
Latitude
Longitude
41.910
41.908
41.906
41.904
41.902
-96.355
-96.350
-96.34
Rotation 2, Field 4
Latitude
Longitude
41.952
41.950
41.948
41.946
41.944
-96.504
-96.500
-96.496
Rotation 1, Field 5
Latitude
Longitude
41.8675
41.8650
41.8625
41.8600
41.8575
-96.150
-96.145
-96.14
Rotation 1, Field 6
Latitude
Longitude
41.759
41.757
41.755
41.753
41.751
41.749
-96.650
-96.645
Rotation 1, Field 7
Latitude
Longitude
40.1625
40.1600
40.1575
40.1550
40.1525
-95.6350
-95.6325
-95.6300
-95.6275
-95.6250
-95.6225
Rotation 2, Field 5
Latitude
Longitude
41.8675
41.8650
41.8625
41.8600
41.8575
-96.150
-96.145
-96.14
Rotation 2, Field 6
Latitude
Longitude
41.759
41.757
41.755
41.753
41.751
41.749
-96.650
-96.645
Rotation 2, Field 7
Latitude
Longitude
40.1625
40.1600
40.1575
40.1550
40.1525
-95.6350
-95.6325
-95.6300
-95.6275
-95.6250
-95.6225

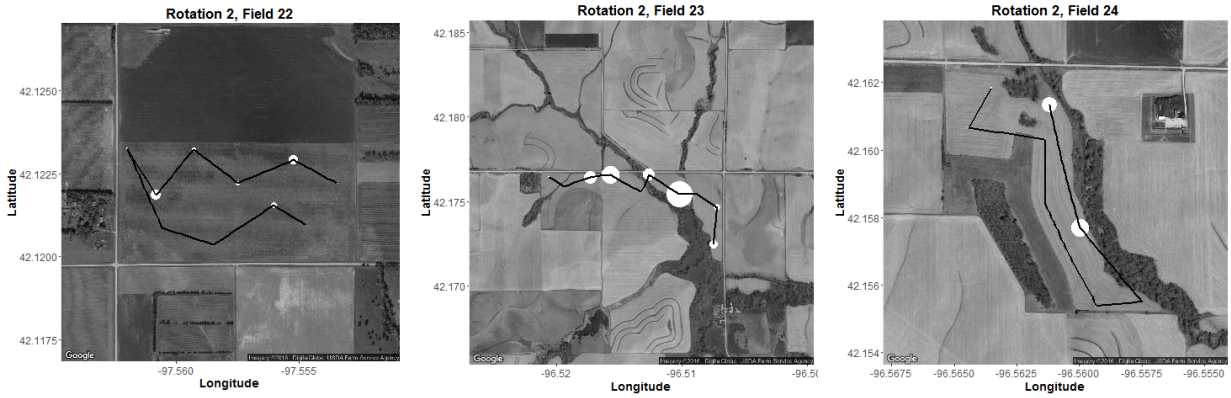

Rotation 1, Field 9
Rotation 1, Field 11
Rotation 1, Field 13
Latitude
Longitude
Google

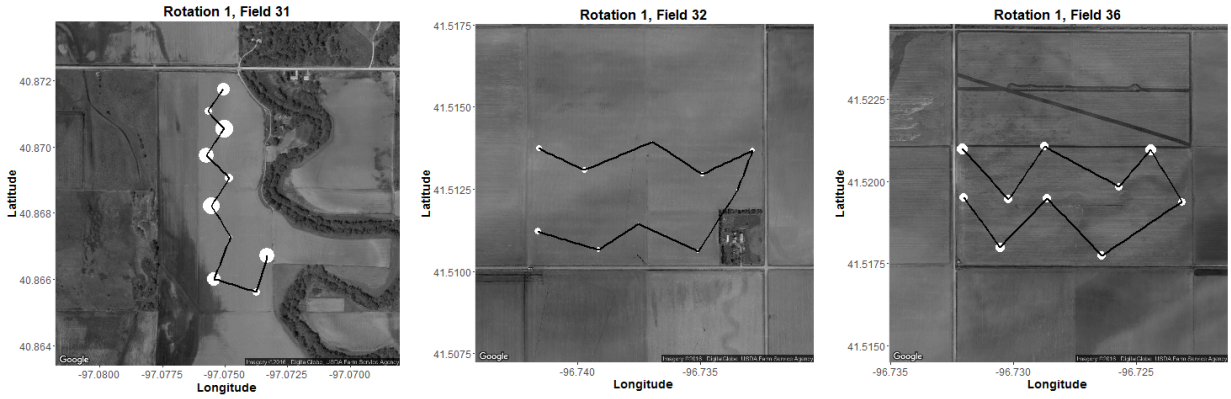

Rotation 2, Field 9
Rotation 2, Field 11
Rotation 2, Field 13
Latitude
Longitude
Google

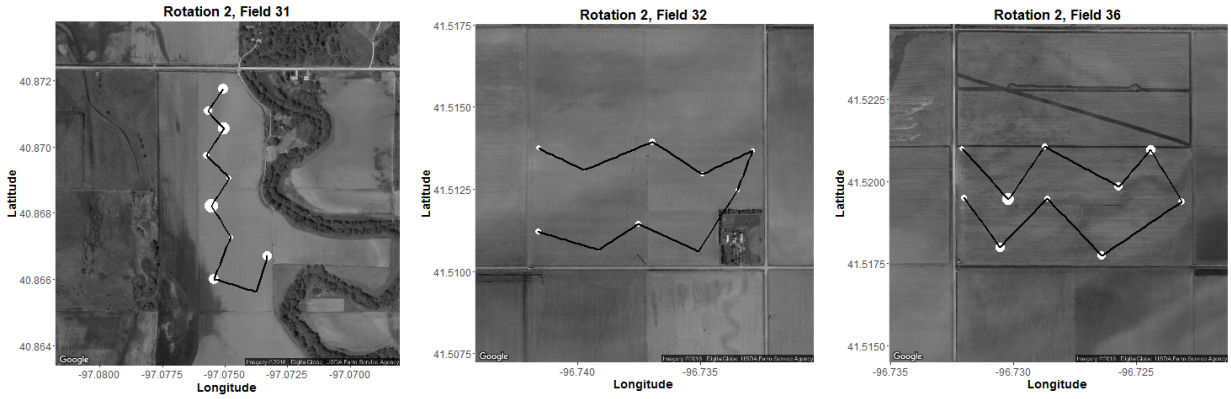

Rotation 1, Field 14
Rotation 1, Field 16
Rotation 1, Field 18
Latitude
Longitude
Google

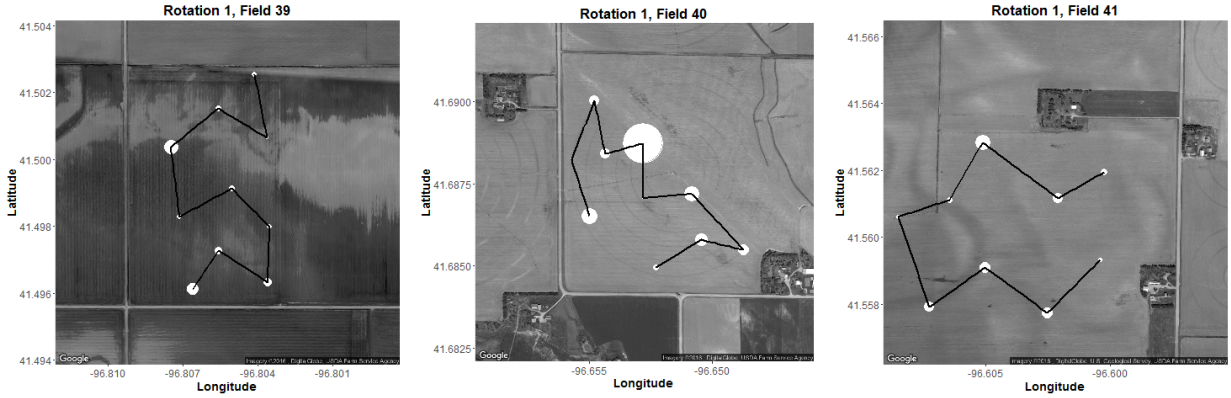

Rotation 2, Field 14
Latitude
41.675
41.670
41.665
41.660
-96.170
-96.165
-96.160
-96.155
-96.150
Longitude
Rotation 2, Field 16
Latitude
41.682
41.680
41.678
41.676
41.674
-96.136
-96.132
-96.128
-96.124
Longitude
Rotation 2, Field 18
Latitude
41.7050
41.7025
41.7000
41.6975
-96.144
-96.140
-96.136
Longitude
Rotation 1, Field 19
Latitude
42.3600
42.3575
42.3550
42.3525
42.3500
-97.735
-97.730
Longitude
Rotation 1, Field 20
Latitude
42.2850
42.2825
42.2800
42.2775
-97.885
-97.880
-97.875
Longitude
Rotation 1, Field 21
Latitude
42.2775
42.2750
42.2725
42.2700
-97.895
-97.890
-97.885
Longitude
Rotation 2, Field 19
Latitude
42.3600
42.3575
42.3550
42.3525
42.3500
-97.735
-97.730
Longitude
Rotation 2, Field 20
Latitude
42.2850
42.2825
42.2800
42.2775
-97.885
-97.880
-97.875
Longitude
Rotation 2, Field 21
Latitude
42.2775
42.2750
42.2725
42.2700
-97.895
-97.890
-97.885
Longitude

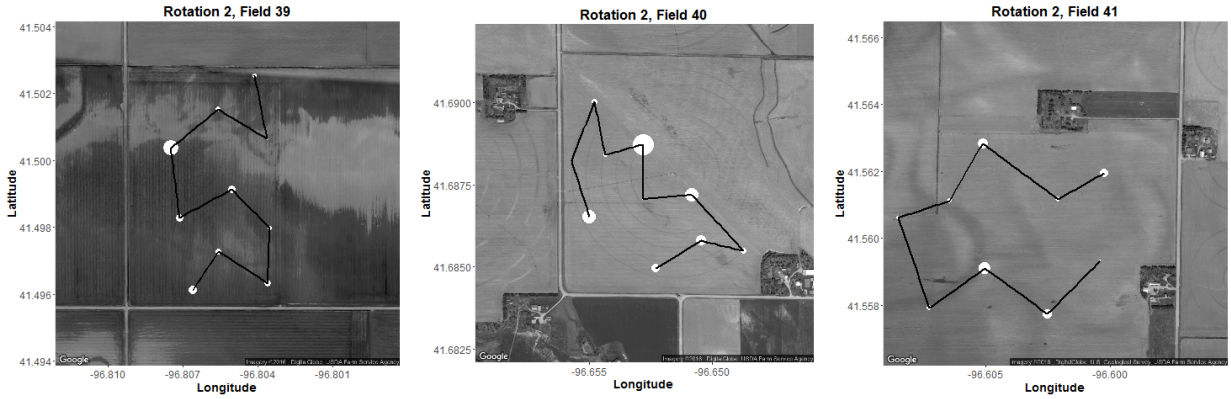

Rotation 1, Field 22
Latitude
42.1250
42.1225
42.1200
42.1175
-97.560
-97.555
Longitude
Rotation 1, Field 23
Latitude
42.185
42.180
42.175
42.170
-96.52
-96.51
Longitude
Rotation 1, Field 24
Latitude
42.162
42.160
42.158
42.156
42.154
-96.5675
-96.5650
-96.5625
-96.5600
-96.5575
-96.5550
Longitude

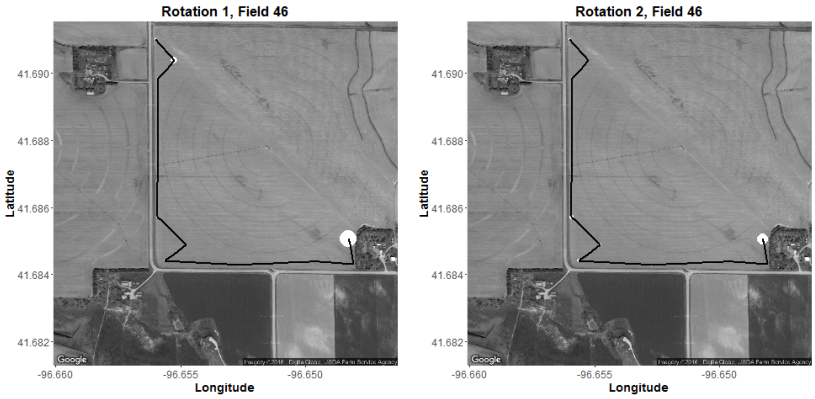

Rotation 2, Field 22
Latitude
42.1250
42.1225
42.1200
42.1175
-97.560
-97.555
Longitude
Rotation 2, Field 23
Latitude
42.185
42.180
42.175
42.170
-96.52
-96.51
Longitude
Rotation 2, Field 24
Latitude
42.162
42.160
42.158
42.156
42.154
-96.5675
-96.5650
-96.5625
-96.5600
-96.5575
-96.5550
Longitude


Rotation 1, Field 31
Latitude
40.872
40.870
40.868
40.866
40.864
-97.0800
-97.0775
-97.0750
-97.0725
-97.0700
Longitude
Rotation 1, Field 32
Latitude
41.5175
41.5150
41.5125
41.5100
41.5075
-96.740
-96.735
Longitude
Rotation 1, Field 36
Latitude
41.5225
41.5200
41.5175
41.5150
-96.735
-96.730
-96.725
Longitude


Rotation 2, Field 31
Latitude
40.872
40.870
40.868
40.866
40.864
-97.0800
-97.0775
-97.0750
-97.0725
-97.0700
Longitude
Rotation 2, Field 32
Latitude
41.5175
41.5150
41.5125
41.5100
41.5075
-96.740
-96.735
Longitude
Rotation 2, Field 36
Latitude
41.5225
41.5200
41.5175
41.5150
-96.735
-96.730
-96.725
Longitude


Rotation 1, Field 39
Latitude
41.504
41.502
41.500
41.498
41.496
41.494
-96.810
-96.807
-96.804
-96.801
Longitude
Rotation 1, Field 40
Latitude
41.6900
41.6875
41.6850
41.6825
-96.655
-96.650
Longitude
Rotation 1, Field 41
Latitude
41.566
41.564
41.562
41.560
41.558
-96.605
-96.600
Longitude

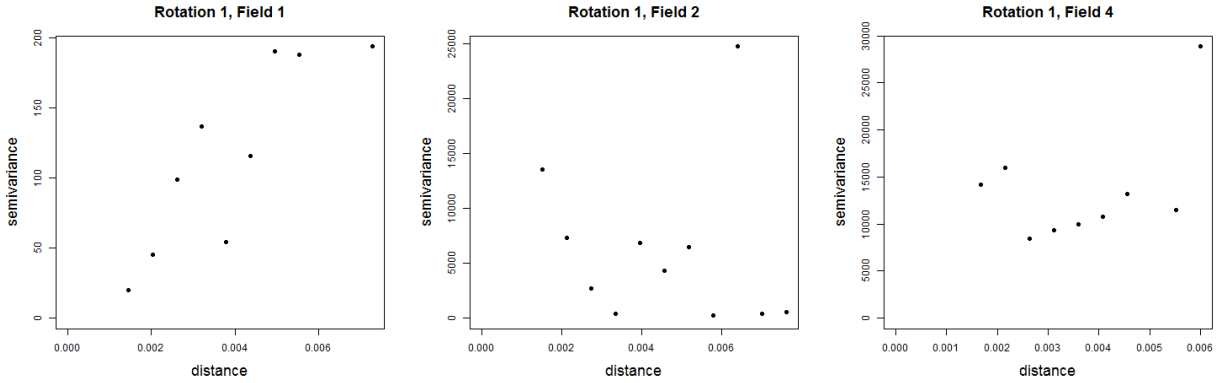

Rotation 2, Field 39
Latitude
41.504
41.502
41.500
41.498
41.496
41.494
-96.810
-96.807
-96.804
-96.801
Longitude
Rotation 2, Field 40
Latitude
41.6900
41.6875
41.6850
41.6825
-96.655
-96.650
Longitude
Rotation 2, Field 41
Latitude
41.566
41.564
41.562
41.560
41.558
-96.605
-96.600
Longitude

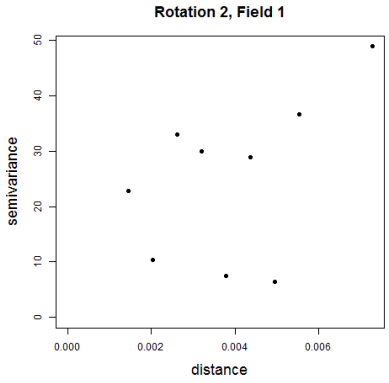

Rotation 1, Field 46
Rotation 2, Field 46
Latitude
Longitude
41.690
41.688
41.686
41.684
41.682
-96.660
-96.655
-96.650
Google

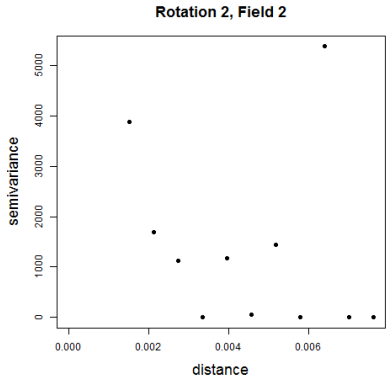
Appendix B

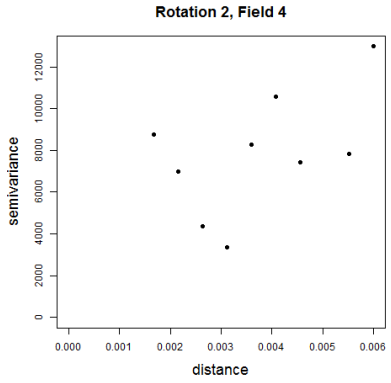
Variograms

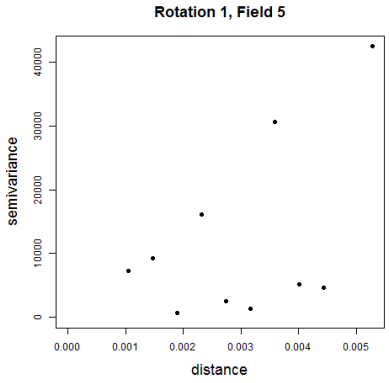
Sample variograms calculated for the fields that were sampled. The field numbers correspond to the fields shown in Appendix A. 'Rotation 1' refers to before crop rotation and 'Rotation 2' refers to after crop rotation. These were produced using the geoR[5] package with R software.[2]

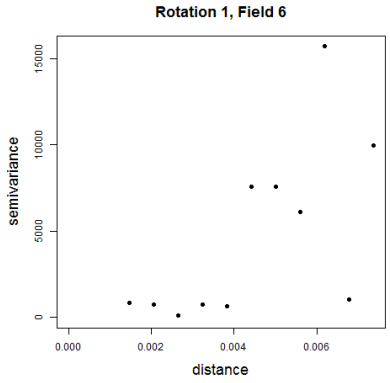


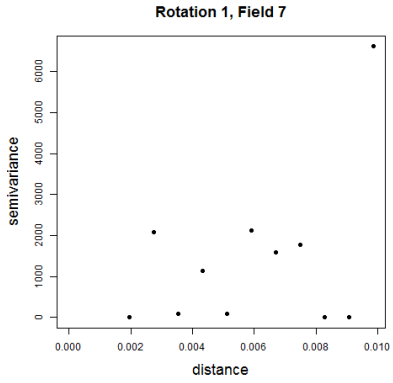

Rotation 2, Field 1
semivariance
distance

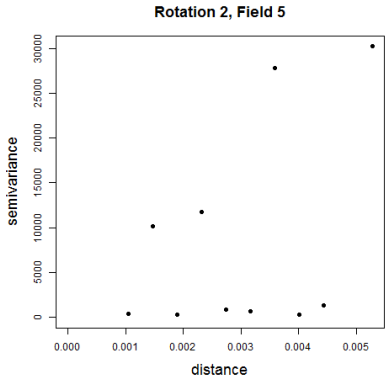

Rotation 2, Field 2
semivariance
distance

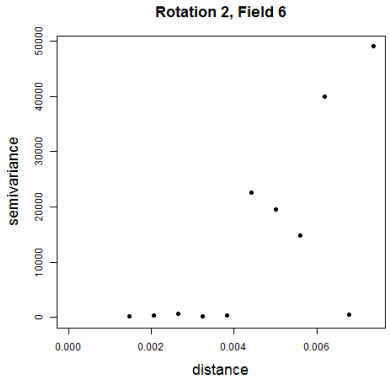

Rotation 2, Field 4
semivariance
distance

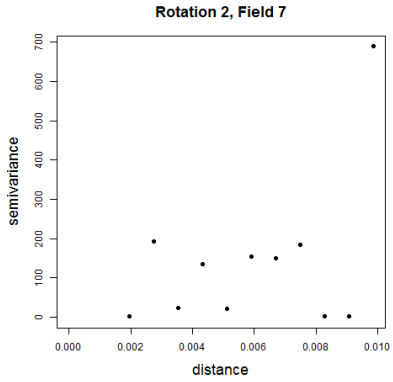

Rotation 1, Field 5
semivariance
distance

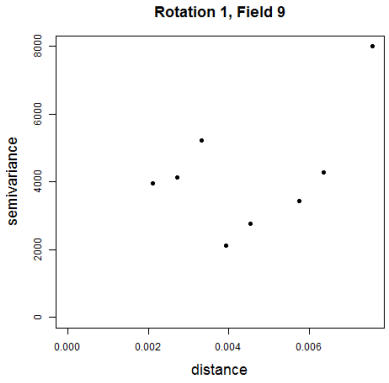

Rotation 1, Field 6
semivariance
distance

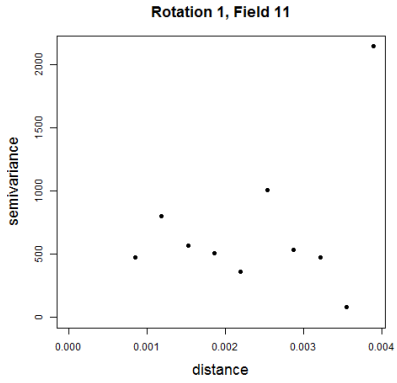

Rotation 1, Field 7
semivariance
distance

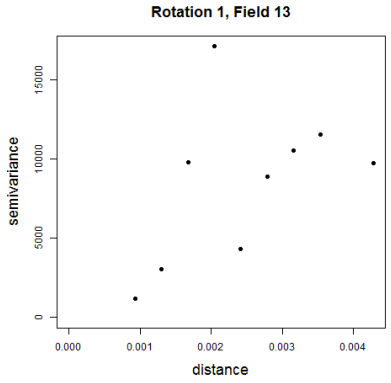

Rotation 2, Field 5
semivariance
distance

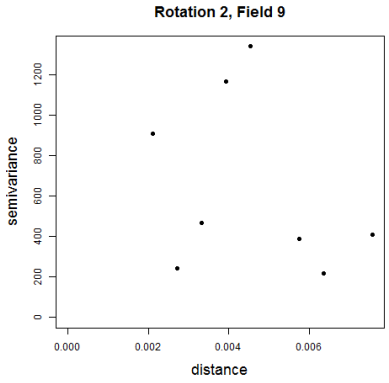

Rotation 2, Field 6
semivariance
distance

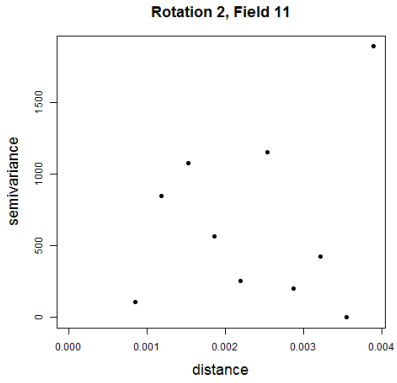

Rotation 2, Field 7
semivariance
distance

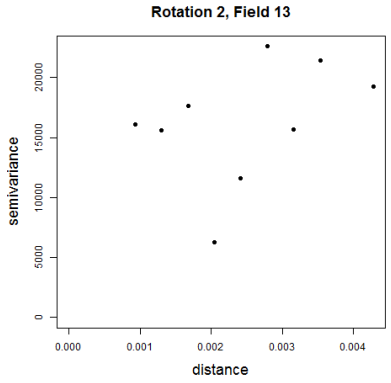

Rotation 1, Field 9
semivariance
distance

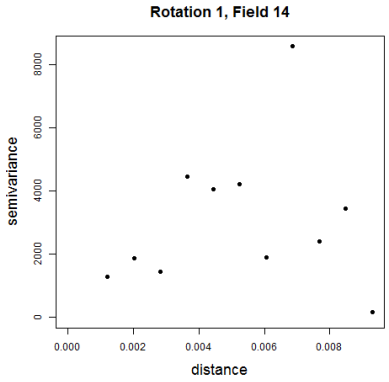

Rotation 1, Field 11
semivariance
distance

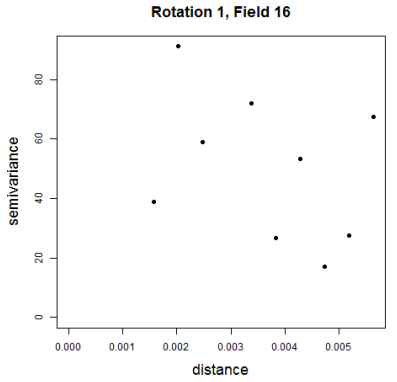

Rotation 1, Field 13
semivariance
distance

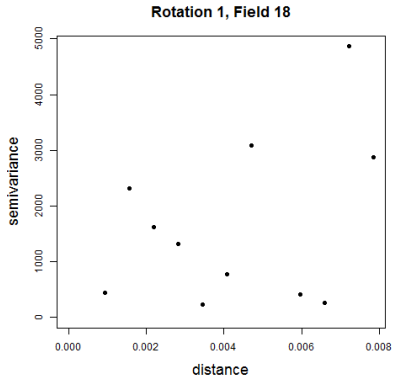

Rotation 2, Field 9
semivariance
distance

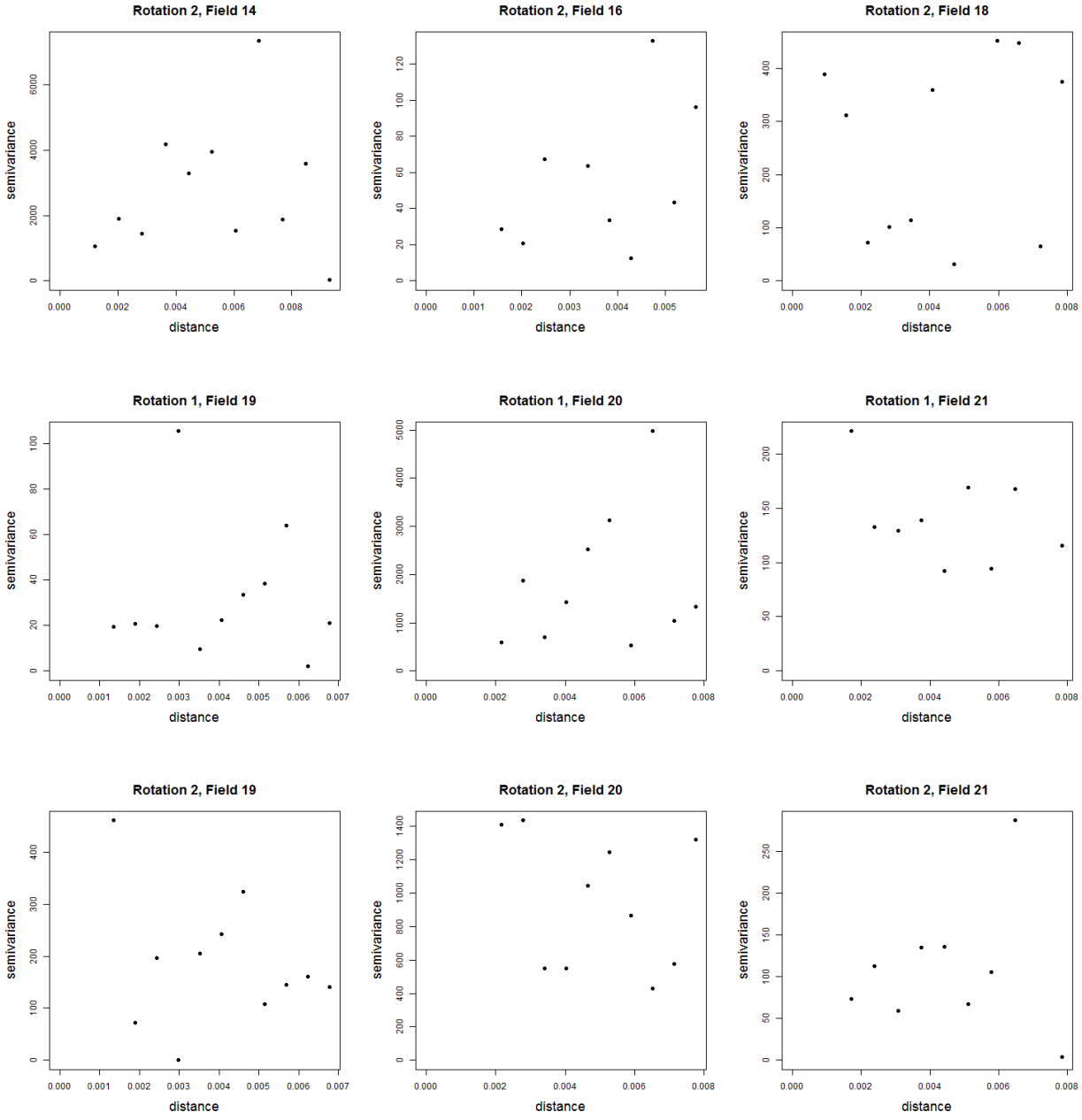

Rotation 2, Field 11
semivariance
distance

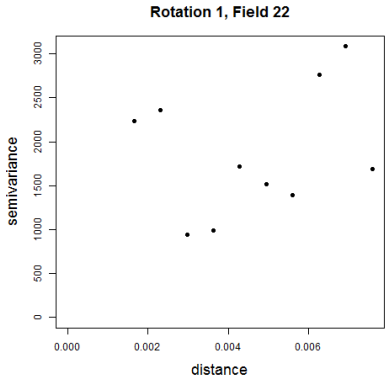

Rotation 2, Field 13
semivariance
distance

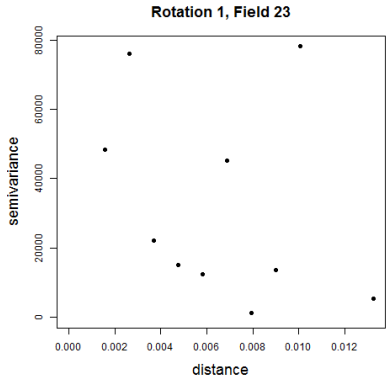

Rotation 1, Field 14
semivariance
distance

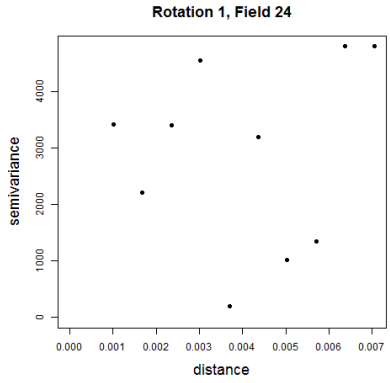

Rotation 1, Field 16
semivariance
distance

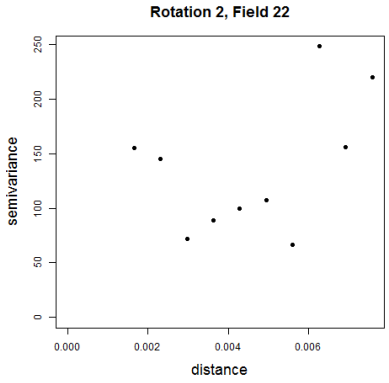

Rotation 1, Field 18
semivariance
distance

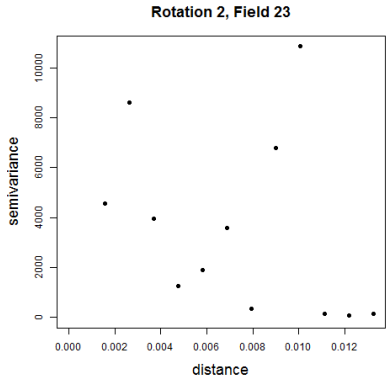

Rotation 2, Field 14
Rotation 2, Field 16
Rotation 2, Field 18
Rotation 1, Field 19
Rotation 1, Field 20
Rotation 1, Field 21
Rotation 2, Field 19
Rotation 2, Field 20
Rotation 2, Field 21
semivariance
distance

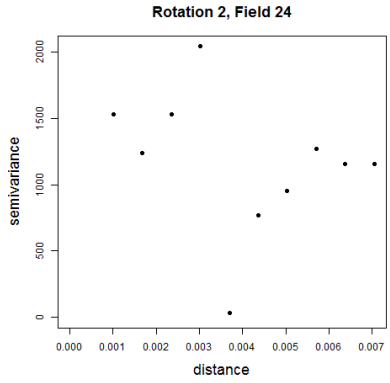

Rotation 1, Field 22
semivariance
distance

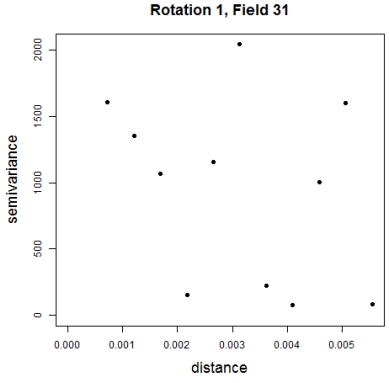

Rotation 1, Field 23
semivariance
distance

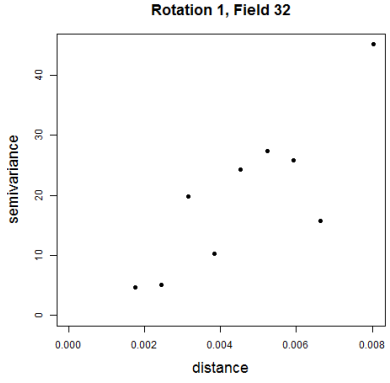

Rotation 1, Field 24
semivariance
distance

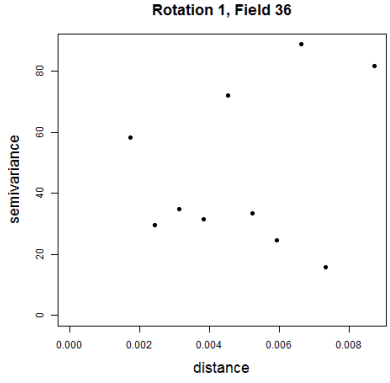

Rotation 2, Field 22
semivariance
distance

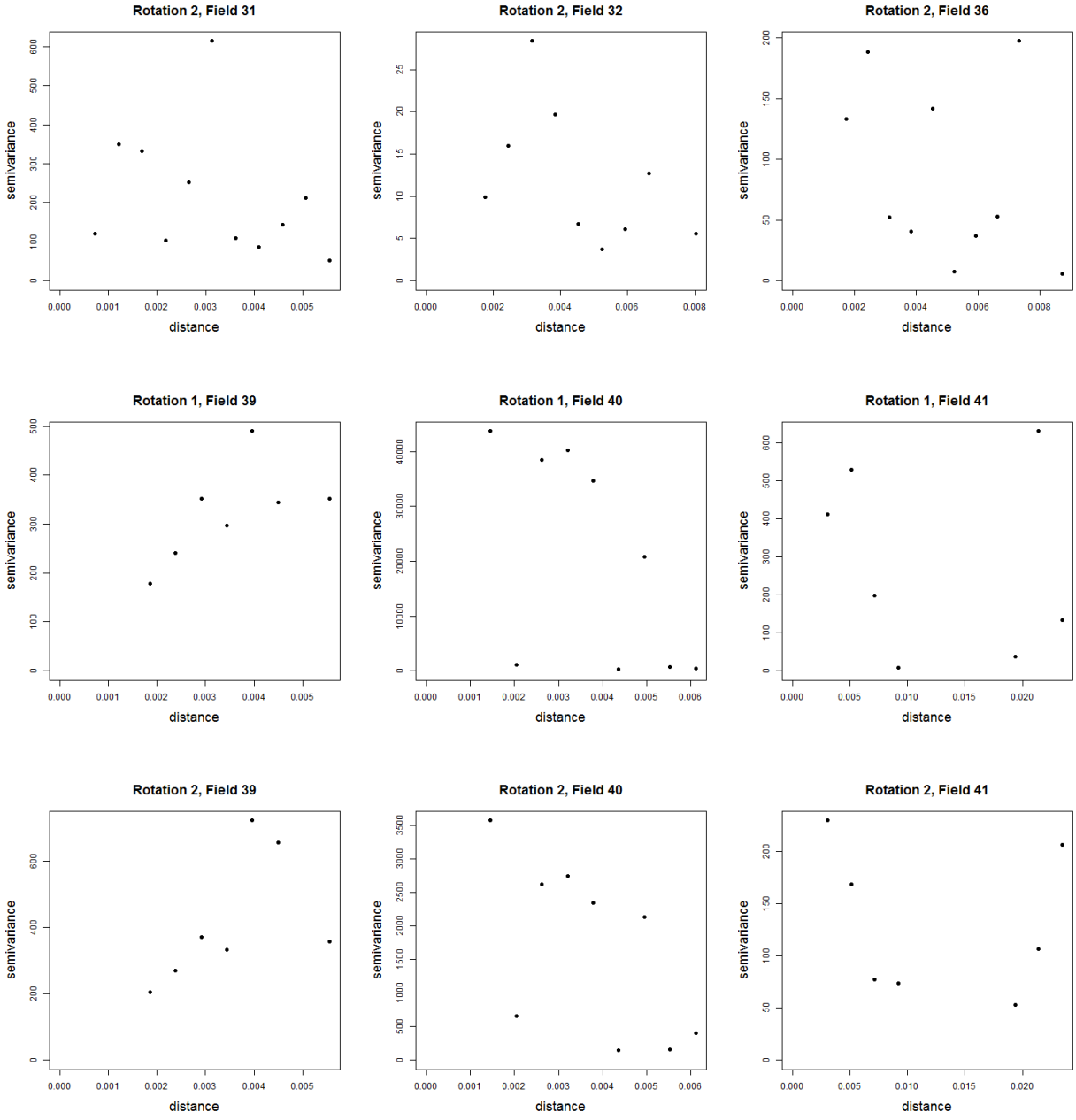

Rotation 2, Field 23
semivariance
distance

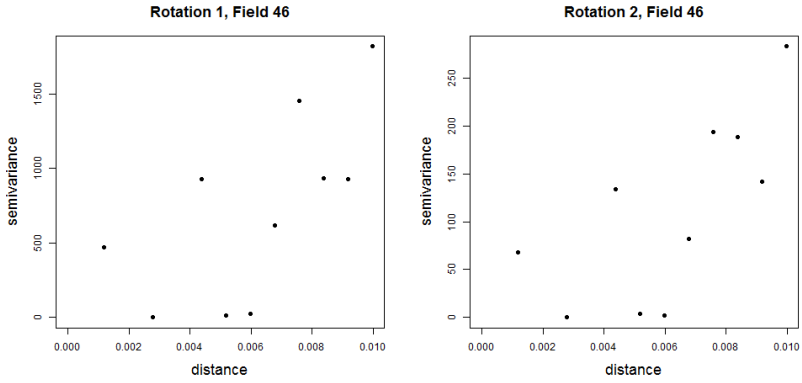

Rotation 2, Field 24
semivariance
distance


Rotation 1, Field 31
semivariance
distance


Rotation 1, Field 32
semivariance
distance


Rotation 1, Field 36
semivariance
distance


Rotation 2, Field 31
Rotation 2, Field 32
Rotation 2, Field 36
Rotation 1, Field 39
Rotation 1, Field 40
Rotation 1, Field 41
Rotation 2, Field 39
Rotation 2, Field 40
Rotation 2, Field 41
semivariance
distance


Rotation 1, Field 46
semivariance
0
500
1000
1500
0.000
0.002
0.004
0.006
0.008
0.010
distance
Rotation 2, Field 46
semivariance
0
50
100
150
200
250
0.000
0.002
0.004
0.006
0.008
0.010
distance

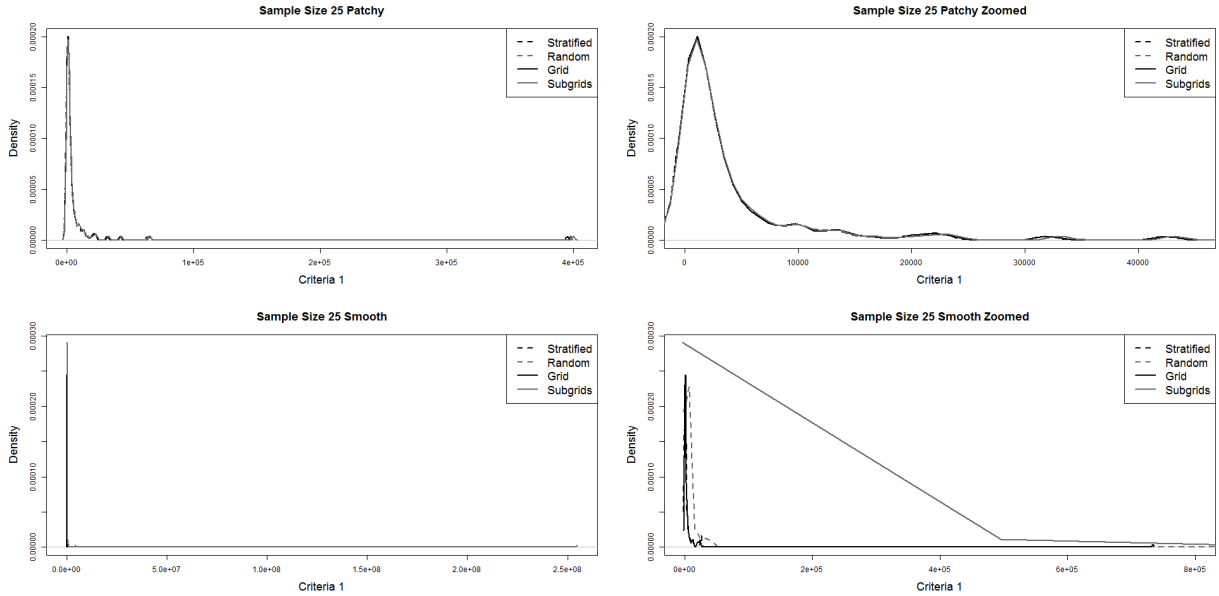
Appendix C

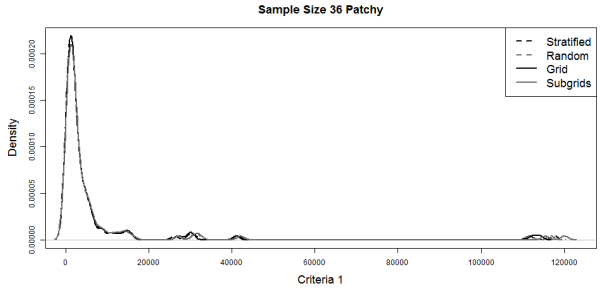
Simulation Result Plots

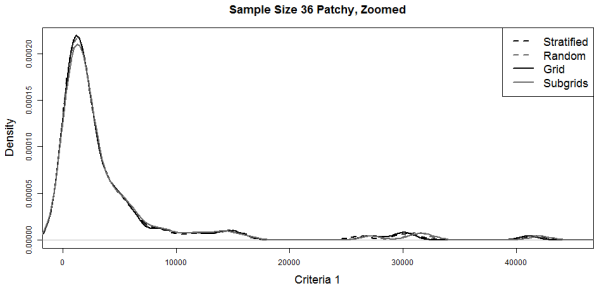
C.1 Criteria 1: Maximum Estimated MSE

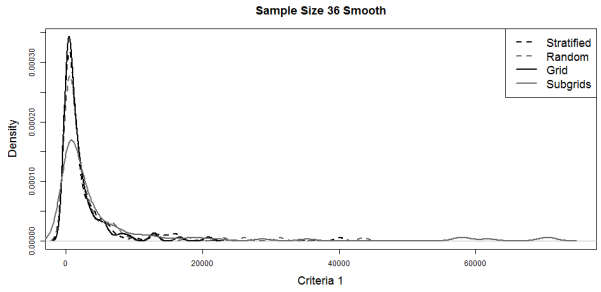
$$\underset{i}{maximum}\ \widehat{MSE}(x_i) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\lambda}_b(x_i) - \lambda(x_i)\right]^2 \tag{C.1}$$

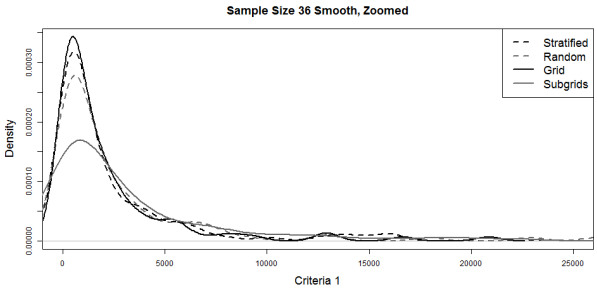
Kernel Density Estimators. Full size, including extreme values, on the left and adjusted scale on the right, to show more detail in the bulk of the distribution.

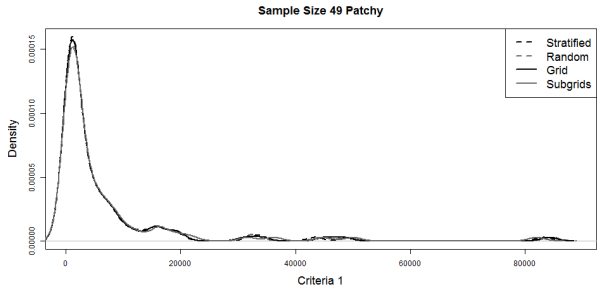


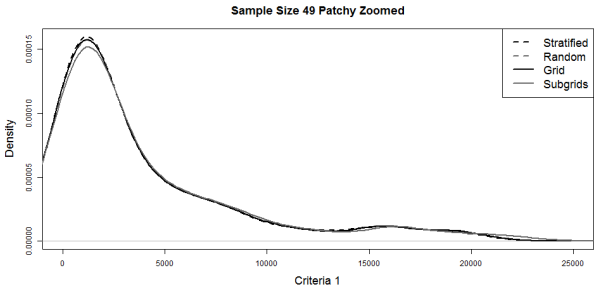

Sample Size 36 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

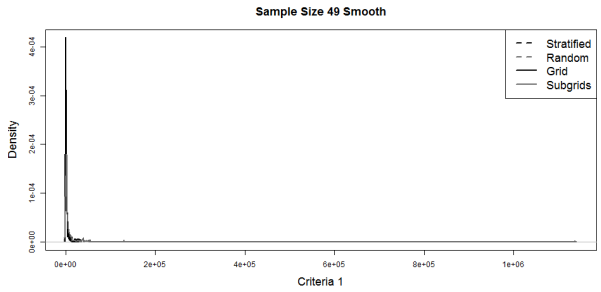

Sample Size 36 Patchy, Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

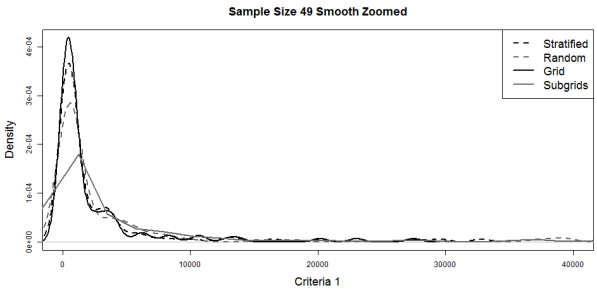

Sample Size 36 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1

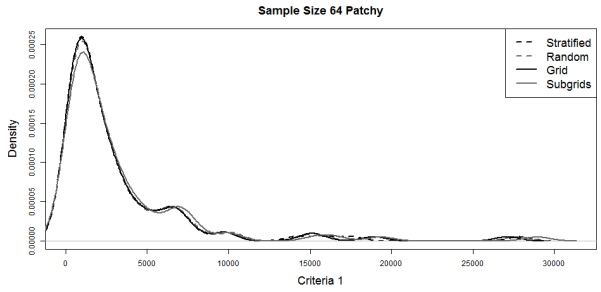

Sample Size 36 Smooth, Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

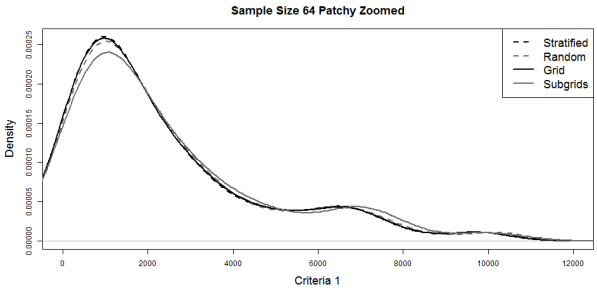

Sample Size 49 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

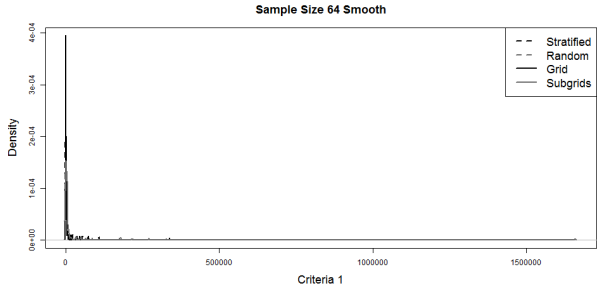

Sample Size 49 Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

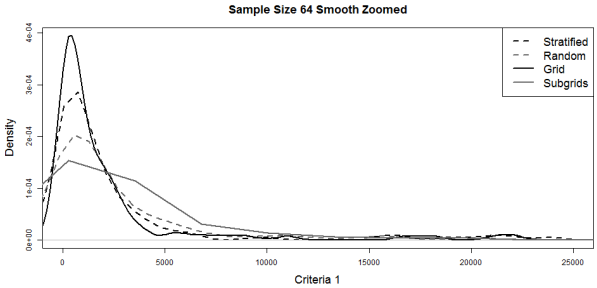

Sample Size 49 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1

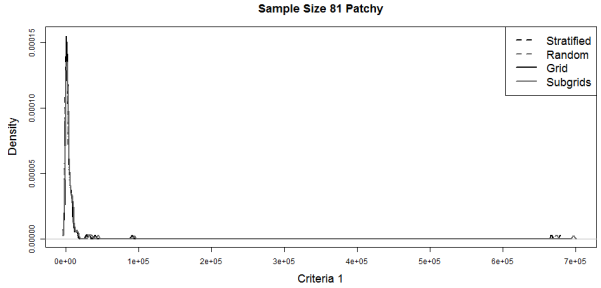

Sample Size 49 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

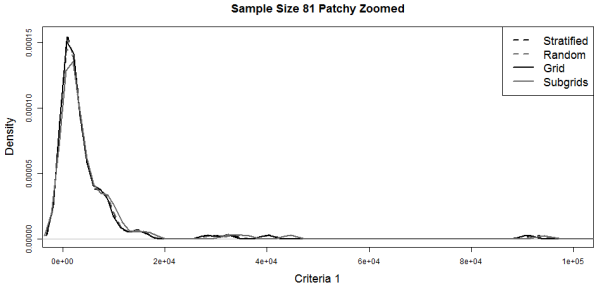

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

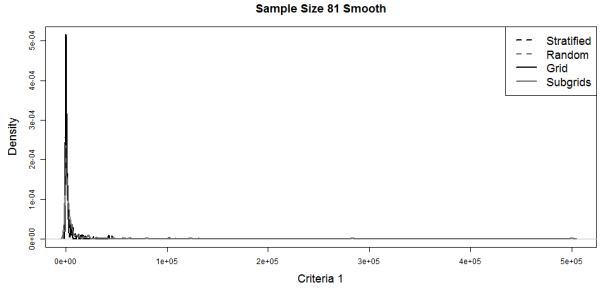

Sample Size 64 Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

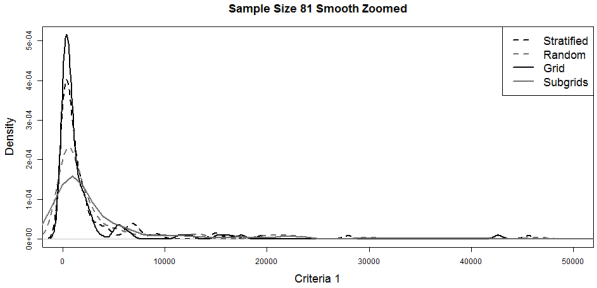

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1

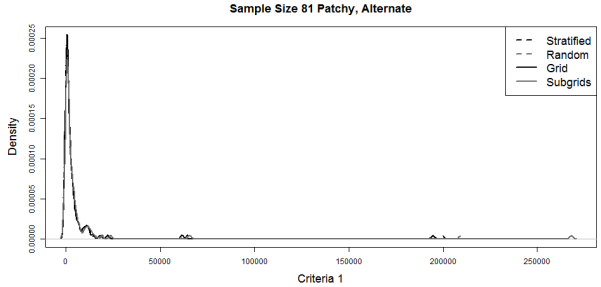

Sample Size 64 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

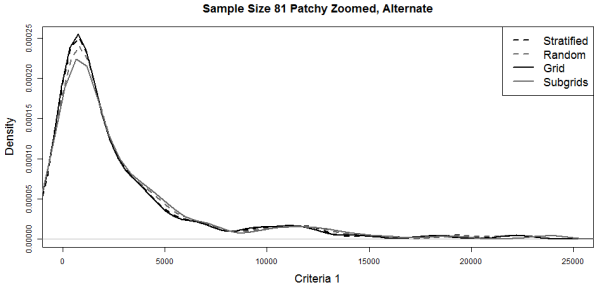

Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

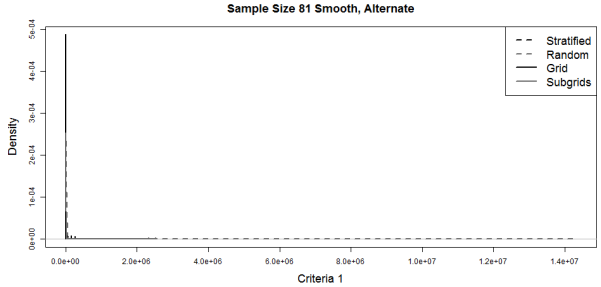

Sample Size 81 Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

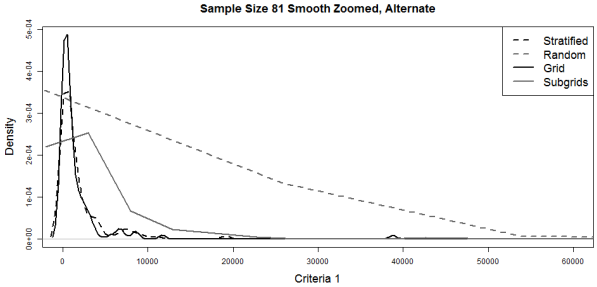

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1


Sample Size 81 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

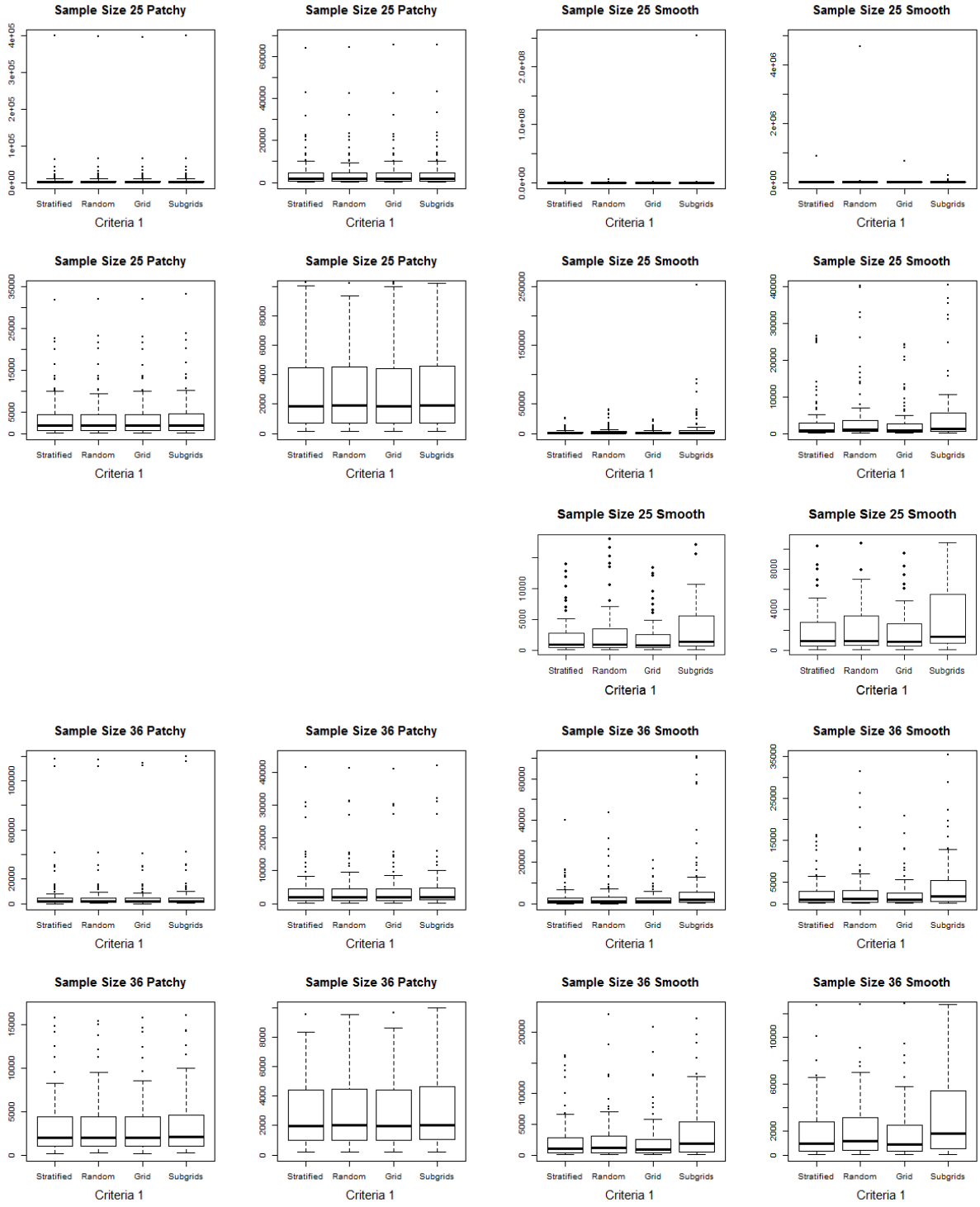

Sample Size 81 Patchy, Alternate
Stratified
Random
Grid
Subgrids
Density
Criteria 1

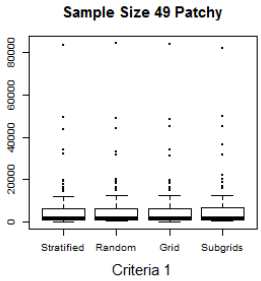

Sample Size 81 Patchy Zoomed, Alternate
Stratified
Random
Grid
Subgrids
Density
Criteria 1

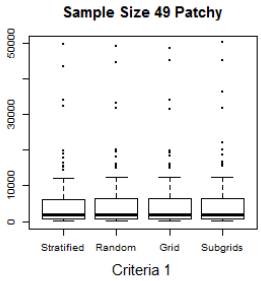

Sample Size 81 Smooth, Alternate
Stratified
Random
Grid
Subgrids
Density
Criteria 1

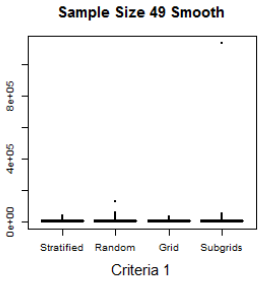

Sample Size 81 Smooth Zoomed, Alternate
Stratified
Random
Grid
Subgrids
Density
Criteria 1

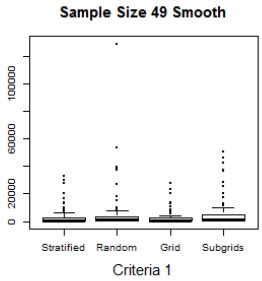
Box-plots. Scales are gradually adjusted to better display the center of the distributions.

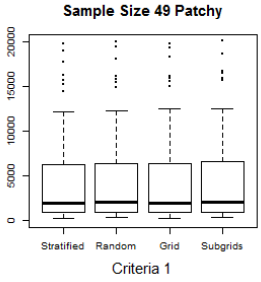


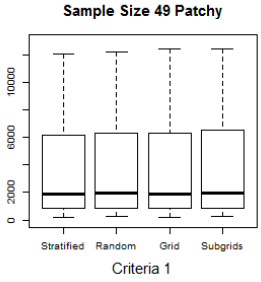

Sample Size 49 Patchy
0
20000
40000
60000
80000
Stratified
Random
Grid
Subgrids
Criteria 1

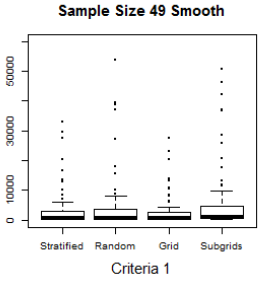

Sample Size 49 Patchy
0
10000
30000
50000
Stratified
Random
Grid
Subgrids
Criteria 1

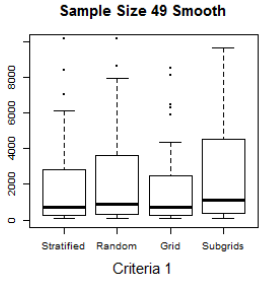

Sample Size 49 Smooth
0e+00
4e+05
8e+05
Stratified
Random
Grid
Subgrids
Criteria 1

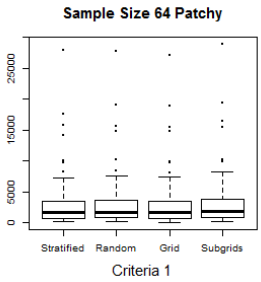

Sample Size 49 Smooth
0
20000
60000
100000
Stratified
Random
Grid
Subgrids
Criteria 1

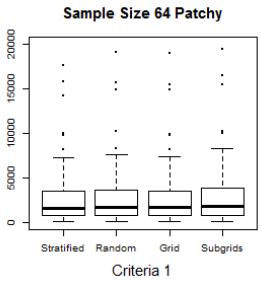

Sample Size 49 Patchy
0
5000
10000
15000
20000
Stratified
Random
Grid
Subgrids
Criteria 1

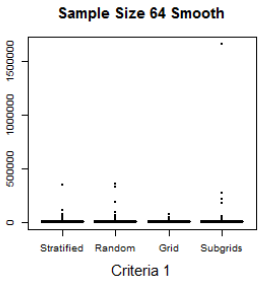

Sample Size 49 Patchy
0
2000
6000
10000
Stratified
Random
Grid
Subgrids
Criteria 1

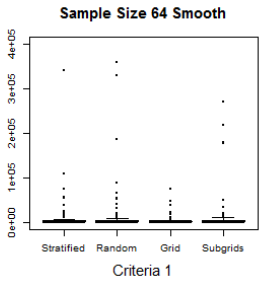

Sample Size 49 Smooth
0
10000
30000
50000
Stratified
Random
Grid
Subgrids
Criteria 1

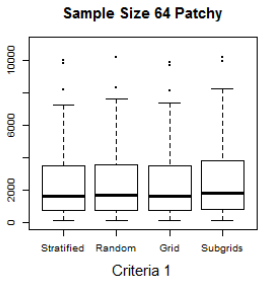

Sample Size 49 Smooth
0
2000
4000
6000
8000
Stratified
Random
Grid
Subgrids
Criteria 1

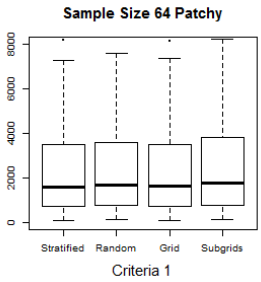

Sample Size 64 Patchy
0
5000
15000
25000
Stratified
Random
Grid
Subgrids
Criteria 1

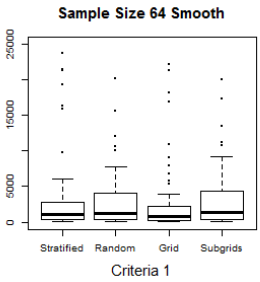

Sample Size 64 Patchy
0
5000
10000
15000
20000
Stratified
Random
Grid
Subgrids
Criteria 1

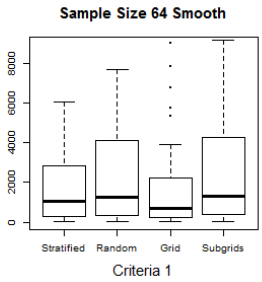

Sample Size 64 Smooth
0
500000
1000000
1500000
Stratified
Random
Grid
Subgrids
Criteria 1

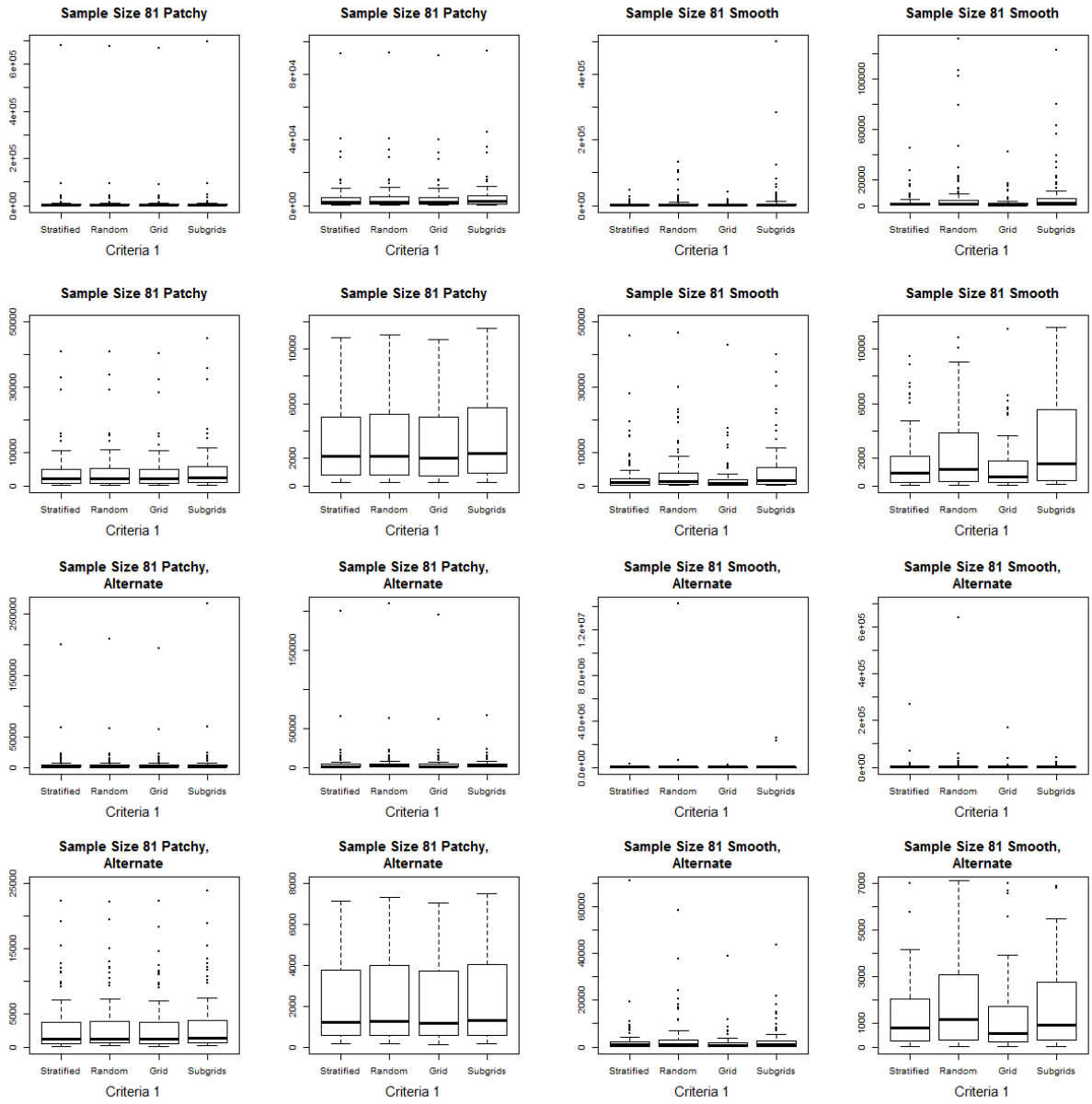

Sample Size 64 Smooth
0e+00
1e+05
2e+05
3e+05
4e+05
Stratified
Random
Grid
Subgrids
Criteria 1


Sample Size 64 Patchy
0
2000
6000
10000
Stratified
Random
Grid
Subgrids
Criteria 1


Sample Size 64 Patchy
0
2000
4000
6000
8000
Stratified
Random
Grid
Subgrids
Criteria 1


Sample Size 64 Smooth
0
5000
15000
25000
Stratified
Random
Grid
Subgrids
Criteria 1

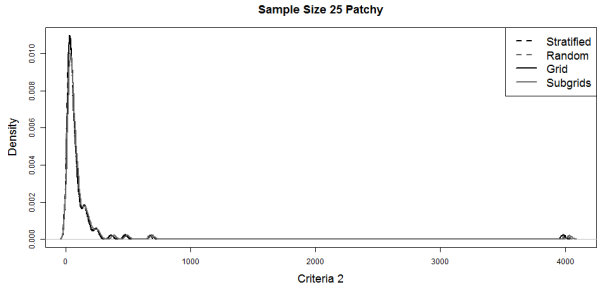

Sample Size 64 Smooth
0
2000
4000
6000
8000
Stratified
Random
Grid
Subgrids
Criteria 1

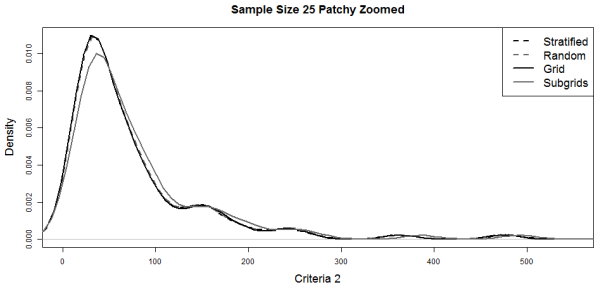


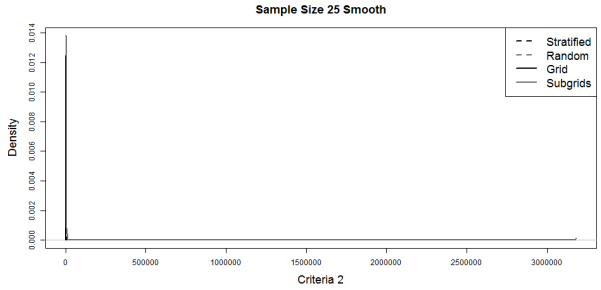
C.2 Criteria 2: Mean Estimated MSE

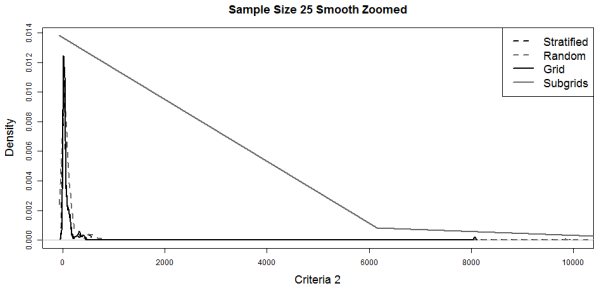
$$\underset{i}{mean}\ \widehat{MSE}(x_i) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\lambda}_b(x_i) - \lambda(x_i)\right]^2 \tag{C.2}$$

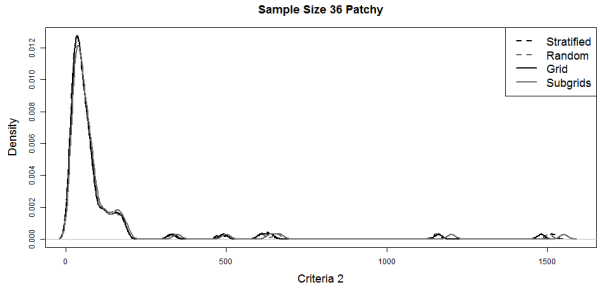
Kernel Density Estimators. Full size, including extreme values, on the left and adjusted scale on the right, to show more detail in the bulk of the distribution.

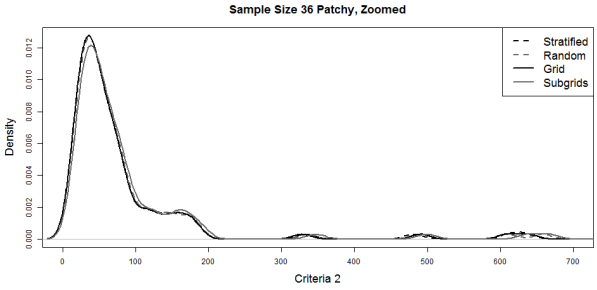


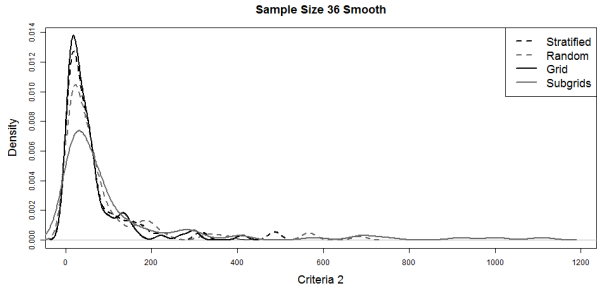


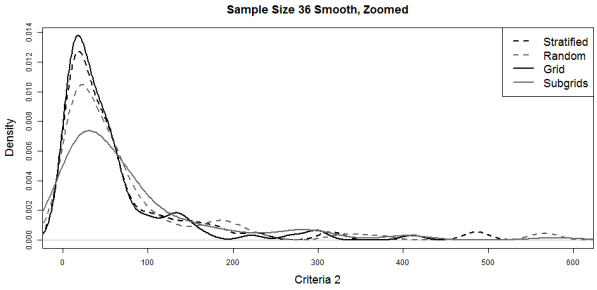


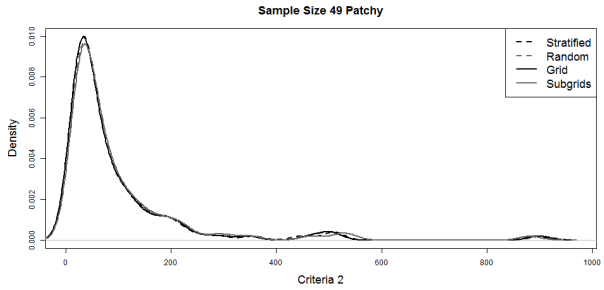


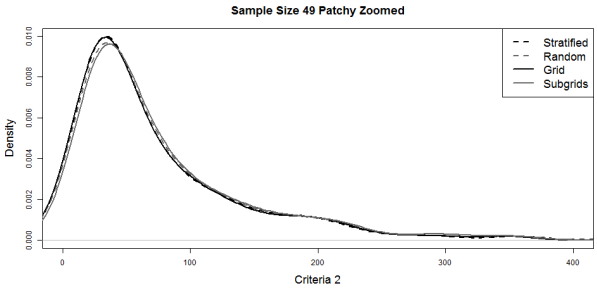


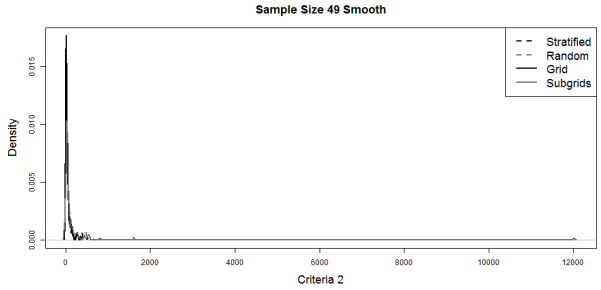


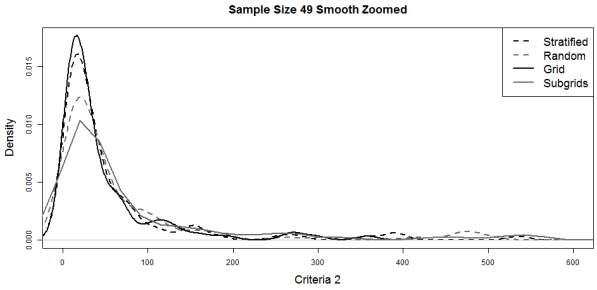


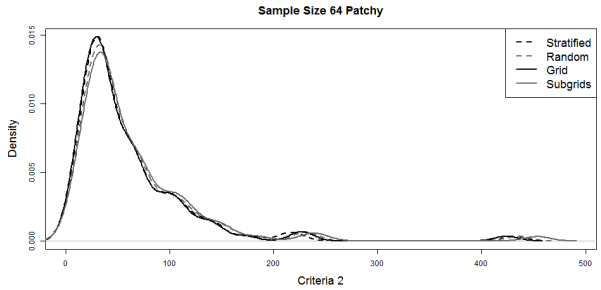


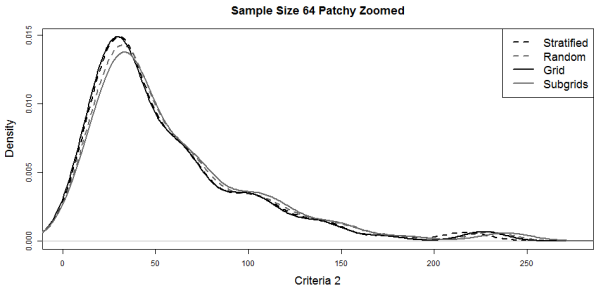


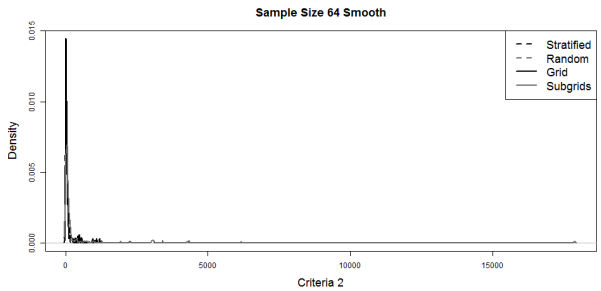


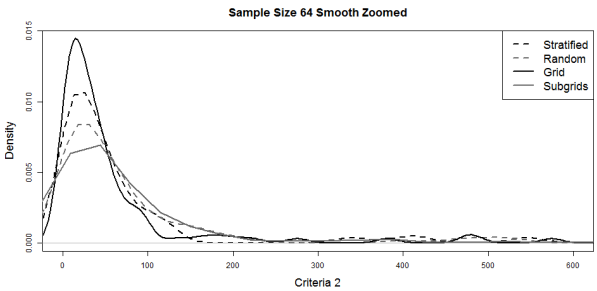

Sample Size 49 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 2

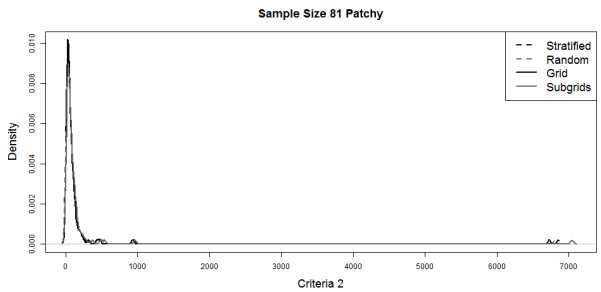

Sample Size 49 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

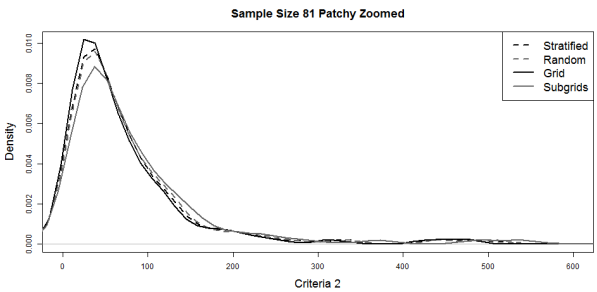

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 2

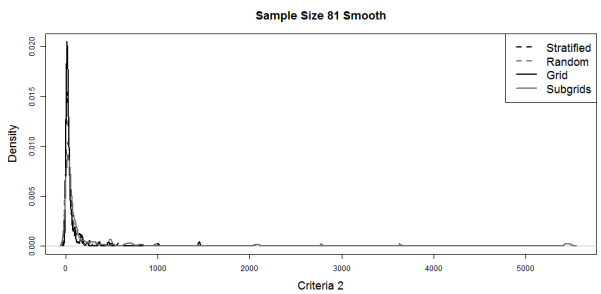

Sample Size 64 Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

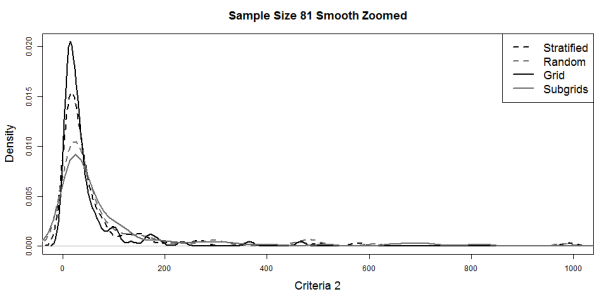

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 2

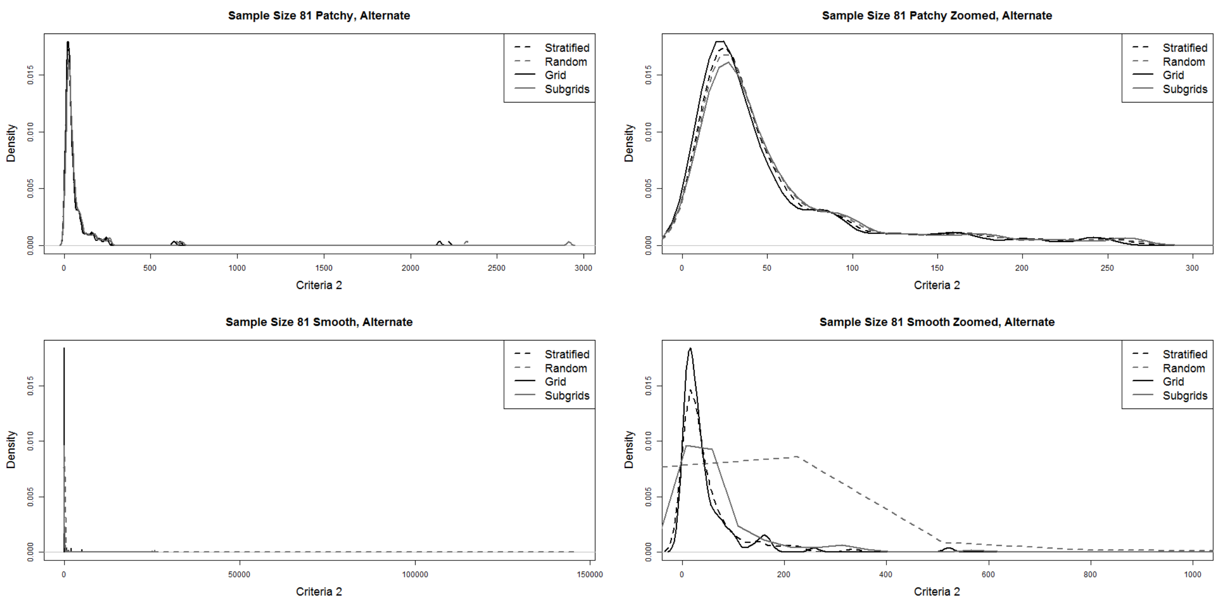

Sample Size 64 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2


Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 2

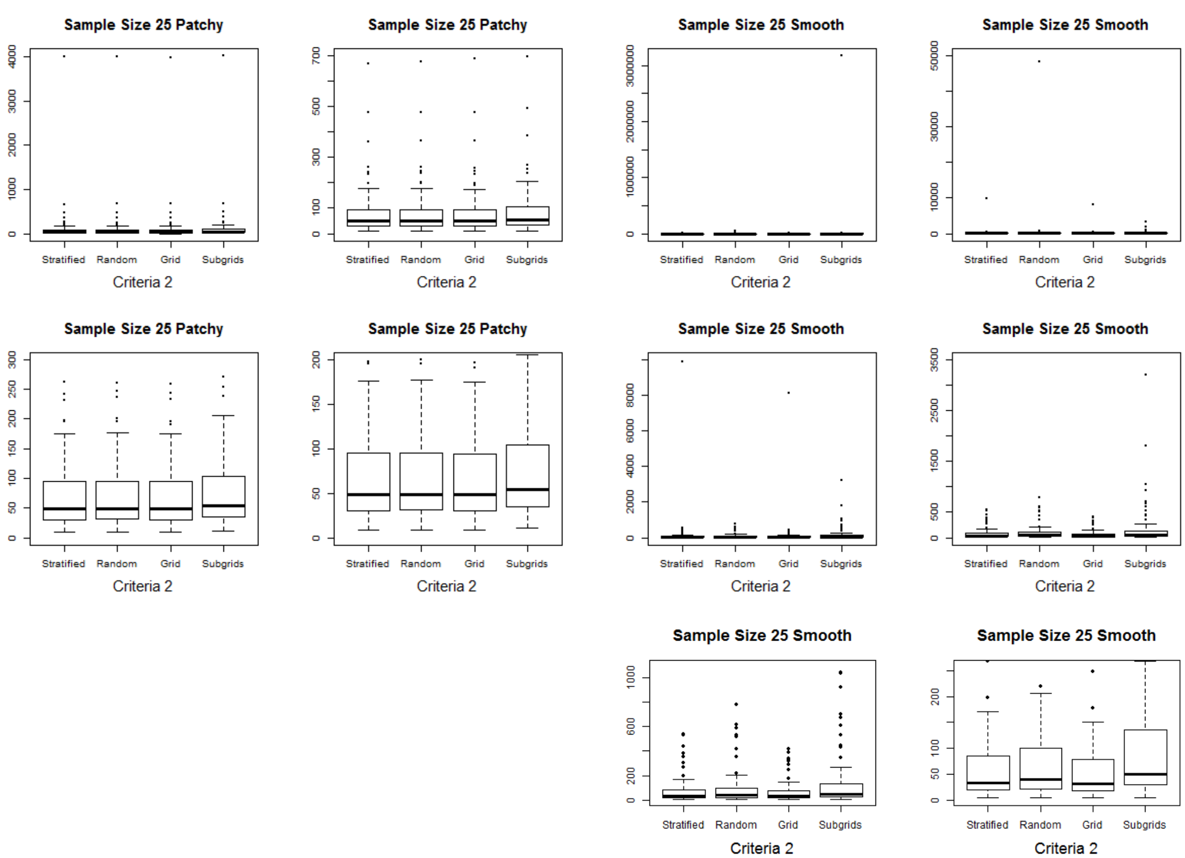

Sample Size 81 Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

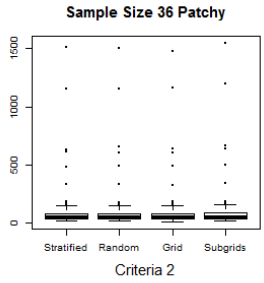

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 2

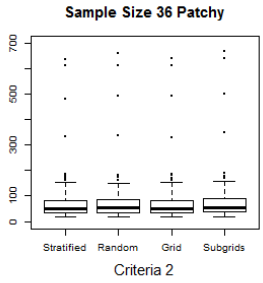

Sample Size 81 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

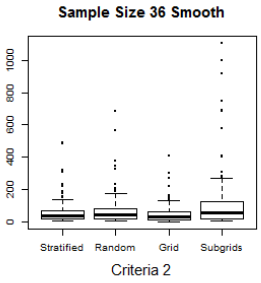


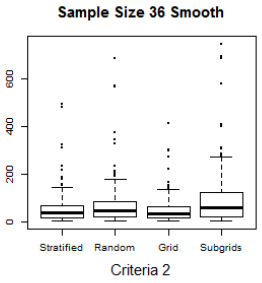
Box-plots. Scales are gradually adjusted to better display the center of the distributions.

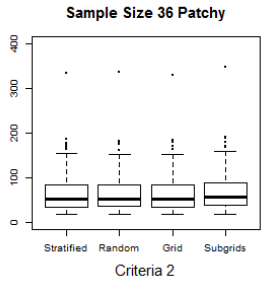


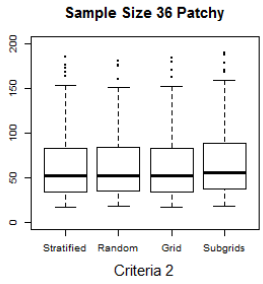

Sample Size 36 Patchy
0
500
1000
1500
Stratified
Random
Grid
Subgrids
Criteria 2

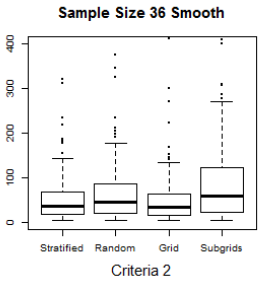

Sample Size 36 Patchy
0
100
300
500
700
Stratified
Random
Grid
Subgrids
Criteria 2

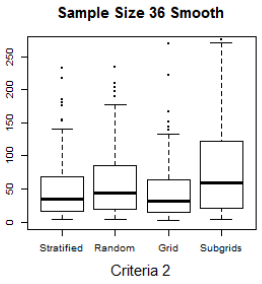

Sample Size 36 Smooth
0
200
400
600
800
1000
Stratified
Random
Grid
Subgrids
Criteria 2

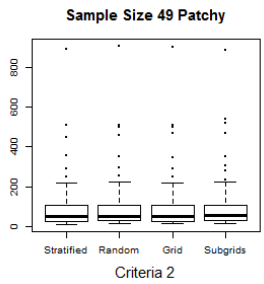

Sample Size 36 Smooth
0
200
400
600
Stratified
Random
Grid
Subgrids
Criteria 2

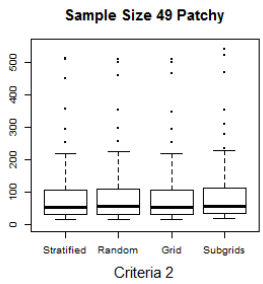

Sample Size 36 Patchy
0
100
200
300
400
Stratified
Random
Grid
Subgrids
Criteria 2

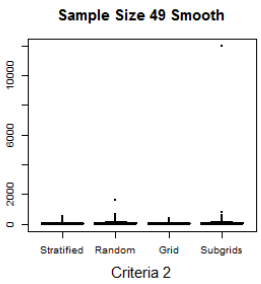

Sample Size 36 Patchy
0
50
100
150
200
Stratified
Random
Grid
Subgrids
Criteria 2

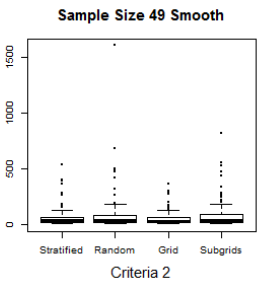

Sample Size 36 Smooth
0
100
200
300
400
Stratified
Random
Grid
Subgrids
Criteria 2

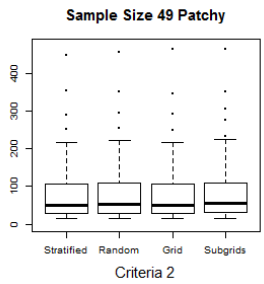

Sample Size 36 Smooth
0
50
100
150
200
250
Stratified
Random
Grid
Subgrids
Criteria 2

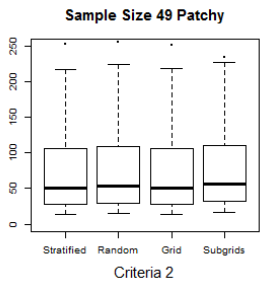

Sample Size 49 Patchy
0
200
400
600
800
Stratified
Random
Grid
Subgrids
Criteria 2

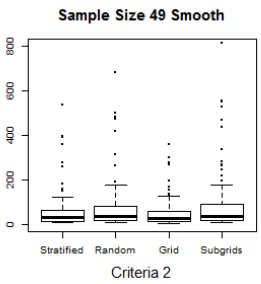

Sample Size 49 Patchy
0
100
200
300
400
500
Stratified
Random
Grid
Subgrids
Criteria 2

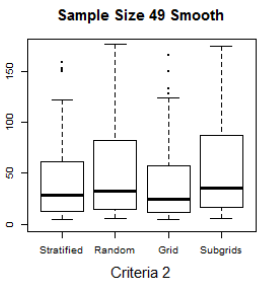

Sample Size 49 Smooth
0
2000
6000
10000
Stratified
Random
Grid
Subgrids
Criteria 2

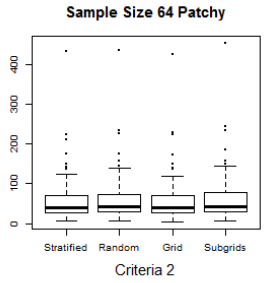

Sample Size 49 Smooth
0
500
1000
1500
Stratified
Random
Grid
Subgrids
Criteria 2

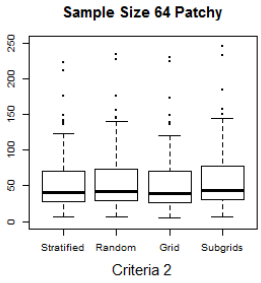

Sample Size 49 Patchy
0
100
200
300
400
Stratified
Random
Grid
Subgrids
Criteria 2

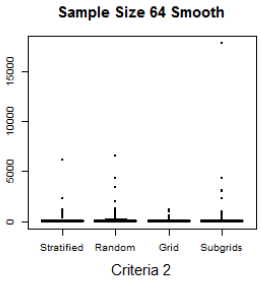

Sample Size 49 Patchy
0
50
100
150
200
250
Stratified
Random
Grid
Subgrids
Criteria 2

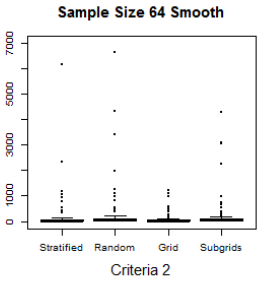

Sample Size 49 Smooth
0
200
400
600
800
Stratified
Random
Grid
Subgrids
Criteria 2

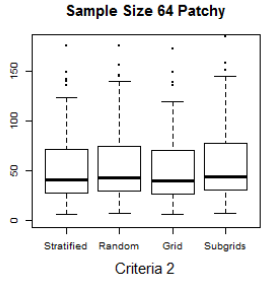

Sample Size 49 Smooth
0
50
100
150
Stratified
Random
Grid
Subgrids
Criteria 2

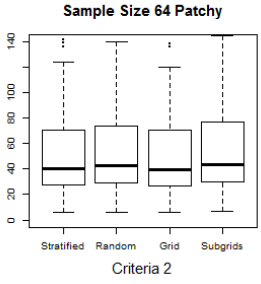

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2

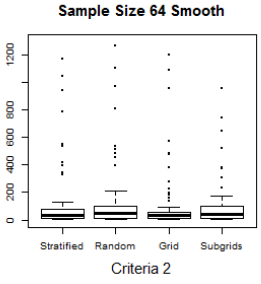

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2

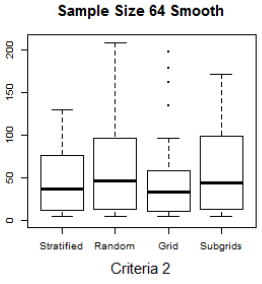

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2

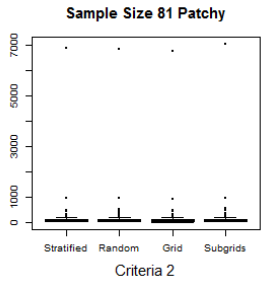

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2

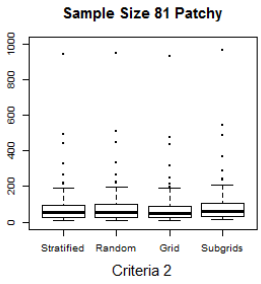

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2

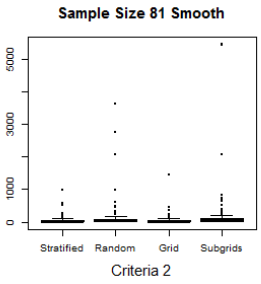

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2

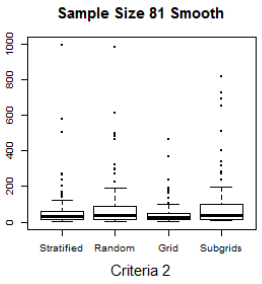

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2

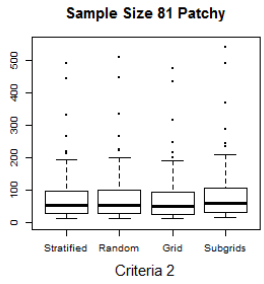

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2

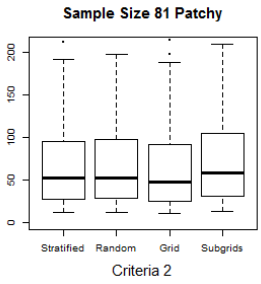

Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2

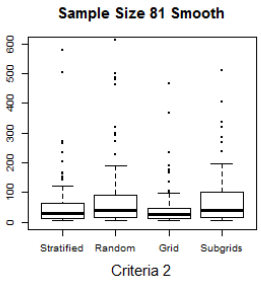

Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2

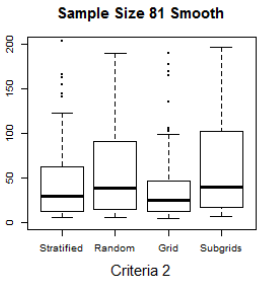

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2

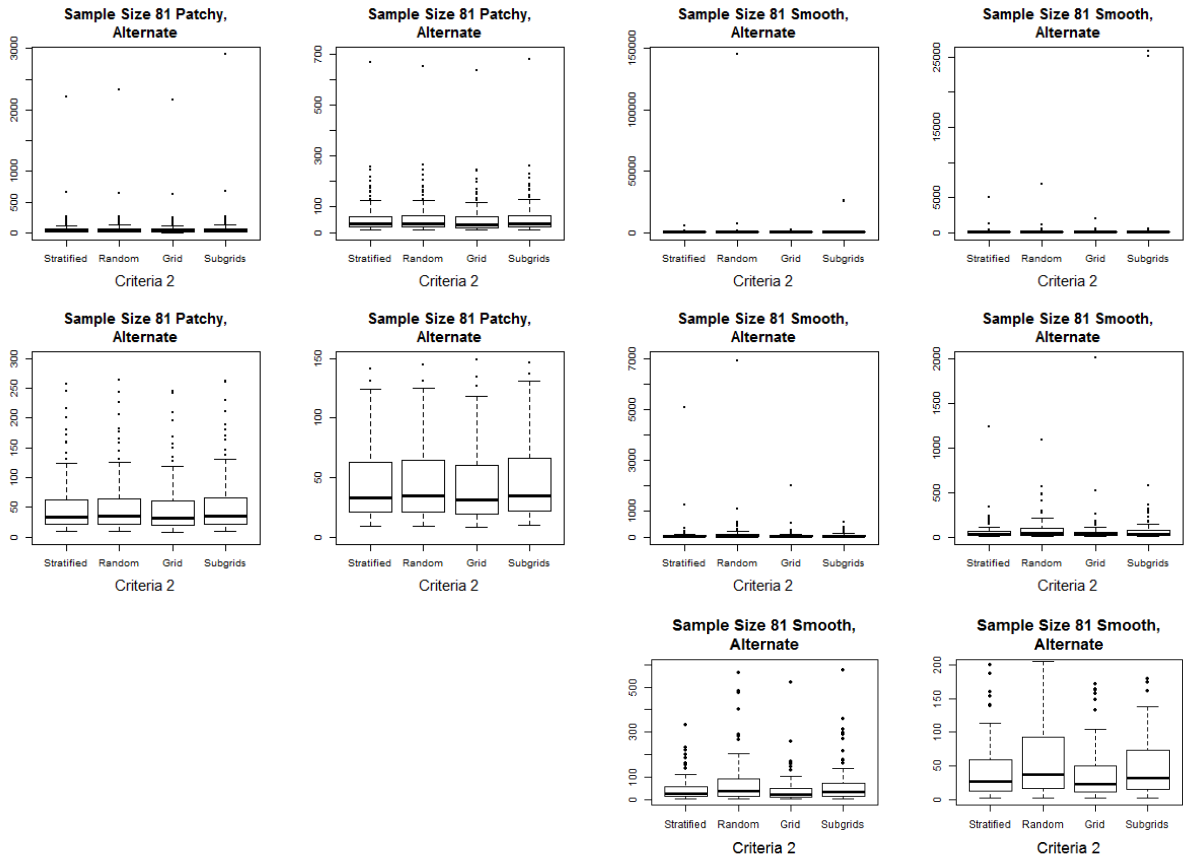

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2


Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2


Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Criteria 2


Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2

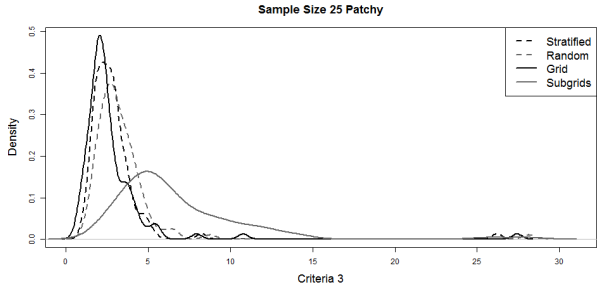

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Criteria 2

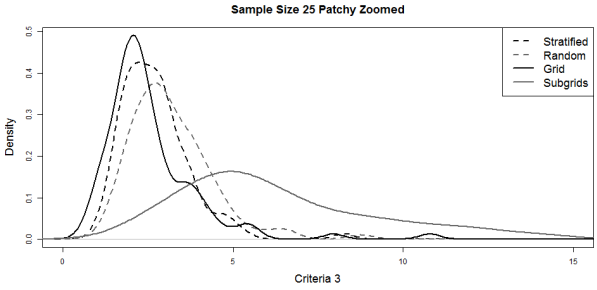


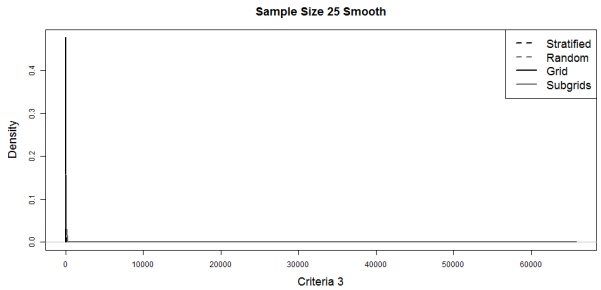
C.3 Criteria 3: Estimated MSE of the Field Mean

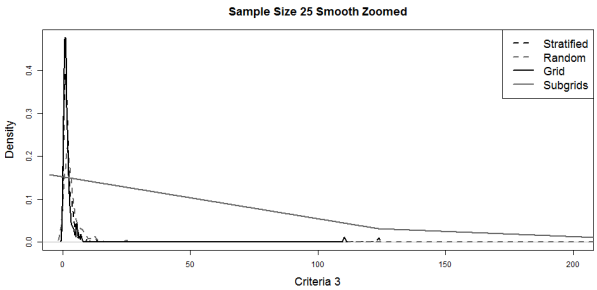
$$\widehat{MSE}(\mu_\lambda) = \left(\frac{1}{B}\sum_{b=1}^{B}\bar{\hat{\lambda}}_b - \bar{\lambda}\right)^2 + \frac{1}{B}\sum_{b=1}^{B}\left[\bar{\hat{\lambda}}_b - \frac{1}{B}\sum_{b=1}^{B}\bar{\hat{\lambda}}_b\right]^2 \tag{C.3}$$

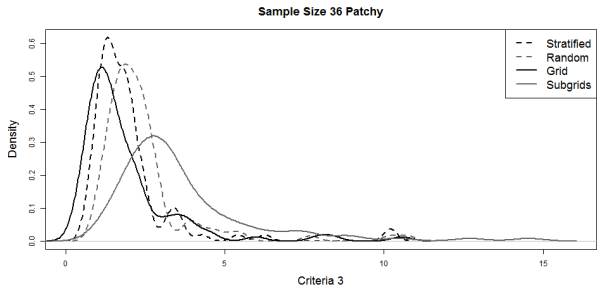
Kernel Density Estimators. Full size, including extreme values, on the left and adjusted scale on the right, to show more detail in the bulk of the distribution.

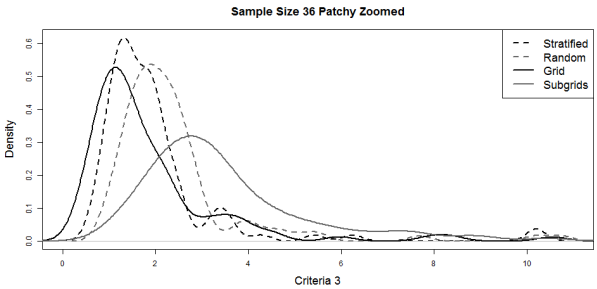


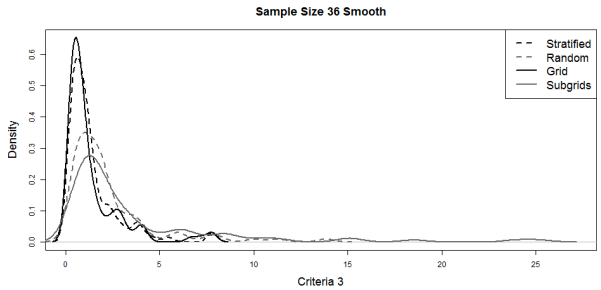


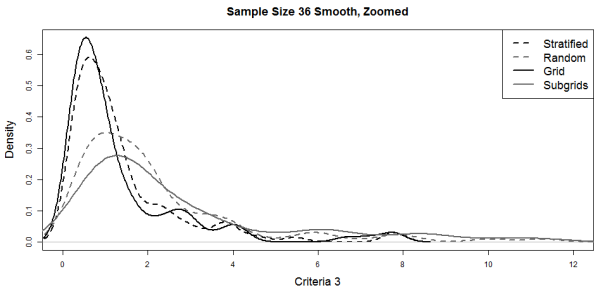


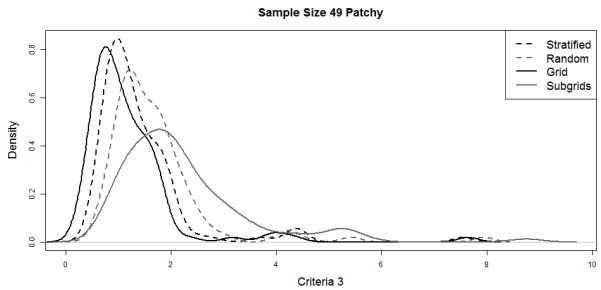


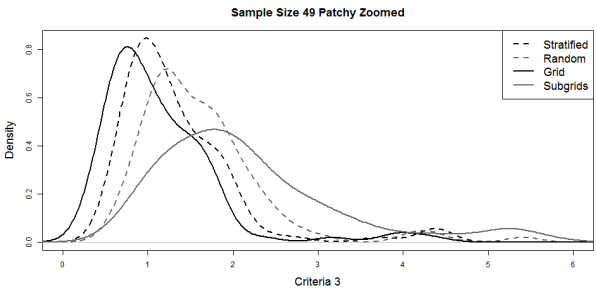


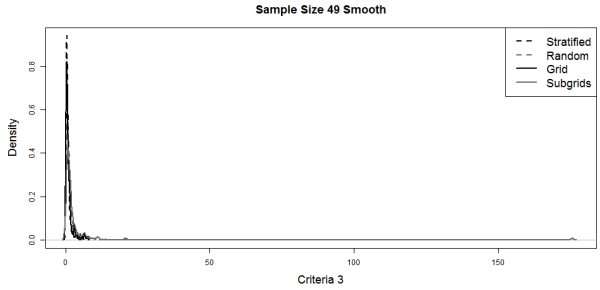


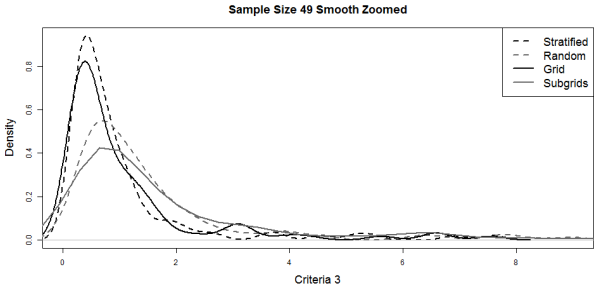


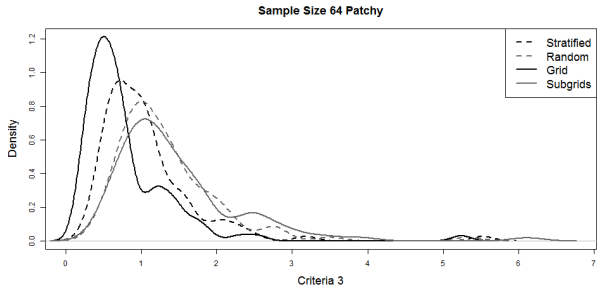


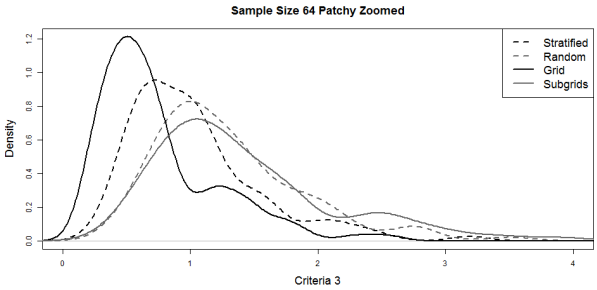


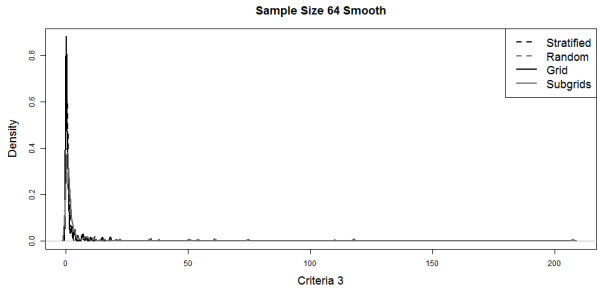


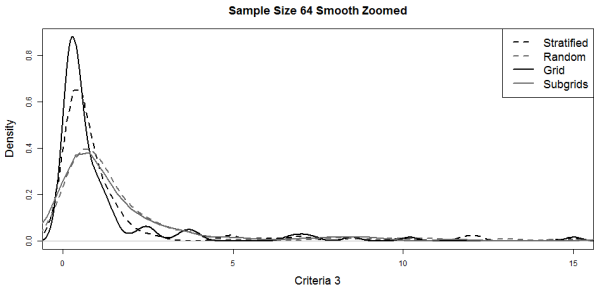

Sample Size 49 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 3

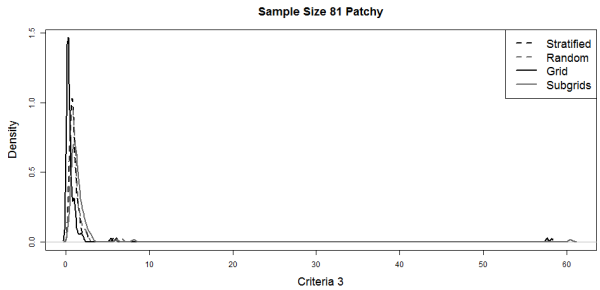

Sample Size 49 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 3

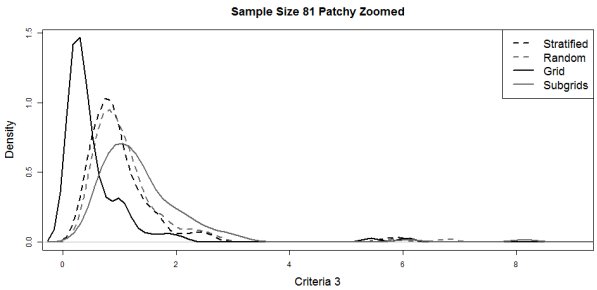

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 3

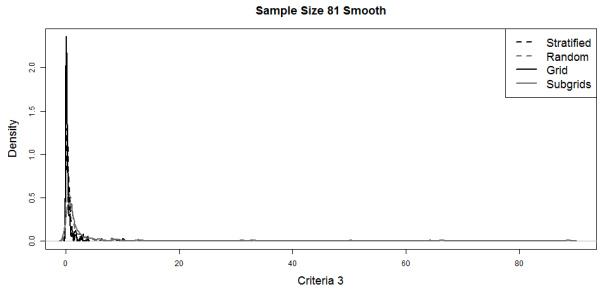

Sample Size 64 Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 3

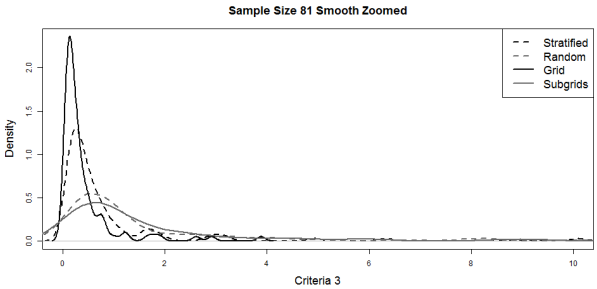

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 3

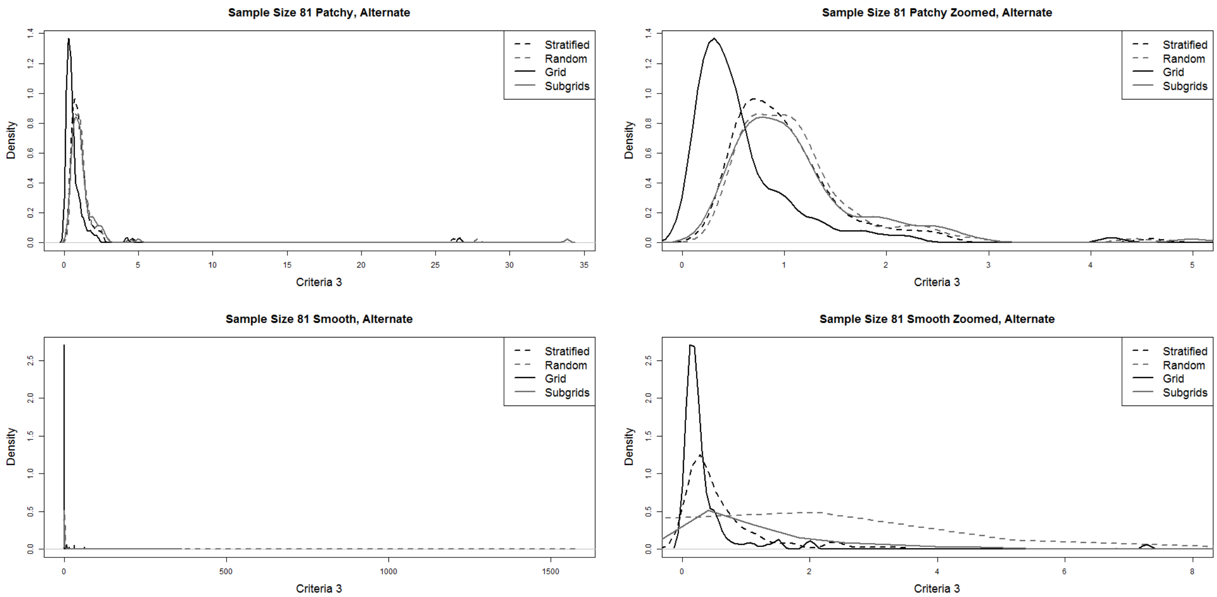

Sample Size 64 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 3


Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 3

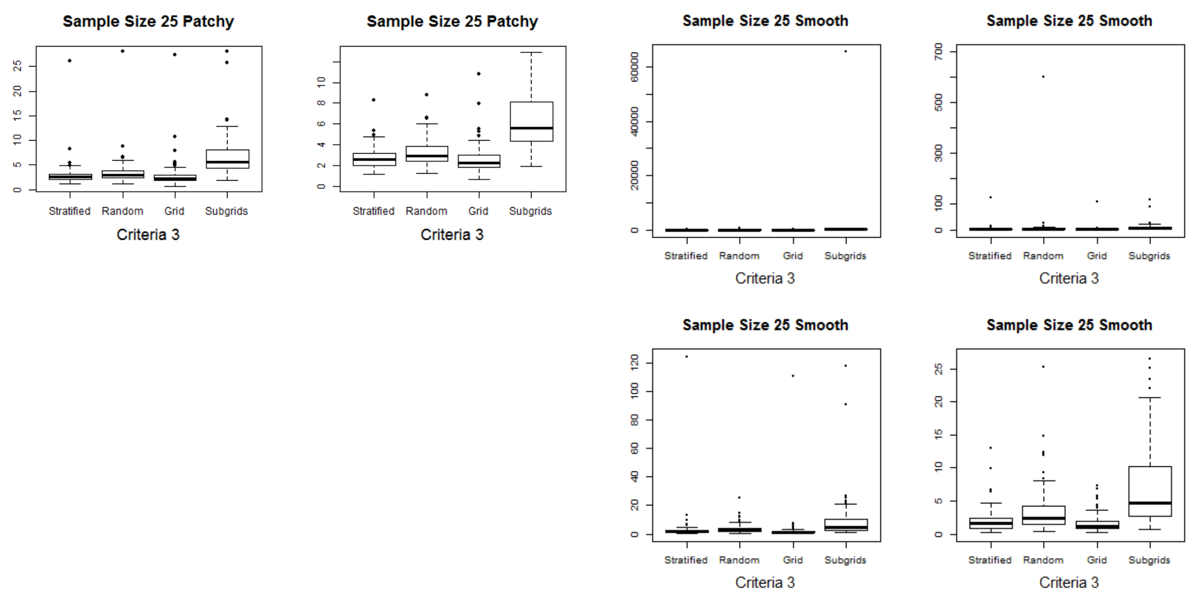

Sample Size 81 Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 3

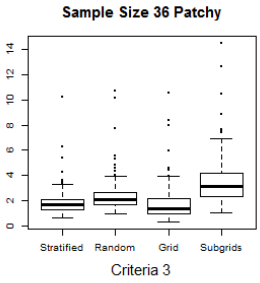

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 3

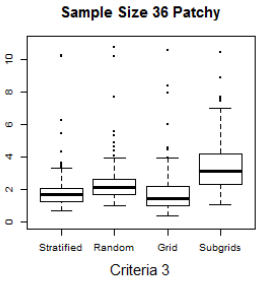

Sample Size 81 Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 3

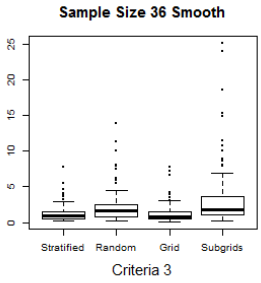


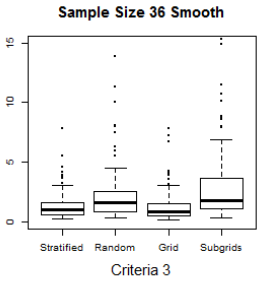
Box-plots. Scales are gradually adjusted to better display the center of the distributions.

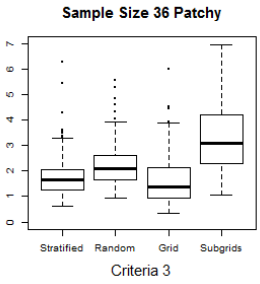


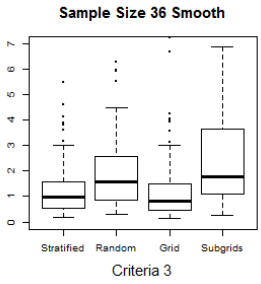

Sample Size 36 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

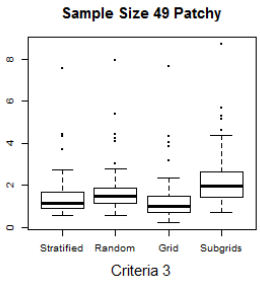

Sample Size 36 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

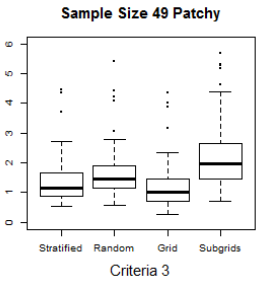

Sample Size 36 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

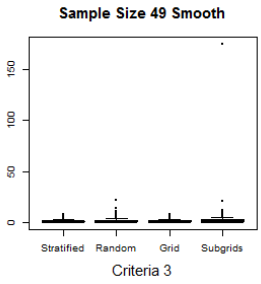

Sample Size 36 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

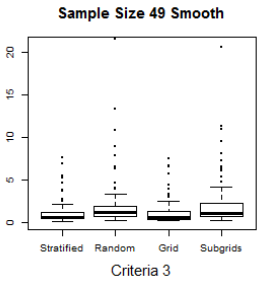

Sample Size 36 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

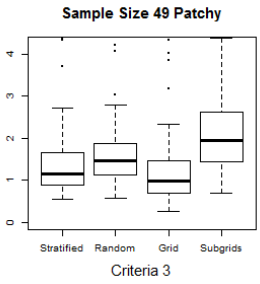

Sample Size 36 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

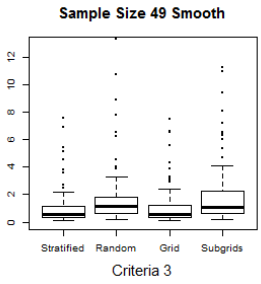

Sample Size 49 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

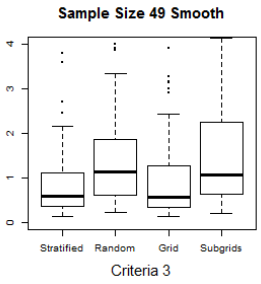

Sample Size 49 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

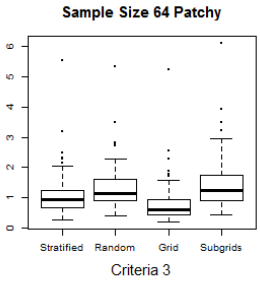

Sample Size 49 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

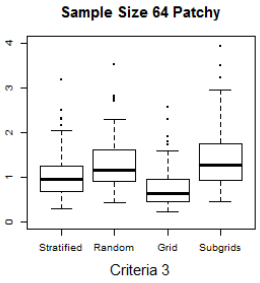

Sample Size 49 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

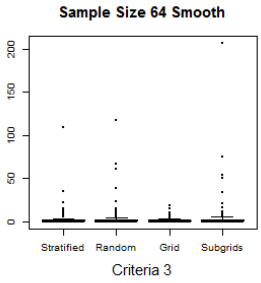

Sample Size 49 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

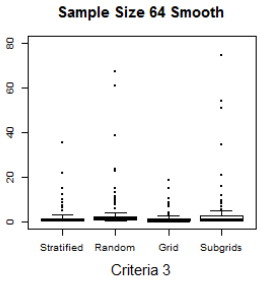

Sample Size 49 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

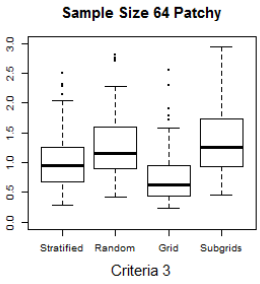

Sample Size 49 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

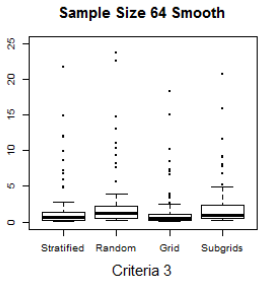

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

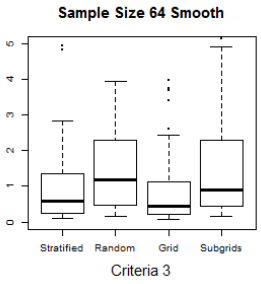

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

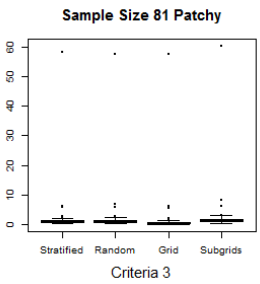

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

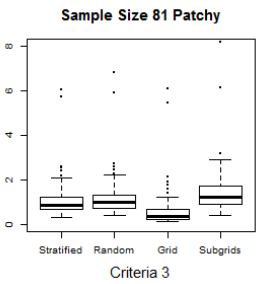

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

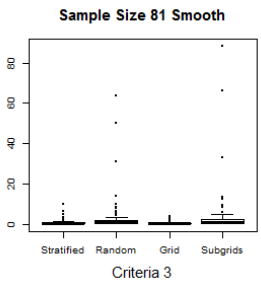

Sample Size 64 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

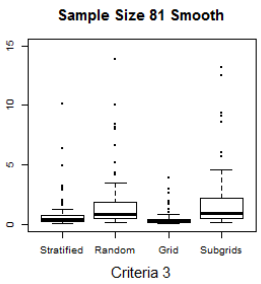

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

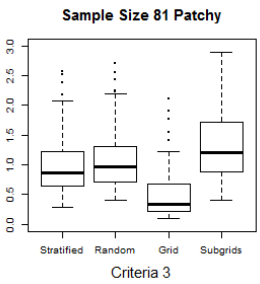

Sample Size 64 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

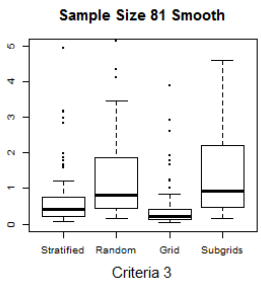

Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

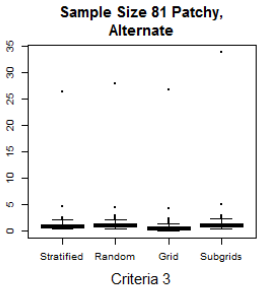

Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

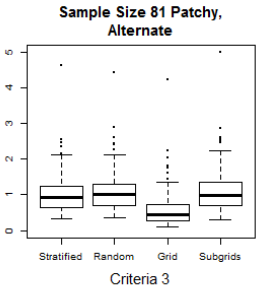

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

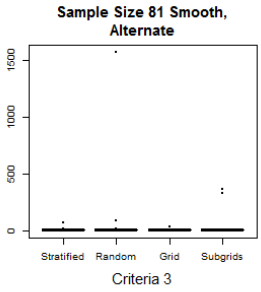

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

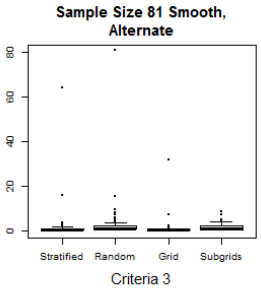

Sample Size 81 Patchy
Stratified
Random
Grid
Subgrids
Criteria 3

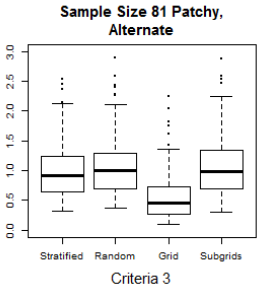

Sample Size 81 Smooth
Stratified
Random
Grid
Subgrids
Criteria 3

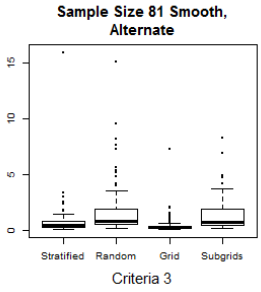

Sample Size 81 Patchy, Alternate
0 5 10 15 20 25 30 35
Stratified Random Grid Subgrids
Criteria 3

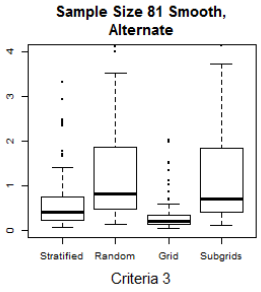

Sample Size 81 Patchy, Alternate
0 1 2 3 4 5
Stratified Random Grid Subgrids
Criteria 3


Sample Size 81 Smooth, Alternate
0 500 1000 1500
Stratified Random Grid Subgrids
Criteria 3


Sample Size 81 Smooth, Alternate
0 20 40 60 80
Stratified Random Grid Subgrids
Criteria 3


Sample Size 81 Patchy, Alternate
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Stratified Random Grid Subgrids
Criteria 3


Sample Size 81 Smooth, Alternate
0 5 10 15
Stratified Random Grid Subgrids
Criteria 3


Sample Size 81 Smooth, Alternate
0 1 2 3 4
Stratified Random Grid Subgrids
Criteria 3

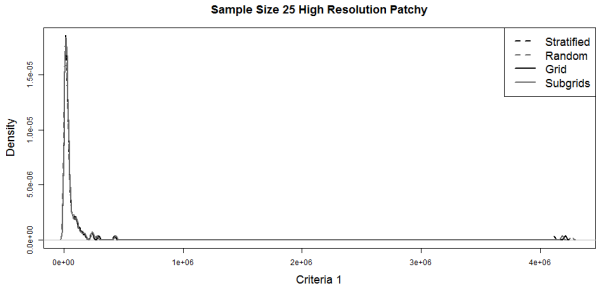
Appendix D

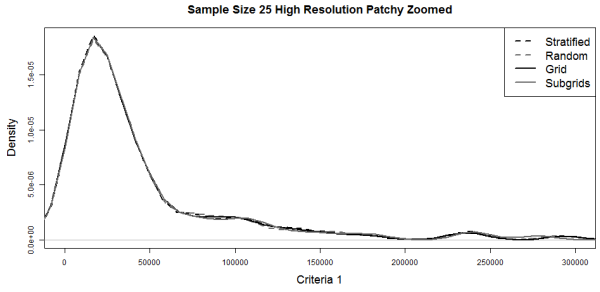
High Resolution Result Plots

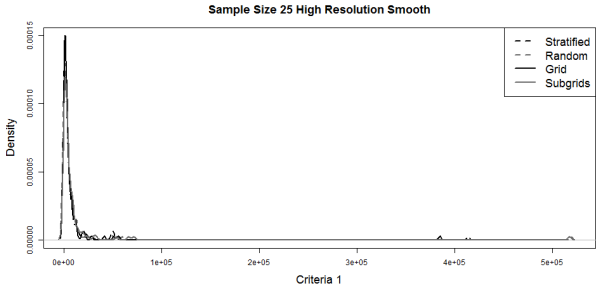
D.1 Criteria 1: Maximum Estimated MSE

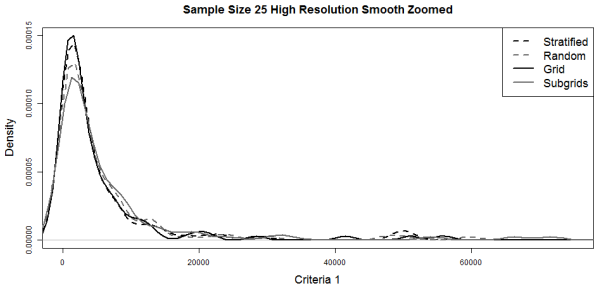
$$\underset{i}{maximum}\ \widehat{MSE}(x_i) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\lambda}_b(x_i) - \lambda(x_i)\right]^2 \tag{D.1}$$

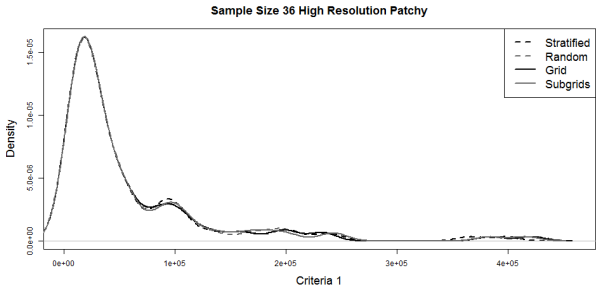
Kernel Density Estimators. Full size, including extreme values, on the left and adjusted scale on the right, to show more detail in the bulk of the distribution. In some cases, the full size plots are sufficient, due to the lack of extreme values, so adjusted scale plots are not included.

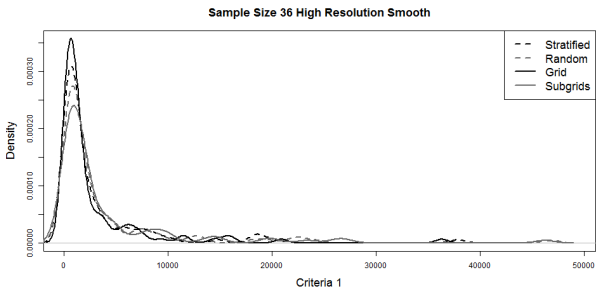


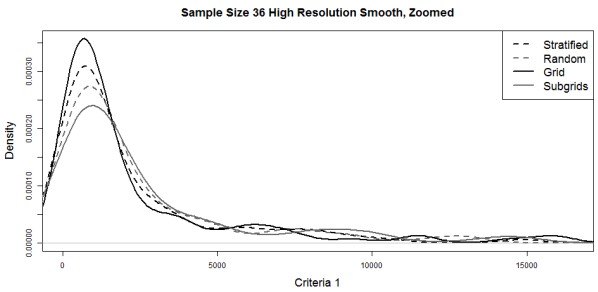


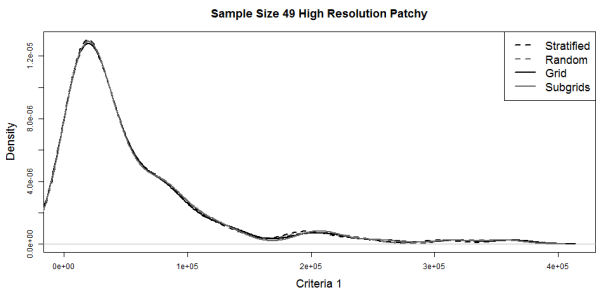

Sample Size 25 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1

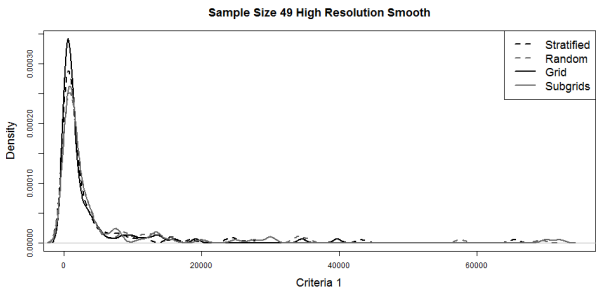

Sample Size 25 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

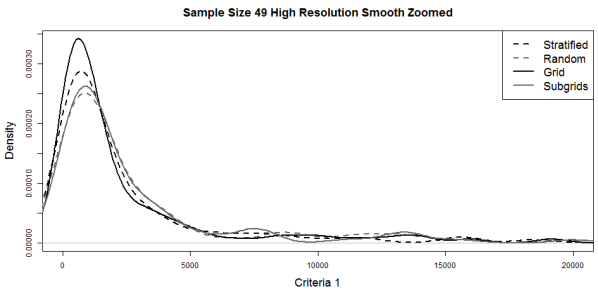

Sample Size 36 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

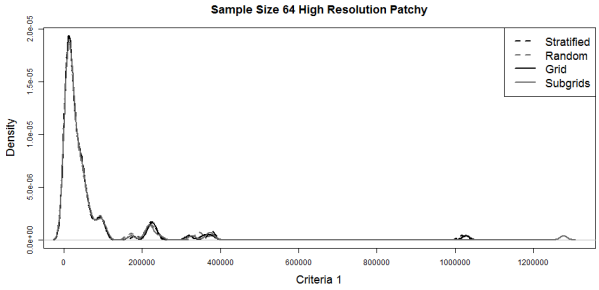

Sample Size 36 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1

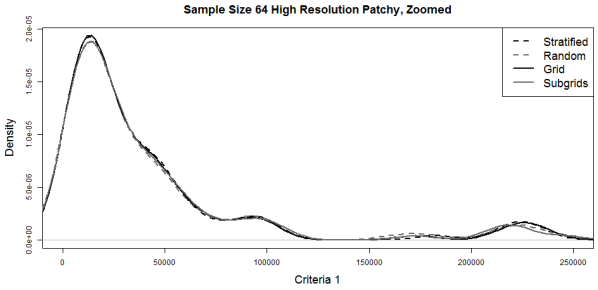

Sample Size 36 High Resolution Smooth, Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

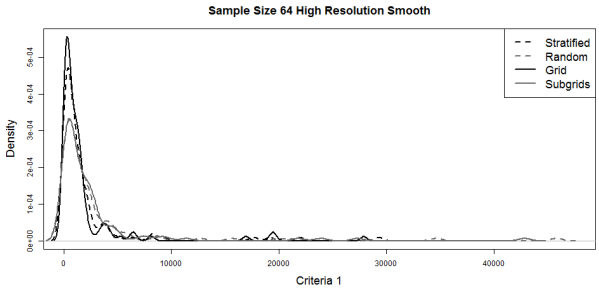

Sample Size 49 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

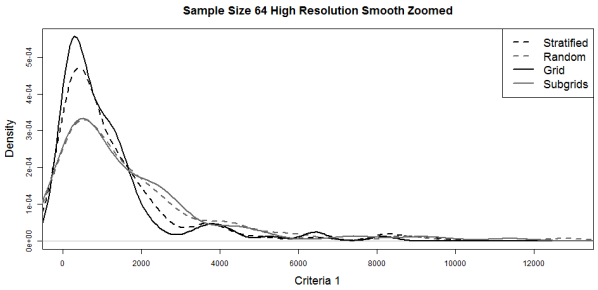

Sample Size 49 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1

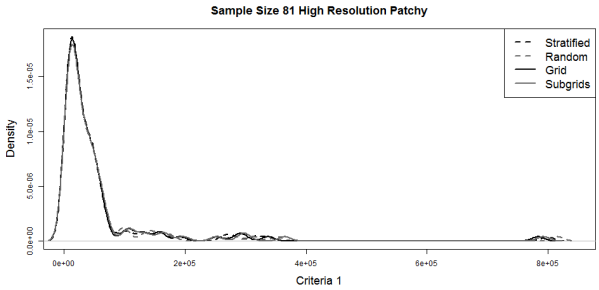

Sample Size 49 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

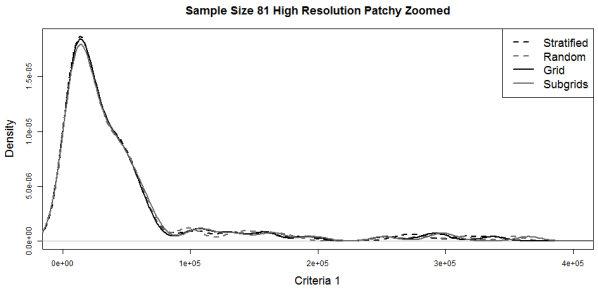

Sample Size 64 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

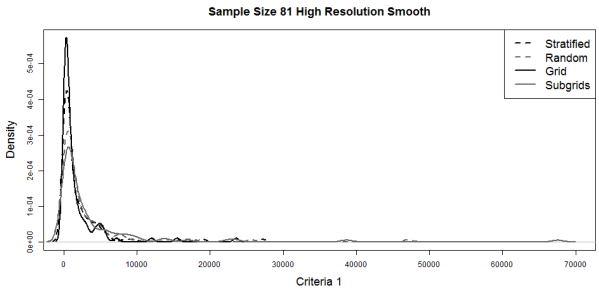

Sample Size 64 High Resolution Patchy, Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1


Sample Size 64 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1

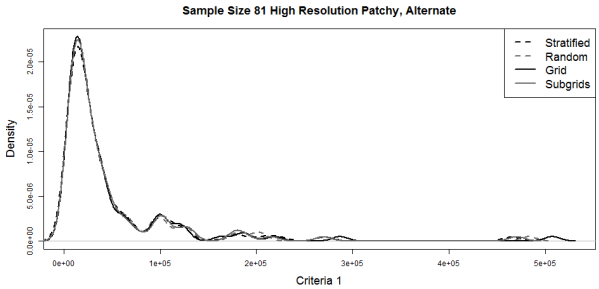

Sample Size 64 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

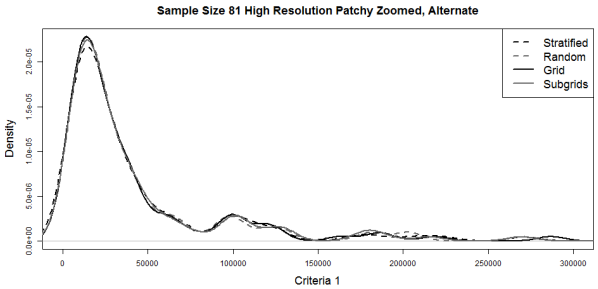

Sample Size 81 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 1

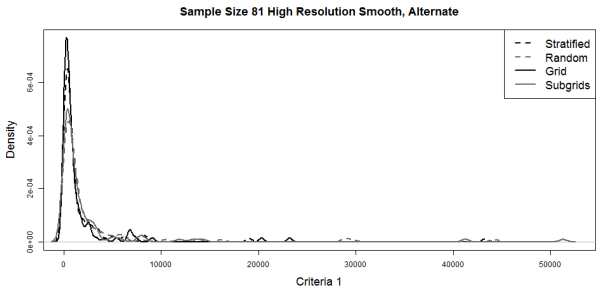

Sample Size 81 High Resolution Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

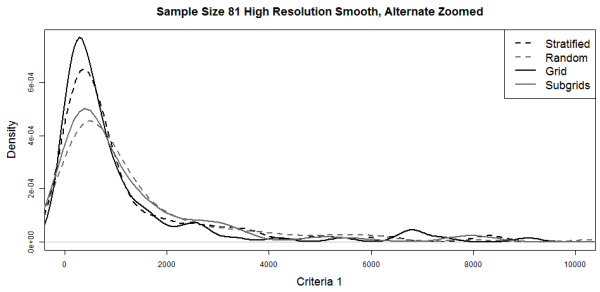

Sample Size 81 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 1


Sample Size 81 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 1

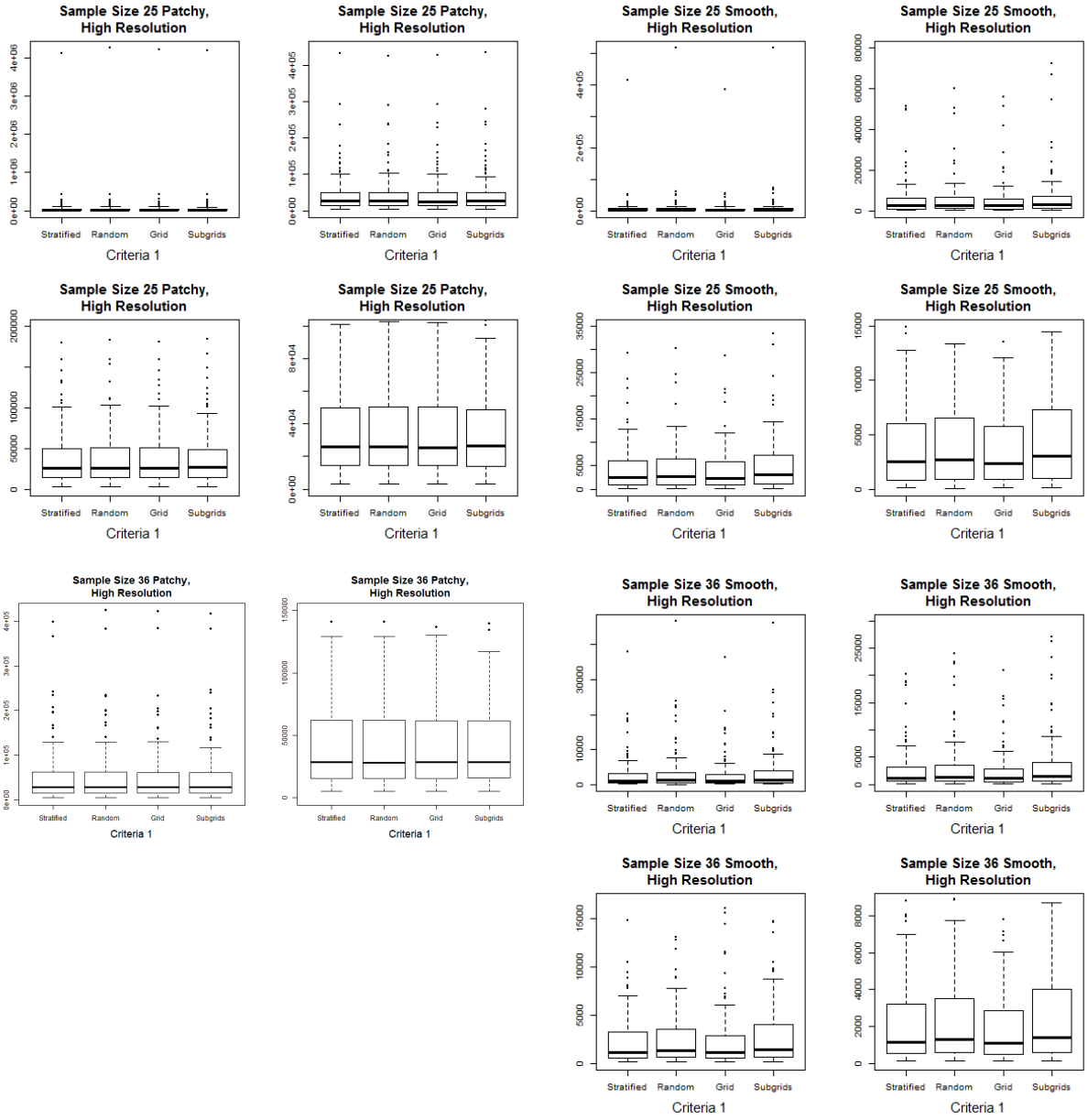

Sample Size 81 High Resolution Patchy, Alternate
Stratified
Random
Grid
Subgrids
Density
Criteria 1

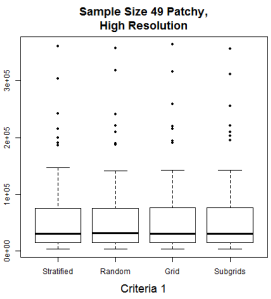

Sample Size 81 High Resolution Patchy Zoomed, Alternate
Stratified
Random
Grid
Subgrids
Density
Criteria 1

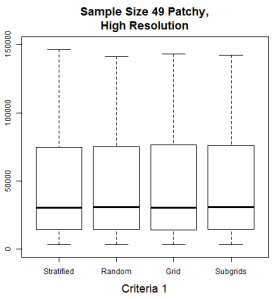

Sample Size 81 High Resolution Smooth, Alternate
Stratified
Random
Grid
Subgrids
Density
0e+00
2e-04
4e-04
6e-04
0
10000
20000
30000
40000
50000
Criteria 1

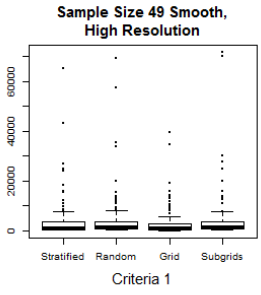

Sample Size 81 High Resolution Smooth, Alternate Zoomed
Stratified
Random
Grid
Subgrids
Density
0e+00
2e-04
4e-04
6e-04
0
2000
4000
6000
8000
10000
Criteria 1

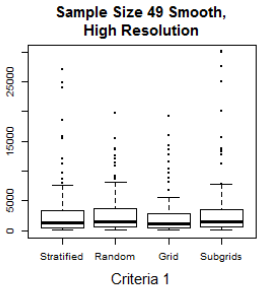
Box-plots. Scales are gradually adjusted to better display the center of the distributions.

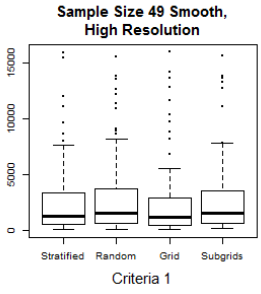


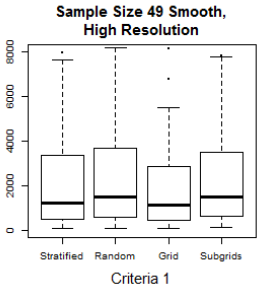

Sample Size 49 Patchy,
High Resolution
0e+00
1e+05
2e+05
3e+05
Stratified
Random
Grid
Subgrids
Criteria 1

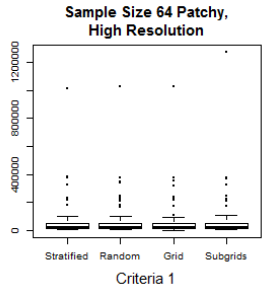

Sample Size 49 Patchy,
High Resolution
0
50000
100000
150000
Stratified
Random
Grid
Subgrids
Criteria 1

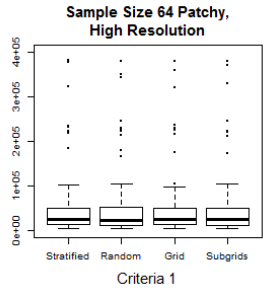

Sample Size 49 Smooth,
High Resolution
0
20000
40000
60000
Stratified
Random
Grid
Subgrids
Criteria 1

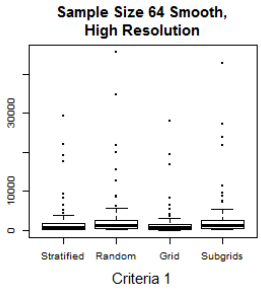

Sample Size 49 Smooth,
High Resolution
0
5000
15000
25000
Stratified
Random
Grid
Subgrids
Criteria 1

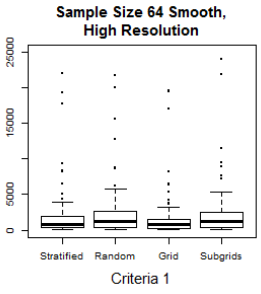

Sample Size 49 Smooth,
High Resolution
0
5000
10000
15000
Stratified
Random
Grid
Subgrids
Criteria 1

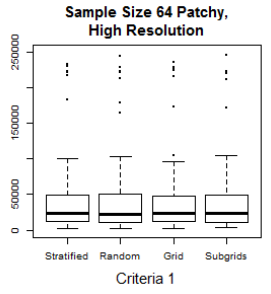

Sample Size 49 Smooth,
High Resolution
0
2000
4000
6000
8000
Stratified
Random
Grid
Subgrids
Criteria 1

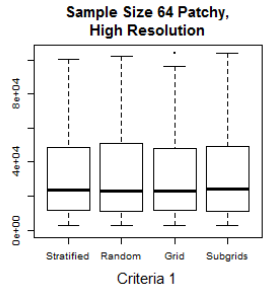

Sample Size 64 Patchy,
High Resolution
0
400000
800000
1200000
Stratified
Random
Grid
Subgrids
Criteria 1

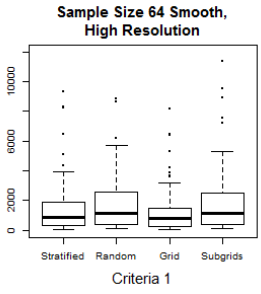

Sample Size 64 Patchy,
High Resolution
0e+00
1e+05
2e+05
3e+05
4e+05
Stratified
Random
Grid
Subgrids
Criteria 1

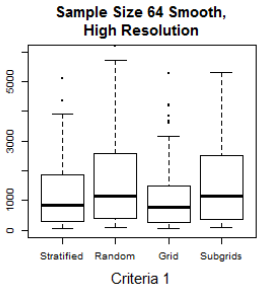

Sample Size 64 Smooth,
High Resolution
0
10000
30000
Stratified
Random
Grid
Subgrids
Criteria 1

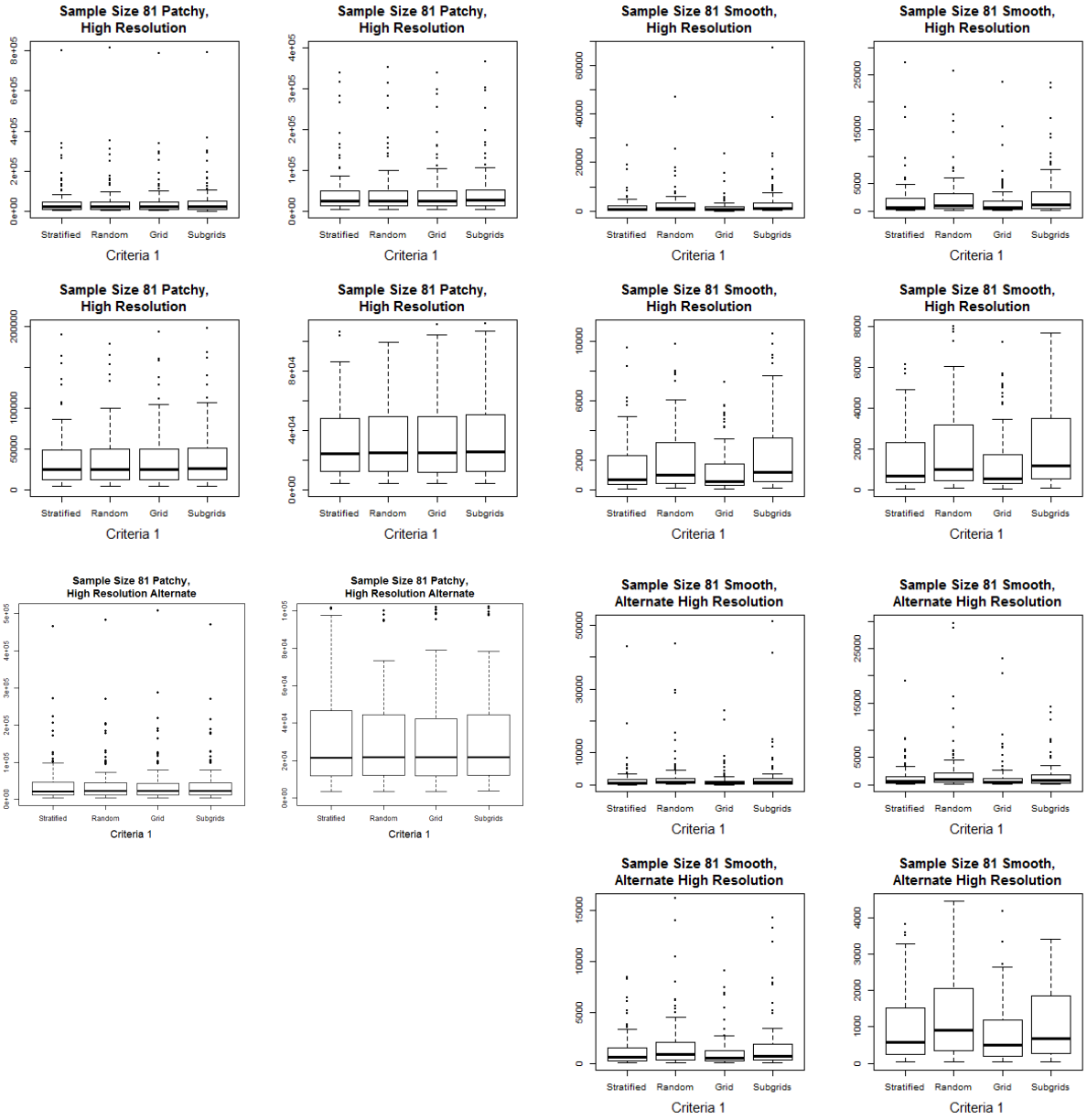

Sample Size 64 Smooth,
High Resolution
0
5000
15000
25000
Stratified
Random
Grid
Subgrids
Criteria 1


Sample Size 64 Patchy,
High Resolution
0
50000
150000
250000
Stratified
Random
Grid
Subgrids
Criteria 1


Sample Size 64 Patchy,
High Resolution
0e+00
4e+04
8e+04
Stratified
Random
Grid
Subgrids
Criteria 1


Sample Size 64 Smooth,
High Resolution
0
2000
6000
10000
Stratified
Random
Grid
Subgrids
Criteria 1

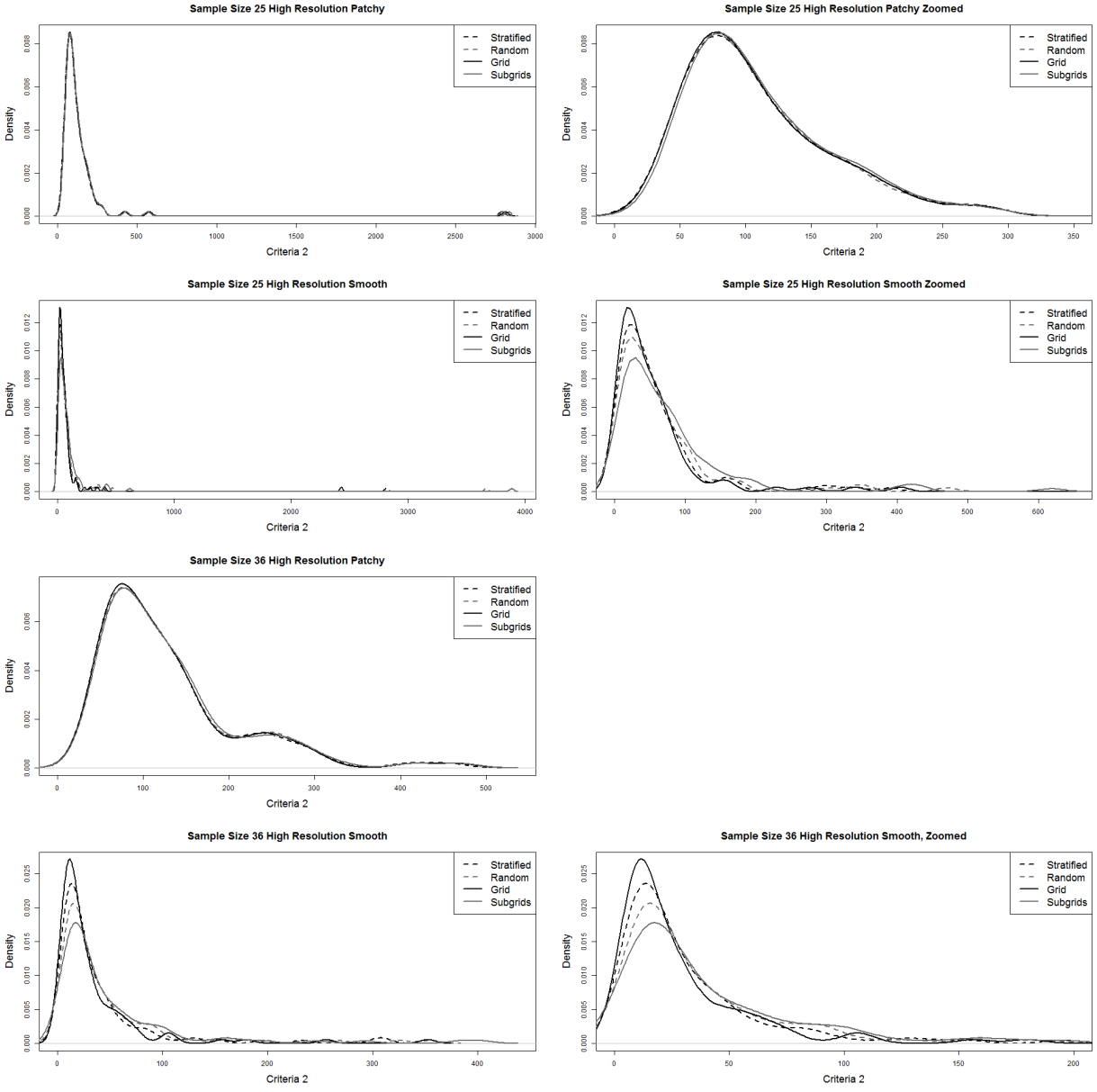

Sample Size 64 Smooth,
High Resolution
0
1000
3000
5000
Stratified
Random
Grid
Subgrids
Criteria 1

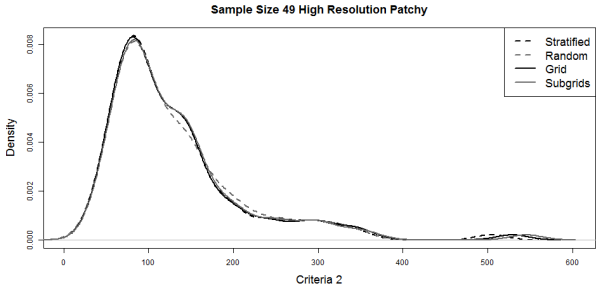


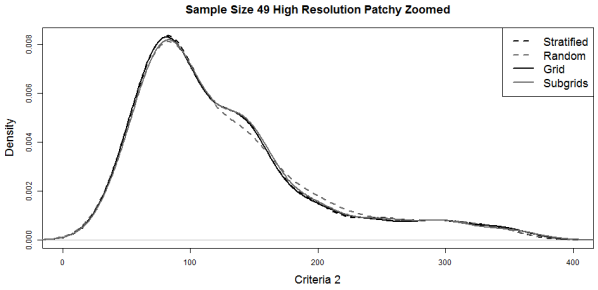
D.2 Criteria 2: Mean Estimated MSE

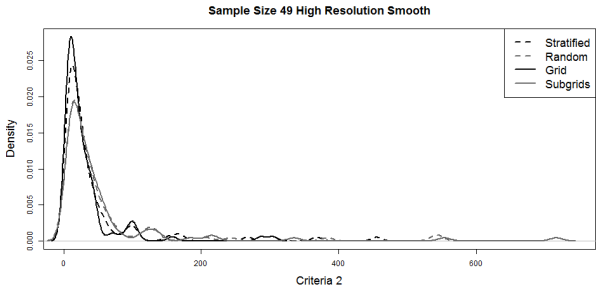
$$\underset{i}{mean}\ \widehat{MSE}(x_i) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\lambda}_b(x_i) - \lambda(x_i)\right]^2 \tag{D.2}$$

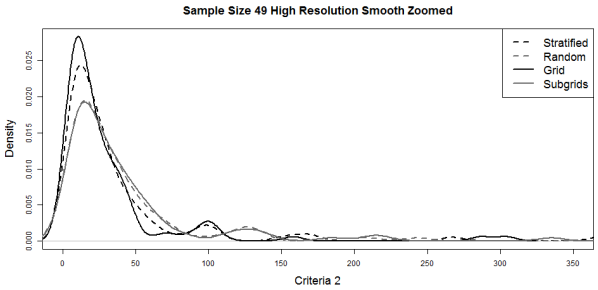
Kernel Density Estimators. Full size, including extreme values, on the left and adjusted scale on the right, to show more detail in the bulk of the distribution. In some cases, the full size plots are sufficient, due to the lack of extreme values, so adjusted scale plots are not included.

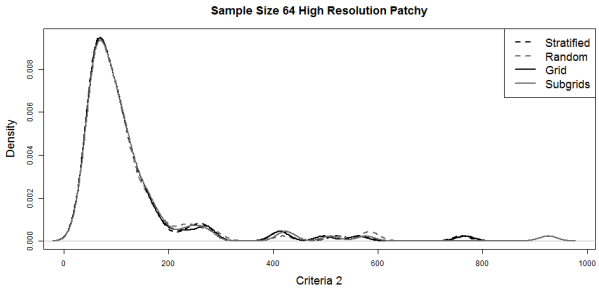


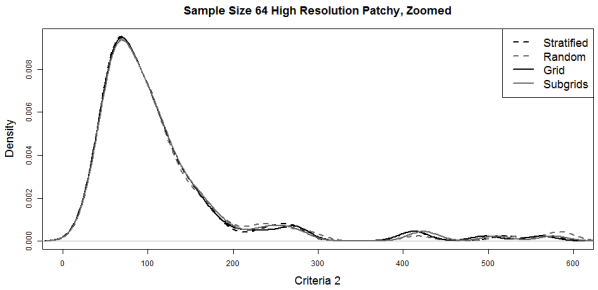

Sample Size 49 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 2

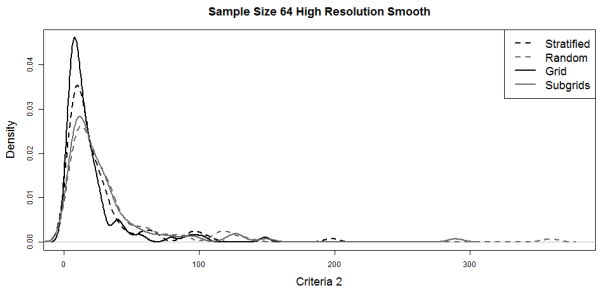

Sample Size 49 High Resolution Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

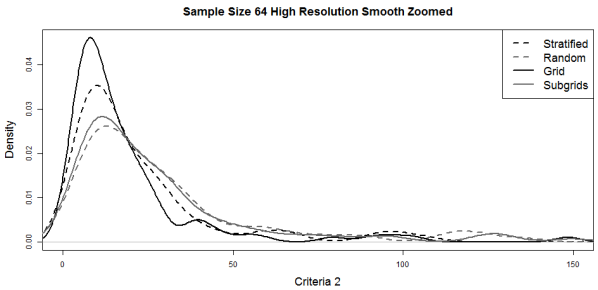

Sample Size 49 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 2

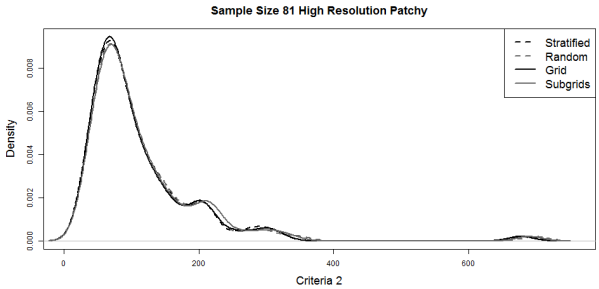

Sample Size 49 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

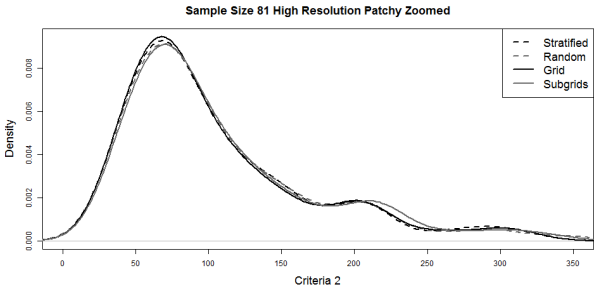

Sample Size 64 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 2

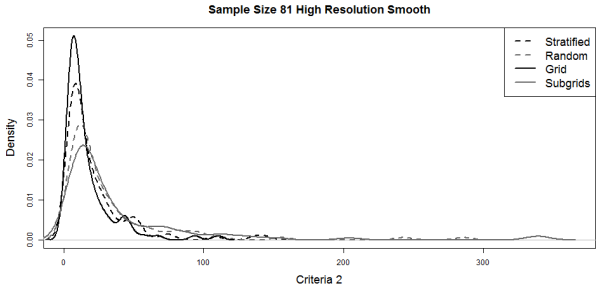

Sample Size 64 High Resolution Patchy, Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

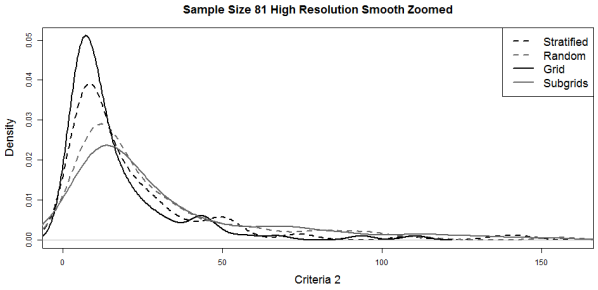

Sample Size 64 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 2

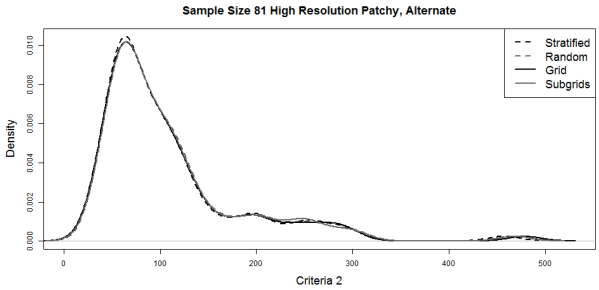

Sample Size 64 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2

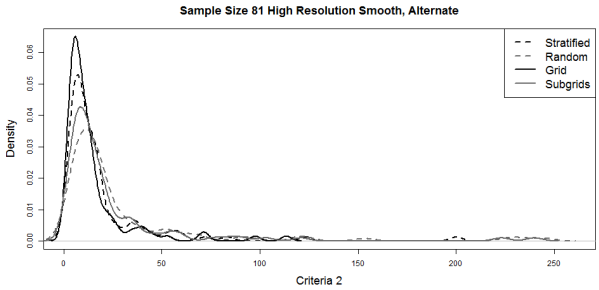

Sample Size 81 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 2

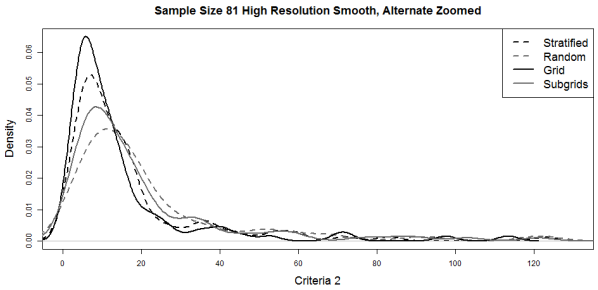

Sample Size 81 High Resolution Patchy Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 2



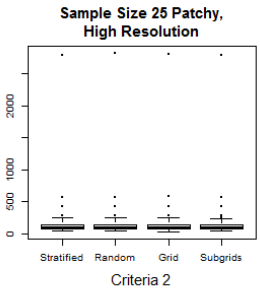


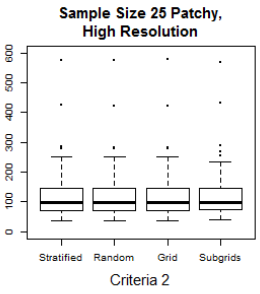


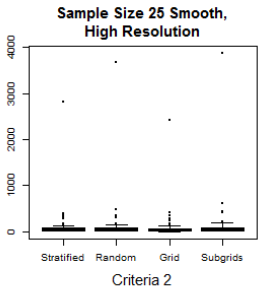


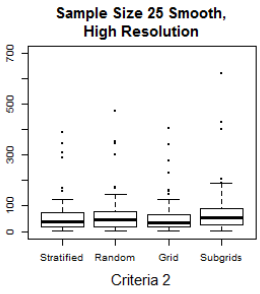


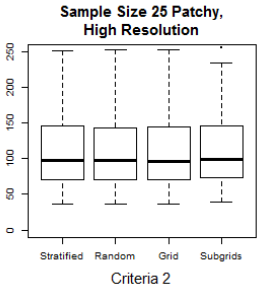
Box-plots. Scales are gradually adjusted to better display the center of the distributions.

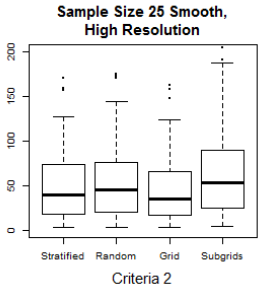


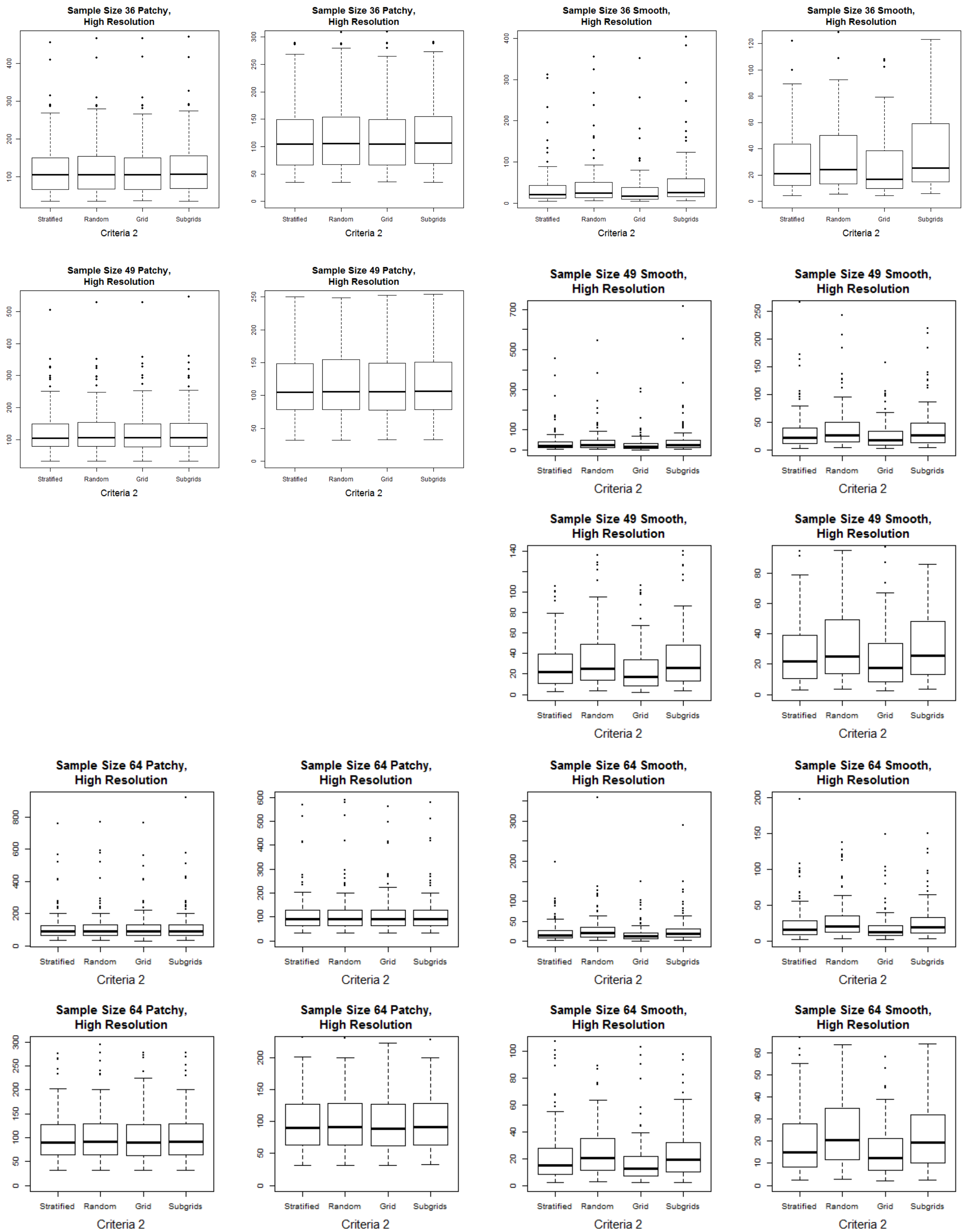


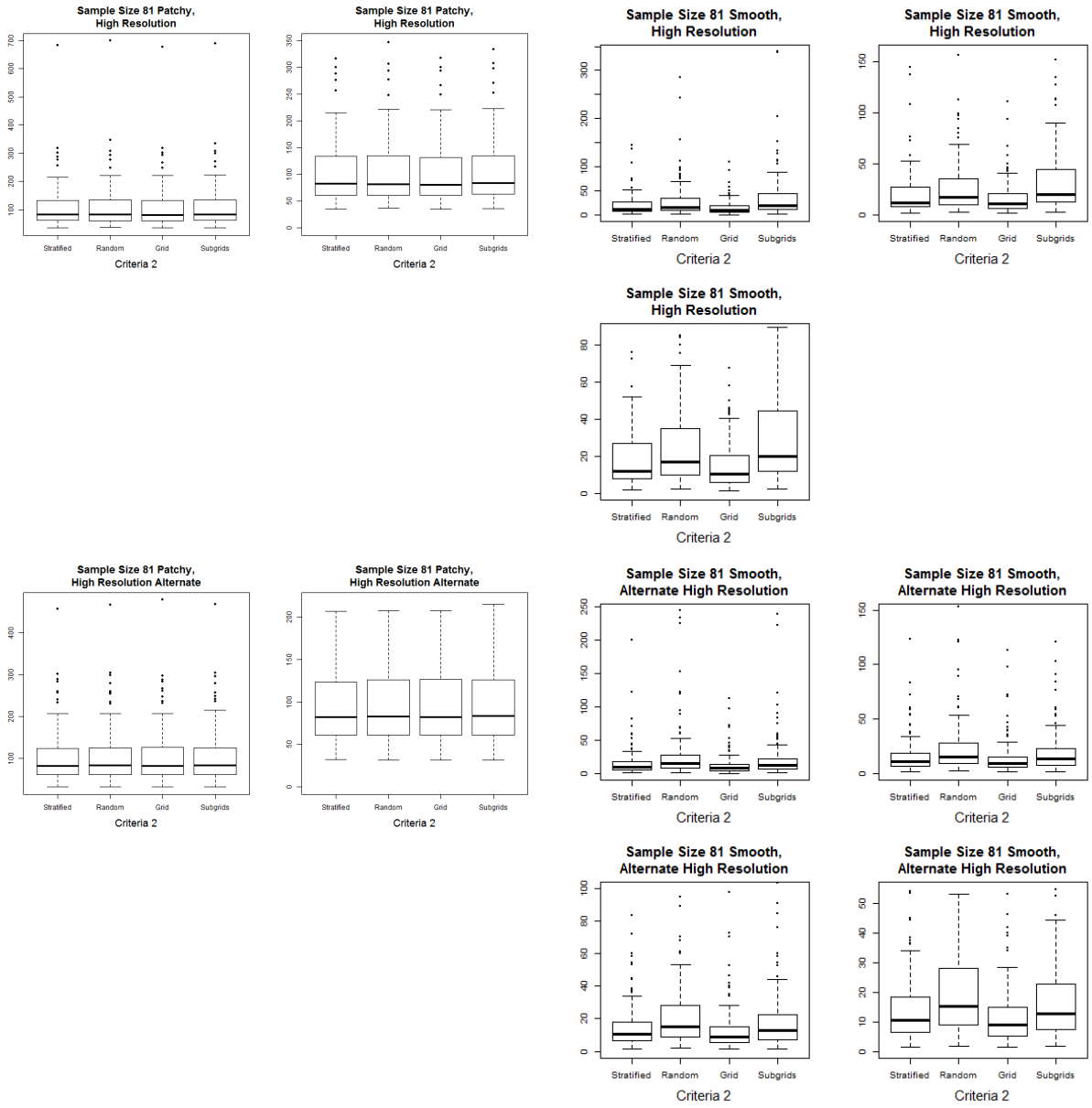








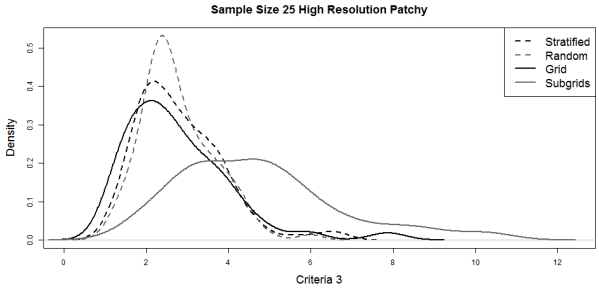

Sample Size 36 Patchy, High Resolution
Sample Size 36 Patchy, High Resolution
Sample Size 36 Smooth, High Resolution
Sample Size 36 Smooth, High Resolution
Sample Size 49 Patchy, High Resolution
Sample Size 49 Patchy, High Resolution
Sample Size 49 Smooth, High Resolution
Sample Size 49 Smooth, High Resolution
Sample Size 49 Smooth, High Resolution
Sample Size 49 Smooth, High Resolution
Sample Size 64 Patchy, High Resolution
Sample Size 64 Patchy, High Resolution
Sample Size 64 Smooth, High Resolution
Sample Size 64 Smooth, High Resolution
Sample Size 64 Patchy, High Resolution
Sample Size 64 Patchy, High Resolution
Sample Size 64 Smooth, High Resolution
Sample Size 64 Smooth, High Resolution
Stratified
Random
Grid
Subgrids
Criteria 2

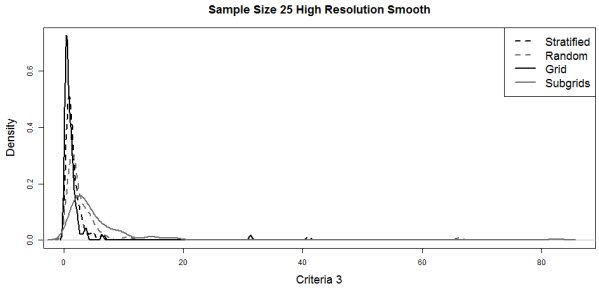


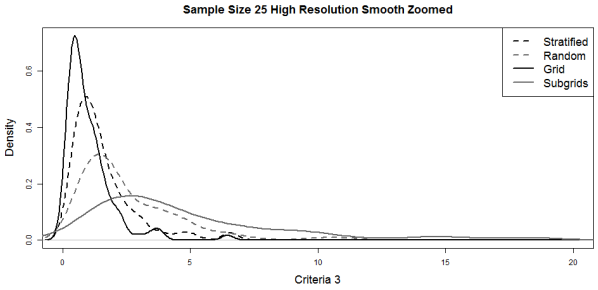
D.3 Criteria 3: Estimated MSE of the Field Mean

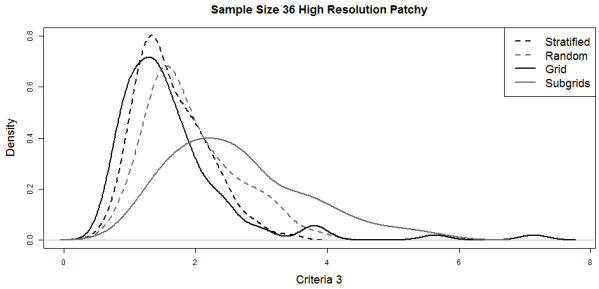
$$\widehat{MSE}(\mu_\lambda) = \left(\frac{1}{B}\sum_{b=1}^{B}\bar{\hat{\lambda}}_b - \bar{\lambda}\right)^2 + \frac{1}{B}\sum_{b=1}^{B}\left[\bar{\hat{\lambda}}_b - \frac{1}{B}\sum_{b=1}^{B}\bar{\hat{\lambda}}_b\right]^2 \tag{D.3}$$

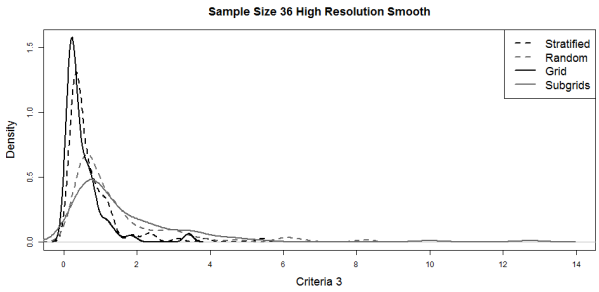
Kernel Density Estimators. Full size, including extreme values, on the left and adjusted scale on the right, to show more detail in the bulk of the distribution. In some cases, the full size plots are sufficient, due to the lack of extreme values, so adjusted scale plots are not included.

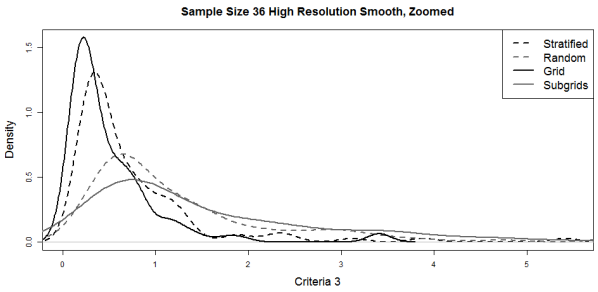


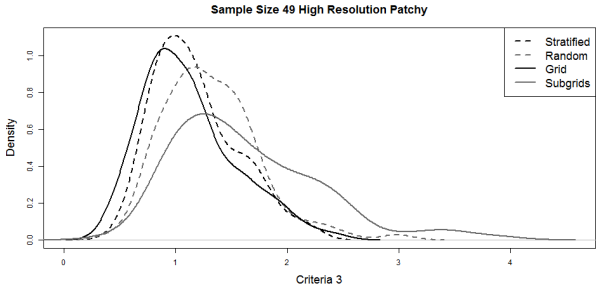


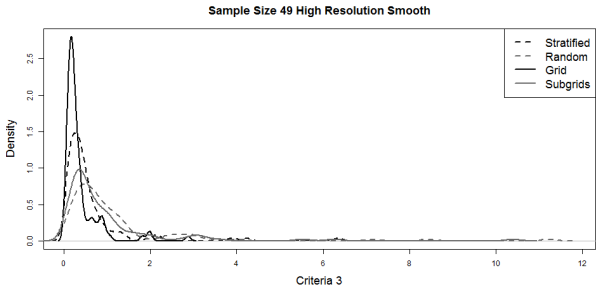


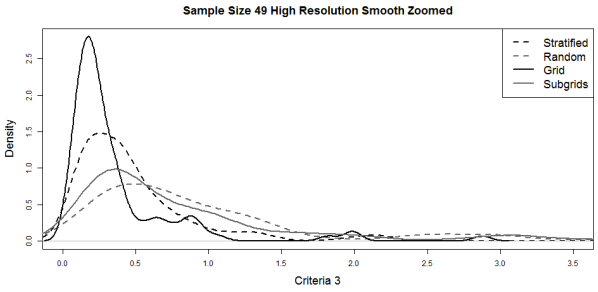


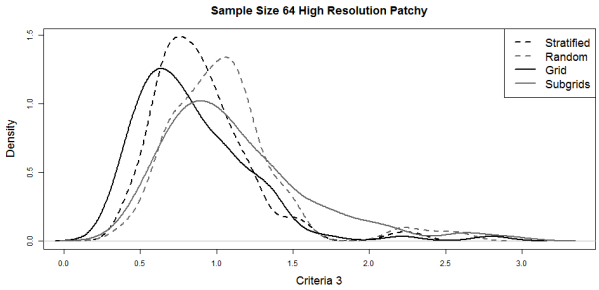


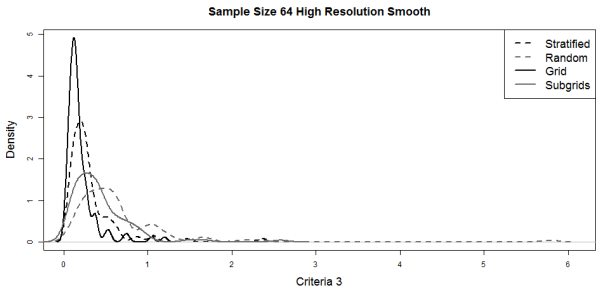


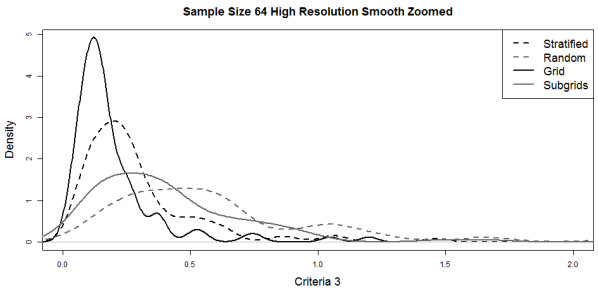

Sample Size 49 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 3

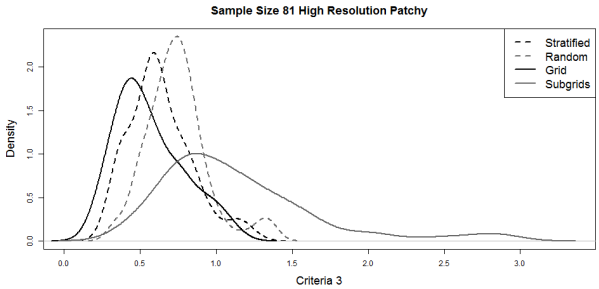

Sample Size 49 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 3

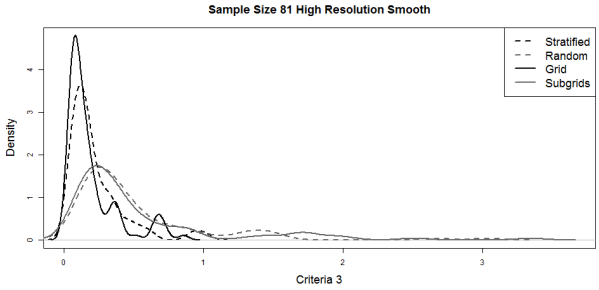

Sample Size 49 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 3

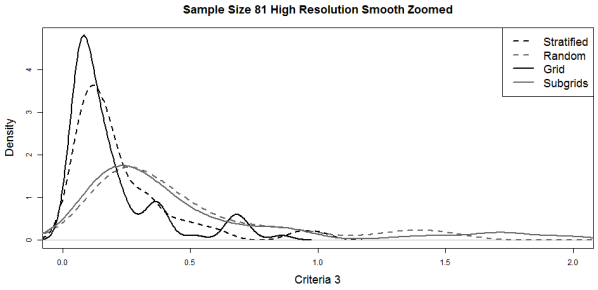

Sample Size 64 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 3

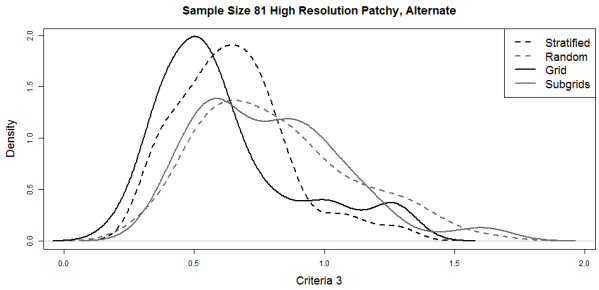

Sample Size 64 High Resolution Smooth
Stratified
Random
Grid
Subgrids
Density
Criteria 3

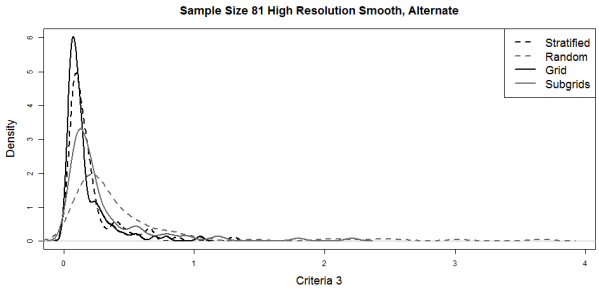

Sample Size 64 High Resolution Smooth Zoomed
Stratified
Random
Grid
Subgrids
Density
Criteria 3

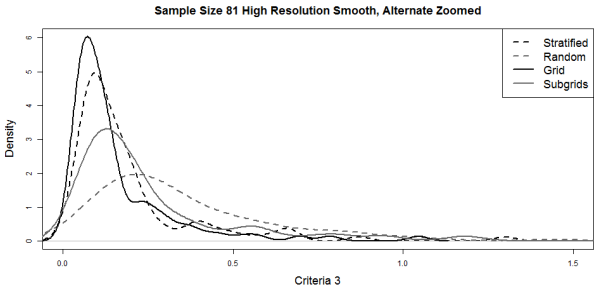

Sample Size 81 High Resolution Patchy
Stratified
Random
Grid
Subgrids
Density
Criteria 3



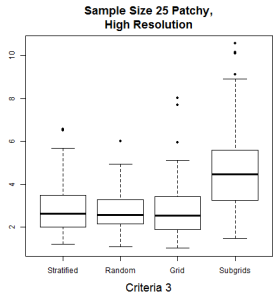


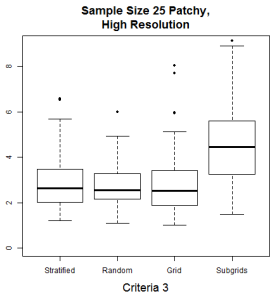


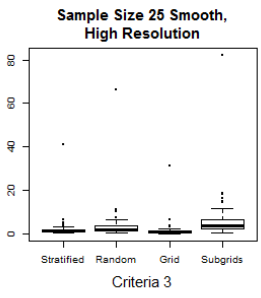


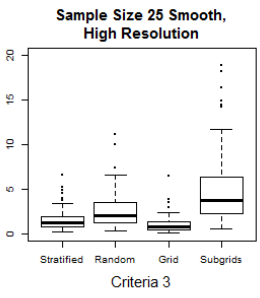


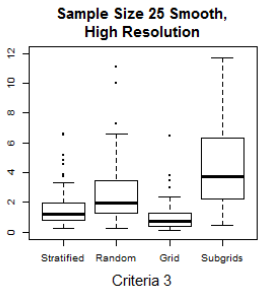
Box-plots. Scales are gradually adjusted to better display the center of the distributions.

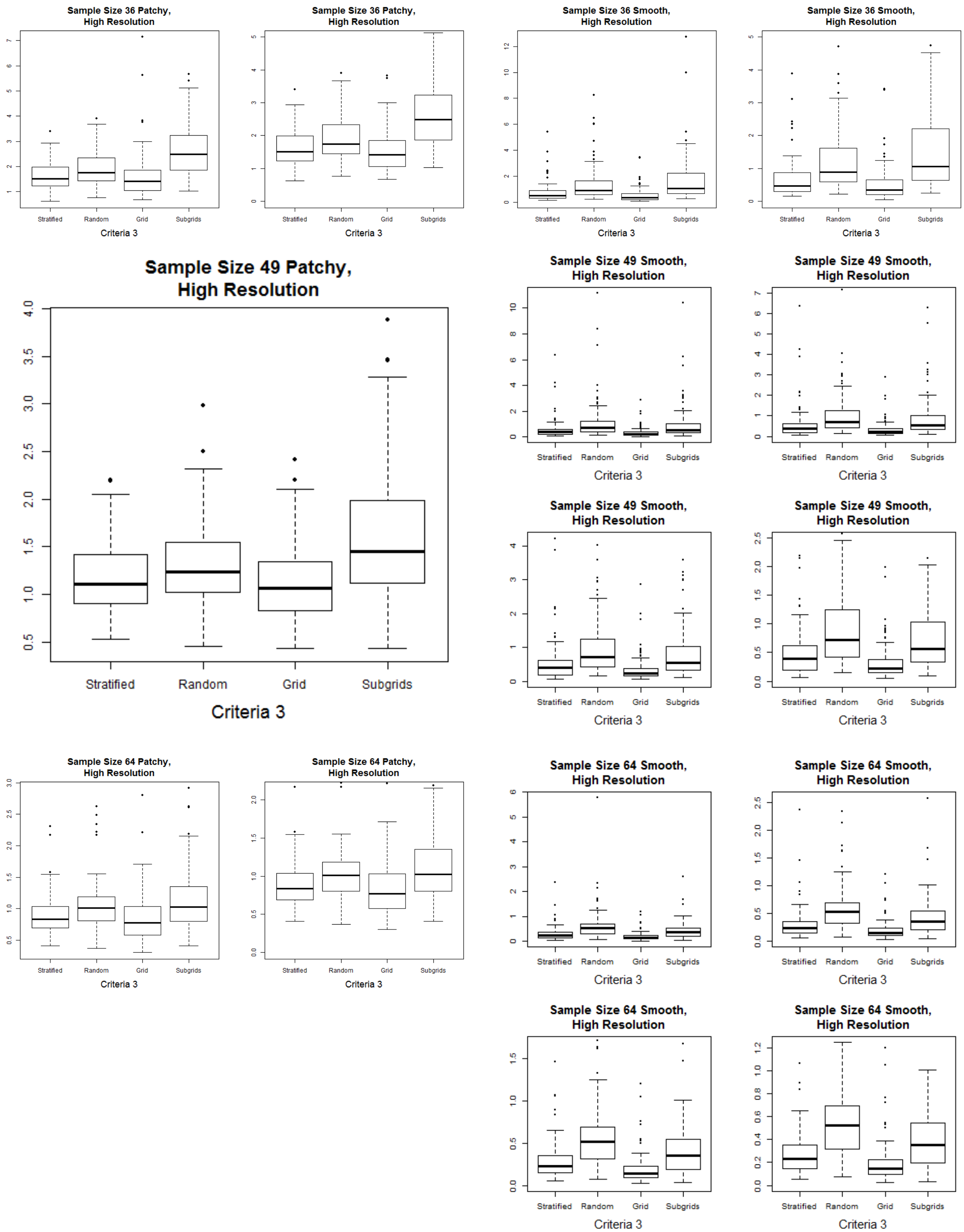


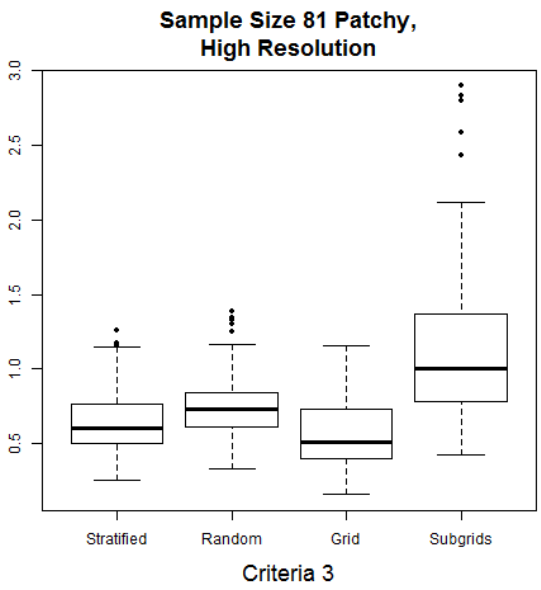


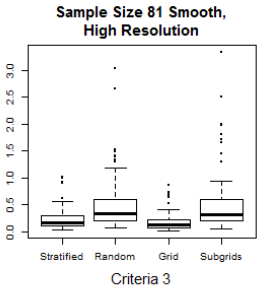


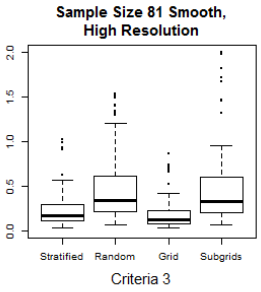


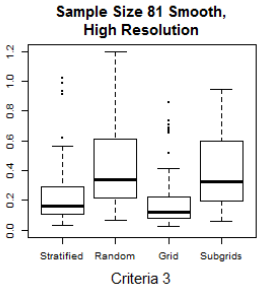


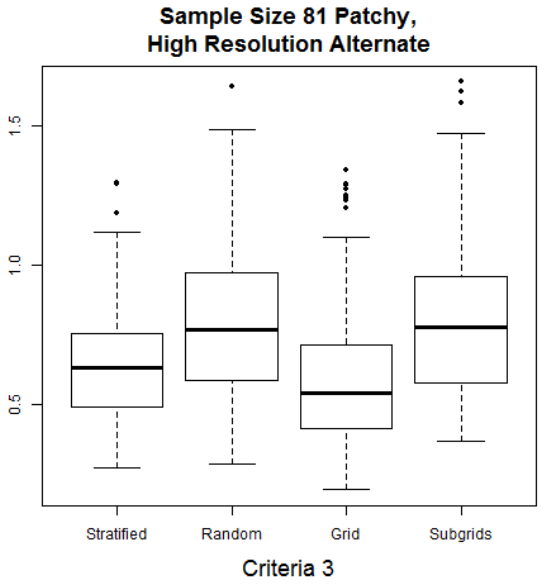

Sample Size 36 Patchy, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 36 Patchy, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 36 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 36 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 49 Patchy, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 49 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 49 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 49 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 49 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 64 Patchy, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 64 Patchy, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 64 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 64 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 64 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3
Sample Size 64 Smooth, High Resolution
Stratified Random Grid Subgrids
Criteria 3

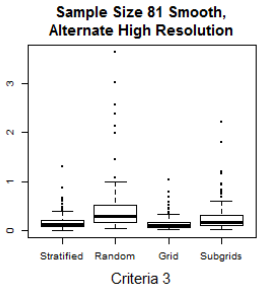

Sample Size 81 Patchy,
High Resolution
Stratified
Random
Grid
Subgrids
Criteria 3

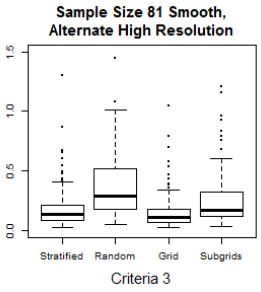

Sample Size 81 Smooth,
High Resolution
Stratified
Random
Grid
Subgrids
Criteria 3

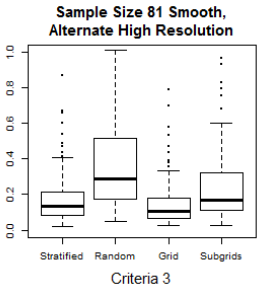

Sample Size 81 Smooth,
High Resolution
Stratified
Random
Grid
Subgrids
Criteria 3


Sample Size 81 Smooth,
High Resolution
Stratified
Random
Grid
Subgrids
Criteria 3


Sample Size 81 Patchy,
High Resolution Alternate
Stratified
Random
Grid
Subgrids
Criteria 3


Sample Size 81 Smooth,
Alternate High Resolution
Stratified
Random
Grid
Subgrids
Criteria 3


Sample Size 81 Smooth,
Alternate High Resolution
Stratified
Random
Grid
Subgrids
Criteria 3


Sample Size 81 Smooth,
Alternate High Resolution
Stratified
Random
Grid
Subgrids
Criteria 3