

A MONTE CARLO SIMULATION OF THE IMPACT OF SAMPLE SIZE AND PERCENTILE METHOD IMPLEMENTATION ON IMAGERY GEOLOCATION ACCURACY ASSESSMENTS

Paul C. Bresnahan, Senior Photogrammetrist

Todd A. Jamison, Chief Scientist

Observera Inc.

3856 Dulles South Court, Suite I

Chantilly, Virginia 20151

pbresnahan@observera.com

tjamison@observera.com

ABSTRACT

Many methods exist to estimate image absolute geolocation accuracy statistics from check-point analyses. Among them is the Percentile Method (PM), which orders the differences between measurements and truth, and estimates the geolocation accuracy statistic at a specific percentile, such as the 90th percentile used for the Circular Error 90% (CE90) statistic. The Percentile Method is elegant because it uses the measured values directly and does not make any assumptions about the population distribution. However, several implementations exist to estimate the value at the desired percentile among the ordered data points. In addition, the number of images available for evaluations is limited, and the precision of a parameter estimate from any method decreases with smaller sample sizes. A Monte Carlo simulation was run using all sample sizes between ten and thirty, inclusive, and using eleven Percentile Method implementations to estimate accuracy values for every decile from 10% to 90%. Sampling distributions were formed for each case using many trials. The difference between the mean of each sampling distribution and the known population mean represents the bias for each estimation method and sample size. The standard deviation of each sampling distribution provides a measure of the variability of the estimation method for each sample size. The results quantify the improvement in precision as the sample size increases. The results also show that some Percentile Method implementations can be heavily biased and that these biases vary with sample size and percentile value.

INTRODUCTION

Commercial imagery providers have established product specifications that state the geolocation accuracy of their products. Consumers are interested in quantifying the geolocation accuracy of the imagery to ensure that it satisfies the intended application and to justify their monetary investment in the purchase of the imagery. Imagery product specifications typically state the accuracy specification in terms of the CE90 statistic for horizontal accuracy and the Linear Error 90% (LE90) statistic for vertical accuracy. Each of these quantities can be defined as the error distance that encompasses 90% of the data points. Alternatively, it is often interpreted to mean that 90% of the time, the geolocation error of points from an image will be less than the stated CE90 distance.

Our simulation is inspired by geolocation accuracy assessments of high-resolution commercial imagery satellites, such as the existing IKONOS, QuickBird, and OrbView-3 satellites (Ager, 2003; Ager 2004b; Dial, 2002b; Dial, 2003) and the future WorldView-1, WorldView-2, and GeoEye-1 satellites. Although commercial imagery providers offer a variety of products, ranging from uncontrolled "Basic" images to ground-controlled orthomosaics (DigitalGlobe, 2006; GeoEye, 2006a and 2006b), this simulation was designed to model the accuracy assessment of uncontrolled Basic images as they are the products most representative of the inherent geolocation accuracy provided by the satellite and its sensor.

Previous geolocation accuracy assessments of uncontrolled, high resolution commercial satellite images have shown that the absolute geolocation error is caused mostly by attitude pointing errors present in the satellite telemetry. Because the sensors also have a narrow field of view, this manifests itself as a nearly constant horizontal translation error when check-points are used to assess the accuracy of a monoscopic image. (Dial 2002a; Dial, 2004) Typically, differences between the errors of individual check-points in an image can be attributed to a combination of pixel measurement error, check-point ground coordinate error, and small relative errors that may be present in the image. In general, these differences are much smaller than the overall translational error. If the

individual errors of the check-points are averaged, then a mean translational error for the image can be estimated with a small dispersion around that mean. (Bethel, 2004) These assumptions and observations should not be applied to other types of sensors without additional analysis and validation.

With the discussion above in mind, there are differing conceptual approaches for designing evaluations for uncontrolled Basic images. In one common approach, individual images are evaluated with the check-point errors used to estimate individual CE90 values for each image. Although such an approach can be effective at estimating individual image CE90 values, it provides no overall CE90 estimate for the population of all uncontrolled Basic images, which is of interest to consumers in order to quantify the certainty that subsequent images purchased will be within the advertised specification. Another approach is to pool all of the check-points from multiple images to estimate an overall CE90. However, the primary weakness of this approach is that the test fields used for evaluations often have varying number of check-points, which results in different weightings for each image in the error assessment. For example, an image with 20 check-points would be weighted twice as much as an image with 10 check-points, even though the dispersions of the check-point errors around the mean translational errors are not expected to vary by a factor of two. A more reasonable approach is to take the mean translational error from a set of multiple check-points distributed across an image as a good estimate of the geolocation error of the image and to use this value as a single data point in the overall evaluation of the uncontrolled Basic imagery product type.

We have based our simulation on this approach. Each value from the parent population represents an individual image's mean translational error and is used as a single data point in the sample population. Although the discussion above pertains to horizontal errors and CE90, our analysis also applies to stereo pairs as the pointing error for each stereo mate results in both horizontal and vertical errors. The simulation addresses horizontal and vertical errors separately.

Numerous statistical methods are available for estimating geolocation accuracy for a population of images from a given sensor. Parametric methods assume a statistical model for the parent probability distribution, the parameters of which are estimated from a sample population; whereas, non-parametric methods make no assumptions about the parent population, and the statistical model is estimated directly from the sample population, without assumptions. (Ager, 2004a; Wikipedia, 2007b) The Percentile Method (PM), one such non-parametric method also known as the Rank-order Method, is the basis for many common geolocation error metrics, including CE90, LE90 and Circular Error Probable (CEP) / CE50.

In the PM, the data points from the sample population are rank-ordered from smallest to largest. Then, a value that represents the desired statistical percentile or quantile (Wikipedia, 2007c) of the parent population is estimated. One common, somewhat intuitive approach is to take the 9th ordered value out of 10 or the 18th ordered value out of 20 and so on as being representative of the 90th percentile. Our results using Monte Carlo simulation demonstrate that this common approach is actually a biased estimator that significantly underestimates the 90th percentile. Using our simulation, we compared eleven PM estimation algorithms as to their ability to estimate geolocation error percentiles for every decile between 10% and 90% for both monoscopic images and stereo pairs.

We are particularly interested in the quality of estimators when the sample population is small, since there is very often a practical constraint on the number of independent images that can be included in the assessment of a satellite imaging sensor. Consumers typically have limited budgets to purchase test images, and competition for collection, orbits, and weather set practical limits on the collection time frame for a test. Therefore, such assessments often involve a statistically significant, but small number of images. Hence, we used our simulation to vary the number of images from 10 to 30 in order to consider a range of sample sizes that is typical in satellite imaging assessments. The lower end of this range clearly falls in the "statistically small sample" realm for parametric statistical analysis.

While focused on the imagery geolocation accuracy problem, the results of our analysis are also generally applicable to other similar measurement problems, particularly as they apply to the selection of robust PM estimators.

SETUP OF MONTE CARLO SIMULATION

For the Monte Carlo simulation, a statistical model was chosen to simulate the distribution of many Basic, uncontrolled commercial satellite images. The pointing error of the satellite was assumed to be the dominant factor for the error on the ground, and it was assumed that there are no significant systematic influences on the pointing error of the satellite and sensor system – i.e., that each sample is statistically independent. Each data point in the simulation represented the average error of one image, as would typically be determined via the measurement of multiple check-points. As such, a Gaussian distribution was used to simulate errors in the X, Y, and Z directions,

with no correlation among them. The simulation combines the X and Y directional errors into a horizontal radial error in order to generate the horizontal statistics. The resulting horizontal radial error distribution is a Rayleigh distribution (Hsu, 1999). The vertical directional error is treated separately from the horizontal and is Gaussian in nature. The absolute values of the vertical errors were used in the simulation, since only the error magnitude is relevant to the vertical statistics. Figure 1 shows the horizontal and vertical parent population distributions generated for the simulation.

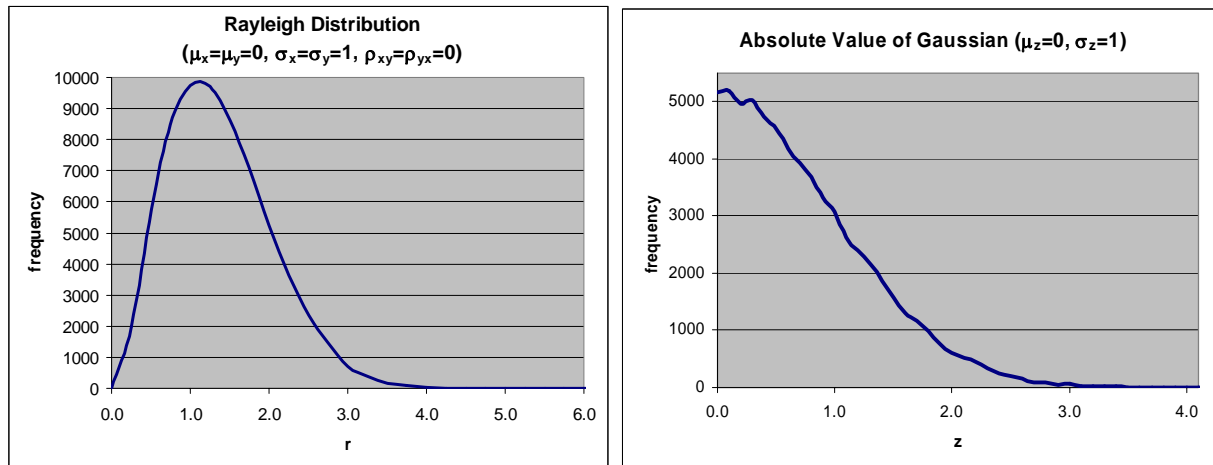


Figure 1. Simulated Population Distributions.

A nominal standard deviation of unity was used for the X, Y, and Z Gaussian distributions. Thus, the theoretical CE90 is the value 2.146 for the horizontal radial dimension, and the theoretical LE90 is 1.6449 for the vertical direction. Actual commercial satellite imagery CE90 and LE90 values vary by satellite and sensor, and the results from this simulation can be scaled to actual cases as appropriate.

The simulation was generated using Microsoft® Excel. The points for the simulated horizontal and vertical parent population were created using the maximum 65,536 rows in Excel. The simulation was run with sample sizes of every integer value between 10 and 30. For each sample size, 20,000 trials were run in which a sample population of size n was randomly selected from the rows containing the parent population. Each sample population was rank-ordered. Percentile values were estimated for each trial using eleven different PM estimators. The decile values were 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. The latter is synonymous with the CE90 and LE90 statistics. The estimates of CE90 and LE90 for each trial represent values that could result from an actual geolocation accuracy evaluation either for the nominal case or scaled for an actual sensor. Overall statistics were computed for each set of 20,000 trial values for each case of sample size, percentile value, and dimension (i.e., horizontal and vertical). These statistics were mean, standard deviation, minimum, maximum, and delta from the simulated population value. The sampling distributions for each percentile statistic were plotted for each PM estimator, dimension, and sample population size, n . As an example, the sampling distributions and statistics for a case using one of the PM estimators are shown in Figure 2 and Table 1, respectively.

Table 1 shows that the simulated population values for each percentile are very close to the theoretical values. The row titled "Method Mean" shows the means of the 20,000 trials for each percentile for this case. Each value is the mean of the respective sampling distribution shown in Figure 2. If, as in this particular case, there are small differences between the Method Mean values and Simulated Population Values, then the method is considered an unbiased estimator. The Standard Deviation Values are indicators of the variability of the estimator for each percentile, which we define as the precision of the estimator.

Table 1. Example Case Statistics

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Theoretical Value	0.459	0.668	0.845	1.011	1.177	1.354	1.552	1.794	2.146
Sim. Pop. Value	0.459	0.665	0.840	1.006	1.174	1.353	1.550	1.793	2.147
Method Mean	0.405	0.626	0.807	0.977	1.146	1.324	1.523	1.765	2.123
Difference	-0.054	-0.039	-0.033	-0.029	-0.028	-0.029	-0.027	-0.028	-0.023
Std. Dev.	0.202	0.217	0.226	0.237	0.250	0.265	0.284	0.313	0.375
Max	1.387	1.628	2.028	2.032	2.250	2.439	2.926	3.060	4.071
Min	0.007	0.087	0.154	0.268	0.304	0.406	0.625	0.719	0.855

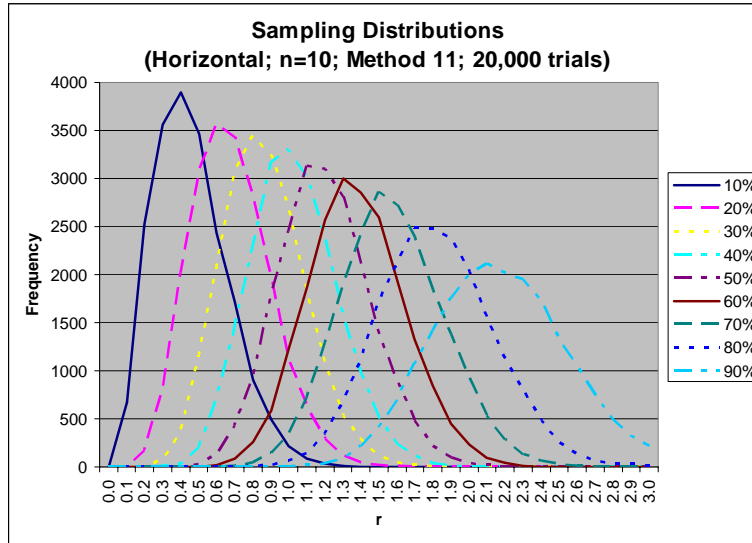


Figure 2. Example Case of Sampling Distributions.

Table 2 defines the eleven PM estimator formulations. The formulations come from four different sources as listed in the table. (xycoon.com, 2006; Mulawa, 2006a and 2006b) The Percentile Value (PV) is computed using the integer part, i , and fractional part, f , of a function defined for each method. The functions are defined using the sample size, n , and the percentile, p , in the range 0.0 to 1.0. The value $x_{(i)}$ is the i th order statistic from the sample population.

For example, given the following rank-ordered sample population with $n=13$:

{0.08, 0.09, 0.15, 0.35, 0.39, 0.45, 0.70, 0.72, 0.89, 1.00, 1.33, 1.97, 2.29}

The 50th percentile gives $p = 0.5$. Thus, for Method 1, $np = 13 \times 0.50 = 6.5$, and $i = 6$, and $f = 0.5$.

So, $PV = (1-0.5) \times 0.45 + 0.5 \times 0.70 = 0.575$.

Table 2. Percentile Method Estimator Formulations

Method	Name	Function	Formula	Source
1	Weighted Average at x_{np}	$np = i + f$	$PV = (1-f)x_{(i)} + f \cdot x_{(i+1)}$	xycoon.com
2	Weighted Average at $x_{(n+1)p}$	$(n+1)p = i + f$	$PV = (1-f)x_{(i)} + f \cdot x_{(i+1)}$	xycoon.com
3	Empirical Distribution Function	$np = i + f$	If $f = 0$: $PV = x_{(i)}$; If $f > 0$: $PV = x_{(i+1)}$	xycoon.com
4	Empirical Distribution Function - Averaging	$np = i + f$	If $f = 0$: $PV = \frac{1}{2}(x_{(i)} + x_{(i+1)})$; If $f > 0$: $PV = x_{(i+1)}$	xycoon.com
5	Empirical Distribution Function - Interpolation	$(n-1)p = i + f$	If $f = 0$: $PV = x_{(i+1)}$; If $f > 0$: $PV = x_{(i+1)} + f(x_{(i+2)} - x_{(i+1)})$	xycoon.com
6	Closest Observation	$(np + \frac{1}{2}) = i + f$	$PV = x_{(i)}$	xycoon.com
7	TrueBasic – Statistics Graphics Toolkit	$(n+1)p = i + f$	If $f = 0$: $PV = x_{(i)}$; If $f > 0$: $PV = f \cdot x_{(i)} + (1-f)x_{(i+1)}$	xycoon.com
8	Older Version of Microsoft Excel	$(n+1)p = i + f$	If $f < 0.5$: $PV = x_{(i)}$; If $f = 0.5$: $PV = \frac{1}{2}(x_{(i)} + x_{(i+1)})$; If $f > 0.5$: $PV = x_{(i+1)}$	xycoon.com
9	Microsoft® Excel 2003 Percentage Function		Built in to Excel	Microsoft® Excel
10	Weighted Average at $x_{(np+1/2)}$	$np + \frac{1}{2} = i + f$	$PV = (1-f)x_{(i)} + f \cdot x_{(i+1)}$	Mulawa
11	Weighted Average at $x_{(n+1/2)p}$	$(n+\frac{1}{2})p = i + f$	$PV = (1-f)x_{(i)} + f \cdot x_{(i+1)}$	Authors

RESULTS AND ANALYSIS OF MONTE CARLO SIMULATION

The variables of the Monte Carlo simulation are the eleven PM implementations, the sample sizes from 10 to 30, the deciles from 10% to 90%, and the horizontal and vertical dimensions. Each case was analyzed as to the performance of the PM implementation across the ranges of sample sizes and deciles. Particular attention is given to

the amount of bias in each estimator and variability of the bias over the ranges of sample sizes and deciles. In addition, the precision of the PM estimators is analyzed across the ranges of sample sizes and deciles.

Closeness of Estimators to Parent Population Values

The methods were compared to determine their degree of bias and the consistency of their bias across the ranges of both sample sizes and percentiles. We found that the methods could be categorized into five groups as follows: (1) Method 1 stood alone as a straightforward, but highly-biased method, (2) other highly-biased methods, (3) erratic methods, (4) erratic methods useful for median estimation only, and (5) generally consistent, slightly biased methods. The term “erratic” is used here to describe estimation methods for which the bias varies greatly with both percentile and sample size.

Table 3 lists our categories and the methods in each with some general qualitative characteristics for each method based on the analysis of results of the simulation. If the method produces a nearly unbiased estimate near a specific percentile, it is listed in an “Unbiased %” column in the table. The characteristics at percentiles above and below the unbiased percentile are described as underestimating, overestimating, or being erratic. One or more methods are nearly unbiased at percentiles of 10%, 40%, 50%, 55%, 60%, 70%, and 75%. However, of these percentiles only the 50th percentile is commonly used as a statistical metric for geolocation error, namely being the median or CEP.

Table 3. Qualitative Summary of Bias Characteristics for Percentile Method Implementations

Category	Method	Horizontal			Vertical		
		%’s < Unbiased %	Unbiased %	%’s > Unbiased %	%’s < Unbiased %	Unbiased %	%’s > Unbiased %
1	1	Always Underestimates			N/A	10%	Underestimates
2	2	Underestimates	40%	Overestimates	N/A	10%	Overestimates
	5	Overestimates	55%	Underestimates	Overestimates	60%	Underestimates
	9	Identical to Method 5					
3	3	Erratic			Erratic		
	6						
4	4	Erratic	50%	Erratic	Erratic	50% (slightly biased)	Erratic
	7						
	8						
5	10	Slight Overestimate	70%	Slight Underestimate	Slight Overestimate	75%	Slight Underestimate
	11	Slight Underestimate			N/A	10%	Slight Underestimate

Figures 3 through 7 show plots of estimator bias for representative cases selected from Table 3. Each figure shows the bias results for one estimator over the range of deciles, with each sample size, $n = 10$ to 30 , represented by one curve. Each data point represents the differences between the parent population decile and the mean value of the 20,000 estimations for the deciles from the sample populations. Method 9, the Percentile function in Microsoft[®] Excel 2003, is not depicted and is not referred to in our results because the simulation and subsequent mathematical analysis showed that it is identical to Method 5. Methods 4, 6, and 7 are not depicted because they are categorically similar to Methods 3 and 8, and Method 2 is not depicted because it is categorically similar to Method 5.

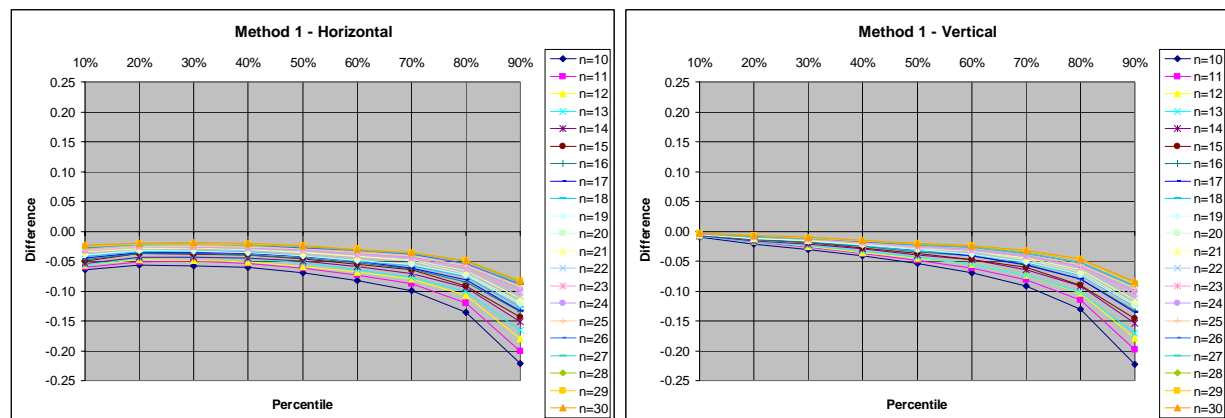


Figure 3. Method 1 (Common but Highly-Biased Approach).

Category 1: Method 1. Figure 3 shows the plots for Method 1, which was placed in its own category because it is perhaps the most common approach to determining a percentile position among rank ordered data points. It underestimates at all deciles for the horizontal dimension. It is interesting to note that, for the vertical dimension, it underestimates the accuracy except at the 10% decile. Furthermore, the amount of underestimation increases with decile and inversely with sample size. The underestimation is worse with smaller sample sizes and only slowly improves as the sample size increases. The slow improvement as sample size increases is shown well beyond $n=30$ at the 90th percentile by Mulawa. (Mulawa, 2006a and 2006b) It is also important to note that the underestimation is largest at the 90th percentile, which is a percentile commonly used for imagery geolocation accuracy evaluations, reaching underestimations of 10% and 13% for CE90 and LE90, respectively. Method 1 may be suitable for applications requiring highly efficient estimation of low-rank deciles ($< 20\%$) from normally distributed populations using large sample sizes; however, for other cases, we believe this method is not appropriate.

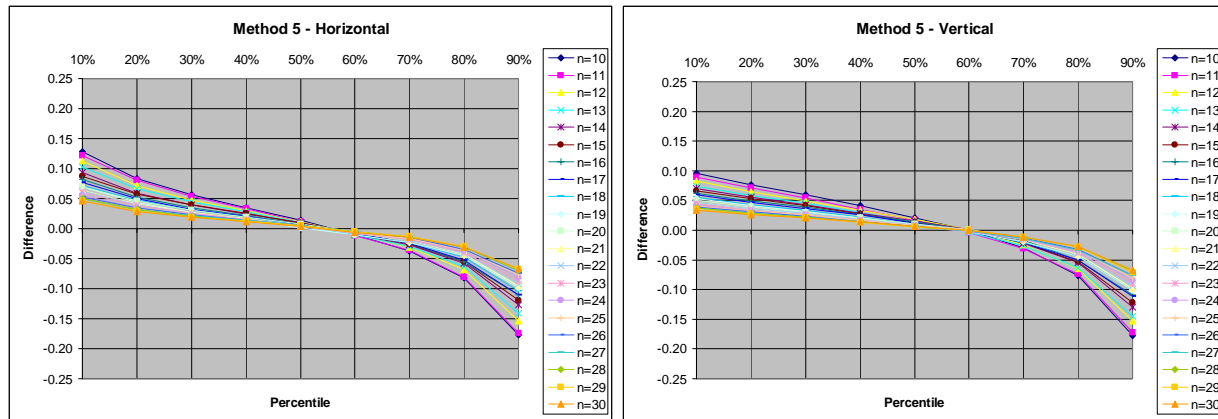


Figure 4. Other Highly-Biased Method.

Category 2: Other Highly-Biased Methods. Figure 4 shows Method 5 which has a smooth, non-erratic trend. It is unbiased around the 55th percentile for the horizontal and the 60th percentile for the vertical, but these percentiles are not generally used for imagery geolocation evaluations. It is clear that Method 5 overestimates at lower percentiles and underestimates at higher percentiles. The estimation is the worst at the smallest sample size, and it only slowly improves as the sample size increases. It is also interesting to note, that as a percentage of value, the overestimates in the lower percentiles can become quite large. For example, the 0.13 overestimate of the 10th percentile represents a 30% bias of that decile. Although not depicted, the trend for Method 2 is opposite to that of Method 5. As indicated in Table 3, Method 2 has a similar but inverted trend for CE90, with underestimates at the lower percentiles and overestimates at the higher percentiles.

The methods in this category may be useful for applications where mid-range deciles (40% - 70%) are estimated using larger sample sizes.

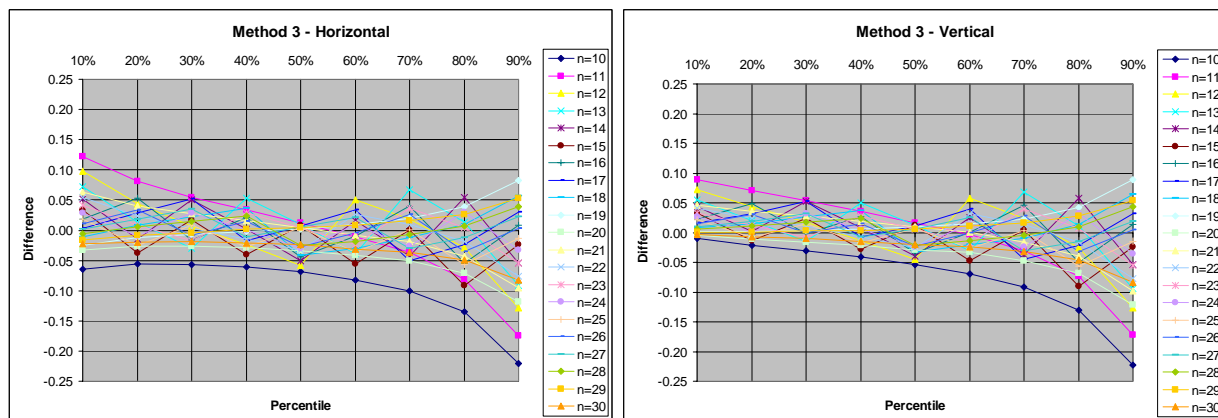


Figure 5. Erratic Method.

Category 3: Erratic Methods. Figure 5 shows Method 3 as an example of an erratic method in which biases occur for nearly every case and there is little consistency across the ranges of sample sizes and percentile values. At some sample sizes, such as at $n=10$, the trend may be a smooth curve. However, at other sample sizes, the curves often jump from overestimation to underestimation at neighboring deciles. This is due to the quantized nature of this class of estimators; wherein a value is selected based upon a rule. We believe that the methods in this category are unsuitable for any application.

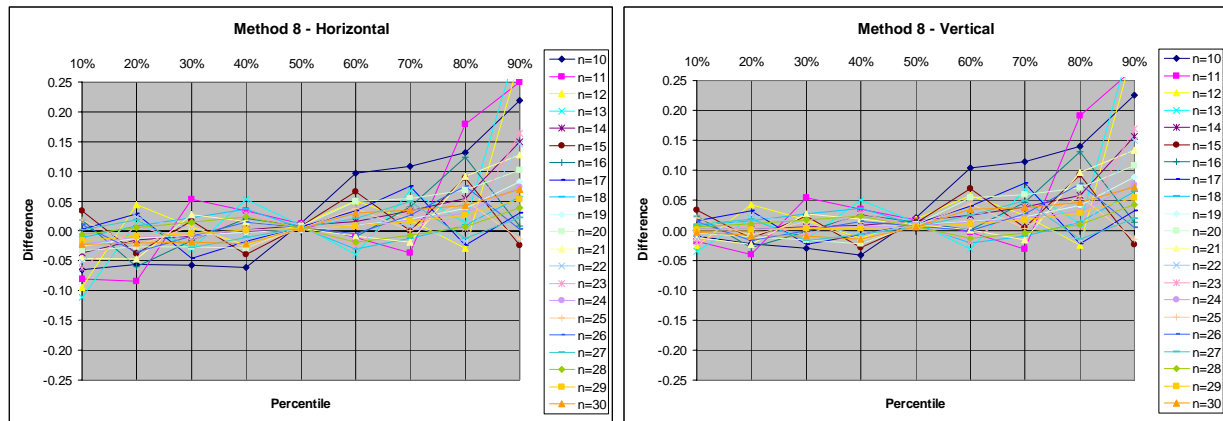


Figure 6. Erratic Method Useful for Median Only.

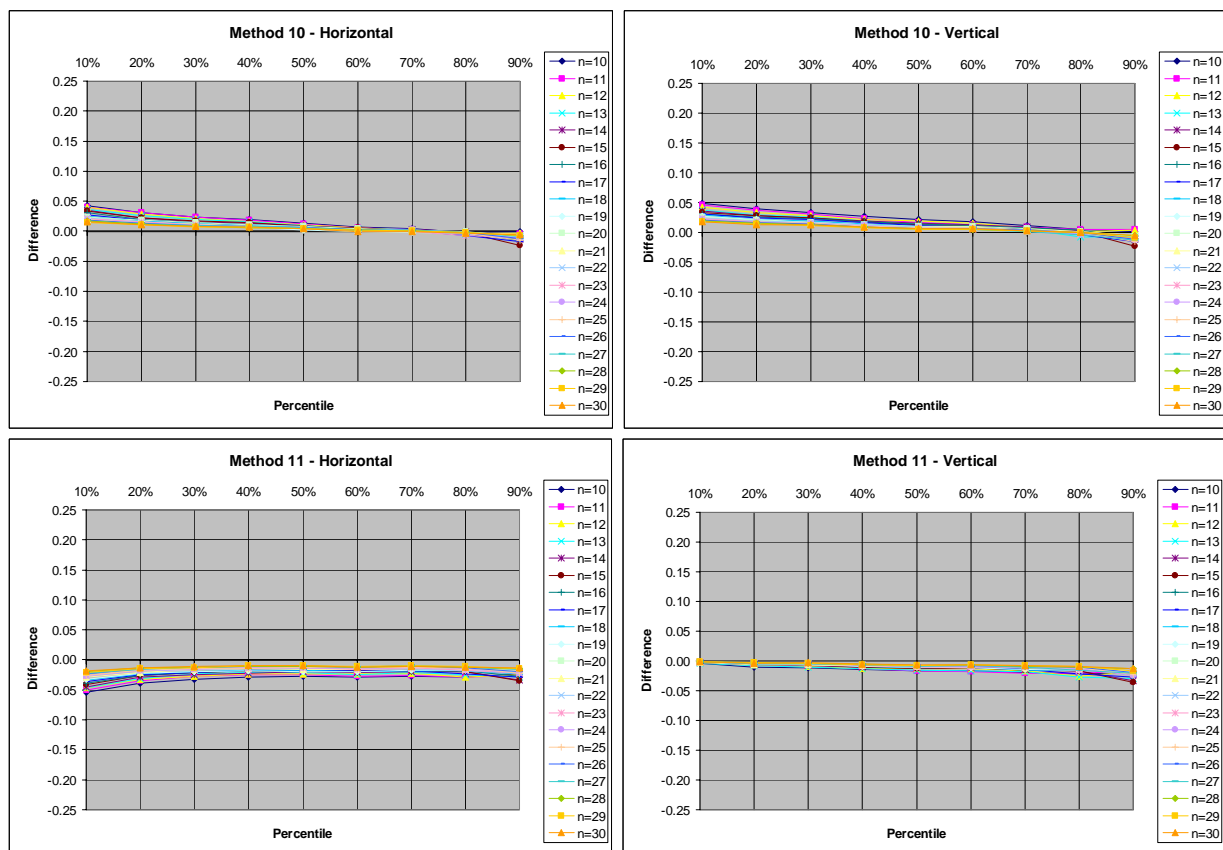


Figure 7. Generally Consistent, Slightly Biased Methods.

Category 4: Erratic Methods Useful for Median Estimation. Figure 6 shows Method 8 as an example of an erratic method in which biases occur everywhere except at 50%, where there is only a small bias. In this category, the 50% value is slightly more biased in the vertical compared to the horizontal.

Methods in this category may be useful only for applications in which a highly efficient estimator for the median statistic is needed.

Category 5: Generally Consistent, Slightly Biased Methods. Of the eleven methods analyzed, Methods 10 and 11, shown in Figure 7, provide the most consistent and unbiased estimates across the ranges of sample sizes and deciles. Method 10 slightly overestimates at lower percentiles for both the horizontal and vertical directions. Method 11 slightly underestimates at lower percentiles for the horizontal direction and is unbiased at 10% for the vertical direction. Although both methods slightly underestimate the 90th percentile, the amount of bias is generally insignificant even when the values are scaled from the nominal values used for the simulation to larger values more commensurate with the accuracy of a particular sensor being evaluated.

Either of these estimators make good candidates for general percentile estimation. They are clearly much better than the other methods, both in minimizing bias and in consistency across deciles and sample sizes. Method 10 has a slight advantage for Rayleigh distributed populations and mid- to high-range percentiles, as its bias trends closer to zero. Method 11 has an advantage for normally distributed populations and is particularly good at low- to mid-range percentiles for that population.

Percentile versus Sample Size

Our analysis also included assessing the variability of results for each decile as a function of sample size. Because the 90th percentile is a primary focus of our geolocation evaluation activities, we will focus on that decile. Similar results were found for the other deciles. Figure 8 shows the estimator biases for the 90th percentile as a function of sample size. The data points are the same as those used in the prior analysis, but presented differently – each curve represents a method and each data point represents a sample size. Figure 8 illustrates similar trends as those seen in Figures 3 through 7. Method 1 underestimates at all sample sizes, with the worst underestimation at the smallest sample size, and it only shows a slow improvement as the sample size increases. Methods 2 and 5 are also similarly biased with Method 5 having a negative bias for all sample sizes and Method 2 having a positive bias. The Erratic Methods (3, 4, 6, 7, and 8) all exhibit discontinuities due to the quantization shifts occurring at specific sample sizes. Methods 10 and 11 are nearly unbiased and are the most consistent estimators across the range of sample sizes. At the 90th percentile, Method 10 is slightly less biased than Method 11.

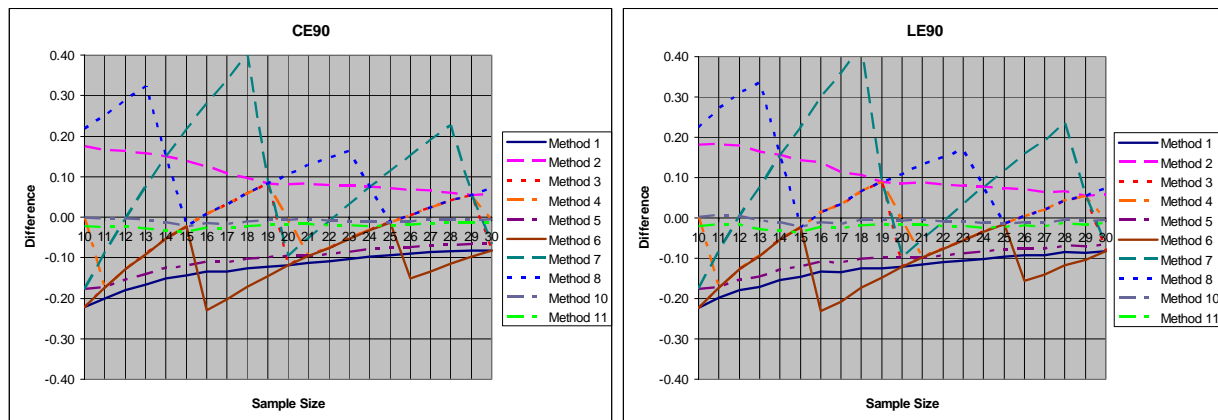


Figure 8. Bias of CE90 and LE90 Percentile Method Estimators By Sample Size.

Precision of Estimators

The standard deviations of the sampling distributions formed by the 20,000 trials for each case are representative of the precision of the PM estimator. Figure 9 charts the Standard Deviation of the estimators as a function of the percentiles, methods, and sample size. In order to reduce clutter on the graphs, Figure 9 shows the horizontal and vertical standard deviations for all methods across the range of deciles only for the sample sizes of 10 and 30. One would expect that precision would increase with both sample size and percentile, and, in fact, these are clear trends. However, the standard deviation values do not vary as much among methods as the sampling distribution means.

Figure 9 also shows the improvement in estimator precision as the sample size increases. At $n=10$, some differences in precision can be observed among the methods. Methods 2 and 8 exhibit less precision than the other methods. It is most pronounced at $n=10$ although it is also noticeable at $n=30$. The remaining methods are more closely clustered at both $n=10$ and $n=30$.

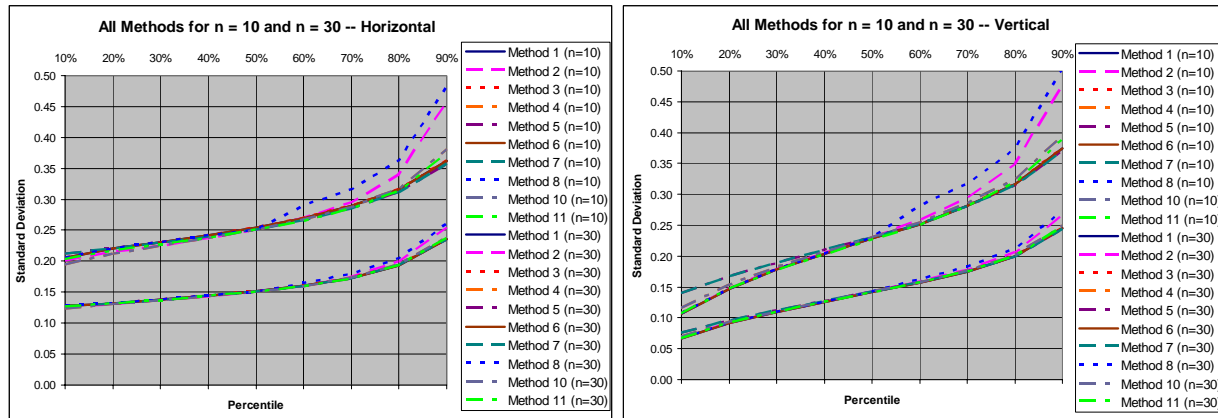


Figure 9. Precision of Percentile Method Implementations for $n=10$ and $n=30$.

Figure 10 shows the standard deviations for the eleven PM estimators across the range of sample sizes for the horizontal and vertical 90th percentiles. Most methods show the expected improvement in precision as the sample size increases, although Methods 7 and 8 exhibit noticeable undulations that are very undesirable. Method 2 exhibits the general trend of improvement with sample size, but is less precise than the other consistent methods. Methods 1 and 5 exhibit the best precision. Methods 10 and 11 are not the most precise estimators, especially at smaller sample sizes; however, the difference in precision between Methods 10 and 11 and other more precise estimators is only slight and the difference is qualitatively outweighed by the superior performance of Methods 10 and 11 for providing nearly unbiased estimates across the range of sample sizes. To put this into perspective, it means that for Methods 10 and 11, there is a slightly wider variation around a more accurate mean, and for the other methods, there is a slightly narrower variation around a less accurate mean.

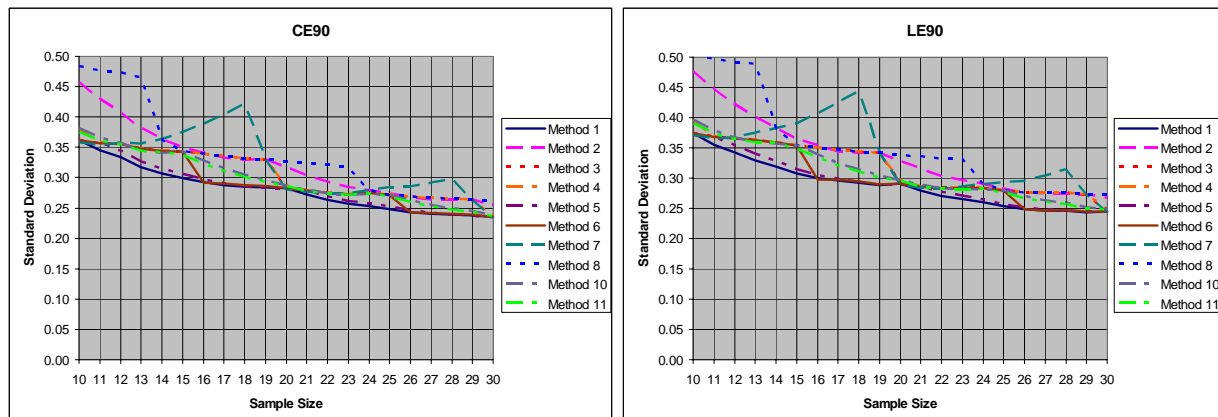


Figure 10. Precision of CE90 and LE90 Percentile Method Implementations.

Visualizing Estimator Characteristics

One way to visualize both estimator characteristics is given in Figure 11. It compares the sampling distributions for five sample sizes at the 90th percentile for Methods 1 and 11. Each sampling distribution curve depicts the shape of the density function of the estimator for the given case. The mean values of each sampling distribution are shown in tables inset into Figure 11, along with the parent population CE90 and LE90 values of 2.15 and 1.64, respectively. The parent population values are also shown on the graph as a vertical line. For the 90th percentile, the sampling distribution for an unbiased estimator should have a mean value that is very close to the simulated CE90 and LE90 values. The standard deviation of the sampling distribution indicates the precision of the estimator – or how much

variability the estimator has around the mean. An estimator with a narrower sampling distribution has a smaller standard deviation and a greater precision, which means that any given estimate is going to be closer to its mean than an estimate from a less precise estimator.

Like Figures 3 and 8, Figure 11 shows that Method 1 is biased and will tend to underestimate CE90 and LE90 at all sample sizes. Furthermore, the improvement is slow as the sample size increases. The increased precision of the estimator is also shown as the sampling distributions get narrower as the sample size increases. Like Figures 7 and 8, Figure 11 shows Method 11 to be an estimator that closely recovers the simulated values, and the estimator has minimal bias at all sample sizes. Like the other cases, the precision of the estimator also improves as the sample size increases. Figure 12 superimposes the $n=15$ sampling distributions for Method 1 and Method 11. It can be clearly seen that although Method 1 has slightly higher precision, the probability of underestimation is significant compared to that of Method 11.

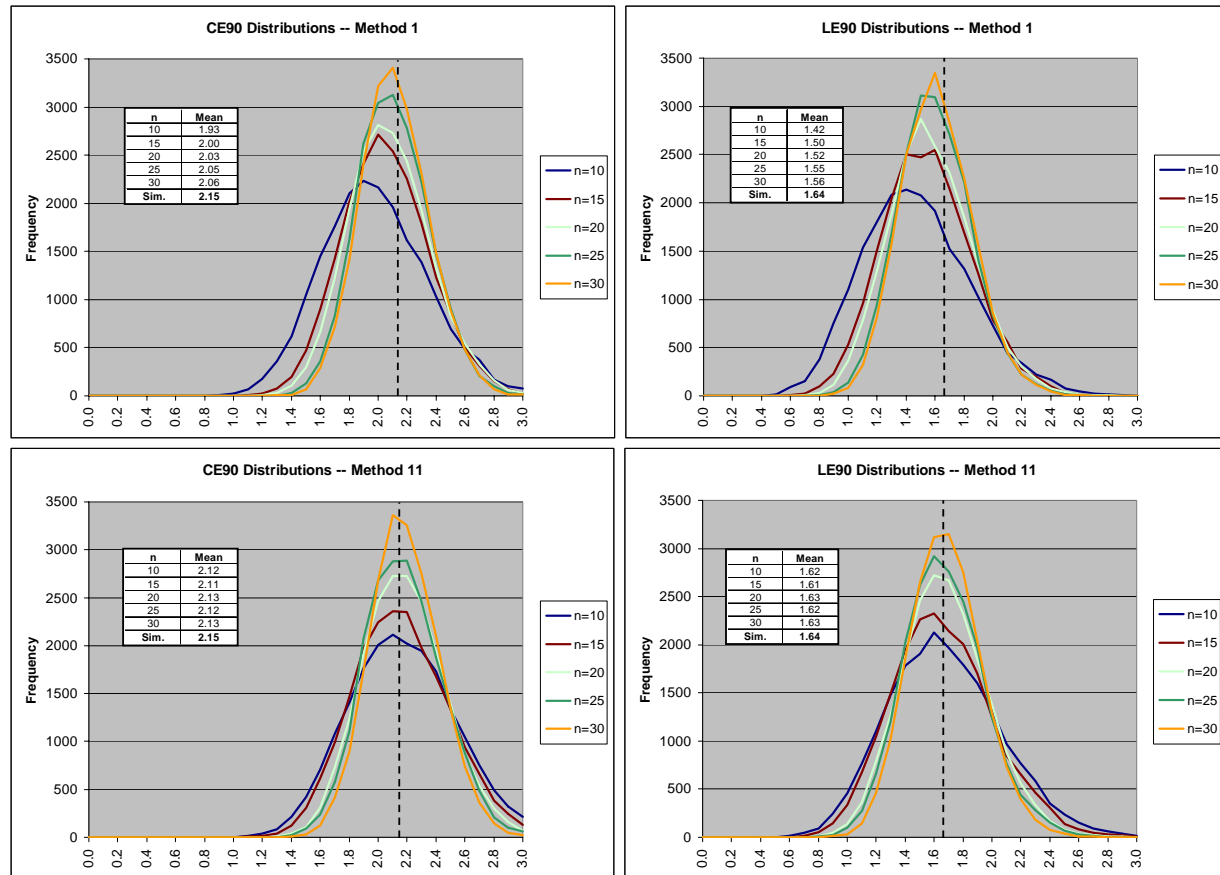


Figure 11. Improvement in Percentile Method Estimator Precision as Sample Size Increases.

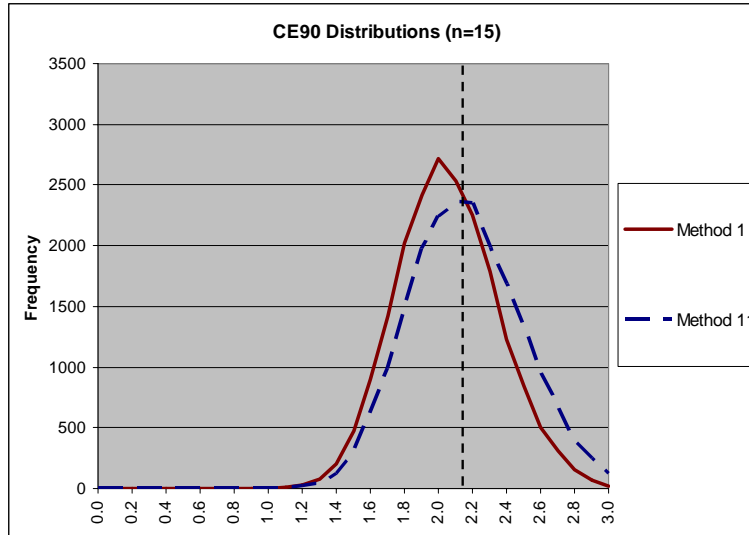


Figure 12. Qualitative Advantage of Method 11 over Method 1.

SUMMARY

The purpose of the work reported on here was to characterize different methods for estimating rank-order statistics from data sets of small- to medium-sized samples. Specifically, the application was to determine which methods are appropriate for use in estimating satellite image geolocation accuracy, both horizontal and vertical, from sample sizes ranging from ten to thirty images. Horizontal errors are Rayleigh distributed and vertical errors are single-tailed normal. For this analysis, a Monte Carlo simulation was created that generated a parent population of over 65,000 simulated horizontal and vertical geolocation errors. Sample populations of all sizes between ten and thirty were taken from the parent population and accuracy statistics for every decile between 10% and 90% were estimated using eleven different Percentile Method formulas. Sampling distributions were formed for each case using 20,000 trials. The difference between the mean of each sampling distribution and the known population value was computed as the estimator bias for each case. The standard deviation of each sampling distribution was also computed as a measure of the precision of the estimator for each case.

One of the estimators was found to be identical to another, so it was eliminated from further analysis. Eight of the remaining ten estimators exhibited significant bias characteristics. For three methods, the bias varies smoothly as the percentile and sample size change. For five methods, the variation is erratic with no consistency across the ranges of percentiles and sample sizes. Only two methods, Methods 10 and 11, exhibited nearly unbiased estimates across the ranges of percentiles and sample sizes.

The estimators for all methods exhibit an overall trend toward improvement in precision as the sample size increases, which was expected; however, the precision of four of the estimators did not consistently decrease with sample size, which raises concerns about their reliability. These methods were among those that exhibited erratic bias behavior, as well.

All of the methods that exhibited erratic bias behavior have limited applicability in small- and medium-sample size applications and should not be used, except under special circumstances (e.g., low accuracy, high performance applications). Several of the Methods are good estimators at specific percentiles and may have applicability to specific classes of problem. Because Methods 10 and 11 are nearly-unbiased estimators across the ranges of percentiles, samples sizes, and for both Rayleigh- and Normally-distributed parent populations, and their precision is comparable to the best of the other methods, we recommend either of these for general rank-order statistic estimation problems, especially for those with small sample sizes. Hence, we conclude that both Methods 10 and 11 are appropriate for imagery geolocation accuracy evaluations similar to those modeled by this simulation.

REFERENCES

- Ager, T. (2003). *Evaluation of the geometric accuracy of Ikonos imagery*. Proc. SPIE, Vol. 5093, pp 613-620.
- Ager, T. (2004a). *An analysis of metric accuracy definitions and methods of computation*. NIMA InnoVision whitepaper, 29 March 2004.
- Ager, T., T. A. Jamison, P. C. Bresnahan, and P. Cyphert. (2004b). *Geopositional accuracy evaluation of QuickBird ortho-ready standard 2A multispectral imagery*. Proc. SPIE, Vol. 5425, pp 488-499.
- Bethel, J. (2004). *Evaluation of projection errors using commercial satellite imagery*. 2004 Joint Remote Sensing Seminar, 15 September 2004. http://www.lars.purdue.edu/seminar/presentations/2004_seminar3.pdf (accessed 25 January 2007).
- Bresnahan, P. C. (2006). *Geolocation Accuracy Evaluations of OrbView-3, EROS-A, and SPOT-5 Imagery*. Proceedings of Civil Commercial Imagery Evaluation Workshop, 14-16 March 2006, Laurel, Maryland.
- Bresnahan, P. C. (2004). *Geopositional Accuracy Evaluations of QuickBird and OrbView-3*. Proceedings of High Spatial Resolution Commercial Imagery Workshop, 8-10 November 2004, Reston, Virginia.
- Conover, W. (1999). *Practical Nonparametric Statistics*, 3rd Edition, John Wiley & Sons, New York.
- Dial, G., and J. Grodecki (2002a). *Block adjustment with rational polynomial camera models*. Proceedings of ASPRS 2002 Conference, Washington, DC, April 22-26, 2002.
- Dial, G., and J. Grodecki (2002b). *IKONOS accuracy without ground control*. Proceedings of ISPRS Commission I Mid-Term Symposium, Denver, Colorado, November 10-15, 2002.
- Dial, G., and J. Grodecki (2003). *IKONOS stereo accuracy without ground control*. Proceedings of ASPRS 2003 Conference, Anchorage, Alaska, May 5-9, 2003.
- Dial, G., and J. Grodecki (2004). *Satellite image block adjustment simulations with physical and RPC camera models*. Proceedings of ASPRS 2004 Conference, Denver, Colorado, May 23-28, 2004.
- DigitalGlobe (2006). *QuickBird Imagery Products – Product Guide*. Revision 4.7.2, 18 October 2006.
- GeoEye (2006a). *IKONOS Imagery Products Guide*. Version 1.5, January 2006.
- GeoEye (2006b). *OrbView-3 Commercial Satellite Imagery Product Catalog*. Version 1.0, January 2006.
- Hsu, D. (1999). *Spatial Error Analysis: A Unified Application-Oriented Treatment*. IEEE Press, New York, p. 41.
- Mulawa, D. (2006a). *CE sampling statistic*, MathCAD document, June 2006.
- Mulawa, D. (2006b). *LE sampling statistic*, MathCAD document, June 2006.
- Ross, K. (2004). *Geopositional Statistical Methods*. Proceedings of High Spatial Resolution Commercial Imagery Workshop, 8-10 November 2004, Reston, Virginia.
- Wikipedia (2007a). Parametric Statistics. http://en.wikipedia.org/wiki/Parametric_statistics. (accessed 25 January 2007).
- Wikipedia (2007b). Non-Parametric Statistics. http://en.wikipedia.org/wiki/Non-parametric_statistics. (accessed 25 January 2007).
- Wikipedia (2007c). Quantile. <http://en.wikipedia.org/wiki/Quantile>. (accessed 31 January 2007).
- Xycoon (2006). *Descriptive Statistics*. <http://www.xycoon.com>. (accessed 2 May 2006).