2008

Development of digital elevation models (DEMs) for agricultural applications

Samsuzana Abd Aziz
Iowa State University

Development of digital elevation models (DEMs) for agricultural applications

by

Samsuzana Abd Aziz

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Agricultural Engineering

Program of Study Committee:
Brian L. Steward, Major Professor
Stuart J. Birrell
Lie Tang
Amy L. Kaleita
Chris Harding

Iowa State University

Ames, Iowa

2008

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this dissertation. First and foremost, Dr. Brian L. Steward for his guidance, patience and support throughout this research and the writing o f this dissertation. His insights and words of encouragement have often inspired me and renewed my hopes for completing my graduate education. I would also like to thank my committee members for their efforts and contributions to this work: Dr. Stuart Birrell, Dr. Lie Tang, Dr. Amy Kaleita and Dr. Chris Harding. I would additionally like to thank graduate assistant Manoj Karkee for his inputs in completing this dissertation. Finally, my husband Fakhrul Zaman Rokhani and my family have provided immeasurable love and encouragement through my studies at Iowa State University.

CHAPTER 1. GENERAL INTRODUCTION

Introduction

Terrain modeling is one of the prime approaches that can be used to assess the spatial variability of agricultural fields and their surrounding ecosystems (Franklin et al., 2000; Mackey et al., 2000; Fried et al., 2000). The representation of terrain in the form of digital elevation models (DEMs) can be used to help the implementation of the applications of precision conservation management practices. For example, terrain analysis models can use DEM-based topography to identify runoff-contributing areas and calculate slopes for use in field-runoff and buffer-filtration models (Dosskey et al. 2005). DEMs also were used to calculate topographic factor for soil erosion (Renschler et al., 2002; Renschler and Flanagan, 2008) and pollution predictions in watersheds (Binger and Theurer, 2001). Similarly, Murthy et al. (2004) used DEMs to derive the path taken by surface runoff at each DEM grid to develop rainfall-runoff models to provide data for un-gauged catchments, and for scenario analysis. They demonstrated that, with effective utilization of spatial information technologies such as geographic information systems (GIS), the understanding of the spatial distribution of run-off derived from DEMs can be used to effectively conserve soil and water in their study area.

Given the important role of DEMs in precision conservation, the availability of sufficiently-accurate topographic data, however, can be a particularly important and challenging problem to address. Generally, raw elevation data or field surveys and the equipment necessary to process these data are not readily available to potential users of a DEM. Commonly available data sources such as 30-m digital elevation models from

government agencies such as United State Geological Survey (USGS) provide a starting point, but more accurate data can yield much better results for a broader range of applications. Currently, aircraft−or satellite−based remote sensing techniques such as photogrammetry, synthetic aperture radar (SAR; Evans and Apel, 1995), and light detection and ranging (LiDAR) are used more often for topographic development. Although new remote sensing technologies like LiDAR could provide high accuracy measurements, aerial survey techniques are cost-effective only over large areas. In addition, remotely-sensed data in raw format not only contains geometrical distortion requiring complicated georeferencing correction processes, but such massive data set will also require substantial computational resources and processing time (Agarwal et al., 2006; Kim et al., 2006, Maune, 2007).

The publicly-available Global Positioning System (GPS) provides new and affordable opportunities for not only researchers but also agricultural producers to gather elevation data to develop agricultural field DEMs. The widespread advancement of GPS technology such as real-time kinematic differential GPS (RTK-DGPS), which offers higher accuracy of topographic measurements (centimeter level position accuracy) makes the effort of using this technology for field DEM generation more reliable (Clark and Lee, 1998; Renschler et al, 2002; Westphalen et al., 2004).

With an increasing proportion of agricultural vehicles equipped with GPS systems for such applications as yield monitoring (Pelletier and Upadhyaya, 1999; Witney et al., 2001; Vellidis et al., 2001) and auto guidance (Will et al, 1998; Billingsley, 2000; Reid et al., 2000), measurement of elevation data during normal farm operations has become more practical for producers. GPS equipped farm vehicles enable farm managers to gather more

accurate elevation measurements which have the potential to be used in addition to or as substitute for commonly available USGS topographic data sources.

The development of a DEM is likely to involve spatial interpolation techniques to estimate values at unsampled locations. It is important to keep in mind that these estimates contain deviation from the ground truth or error. Commonly, higher accuracy data sources were used as the 'true values' to calculate error and quantify the accuracy of DEMs. By comparing with higher accuracy data, measures such as standard deviation or Root mean square error (RMSE) are typically used to represent the DEM quality. However, such non-spatial statistics are global measures and specifically do not provide an accurate assessment of how precise each grid in a DEM represents a true topographical parameters (Hunter and Goodchild, 1995; Wise, 1998; Wechsler, 2007). For example a 10-m US Geological Survey (USGS) DEM consists of 160 thousands elevation points has a stated vertical accuracy (RMSE) of 15 m. This value was computed from an external ground survey of 28 co-located points (representing 0.0175% of the dataset). USGS assures 95% of the data deviates from the surveyed elevation by less than ± 15 m. However, five percent of the values (eight thousands points) could deviate by ±15 m to ±30 m. If for example a farmer would like to use just a small area of the DEM representing a 63 ha field (6241 points from the 10-m USGS DEM that coincide with his agricultural field), the adequacy of the RMSE is questionable.

Despite the fact that the DEMs contain errors, in practice, DEMs are often assumed to be the true representation of elevation. In agricultural applications, errors in DEM propagate through its derived parameters and provide uncertainty in the implementation results (Oksanen and Sarjakoski, 2005). To date, information about DEM uncertainty is not readily

available and often neglected by the DEM users (Weschler, 2007). While the development of more accurate field DEMs are possible with the help of GPS, the uncertainty in DEM estimates must be assessed so that the propagation of these errors can be accounted for in the subsequent studies (Hunter et al., 1995; Holmes et al., 2000; Endreny and Wood, 2001; Wang et al., 2006; Weschler and Kroll, 2006).

The papers contained in this dissertation will investigate procedures of utilizing elevation data for the development of agricultural field DEMs and assessing the DEM uncertainty. Overall, through this work, my study can be useful in a variety of ways. In one case, it could help land users, such as farmers adopting precision agricultural practices, to utilize more accurate topographic information available from typical farm operations to improve their farming applications. Additionally, a better understanding in the uncertainty in elevation estimates could be used for designing more efficient sampling strategies that could benefit agricultural producers. Last of all, this study could help decision makers improve their predictions in environmental management by better understanding the effect of DEM uncertainty on applications.

Objectives

The overall goal of this study was to provide methods for landowners and producers to develop field DEMs to help agricultural applications. In order to accomplish this goal the following objectives were set:

1. Develop a methodology of combining repeated GPS surveys from field operations
2. Develop a targeted sampling method based on spatial uncertainty of prior measurements for topographic mapping.

3. Assess the uncertainty in field DEMs elevation estimates and their effect on soil loss prediction.

Dissertation Organization

The process of developing DEMs is likely to involve estimating elevation at un-sampled locations and assessing the uncertainties of the estimates. Summaries of the theory needed for understanding the implementation of the kriging technique for interpolation, and the geostatistical technique of stochastic simulation for uncertainty assessment used in this work were described in the Chapter 2.

Chapter 3 of this thesis contains a paper entitled *Utilizing Repeated GPS Surveys from Field Operations for Development of Agricultural Field DEMs.* In the paper, the baseline methods for combining repeated elevation surveys of agricultural fields obtained from field operations for DEM development were presented. Kriging interpolation was used to generate field DEMs from different surveys. The methods use the kriging variance as the measure of confidence in the elevation estimates before combining multiple estimates calculated from different surveys. Fuzzy logic, weighted averaging and grid-wise averaging techniques were investigated for combining survey measurements. This paper has been submitted to Transactions of the ASABE for publication and is currently in the review process.

Chapter 4, entitled *Targeted Sampling of Elevation Data Based on Spatial Uncertainty of Prior Measurements* reports on investigations into the design of targeted sampling of field elevation based on the DEMs uncertainties resulting from prior field measurements. The uncertainty and spatial distribution of elevation estimates from prior

measurements was used as a rational basis for a future sampling plan to improve the accuracy of field DEMs. The uncertainty of elevation estimates across the DEMs was assessed using a geostatistical simulation technique to delineate the regions in the field that needed to be re-sampled. Additional samples were targeted and obtained from specified locations rather than re-sampling the whole field which reduced the time required for data collection and resulted in DEMs with relatively low RMSE. The preliminary results of this paper were presented at ASABE Annual International Meeting, 2007 in Providence, RI. This paper will be submitted to Transactions of the ASAE in the near term.

Chapter 5, *Assessing the Effects of DEM Uncertainty on Soil Loss Estimation in an Agricultural Field* is focused on DEM uncertainty and analyzing its effect on soil loss prediction in the field. This study compared the soil loss prediction of an agricultural field using a 7.5-minute USGS DEM, and DEMs developed using RTK-DGPS and dual frequency (DF)-DGPS field surveys. Spatial prediction and DEMs uncertainty analysis was carried out using a sequential Gaussian simulation technique to assess the effects on the predicted soil loss across the study area. This paper will be submitted to Transactions of the ASAE in the near term.

The final chapter of this dissertation is a summary of the work contained in Chapters 3, 4 and 5. This section outlines the findings and general conclusions from the studies and ends with suggestions for future work.

CHAPTER 2. THEORY AND BACKGROUND

Geostatistics is a means to describe spatial patterns and to predict the values of spatial attributes at unsampled locations, where a sample is expected to be affected by its position and relationships with its neighbors. The basic concept in geostatistics is that the spatial correlation between two sample points depends only on the distance of the sample points, not their locations. It is an intuitive concept that locations close to one another to have values more alike than locations that are farther away. The goal of this chapter is to provide a summary of the theory needed for understanding the implementation of the geostatistical methods, realized as part of this work. Detailed descriptions of the theory of geostatistics are given, for example, by Cressie (1989); Isaaks and Srivastava (1989); Journel and Huijbregts (1978); and Goovaerts (1997) and Olea (1999).

The development of a DEM involves spatial interpolation techniques to estimate values at unsampled locations using the elevation measurements taken at surveyed locations. A key feature of topographical data is each observation relates to a particular location in space. Geostatistics is suitable for elevation data because elevation is generally expected to present a high spatial dependence that has a high data similarity at closer locations than distant locations. The semivariogram is a widely used-tool for investigating the spatial continuity of spatial data in geostatistics. Details description about semivariogram analysis is discussed in this chapter.

Mostly, the contents of this research are confined to the application of geostatistics using an interpolation technique known as kriging to predict the value at unsampled location for DEM development. Kriging is a powerful spatial interpolation technique and is widely

used throughout the earth and environmental sciences. The estimation at an unsampled location is given as the linearly weighted sum of its surrounding points. Calculation of the weighting factors is done by minimizing the error variance of a given or assumed model of the spatial continuity for the data with regard to the spatial distribution of the observed data points. Other than kriging, geostatistics also provide a method of determining alternative, but equally probable, realizations of elevation estimates in a DEM. This method is known as conditional stochastic simulation which was used to assess the uncertainty in the DEM estimates. In this chapter, the theory of basic kriging and conditional stochastic simulation methods was discussed.

Spatial Continuity

Geostatistical interpolation estimates elevation values at unsampled locations based on available sample data and a model of spatial continuity of the data. Consider x and $x+h$ to be locations where measurements, Z(x) and Z($x+h$), were taken. The locations $x+h$ and x are separated spatially by a distance, called the lag distance, h. Under first order stationarity, the mean of the sample data is assumed to be constant over the region. A covariance function describes the relation between the samples variance and distance, h as:

$$C(h) = \frac{1}{\mathrm{N}(h)} \sum_{i=1}^{\mathrm{N(h)}} \{Z(x_i) - \mu\}\{Z(x_i + h) - \mu\} \qquad (1)$$

where N(h) is the number of data pairs separated by a distance h, and μ is the mean of the data. The spatial relationship of the samples also can be described using a semivariogram. A semivariogram describes the average dissimilarity between data separated by a distance h as:

$$\gamma(h) = \frac{1}{2\mathrm{N}(h)} \sum_{\mathrm{i}=1}^{\mathrm{N}(h)} \left[Z(x_{\mathrm{i}} + h) - Z(x_i) \right]^2 \qquad (2)$$

where $\gamma(h)$ is called semivariance computed as half the average squared difference between the components of every data pair. The relation between covariance function and semivariogram $\gamma(h)$ is

$$\gamma(h) = C(0) - C(h) \qquad (3)$$

where C(0) is the variance of the data and C(h) is the covariance at distance h. The covariance function requires an assumption of stationarity in the data while semivariance function does not require the stationarity assumption (Metheron, 1973). Although geostatisticans eventually solve the basic kriging equation in terms of covariance, most of initial calculations are done in term of semivariance (Isaaks and Srivastava, 1989).

A covariance function starts at the variance σ^2 of the data and decreases with increasing distance h. Theoretically, a semivariogram starts at zero and increases as distance h increases; however in practice, the semivariogram does not necessarily starts zero where often there is often a discontinuity at the origin of the semivariogram (a non-zero value at lag distance zero) (Fig. 1). This phenomenon is called nugget effect and is due to sampling or measurement error and short scale variability which cause sample values separated by extremely small distance to be dissimilar.

As shown in Fig. 1(a), pairs of locations which are closer have smaller semivariance as compared to pairs of locations which are farther apart. The variance gradually increases until it plateaus at a particular separation distance called the range. The range is the maximum distance at which significant spatial correlation exists between two points. Once the distance between two points is beyond the range, the variance becomes independent of

the distance and maintains a constant value. The maximum semivariance value is called the sill (Fig. 1).

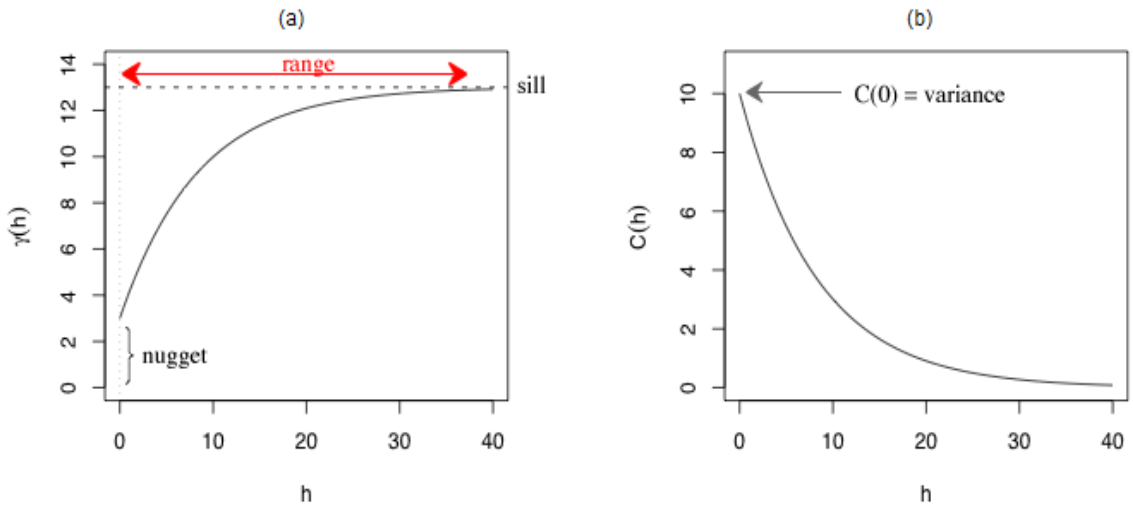

Figure 1: Graph of a typical (a) semivariogram and (b) covariogram.

Usually, data that varies smoothly like elevation are generally expected to present a high spatial dependence with semivariograms that have a shallow slope near zero distance (high data similarity at short distances) (Burrough , 1987; and Valeriano et al., 2006). Topographic data also typically have an increase in semivariance with increasing distances because elevation has the potential to become more and more variable over larger distance due to surficial process (i.e. the semivariogram hardly reach a sill) (Holmes et al, 2000).

The semivariogram developed by calculating the semivariance from the sampled data is called the experimental semivariogram and exists as a discrete and irregularly sampled function. The experimental semivariogram need to be fit with a continuous parameter model for applications. The following are the four most frequently used basic semivariogram models:

Linear model:

$$\gamma(h) = c_0 + ph, \quad if \;\; h < a \tag{4}$$

Exponential model:

$$\gamma(h) = c_0 + c\left[1 - \exp(-3h/a)\right] \tag{5}$$

Spherical model:

$$\gamma(h) = \begin{cases} c\left[1.5(h/a) - 0.5[h/a]^3\right], & if \;\; h \le a \\ c, & if \;\; h \ge a \end{cases} \tag{6}$$

Gaussian model:

$$\gamma(h) = c_0 + c\left[1 - \exp\left(-3h^2/a^2\right)\right] \tag{7}$$

where p is the slope of the linear model, a is the semivariogram range, c_0 the nugget component of the semivariogram, and c is the positive variance contribution or sill.

In terrain analysis, a Gaussian semivariogram model (Fig. 2) is often used to model the spatial continuity in elevation because a Gaussian model presents a region of low slope near the zero distance, which is suitable for data that varies smoothly like elevation.

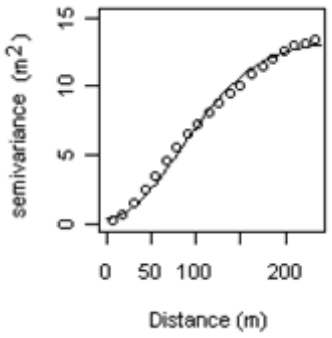

Figure 2: An experimental semivariogram of elevation data from a study field (dotted) fit with a Gaussian model (solid line).

Kriging

The most well-known family of interpolation methods in geostatistics is based on kriging. Kriging is an interpolation method named after a South African mining engineer named D. G. Krige who developed the technique in an attempt to more accurately predict ore reserves. The advantage of kriging over other interpolation methods such as Inverse Distance Weighting (IDW) is that kriging estimates a value at unsampled locations using a minimized estimation variance derived from a semivariogram model, accounting for spatial correlation in the samples (Deutsch and Journel 1998). Over the past several decades kriging has become a fundamental tool in the field of geostatistics.

In general, kriging is a linear interpolation technique, where the estimated value is a weighted average of the surrounding sampled values. The value of variable *Z* at an unmeasured location x_0, $Z^*(x_0)$ is estimated as

$$Z^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i) \qquad (8)$$

where $Z(x_i)$ are the values at neighboring sampled locations and w_i are the weights assigned to these values. The weights w_i are determined to minimize the error variance

$$\sigma_E^2 = Var\left\{Z^*(x_0) - Z(x_0)\right\} \qquad (9)$$

under the constraint of unbiasedness (i.e. mean prediction error is zero):

$$E\left[Z^*(x_0) - Z(x_0)\right] = 0 \qquad (10)$$

Generally, the kriging weights depend on the semivariance model, which is derived from the available data. As expected, the kriging weights decrease as the sample locations get farther from the estimated location. However, for a particular data configuration or sampling

pattern, the changes may drastically vary, depending on the shape of the semivariogram model. Under an undesirable condition, when a semivariogram entails a complete lack of spatial correlation (pure *nugget effect* or very small range), the weights all become 1/n and the estimation procedure becomes the mean of the available data samples. The aspect of kriging weights and its relation to semivariogram model is thoroughly discussed in Goovaerts (1997) and Isaaks and Srivastava (1989).

The values at the neighboring sampled location used for estimation are restricted by the search neighborhood. Intuitively, maximizing the radius of the search neighborhood, hence taking as many sample values as possible would lead to the most accurate solution. In practice, however, limiting the search neighborhood into a smaller radius is necessary because:

- For large distances, the spatial correlation between data as defined by the semivariogram function is usually unreliable because of the low number of data pairs available for inference at such large distance. Moreover, with increasing distance, the spatial correlation is decreasing. Therefore the associated kriging weights are very small and hence do not necessarily improve the estimation.
- Restricting the search neighborhood allows one to account for local departure from stationarity over the area and the estimation is therefore more representative.
- The size of the matrix in computing the kriging weights increases with the number of points in the search neighborhood. The computation time drastically increase with the number of data retained, approximately in proportion to $[N(h)]^3$.

Basic types of kriging

There are a number of different versions of kriging. The most important ones are: simple kriging, ordinary kriging, cokriging and universal kriging. In cokriging, the estimation of a variable is not only based on its own auto-covariance function, but also on its spatial relationship to another variable or parameter. This can be useful if a variable is sparsely sampled but has a similar spatial relationship as extensively sampled variable. Another common kriging procedure is universal kriging, in which the sample data are assumed not to be stationary, but to follow a trend. The drawback of universal kriging is the need to specify the model of the trend, in which there is no direct statistical test to guide the analysis (Olea, 1999). The most common kriging versions are simple kriging and ordinary kriging, which are both discussed in some detail below and have been implemented in this study.

Simple Kriging

Simple kriging estimates an un-sampled value Z at location x_0, as a linear combination of n sampled values $Z(x_i)$ plus a regional mean m which considered to be constant throughout the estimation region:

$$Z_S^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i) + \left[1 - \sum_{i=1}^{n} w_i\right] m \qquad (11)$$

where w_i is the kriging weight at location x_i with the sampled value $Z(x_i)$. The simple kriging weights w_i can be solved in term of data covariances so that:

$$\sum_{i=1}^{n} w_i C(x_i - x_j) = C(x_i - x_0) \qquad j = 1,...,n. \qquad (12)$$

where $C(x_i - x_0)$ is the data-to-estimation covariance and $C(x_i - x_j)$ is the data-to-data covariance. As the location x_0 being estimated gets further away from sample locations, the data-to-estimation covariances decrease and the data-to-data covariances remain unchanged. Consequently the simple kriging weights w_i tend to decrease, hence the estimates $Z_S^*(x_0)$ gets closer to the stationary mean m. This equation (Eqn. 12) can be written in a matrix form as:

$$\mathrm{C} \,.\, w = \mathrm{D} \qquad (13)$$

where **C** is the matrix of covariances between data points, with elements $C_{i,j} = C(x_i - x_j)$, **D** is the vector of covariances between the data points and the estimation point, with elements given by $D_i = C(x_i - x_0)$, and ***w*** is the vector of simple kriging weights for the surrounding data points. Then the kriging weights are obtained by:

$$w = \mathrm{C}^{-1}.\mathrm{D} \qquad (14)$$

A kriging system not only accounts for *distance* in term of covariances (matrix **D**), but also accounts for clustering among the points. The clustering information is represented in the data-to-data covariances in the matrix **C**. Multiplying **D** by $\mathbf{C}^{-1}$ will reduce the influence of points falling in clusters relative to isolated points at the same distance. Once the weights are retrieved, the minimum error variance of the estimation can be expressed by:

$$\sigma_S^2(x_0) = C(0) - \sum_{i=1}^{n} w_i \, C(x_i - x_0) \qquad (15)$$

where $C(0)$ is the covariance of the random variable $Z(x_0)$ with itself which equal to the variance of the data. In the case where the location x_0 being estimated coincides with a data location x_i, the simple kriging estimator honors the data value $Z(x_i)$ at that location. This can easily be shown from equation 12, where as $x_i = x_0$, the corresponding weight at that location

is uniquely solved as 1; entails that the weights elsewhere are zero. Hence, solving the equation 11, the estimate $Z_S^*(x_0)$ will be equal to $Z(x_i)$. Then from equation 15, the estimation variance at that location will be zero as $C(x_i - x_0) = C(0)$. This exactitude property leads to kriging being known as an exact interpolator.

Ordinary Kriging

For ordinary kriging, rather than assuming that the mean is constant over the estimation region, the mean is assumed constant in the local neighborhood of each estimation point. From equation 11, the mean *m* is filtered from the linear estimator by requiring that the kriging weights sum to 1 to ensure the unbiassedness. The ordinary kriging estimator is thus written as:

$$Z_R^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i) \text{ with } \sum_{i=1}^{n} w_i = 1 \qquad (16)$$

In order to minimize the error variance σ_E^2 subject to the unit-sum constraint on the weights, the system calls for the definition of Lagrangian, which is a function of data weights w_i and an additional term involving a Lagrange parameter, μ:

$$L(w_i, i = 1,..., n; 2\mu) = \sigma_E^2 + 2\mu\left[\sum_{i=1}^{n} w_i - 1\right] \qquad (17)$$

The minimization with respect to the Lagrange parameter forces the constraint to be obeyed:

$$\frac{1}{2}\frac{\partial L}{\partial \mu} = 1 - \sum_{i=1}^{n} w_i = 0 \qquad (18)$$

In this case, the system of equations for the kriging weights turns out to be:

$$\begin{cases} \sum_{i=1}^{n} w_i C(x_i - x_j) + \mu = C(x_i - x_0) \\ \sum_{i=1}^{n} w_i = 1 \end{cases} \qquad j = 1,...,n. \qquad (19)$$

Similar to simple kriging, a set of ordinary kriging weights w_i is also designed to account for distance in term of data-to-estimation covariances $C(x_i - x_0)$; and clustering defined by covariances among the points $C(x_i - x_j)$. Once the kriging weights (and Lagrange parameter) are obtained, the ordinary kriging variance is given by:

$$\sigma_R^2(x_0) = C(0) - \sum_{i=1}^{n} w_i C(x_i - x_0) - \mu \qquad (20)$$

Like simple kriging, ordinary kriging also hold the exactitude property where the estimator honors the data values at their location. The difference between the two kriging is only the assumption about the mean. Intuitively, simple kriging yields estimates that are close to a constant mean, away from the data, because of the assumption of stationary mean. In contrast, ordinary kriging yields estimates that better follow the data fluctuations because of the use of local estimation of the mean within search neighborhood (Goovaerts, 1997).

Kriging Variance

Kriging produces the estimates of the unknown value at un-sampled locations with minimum error- variance (kriging variance) at each location based on the underlying semivariogram model. A clear understanding on the fundamental features of kriging variance is necessary for the application of this study.

First, the kriging variance only accounts for relative geometry of data locations and their distances to the location being estimated. Hence the kriging variance is zero at data locations and increases away from the data (Fig. 3b). Second, the kriging variance is independent of data values. For a given semivariogram model, any locations surrounded by data at similar distances will have similar error variances no matter what the data are. Where intuitively, however, we would expect location x_1 surrounded by similarly valued data has lower variance than x_2 which surrounded by data that are very different (Fig. 3). The link between kriging variance and data values is just through the semivariogram which only gives a ranking index of data geometry and not a measure of the local spread of error of the predictions (Goovaerts, 1997).

Thus, a smaller kriging variance indicates that the kriging estimate at unsampled location is more strongly supported by the sample data and higher kriging variance indicates that the estimate is less supported by the sample data. In this study, the information about kriging variance were used as a measure of confidence in estimates when generating field DEM using different field surveys.

Conditional Stochastic Simulation

In predicting a value at un-sampled locations, kriging provides an estimate with minimum kriging variance at each location. However, kriging typically tends to smooth out local detail of the spatial variation of the estimate, with small values typically overestimated and large values underestimated (Fig. 4; Isaaks and Srivastava, 1989). This is a serious shortcoming for example if the estimates of elevation were used for risk assessment in environmental studies such as predicting soil erosion and pollutant concentration.

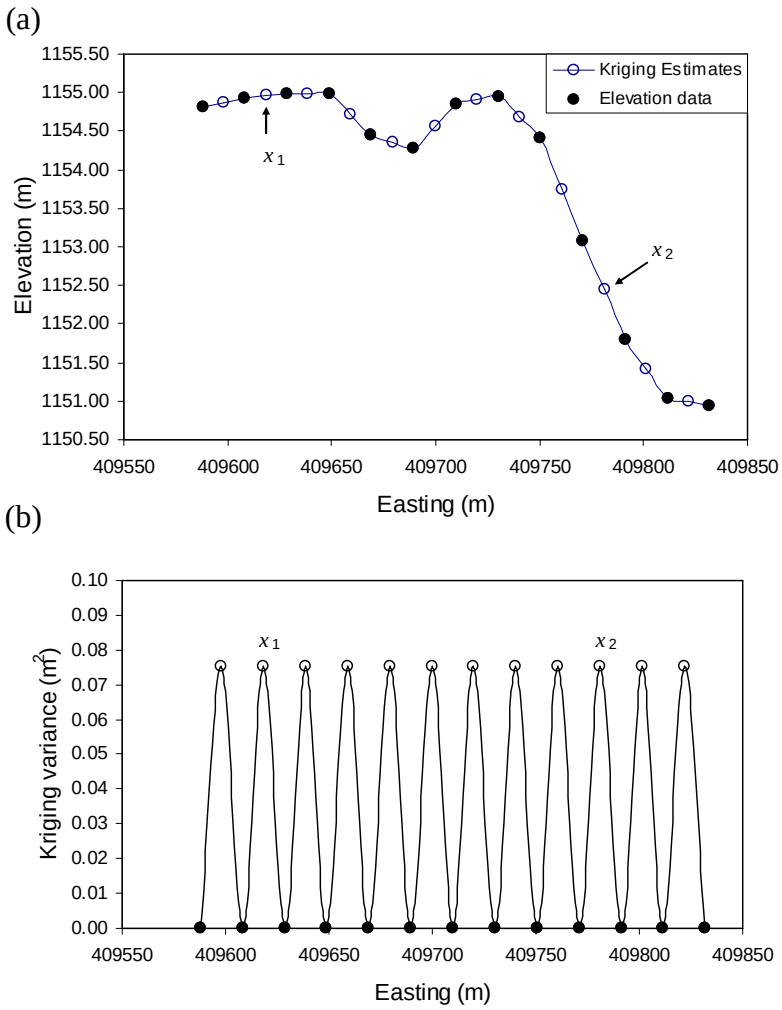

Figure 3: (a) Ordinary kriging estimates and (b) the associated kriging variances. Ordinary kriging was used to estimate value in between uniform sample points along an elevation transect in a USGS DEM of Boxholm, IA.

Geostatisticians address the limitations inherent in kriging by means of conditional stochastic simulation. Conditional stochastic simulation techniques offer the possibility of deriving various simulated estimates (realizations) from a given source model, which is built from the geostatistical information inferred from sampled observations. All realizations have an equal likelihood of being the real value since each simulation employs the same known

data. The multiple equal-probable realizations provide all the information necessary to approximate the distribution of any variable at un-sampled location, consequently allow one to quantify the uncertainty of estimates.

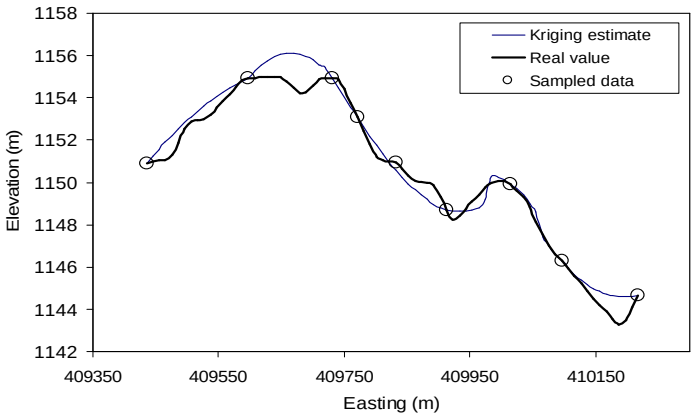

Figure 4: Real and estimated (kriging) profiles of elevation on a USGS DEM of Boxholm, IA. The kriging estimate smoothed the fluctuations of the real data.

Among many other stochastic simulations techniques, sequential Gaussian simulation (SGS) is widely used to estimate continuous variables like elevation; because of the inherent structure of the Gaussian model makes model determination fairly straight forward. The basic idea of SGS is very simple. Recall that kriging gives us an estimate and error variance of the variable at each location of interest. Using the estimate and error variance provided by kriging, we could represent the variable at each location as a random variable following a Gaussian distribution. Rather than choosing the mean as an estimate at each node, SGS chooses a random deviate from this Gaussian distribution, selected according to a uniform random number representing the probability level. For example, in estimating elevation values on a gridded area, kriging gave an estimate of 322.93 m with an error variance of 0.24 m^2 (standard deviation of 0.49 m) at a grid location. By using the kriging estimate, the

cumulative normal probability function (cdf) is generated. In this case, if we happened to generate a uniform random number of p=0.665 for this grid location, the corresponding value of the cdf would be 323.14 m (Fig. 5). This value is then assigned as the simulated elevation value at that location.

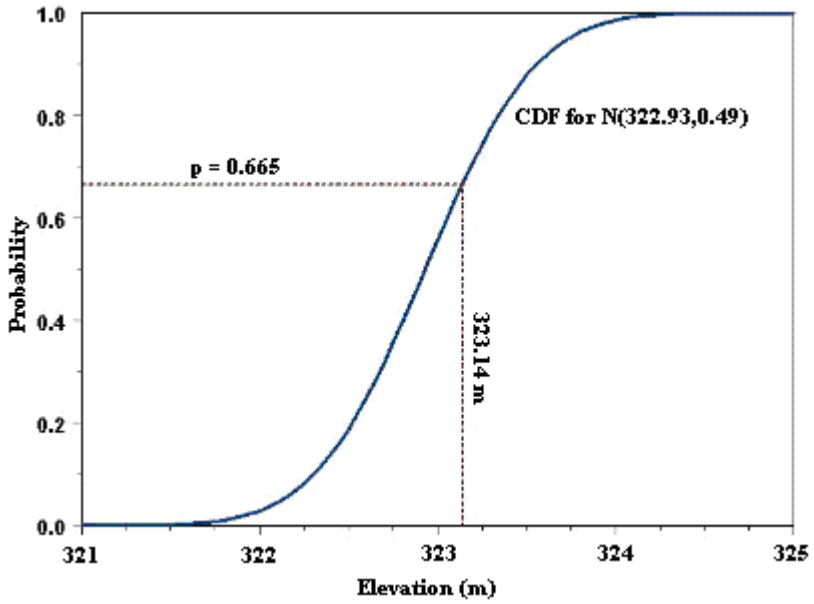

Figure 5: The cumulative normal distribution function of an elevation estimate developed using mean and standard deviation provided by kriging. A uniform random number of p=0.665 was generated to draw a simulated elevation value for that location.

The idea is to use the previously simulated estimate at each location as "data" in order to preserve the proper covariance structure between the simulated values. The procedures to perform SGS involve:

1. Justifying the appropriateness of the standard normal (Gaussian) distribution assumption using normal score transformation on the original data. The normal score transform adjusts each sample value by mapping them from their original cumulative

frequency to the standard-shaped curve corresponding to a normal distribution with mean value of 0 and a standard deviation of 1 (Fig. 6a).

2. Performing the sequential simulation on the normal score data as follows:
    a. A random path was generated so that each location to be estimated is visited only once.
    b. At each location, the parameters (an estimate and its standard deviation) for the Gaussian conditional cdf are found using simple kriging with the normal score semivariogram model. The estimation was based on surrounding normal score data values and values simulated at previously visited grids.
    c. Then a value was selected at random from the corresponding cdf and added to the data set.
    d. The two previous steps were repeated until all locations were visited.
3. Back-transforming the simulated normal scores into simulated values of the original variable by mapping the normal scores to the original cumulative frequency of the variable (Fig. 6b). Detailed description of normal score transformation and back-transformation can be found in Goovaerts, 1997.

Repeating SGS N times will result in N equal-probable realizations at each location. Each time the SGS is carried out, a random path is used to visit all the locations to avoid artifacts induced by walking through the grid in a regular fashion. Using N realizations of equal-probable value at each location, simple statistics such as mean and variance of the estimates can be calculated. The mean at each location calculated by averaging the N realizations, also known as conditional mean or E-type estimate (Journel, 1987), will exhibit similar value predicted using kriging if N is large enough (Fig. 7). Unlike kriging variance,

the variance of N realizations calculated at each location also known as conditional variance will represent the variability of the estimated parameter (Fig. 8). The dispersion of the estimates from the mean also can be quantified such as using standard deviation or confidence interval. This statistics can be used to represent the local uncertainty of the estimated data.

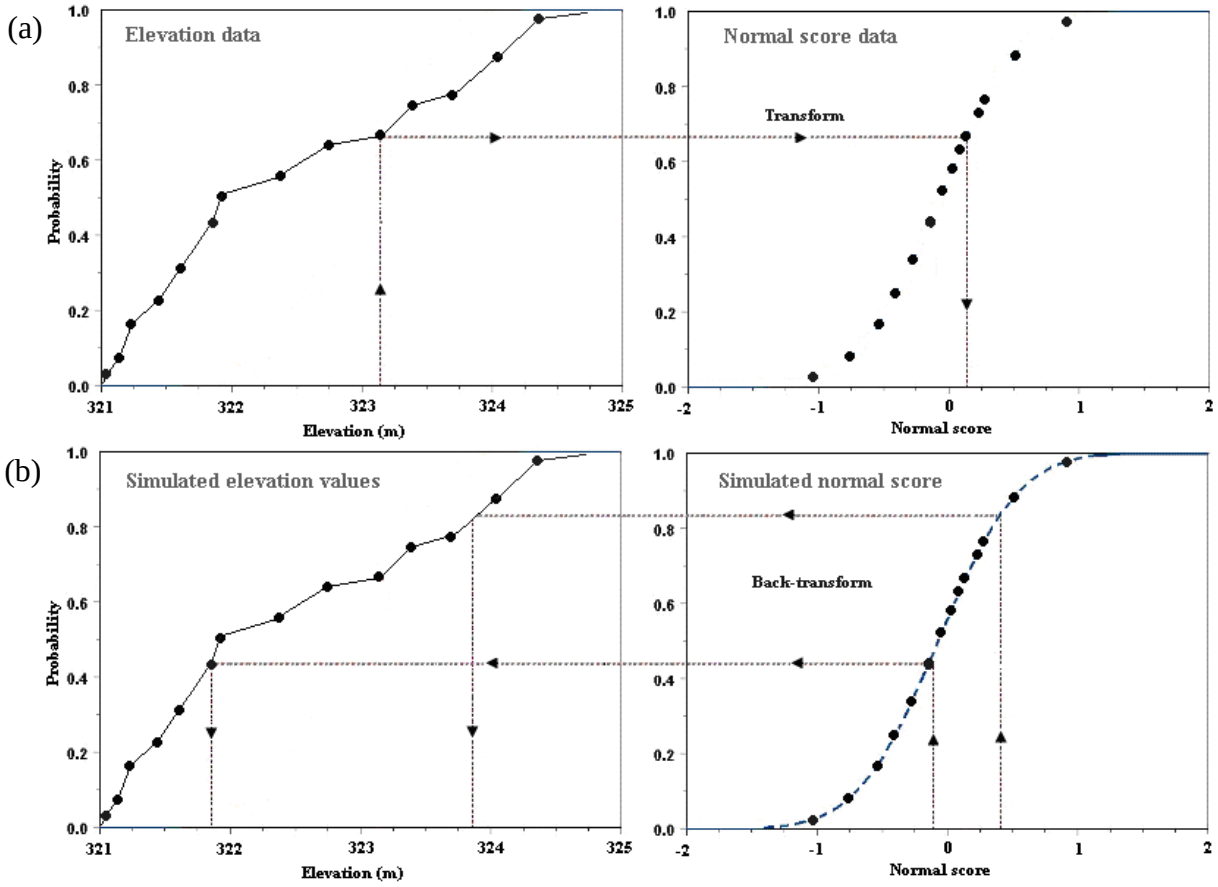

Figure 6: (a) Graphical procedure for transforming the cumulative distribution of original elevation values into standard normal distribution or called normal score and (b) back-transforming the normal score into simulated elevation values. The original elevation data depicted by the black dots are retrieved exactly.

Kriging versus Simulation

Unlike kriging, conditional stochastic simulation does not aim at minimizing a local error variance but focuses on the reproduction of statistics such as sample histogram or the semivariogram model in addition to honoring of data values. For example using elevation data from *gstat* library in R statistical software (Free Software Foundation, Inc., Boston, MA), the histogram of the simulated estimates had a similar shape relative to the sample histogram of elevation data. On the other hand, the histogram of the kriged estimated does not exhibit the sample variogram where the variance of kriged estimates is much smaller than the sample variance (Fig. 9). Also, as illustrated in Fig. 10, the semivariogram of the kriging estimates has a smaller sill than the semivariogram model depicted by solid line, which reflects the underestimation of short-range variability of the estimated values. The semivariogram of the simulated estimate on the other hand is more similar to the model semivariogram.

Simulation techniques provide models that show no smoothing, but it is in the nature of simulations that every repetition gives uncertain estimations, so the repetition of hundreds of simulations will very likely give hundreds of different realizations at each estimation point. For this reason, simulations do not provide good local estimators but they provide good measures to describe spatial uncertainty. If a modeler wants good local estimators, kriging remains the best choice since kriging provides a single numerical value that is “best” in some local sense (Deutsch and Journel, 1998). On the other hand, the simulation technique is preferred if one wants to study the dispersion of spatial variability of the data.

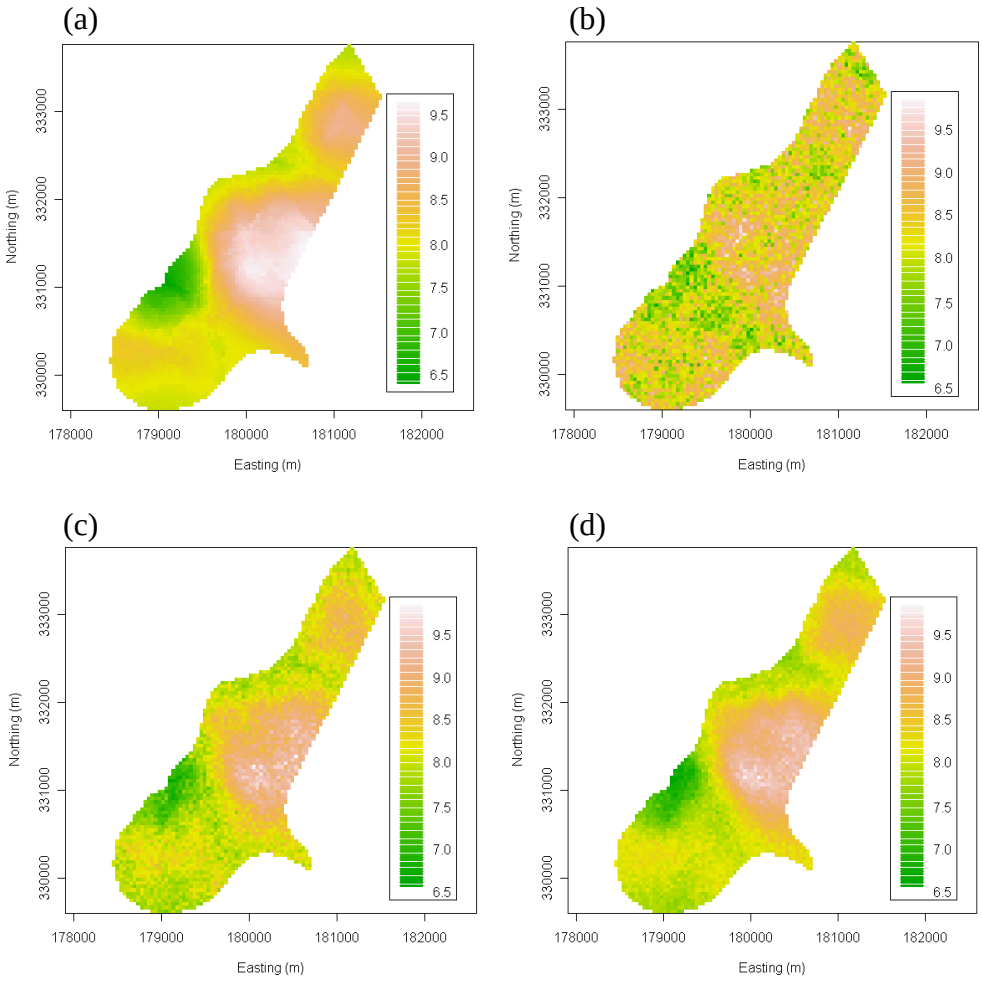

Figure 7: (a)Elevation map of ordinary kriging estimates, (b) a simulated realization; (c) mean of 10 simulated realizations and (d) mean of 100 simulated realizations over a study area from *meuse* data set in *gstat* library in R statistical software library (R statistical software; Free Software Foundation, Inc., Boston, MA). As the number of simulation increased, the simulated map becomes more similar to the kriging map.

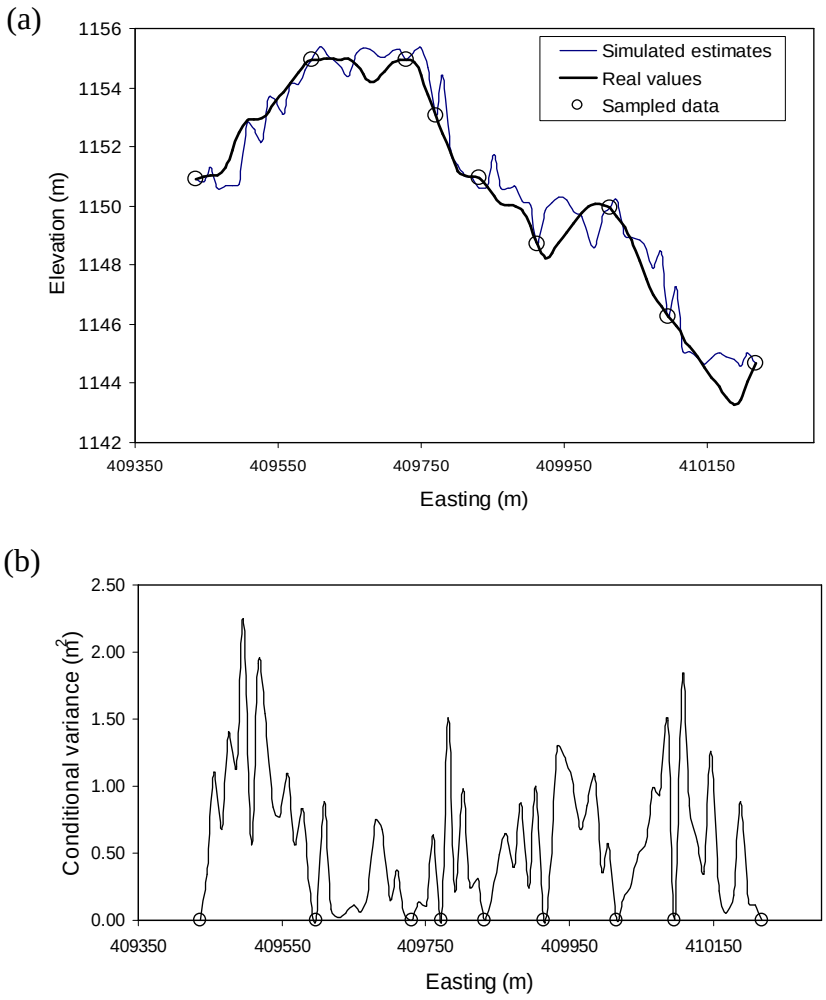

Figure 8: (a) Real elevation profile on a USGS DEM of Boxholm, IA and the 10th simulated elevation along the profiles from 50 SGS. (b) Unlike kriging variance, the associated conditional variances of the 50 SGS elevations describe the variability in the elevation profile. SGS does not provide good local estimators but they provide good measures to describe the spatial variability of the elevation data.

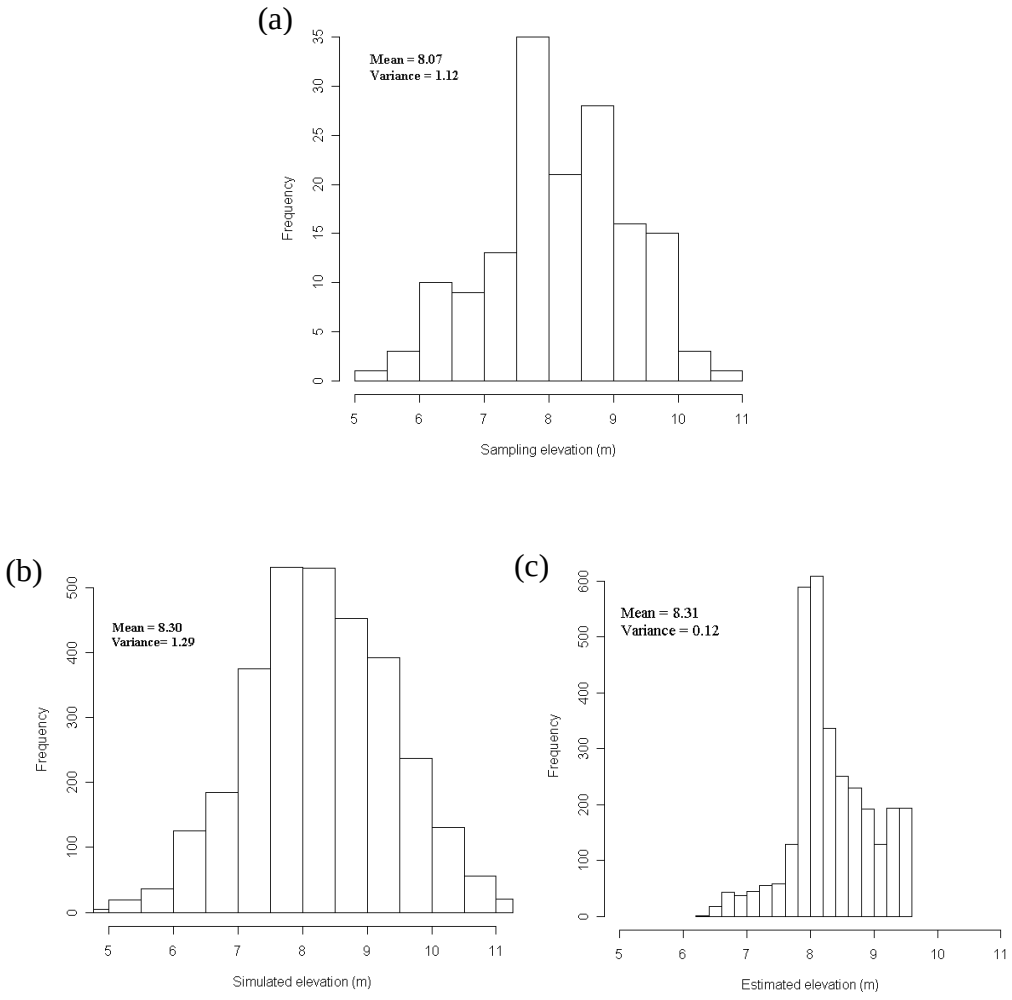

Figure 9: (a) Histogram of elevation data of a study area from *meuse* data set in *gstat* library in R statistical software library (R statistical software; Free Software Foundation, Inc., Boston, MA). Bottom graphs illustrate the (b) histogram of the 50th simulated estimates from 100 SGS and (c) histogram of the kriged estimates (right). The variance of the kriged estimates is much smaller than the variance of the original sample data $\sigma^2 = 1.12$.

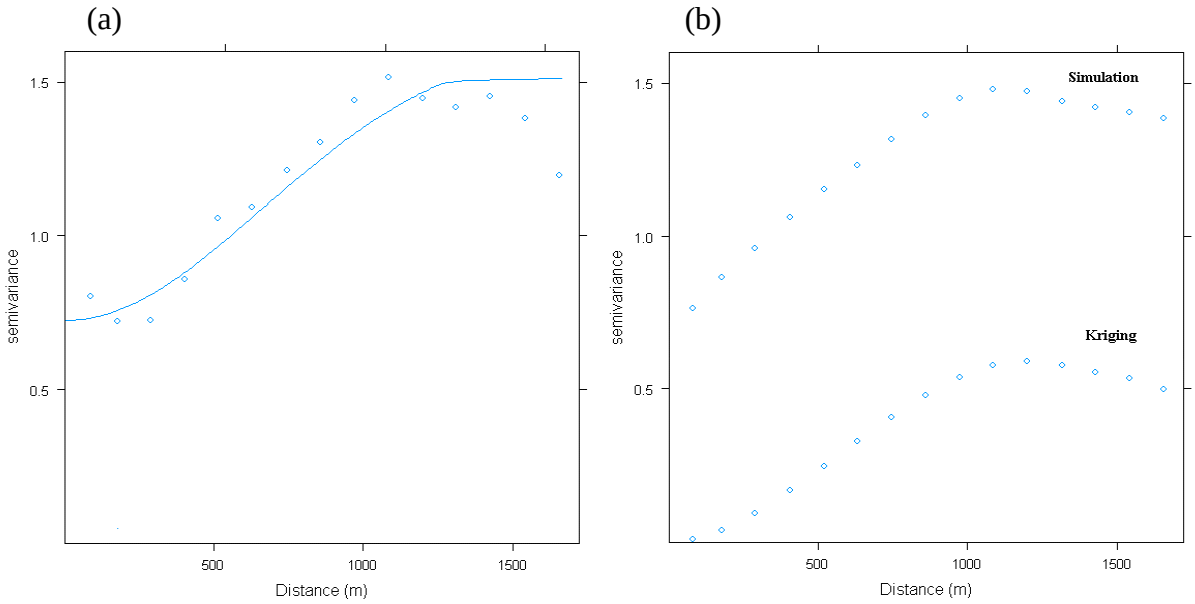

Figure 10: (a) Experimental and model (solid line) semivariogram for elevation data of a study area from *meuse* data set in *gstat* library (R statistical software; Free Software Foundation, Inc., Boston, MA). (b) The semivariograms of their kriged and simulated estimates. Note the smoothing effect of kriging leads to underestimation of the short-range variability of the elevation values.

Chapter Summary

The growing interest of scientists in geostatistics arises because they increasingly realize that quantitative spatial prediction must incorporate the spatial correlation among observations. Also, geostatistics offers an increasingly wide palette of techniques well suited to the diversity of problems and information scientists have to deal with. For interpolation problem, kriging is known as the technique that provides an estimate with a minimum error-variance at each prediction location. In this study, kriging was used to predict elevation value of agricultural field DEMs from repeated field surveys. Kriging variance was used to measure the confidence of the estimates from different surveys, before combining the estimates to develop an accurate field DEM.

There is necessarily uncertainty about the estimated value at an unsampled location, and its assessment is critical for some applications. For uncertainty assessment, SGS is known as the method that could generate multiple possible realizations that honor the data and reproduce aspects of the patterns of spatial dependence in the data. In this study, SGS was used to assess DEMs uncertainty and its effect on agricultural applications such as designing sampling strategy and estimating soil loss.

CHAPTER 3. UTLIZING REPEATED GPS SURVEYS FROM FIELD OPERATIONS FOR DEVELOPMENT OF AGRICULTURAL FIELD DEMS

A paper submitted to the Transaction of ASABE

S. Abd Aziz, B. L. Steward, L. Tang, M. Karkee

Abstract

Topographic data collected using RTK-DGPS-equipped farm vehicles during field operations could add additional benefits to the original capital investment in the equipment through the development of high accuracy field-DEMs. Repeated surveys of elevation data from repeated field operations may improve DEM accuracy over time. However, minimizing the amount of data to be processed and stored is also an important goal for practical implementation. A method was developed to utilize repeated GPS surveys acquired during field operations for generating field-level DEMs. Elevation measurement error was corrected through a continuity analysis. Fuzzy logic (FL) and weighted averaging (WA) methods were used to combine new surveys with past elevation estimates without requiring storage and reprocessing of past survey data. After 20 surveys were included, the DEM of the study area generated with FL and WA methods had an average RMSE of 0.08 m, which was substantially lower than the RMSE of 0.16 m associated with the DEM developed by averaging all data points in each grid. With minimum control of errors in elevation measurements, the effect of these errors can be reduced with appropriate data processing

including continuity analysis, fuzzy logic and weighted averaging. Two years of GPS surveys of elevation data from field operations could reduce elevation error by 50% in field DEMs.

Introduction

In agricultural practices, accurate representation of field topography is required to implement precision agriculture management for more efficient production systems. Topographical information is important because it provides derived parameters such as slope, aspect, topographic index and flow accumulation that are critical for agricultural conservation planning. For example, the movement of sediment, soil particles and agricultural chemicals (Maidment, 1996), and crop residue cover (Brown, 2008) are closely linked to soil topography and slope. In practice, topographic maps in the form of digital elevation models (DEMs) have been used to assess transport of constituents such as sediment and surface runoff from forested and an agricultural watersheds (Ghidey et al.,2001; Ouyang et al., 2005; and Sarangi et al., 2007), derive potential flow accumulation to assess soil moisture patterns and soil texture changes in field (Schmidt and Persson, 2003), and estimate soil erosion for appropriate farm management and soil water conservation planning (de Jong et al., 1999; Oost et al., 2000; Lin and Lin, 2001; and Ritsema et al., 2001). In spite of the importance of DEMs in agriculture, it is nevertheless a challenge to obtain elevation data cost effectively with sufficient accuracy and resolution.

A DEM is a digital representation of land topography representing elevations on the earth's surface. A DEM can be represented by one of three data structures: (1) gridded models, where elevation is estimated for each point on a regular grid; (2) triangulated irregular networks (TIN), where terrain elevation is represented in a network of

nonoverlapping irregular triangles; and (3) contour-based networks, where landscape is divided into small, irregularly shaped polygons based on natural contour lines and their orthogonals (Wilson and Gallant, 2000). The square-grid (gridded) model is the most common form of DEM because of its simplicity and ease of computer implementation (Wise, 1998). This paper is therefore focused on developing gridded DEMs and, for simplicity, the term DEM will be used to refer to them.

Traditionally, DEMs were developed using elevation data collected from conventional surveying techniques such as theodolite and level surveys. Currently, remote sensing techniques, such as traditional aerial photogrammetric surveys, airborne laser scanning (Ackermann, 1999), synthetic aperture radar (SAR; Evans and Apel, 1995) and light detection **a**nd ranging (LiDAR; Vaze and Teng, 2007) are often used. Remote sensing techniques require less labor, but using these data sources to represent the topography of a particular site is often too expensive and may require considerable technical and computer expertise for appropriate data handling and processing. Usually DEMs can be purchased from a service provider such as the U.S. Geological Survey (USGS) which sells DEMs at varying levels of accuracy. USGS 7.5-minute DEMs, with grid spacing of 10 m or 30 m, are the most accurate, with RMSE of 7 m and 15 m, respectively and have been produced by interpolating elevations from vectors or digital line graph hypsographic and hydrographic data.

The advent and widespread use of Global Positioning System (GPS) in agriculture provides new and affordable opportunities for farmers to collect elevation data. Every time GPS-equipped vehicles are operated in the field, elevation data can be recorded. The ability to obtain elevation data using GPS-equipped farm vehicle offers great advantages as surveys

can be done during the course of other field operations, and thus do not require any additional time or labor to collect the data. In 1992, real-time kinematic differential GPS (RTK-DGPS) became commercially available with measurement capabilities within 1- 4 cm accuracy (Buick, 2006). RTK-DGPS is becoming more widely adopted as many applications in precision agriculture require high accuracy and consistent positional data. Auto guidance systems used in row crops, for example, require high accuracy positional measurements where cultivation, strip-tillage, and harvesting must follow the planted rows precisely. In addition, the use of GPS receivers in agriculture will shift toward RTK-DGPS as greater coverage of RTK networks comes available.

Several studies have investigated the feasibility of using vehicle-mounted RTK-DGPS receivers to acquire topography data during typical field operations to generate DEMs. Clark and Lee (1998) compared DEMs produced from stop-and-go measurements with DEMs developed from kinematic measurements collected using an RTK-DGPS receiver mounted on a moving vehicle. They showed that kinematic measurements produced DEMs with slightly higher error (3 to 8 cm), but the increase was minimal relative to the amount of additional effort required to collect stop-and-go (error of 2 to 3 cm) measurements. Westphalen et al. (2004) used RTK-DGPS receivers and an inertial measurement unit (IMU) mounted on an agricultural sprayer to measure vehicle attitude and elevation data to generate DEMs. With the combination of IMU and the kinematic GPS measurements, the root mean square error (RMSE) of the DEMs ranged from 10 to 15 cm.

As a growing proportion of agricultural vehicles are equipped with GPS receivers, elevation data may be gathered continuously during common field operations. The accuracy of elevation data and any derived parameters can be improved using repeated surveys and

averaging GPS point locations over several years (Renschler et al., 2002). Repeated GPS surveys of elevation data from field passes of agricultural vehicles could be advantageous in improving the accuracy of the DEM. However, with repeated surveys comes the challenge of handling increasingly larger amounts of data, particularly if all of the data are required for improving DEM accuracy. Moreover, vertical and horizontal position measurement errors might occur in each survey due to device inaccuracies and human error during data collection.

To address these issues, algorithms were developed to combine repeated GPS surveys improving elevation estimates of agricultural fields. We proposed a process that would minimize user input and intervention and as well as expertise requirements for generating field level DEMs as a by-product of GPS equipped field operations. The goal of this research was to develop a methodology of combining repeated GPS surveys from field operations for the development of agriculture field DEMs. The specific objectives of the research were: 1) to compare fuzzy logic, weighted averaging, and grid-wise averaging techniques for combining repeated GPS surveys and 2) to observe the effect of combining multiple GPS surveys over several years on DEM accuracy.

Data Simulation and Collection

The methods proposed in this study were tested using elevation data from two sources:

- **Simulated RTK-DGPS Elevation Surveys**: RTK-DGPS elevation surveys were simulated to provide datasets for the development of the methodology in the study. Elevation values were interpolated from a USGS DEM along predefined field

operation paths to simulate the GPS surveying process. RTK-DGPS errors were modeled and added to these simulated measurements. Simulated survey data were used to focus on the effects of GPS errors inherent in GPS measurements. With these data, it was assumed that the USGS DEM was the best elevation representation for that area and was thus used as the true surface for validation.

- **Experimental RTK-DGPS Field Surveys**: To test the algorithm on measured data, multiple GPS surveys were conducted on a test field by driving an agricultural vehicle with RTK-DGPS receivers mounted on it. Another set of GPS surveys was collected using a RTK-DGPS receiver mounted on a sled pulled by a utility vehicle. The measurements this latter set of surveys were used as reference measurements for validation. They were collected closer to the ground to minimize the errors due to vehicle dynamics and geometry associated with the test measurements.

Simulated RTK-DGPS Elevation Surveys

A test field was modeled using a 7.5-minute USGS DEM of Winneshiek County, Iowa with 10 m grid spacing. The USGS DEM was acquired from an online GIS data provider (GeoCommunity, 2007). Most of the area in the Winneshiek County consists of farm land (380,034 acres; 86% of the total area). A 120 m by 120 m area with elevation ranged from 1117 to 1124 m (around 8 m elevation difference) was selected (Fig. 1) for the study because it contained some topographical relief, but did not contain features such as streams, rivers or lakes that would prevent contiguous farm operations. To simulate surveys occurring during field operations, vehicle travel paths were predefined based on four field operations (tillage, planting, spraying and harvesting) typical of a corn-soybean rotation in

Iowa (Fig. 2). Then, elevation values were interpolated at each sampling location on the defined paths using inverse distance weighting (IDW) interpolation. Data were sampled along straight North-South paths for planting, spraying, and harvesting operations; and diagonal Northeast-Southwest paths for tillage operations. The swath spacing was 6.1 m for planting, harvesting and tillage operations, and 27.43 m for spraying operations (Table 1). The distance between data points along the path was 0.5 m, based on a 5 Hz measurement rate with 9.7 km/h (6 mph) vehicle speed. For each dataset, the sampling path started near the southwest corner of the field where the first sample point of the path was generated at random distance off of a fixed starting point (normally distributed with $\sigma = 0.5$ m). Thus, the location of the sampling path was in general different for each dataset. This variation was added because the field operations were assumed to be non-controlled traffic operations where the exact positions of the track operations will vary each year. At each simulated sampling location, a five dimensional vector was generated consisting of easting, northing, elevation, DGPS station ID number, and sampling time. A total of 20 simulated GPS measurement surveys (corresponding to 5 years of field operations) from the area were generated with simulated GPS noise added to the datasets. This process was replicated three times for analysis.

Table 1: Machine and track specifications used in simulating the elevation data.

Machines	Track width, m	Track direction
Chisel plow tillage	6.10	Diagonal (Northeast-Southwest)
Planter (8 row, 30 inch spacing)	6.10	Straight (North-South)
High Clearance Sprayer	27.43	Straight (North-South)
Harvester (8 row, 30 inch spacing)	6.10	Straight (North-South)

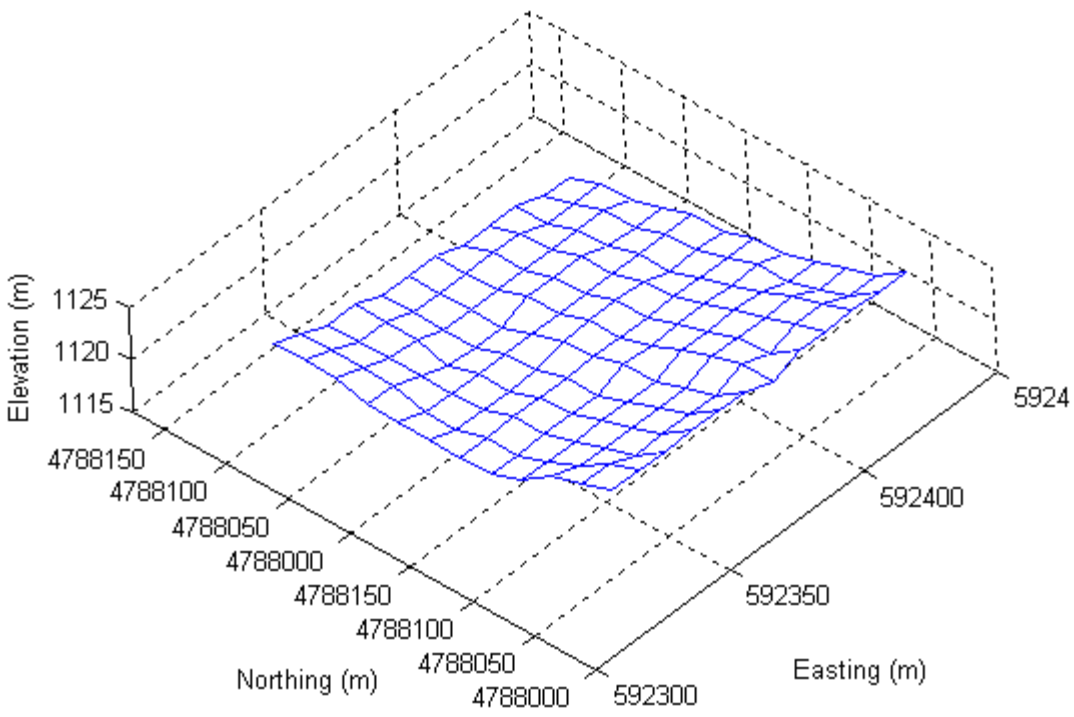

Figure 1. Digital elevation model for the study area from Winneshiek County, Iowa. The standard USGS Universal Transverse Mercator (UTM) format was used with UTM grid zone of 15N for the coordinate projection using North American Datum 1983 (NAD1983).

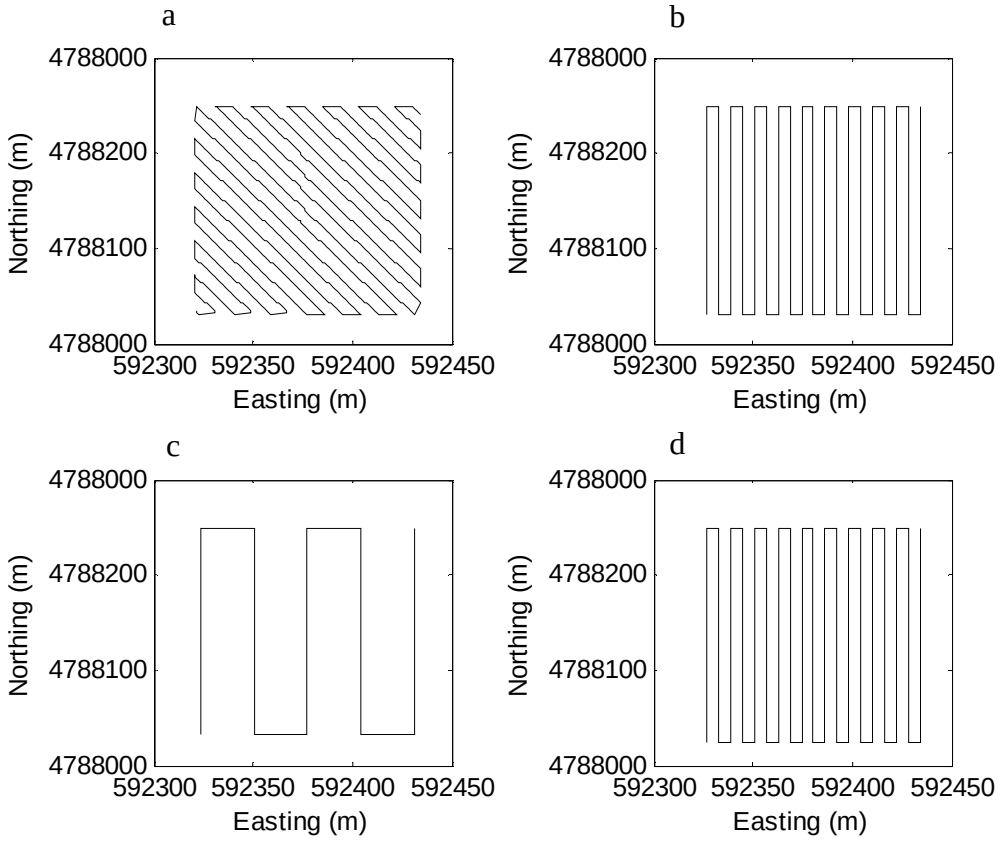

Figure 2. Diagonal (Northeast-Southwest) sampling path for (a) tillage and straight (North-South) sampling patterns for (b) planting, (c) spraying and (d) harvesting operations on the study area.

Vehicle-based RTK-DGPS system accuracy relies on the GPS signal quality and continued availability of differential correction signal. Loss or interruption of the DGPS correction signal will affect the GPS positioning measurement, which introduces errors in the range of centimeters (Scherzinger et al., 2007). Errors may also occur when satellites appear or leave the field of view during the GPS data collection. In our previous work when the RTK-DGPS receiver lost the correction or satellite signal, the receiver mode automatically changed from fixed to float DGPS correction solution (lower accuracy) which introduced large discontinuities in the measurements along the vertical and horizontal planes. We

modeled this noise using pseudorange error statistics for a dual frequency p code DGPS receiver. This noise together with the kinematic DGPS position errors are usually represented as stochastic errors that are correlated in time (Farrell and Barth, 1999). In this study, Gauss-Markov processes (James, 1994) were used to model the errors, because their exponential time-correlation function holds the properties of the errors. The Gauss-Markov terms were modeled as:

$$\varepsilon_i = \varepsilon_{i-1} e^{-\Delta T/\tau} + w_i \quad (1)$$

where;

ε_i is i^{th} error,

ε_{i-1} is the $(i-1)^{th}$ error,

w_i is the RTK-DGPS measurement noise represented as random process drawn from a normal distribution, and

ΔT is the sampling interval.

These processes can be described as an exponential autocorrelation function with variance, σ^2, and time constant, τ (Table 2):

$$R(t) = \sigma^2 e^{-|t|/\tau} \quad (2)$$

Three independent random number generators were used to produce normally distributed random noise. The first random number generator provided the Gauss-Markov noise related to RTK-DGPS errors. The second random number generator produced the Gauss-Markov noise for the discontinuity errors. The discontinuity noise was turned on for six to 10 times at five second intervals by generating random numbers indicating when the noise would occur in the samples (Fig. 3). The third random number generator produces the GPS measurement noise w_i, with standard deviation, $\sigma = 0.316$ m when the discontinuity

noise was turned on and σ = 0.0038 m when the discontinuity noise was turned off (James, 1994). These error models were added to the simulated elevation surveys. Errors were modeled independently along each X, Y and Z measurement axis. The survey simulation algorithm was written in Matlab version 7.0 (The Mathworks, Natick, MA).

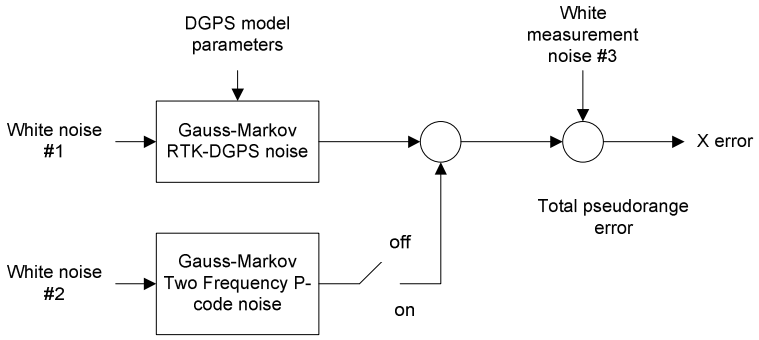

Figure 3: Algorithm for producing error in X axis. (Y and Z axes are similar; James, 1994).

Table 2: Pseudorange error statistics for DGPS error modeling (James, 1994)

	Gauss-Markov Noise		Measurement Noise
	Std. Dev., σ (meter)	Time, τ (sec)	Std. Dev., σ (meters)
RTK-DGPS position error	0.096	600	0.0038
Discontinuity error (Dual frequency)	1.030	600	0.3160

Experimental RTK-DGPS Field Surveys

Multiple GPS surveys were collected from a small portion of a field in Ames, Iowa. The field of interest covered an area of 0.23 ha (36.56 m wide by 60.96 m long) grassy field and had a 0 to 8 degree slope which was oriented at the south-west. The elevation ranged from 323 m to about 326 m (about 3 m elevation difference). Elevation data were collected

using a self-propelled high clearance agricultural sprayer (ASAE standards, 2005) equipped with real-time kinematic differential GPS (RTK DGPS) receivers (StarFire RTK, Deere & Co., Moline, III) operating at 1 Hz with a vertical static root-mean-squared error (RMSE) of less than 1.5 cm. The GPS receivers were mounted at a height of 3.81m above the field surface. The vehicle was driven across the field at a speed between 3.2 to 14.5 km/h (3 to 9 mph) along passes that were 3.05 m (10 ft) apart (Fig. 4). Correction signals were sent from local base station via a radio link (Pacific Crest Corp., Santa Clara, Cal.) The base station was located at 61 m (200 ft) northwest of the test field. A total of 16 datasets of field surveys were collected.

Another set of independent surveys across the entire field were also collected for validation. These reference measurements were acquired using a RTK-DGPS receiver operating at 5 Hz on a custom-built sled. A John Deere utility vehicle (Gator^{TM}, Deere & Co., Moline, Ill) was driven to pull the sled in the field at around 6.4 to 9.7 km/h. These measurements were collected closer to the ground to eliminate the errors caused by the vehicle dynamics contained in the test data. In general, as a vehicle travels over a field's topography, weight transfer leads to changes in vehicle pitch or roll angles relative to the slope based on the suspension stiffness. As such, these elevation measurements have an additional error source associated with the vehicle suspension system and geometry. Since the reference measurements were collected using a GPS receiver mounted on a sled, these errors were minimized.

Since the raw data were in the format of a geographic coordinate system consisting of longitude, latitude, and altitude, the data were converted into a projected coordinate system. Projection was required for spatial data analysis using units of length in the horizontal plane.

The standard USGS Universal Transverse Mercator (UTM) format was used (UTM grid zone 15N; North American Datum 1983; NAD1983).

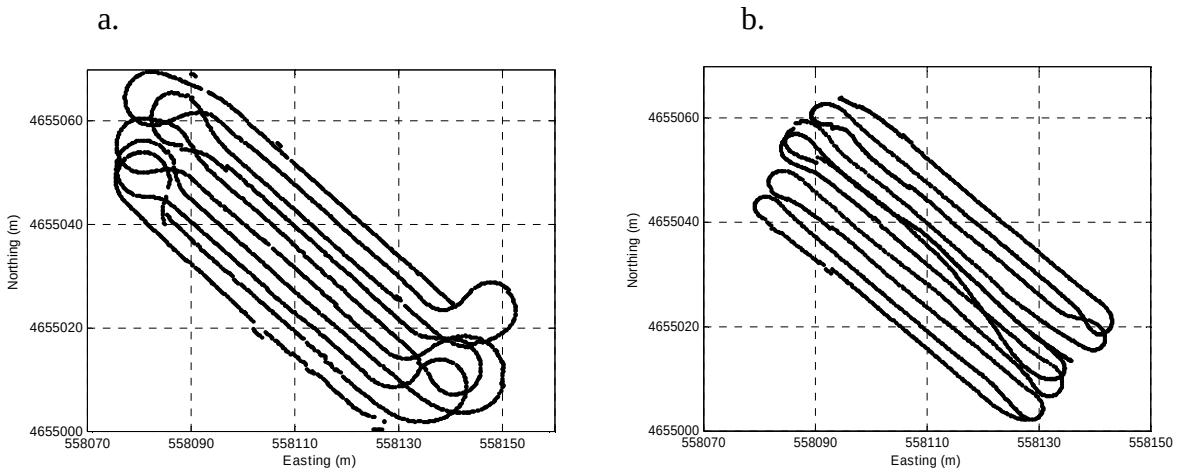

Figure 4. Data collection tracks in the study area. (a) using agricultural sprayer with 1 Hz RTK-DGPS receiver and (b)using custom-developed sled pulled using John Deere utility vehicle with 5 Hz RTK-DGPS receiver which was used as the validation set.

Methods

A custom-developed program was written in Matlab version 7.0 (The Mathworks, Natick, MA) to implement the methodology for generating field DEMs using repeated GPS surveys. The procedure consisted of the following steps:

Discontinuity Detection: When a RTK DGPS receiver lost the correction signal, the receiver mode changed from fixed to float mode solution (lower accuracy) and introduced discontinuous measurements in the dataset. A GPS discontinuity error correction algorithm was developed to correct these discontinuities.

DEM Generation: Next, kriging interpolation was used to interpolate GPS measurements into gridded DEMs. A DEM was developed from each GPS elevation survey.

DEM Combination and Reduction: This study was based on the hypothesis that field DEM accuracy can be improved by combining the DEM estimates over several surveys. However, simply averaging the DEM estimates from different surveys may not be the best approach because one measurement survey may contain more error than the other. Hence, two data combination algorithms were developed; one using a fuzzy logic (FL) approach and the other using a weighted average (WA) approach to combine data. Both methods only kept the current grid elevation estimates and their standard deviation and did not require data from previous surveys to be stored and reprocessed every time new GPS survey data came available. This feature is essential for practical implementation. DEMs were also developed through averaging the elevation at each grid (grid-wise averaging) to compare with the FL and WA results (Fig. 5). Finally, a control method was used to develop a control DEM through grid-wise averaging without discontinuity analysis. Detailed explanations of each step are provided in the following sections.

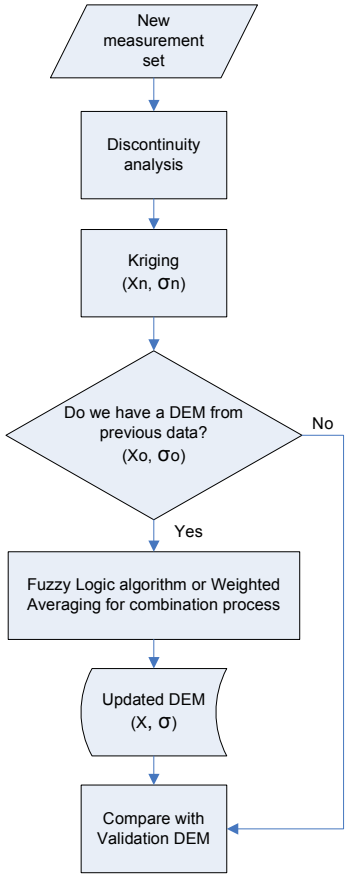

Figure 5. Overview of DEM development process using repeated surveys of elevation measurements. Each new measurement survey is combined with the existing DEM to improve the elevation estimate while reducing the amount of data to be stored.

Discontinuity Detection

An algorithm was developed to detect measurement discontinuities noise for data correction. In previous studies, when the RTK-DGPS receiver introduced large discontinuities in the measurements, the DGPS reference station ID number also changed indicating a change from a fixed solution to a float solution. Hence, discontinuities were identified by finding the changes in the station ID. The points with a float solution ID were characterized as the discontinuities noise. Then the discontinuities in X and Y axes were

corrected by adjusting the coordinates to follow the coordinates direction of the vehicle path. The discontinuities in the elevation were corrected by re-estimating the value using the mean of eight nearest continuous neighboring points (points with fixed solution ID). To ensure the discontinuities were minimized, the differences between adjacent elevation points along the path were computed. The discontinuities were minimized if the differences between the adjacent points along the path were within two standard deviations of the mean elevation differences. The result of this process can be seen through inspection of the elevation plot (Fig. 6).

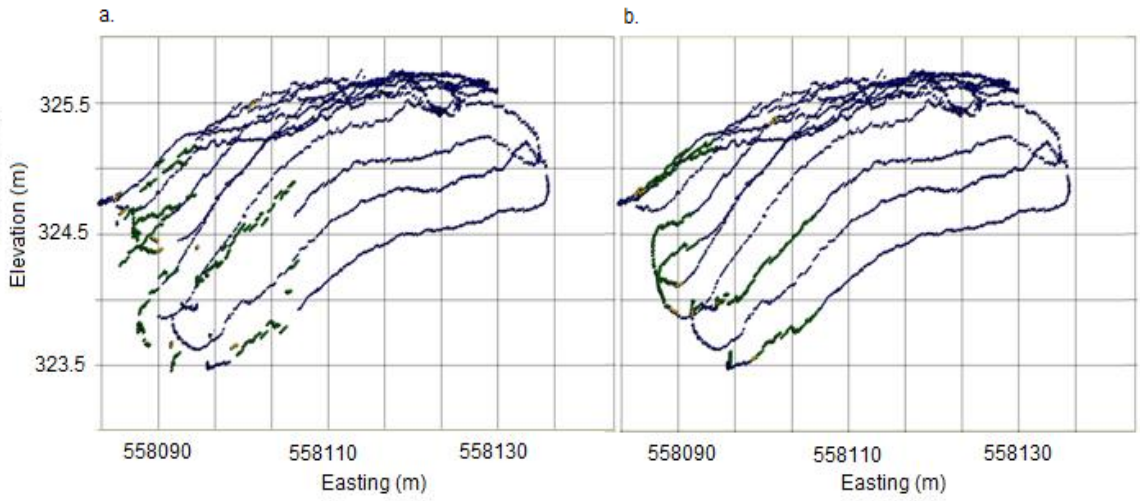

Figure 6. Elevation measurements (a) before and (b) after discontinuity analysis on a measurement survey collected at the study area.

Kriging Interpolation

After discontinuities were removed, cleaned elevation measurements were interpolated to generate a DEM. DEM grid locations were pre-defined so that each DEM developed using measurements from a new survey would use the same grid locations.

Elevation measurements were interpolated into gridded DEM using ordinary kriging, which was chosen because it is commonly used and is shown, based on geostatistical theory, to be an unbiased estimator that minimizes error variance (Isaaks and Srivastava, 1989). In addition, visual inspection of the data indicated no large trends, and ordinary kriging is known to be quite robust (Trangmar et al., 1985). A Matlab kriging toolbox (Sidler, 2003) was used with the von Kármán covariance model (von Kármán, 1948) instead of a common semivariogram model to describe the spatial structure of the data. For our implementation where sampled data were located preferentially and had a high possibility of outliers due to uncertainty in GPS measurements, the covariance model was appropriate. In a deterministic framework, where available sample information is interpolated within the same domain, direct estimation of the covariance model is better than the traditional semivariogram approach because the covariance estimator is less sensitive to extreme values, skewed distributions and clustered sampling than the traditional semivariogram estimator (Isaaks and Srivastava, 1998).

For the simulated GPS surveys, data were fit with an exponential covariance model with a 20 m range. Data points were interpolated to 10 m grids using a minimum of 16 data points. For experimental field surveys, data points were interpolated to 1 m grids with similar kriging parameters. The range distance, grid size, and number of data points represented a trade off between interpolation support and computation time. Anisotropy was taken in account as the search neighborhood was defined as an ellipse centered on location being estimate and rotated with the major axis in the vehicle path direction. The kriging elevation estimate and kriging variance for each grid were stored.

Data Combination and Reduction

After the kriging interpolation, the DEM estimates from different GPS surveys were combined using FL and WA methods. Both methods were developed for improving statistical estimates as new information comes available while not requiring storage of all prior measurements.

Fuzzy Logic Method

The process of combining the elevation estimates from two DEMs used a fuzzy logic algorithm to take into account the uncertainty at each grid represented by the kriging variance at each grid. Specifically, the algorithm adjusted the elevation estimate of a grid with higher kriging variance to be at least within two standard deviations of the grid estimate with a lower kriging variance from the other DEM. Kriging variance is the minimized estimation error variance under the condition of unbiasedness. The error variance is estimated based on the underlying semivariogram model. A smaller kriging variance indicates that the kriging estimate is more strongly supported by elevation measurements and thus more accurately represents the true elevation. Hence, the purpose of adjusting the higher variance grid estimate to within two standard deviation of that with a lower variance is to improve the accuracy of the DEM in the sense that the elevation estimate with lower kriging variance is better supported by measurements than that with a higher kriging variance.

The lower kriging variance grid estimate was used as the base estimate, x_1. Then the grid estimate from the other DEM, x_2, was categorized into *low*, *average* and *high* uncertainty fuzzy classes using a set of fuzzy membership functions developed in a similar

approach (Zhang and Han, 2002). The fuzzy membership functions represented the difference between two grid kriging estimates (Fig. 7):

$$\mu_L(x_2)=\begin{cases}1, & x_2 \le x_1-2\sigma_{x_1}\\ 1-2\left(\dfrac{\left(x_1-2\sigma_{x_1}\right)-x_2}{2\sigma_{x_1}}\right), & x_1-2\sigma_{x_1}<x_2\le x_1-\sigma_{x_1}\\ 2\left(\dfrac{x_1-x_2}{2\sigma_{x_1}}\right)^2, & x_1-\sigma_{x_1}<x_2\le x_1\\ 0, & x_2>x_1\end{cases} \tag{3}$$

$$\mu_A(x_2)=\begin{cases}0, & x_2 \le x_1-2\sigma_{x_1}\\ 2\left(\dfrac{\left(x_1-2\sigma_{x_1}\right)-x_2}{2\sigma_{x_1}}\right)^2, & x_1-2\sigma_{x_1}<x_2\le x_1-\sigma_{x_1}\\ 1-2\left(\dfrac{x_1-x_2}{2\sigma_{x_1}}\right)^2, & x_1-\sigma_{x_1}<x_2\le x_1+\sigma_{x_1}\\ 2\left(\dfrac{\left(x_1+2\sigma_{x_1}\right)-x_2}{2\sigma_{x_1}}\right)^2, & x_1+\sigma_{x_1}<x_2\le x_1+2\sigma_{x_1}\\ 0, & x_2>x_1+2\sigma_{x_1}\end{cases} \tag{4}$$

$$\mu_H(x_2)=\begin{cases}0, & x_2 \le x_1\\ 2\left(\dfrac{x_1-x_2}{2\sigma_{x_1}}\right)^2, & x_1<x_2\le x_1+\sigma_{x_1}\\ 1-2\left(\dfrac{\left(x_1+2\sigma_{x_1}\right)-x_2}{2\sigma_{x_1}}\right)^2, & x_1+\sigma_{x_1}<x_2\le x_1+2\sigma_{x_1}\\ 1, & x_2>x_1+2\sigma_{x_1}\end{cases} \tag{5}$$

where μ_L, μ_A, and μ_H are the degrees of membership (DOM) for the low, average and high fuzzy classes respectively and σ_{x1} and σ_{x2} are the standard deviations associated with the grid

kriging variance from the two DEMs. The *average* class indicated that the estimate is within two standard deviations of the kriging estimate of the lower kriging variance grid and data correction may not be necessary. Elevation estimates with high degrees of membership in the *high* or *low* classes, however, can be corrected by shifting them to be similar to the kriging estimate of the grid with lower variance.

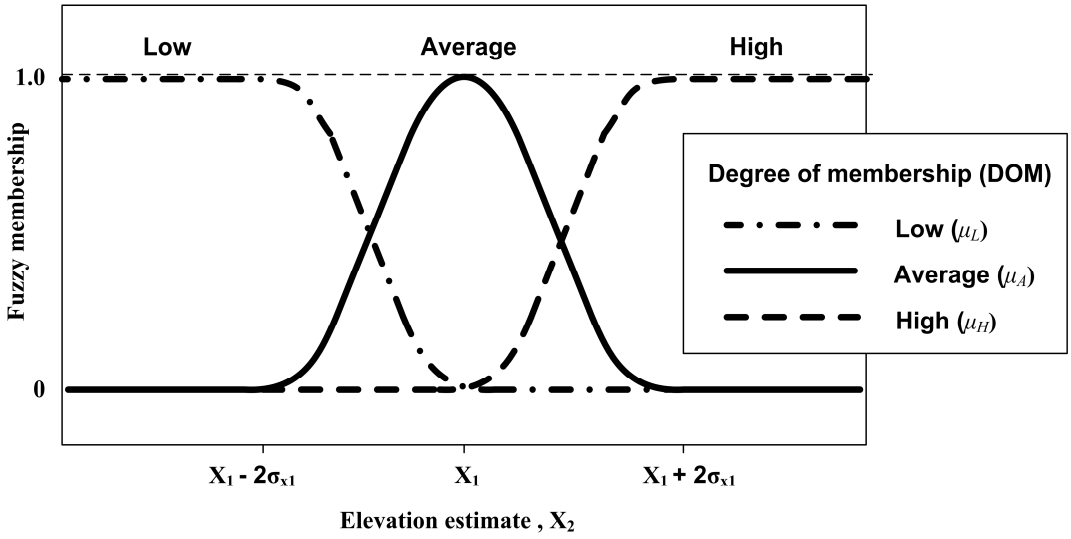

Figure 7. Graphical representation of fuzzy membership functions for low, average and high elevation for estimate x_2 in relation to estimate x_1 in a grid; with standard deviation, σ_{x1}.

The DOMs were used as the inputs to a fuzzy model to produce a crisp output to be used as a weight for data correction. The membership functions of the fuzzy output classes consisted of trapezoid and triangle shapes where the output variable *u* was between -1 and 1 (Fig. 8). The inference of input to the output was based on the following rules:

If elevation estimate, x_2 is Low, then the output, *u* is Negative Shift.

If elevation estimate, x_2 is Average, then the output, *u* is Small Shift.

If elevation estimate, x_2 is High, then the output, *u* is Positive Shift.

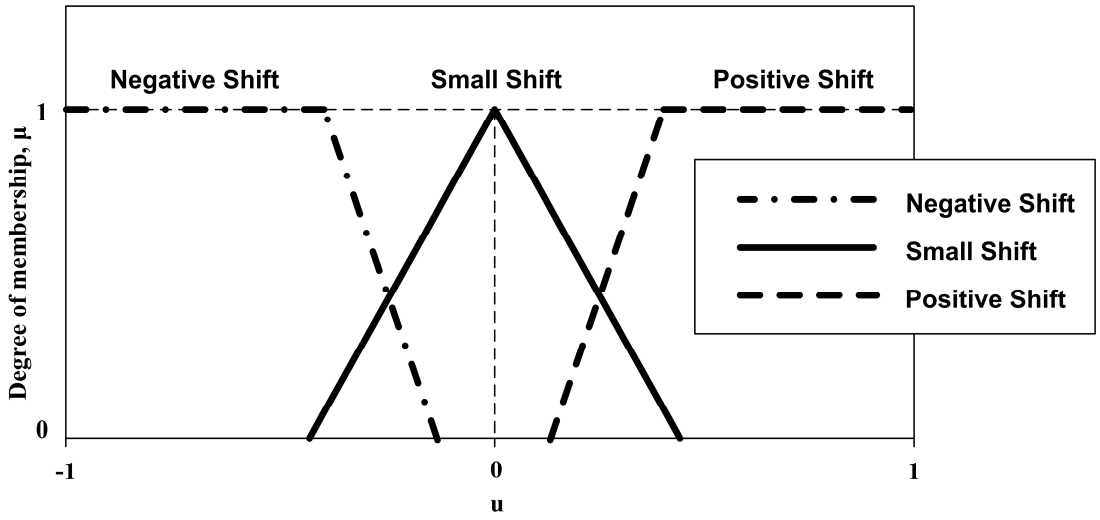

Figure 8. Fuzzy output model for determining the weight for data correction based on the degrees of membership $\mu(u)$ and output variable u.

The crisp output from this fuzzy model was determined using the centroid defuzzification technique defined as:

$$u^* = \frac{\sum_{i=L,A,H} \int_{-1}^{1} \mu_i(u) \cdot u du}{\sum_{i=L,A,H} \int_{-1}^{1} \mu_i(u) du} \tag{6}$$

where $\mu_i(u)$ is the DOM of the output membership function. The fuzzy output was then used as a weight in the correction function defined as:

$$x'_2 = x_2 - u^* \cdot D_{x_2 - x_1} \tag{7}$$

where;

x'_2 is the corrected estimate,

x_2 is the kriging estimate of a grid,

u^* is the weight obtained from fuzzy logic algorithm,

$D_{x2 - x1}$ is the absolute differences between x_2 and x_1.

After data correction using fuzzy logic, the estimates x'_2 and x_1 were averaged and stored. Since the accuracy of x_2 estimates were improved using the fuzzy algorithm relative to x_1 estimate, the variance of the combined estimates was represented by σ_{x1}, the kriging variance associated with x_1. This variance was passed along to be used for analysis when the next DEM was available. This process was repeated for all the grids of the study area. The process kept the current estimate of the elevation in each grid and its associated variance, and thus no additional prior data were required for future DEM recombination with new survey measurements.

Weighted Averaging Method

In the WA method, DEMs were combined grid-wise using a weighted average function defined as:

$$x_\mu = \left(\frac{\sigma_{x_2}^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2} \right) x_1 + \left(\frac{\sigma_{x_1}^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2} \right) x_2 \tag{8}$$

where x_μ is the new estimate (combination) of the elevation of the grid, x_1 and x_2 are the estimates from the two DEMs.

The variance of the combined estimate, σ_μ^2, was then defined as:

$$\sigma_\mu^2 = \left(\frac{\sigma_{x_1}^2 \sigma_{x_2}^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2} \right) \tag{9}$$

From equation (9), the updated variance is less than the smallest input variance since:

$$\frac{1}{\sigma_{\mu}^{2}} = \left(\frac{1}{\sigma_{x_1}^{2}}\right) + \left(\frac{1}{\sigma_{x_2}^{2}}\right) \tag{10}$$

The averaging function weighted the estimate based on the kriging variance, so if the kriging variance of a particular grid from DEM 1, x_1, is greater than that from DEM 2, x_2, then x_2 contributes more than x_1 to x_μ. From equation (10), the standard deviation σ_μ is less than either σ_{x1} or σ_{x2}, which implies the uncertainty in the estimate decreases by combining the two pieces of information. This process was repeated for all the grids of the study area, and the same procedure was used when new surveys were acquired. The process only kept the current estimate of the elevation in each grid and its associated variance. It did not require all previous data to be stored and reprocessed every time new surveys were acquired.

Accuracy of DEM Elevations

Root mean squared error (RMSE), a typical measure of DEM error (Wise, 1998; Bishop and McBratney, 2002; Wilson et al., 2005; Westphalen et al., 2004), was used to measure the performance of the various methods in producing accurate DEMs. For the simulated GPS surveys, the original USGS DEM data were used as the validation value for each DEM grid, and error was calculated by subtracting the elevation estimates from the USGS DEM values. For the DEMs developed using multiple GPS surveys, error was calculated by subtracting the DEM estimates from the nearest reference measurement value.

Results and Discussions

DEMs Developed From Simulated RTK-DGPS Elevation Surveys

The contour map of the DEMs produced using the control method exhibited many artifacts (Fig. 9a). The artifacts were mainly due to the discontinuity noise in the measurements, which was not removed in the control process. The RMSE of the DEM developed using the control method was substantially high with average maximum value of 1.38 m when a single simulated elevation survey was used to generate the DEM. The average RMSE from three replications decreased to 0.40 m after 20 simulated surveys were used (Fig. 9b). With proper data processing techniques, errors in the field measurements could be reduced to improve the DEMs accuracy.

a. b.

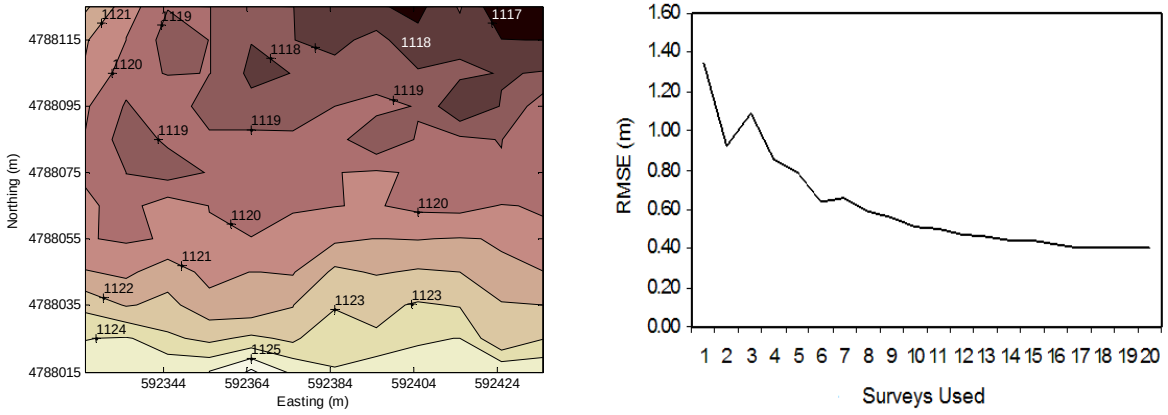

Figure 9. (a) Contour map of 10-m DEM of the test area from Winneshiek County produced by regularly averaging all data points in each grid without discontinuity error detection (control method) (b) RMSE of the DEM as multiple GPS surveys were combined as they become available using the control process.

The contour maps generated with data from the original USGS DEM of the test field (Fig. 10a) was compared with contour maps of the DEMs produced from the simulated surveys (Fig. 10b, c, and d). The contour map of the grid-wise averaging DEM (Fig. 10b) exhibited similar contour lines with a few artifacts. However, with the FL and WA DEMs,, most of the artifacts were removed (Fig. 10d and c). These topographic maps had contour lines similar to the original USGS DEM contour map. It is obvious that data processing is needed in order to generate acceptable topographic maps from these simulated surveys.

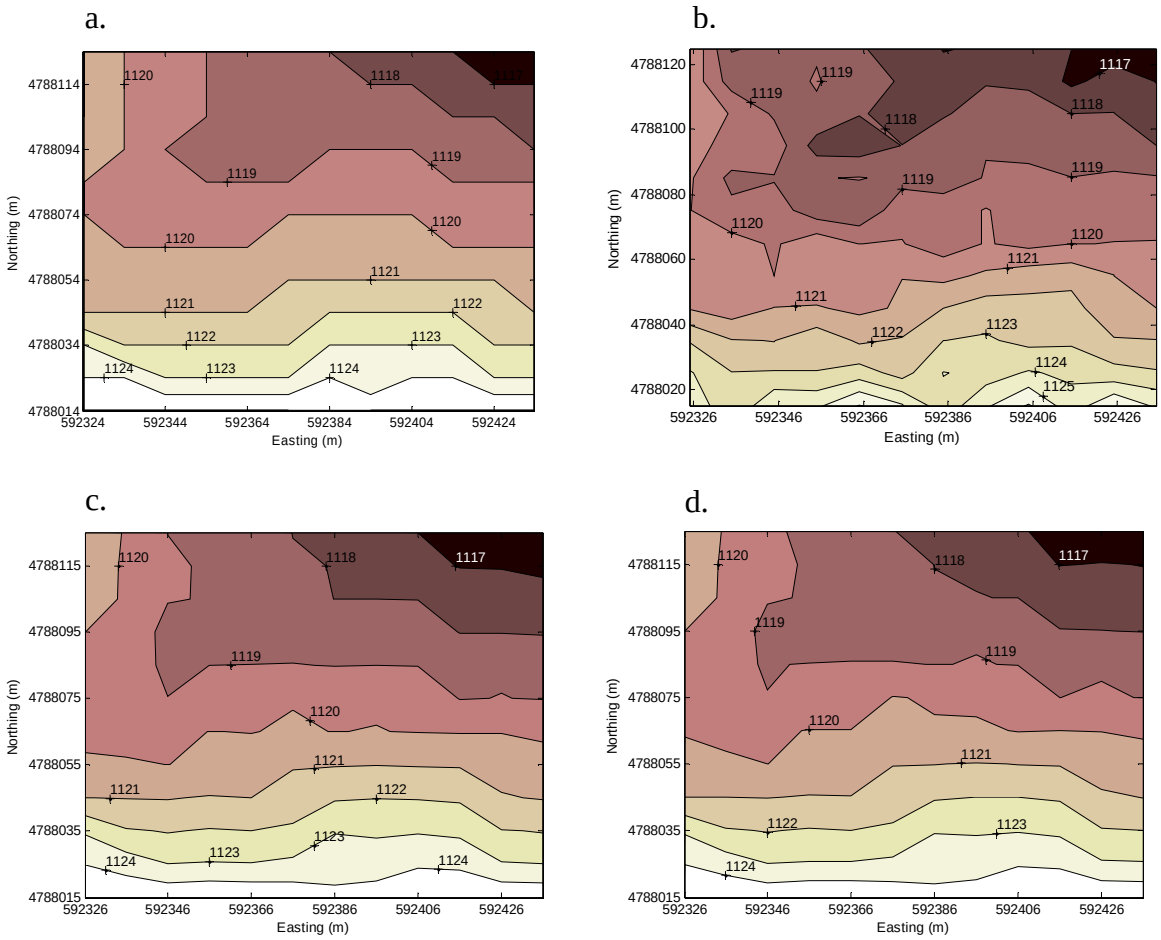

Figure 10. (a) Reference contour map compared to contour maps from DEMs produced by (b) grid-wise averaging method, (c) WA method, and (d) FL method.

The mean RMSE obtained from the DEM accuracy analysis for over three simulated measurement surveys decreased as the number of elevation surveys increased (Fig. 11). DEMs developed using the grid-wise averaging method with discontinuity error detection had higher RMSE compared to DEMs developed using the FL and WA methods. Overall, from this plot, the RMSE for all methods decreased as the number of simulated elevation surveys increases. For the grid-wise averaging method, the average RMSE from three replications decreased from 0.34 m to 0.14 m after 20 surveys were used. For FL and WA methods, the average RMSE decreased from 0.28 m to 0.07 and 0.08 m respectively as the number of elevation surveys used increased. For the first three surveys, the differences in RMSE between the FL and WA methods were very small. However, the RMSE from the WA method dropped lower than the FL method as more simulated elevation surveys were added. This lower error was due to how the variance was handled in each method. As more surveys were added, the WA method reduced the estimation variance through Eqn. 9, while FL method retained the minimum variance.

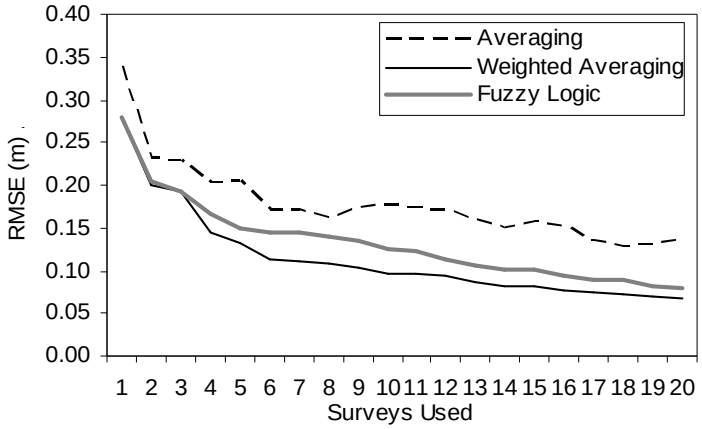

Figure 11. RMSE of DEMs as multiple simulated RTK-DGPS elevation surveys were combined using the methods developed for this study. RMSE were averaged from three independent replications.

DEMs Developed From Experimental RTK-GPS Field Surveys

Similar analysis was done with the GPS surveys collected from the study area. Due to the discontinuity noise in the measurements, which was not removed in the control process, the contour map of the DEMs produced using the control method exhibited many artifacts (Fig. 12a). The RMSE of the DEM developed using the control method was substantially high with average maximum value of 0.83 m. The RMSE varied substantially with a lower number of elevation surveys and became more stable at approximately 0.6 m as the number of surveys increased to five and above (Fig. 12b).

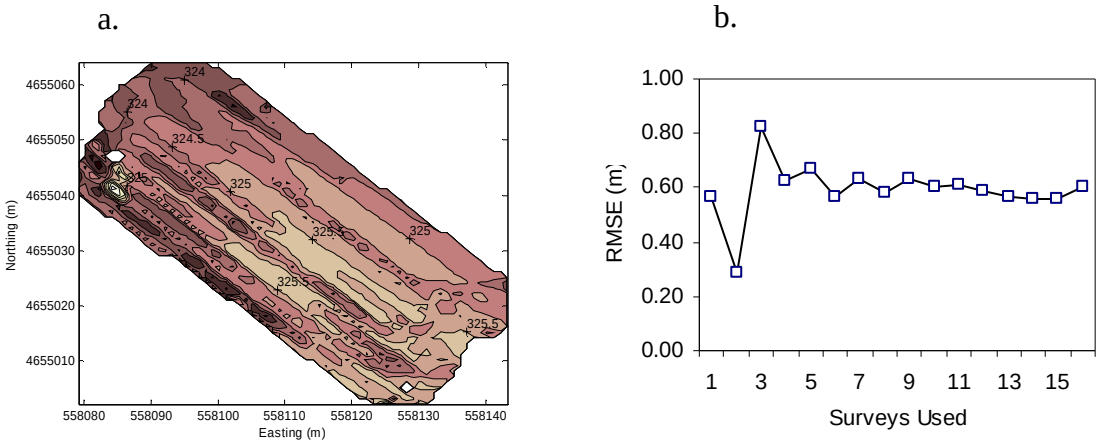

Figure 12. (a) Contour map of 1-m DEM of the study field by averaging all data points in each grid without discontinuity error detection (control process) (b) RMSE of the DEM as multiple RTK-DGPS surveys were combined as they become available using the control process.

DEMs developed using grid-wise averaging method (Fig. 13b) had a smoother contour lines compared to the contour map produced using the reference measurements (Fig. 13a). This may indicate that some of the topographical details in the terrain were not captured in the DEM developed using the grid-wise averaging method. On the other hand, the contour

maps from DEMs developed using the FL and WA methods (Fig.13c and b) were very similar to that of the reference measurements.

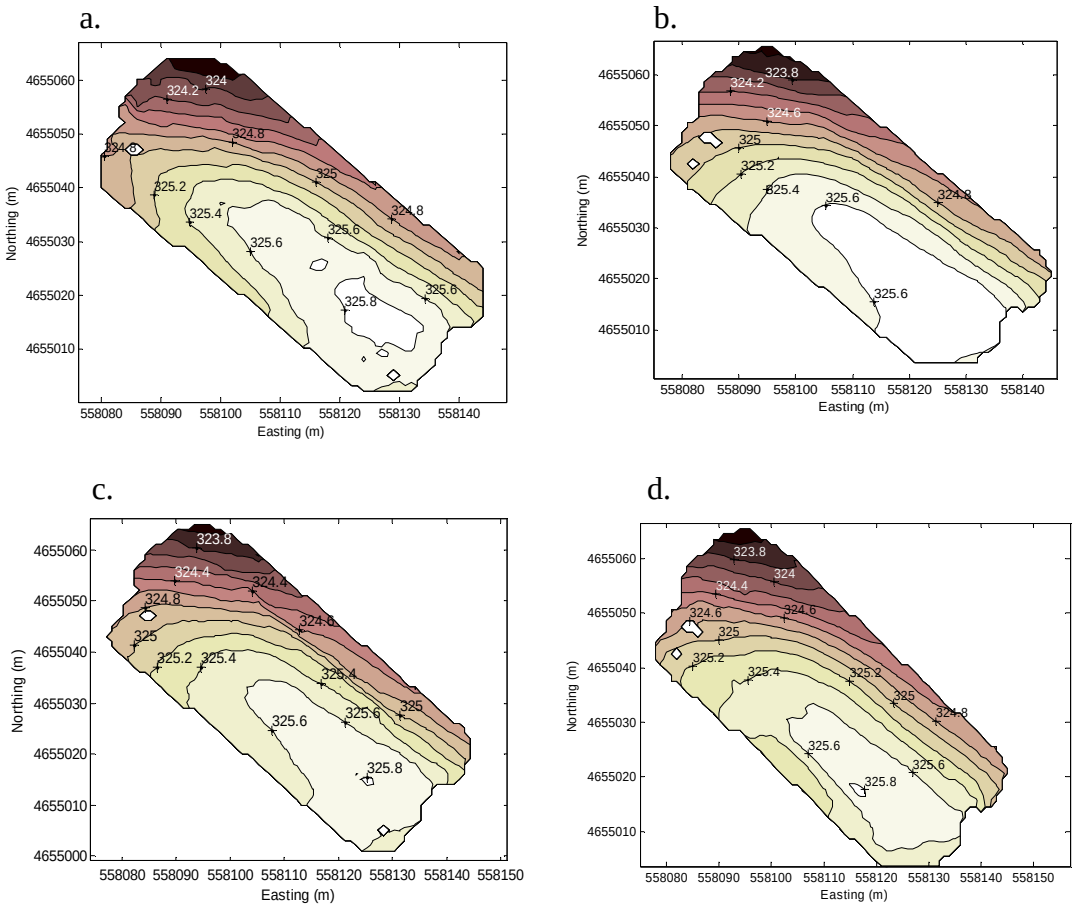

Figure 13. (a) Reference contour map compared to contour maps from DEMs produced by (b) grid-wise averaging method, (c) WA method, and (d) FL method.

The algorithm robustly handled the error inherent in GPS measurements, the three methods had similar performance (Fig. 14) as that observed with the simulated surveys. The DEMs developed using the grid-wise averaging method had substantially higher RMSE compare to DEMs developed with the FL and WA methods. The RMSE for the grid-wise averaging method decreased from 0.21 m to 0.13 m after 20 surveys were used. The RMSE of DEMs developed using the FL and WA methods decreased from 0.20 m to 0.08 m as the

number of surveys used increased (Fig. 14). The differences in RMSE between these two methods were very small. However, the RMSE of FL method DEM varied more than that of the WA method as new surveys were used.

Possible causes of other errors were the field conditions and vehicle dynamics. As the vehicle traveled over the field surface, it interacted with micro-topography – small scale variance in the field surface. There are also variations in weight distribution of the vehicle from test to test as data collections were conducted on the same vehicle path repeatedly. The soil became more compacted and compressed lower than the soil around it as more passes were made. The temperature variation during data collection might also have caused changes in the air suspension system stiffness.

The effects of these errors were reduced when the DEMs were developed using FL and WA methods. For the FL and WA methods, kriging produced an estimate by optimally weighting surrounding measurements. The grid-wise averaging method, however, is sensitive to outliers because it produced estimates by giving each measurement in a grid the same weight. Beyond initial estimation, the FL and WA methods were robust to outliers because in the data combination process, the kriging variance was passed along as measure of confidence in the estimates based on prior sampling configurations. A smaller kriging variance indicated that the elevation estimate was more strongly supported by elevation measurements and thus should be more representative of the true elevation. When the estimates from different DEMs were combined, the FL method adjusted the estimate with less measurement support relative to that with more support before combining. The WA method weighted the estimates based on the confidence in the estimates during combination. Hence, these two methods were more robust to measurement errors and resulted in improved

performance over the grid-wise averaging method. This study demonstrates the importance of passing along a measure of estimate confidence in the process as measurements from new surveys are added.

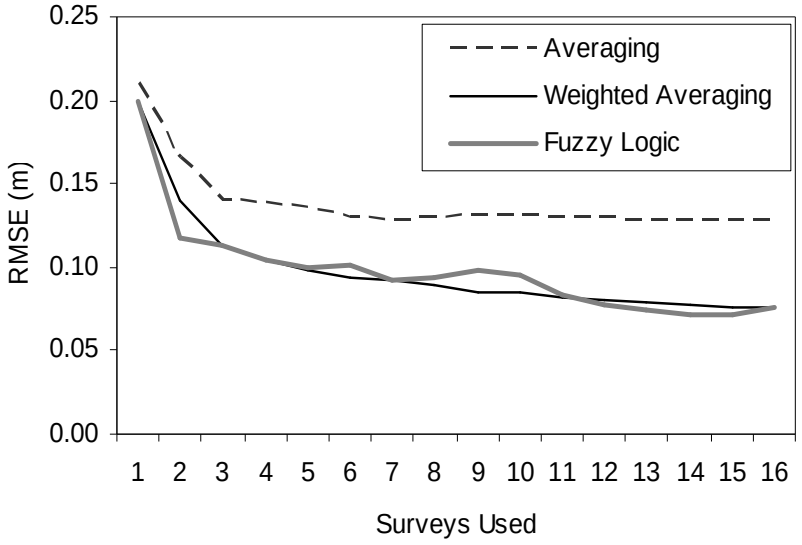

Figure 14. RMSE of DEMs as repeated RTK-DGPS surveys were combined using the methods developed for this study.

Conclusions

GPS data surveys for the development of the field DEMs were simulated using publically available USGS DEM and acquired using GPS-equipped farm vehicles. Repeated GPS surveys of elevation data improved the DEM accuracy over time. This paper presented two methods for the development of field DEMs as a by-product of GPS-aided farm operations. These methods provided means to reduce the amount of raw elevation data passed on between measurements and combined them for improved elevation estimate. From this work, the following conclusions can be drawn:

- The DEMs from the experimental RTK-DGPS field surveys developed using the FL and WA methods had average RMSE of 0.08 m after using 16 surveys which was substantially lower than RMSE of 0.60 m associated with DEM developed by averaging all data points in each grid without discontinuity error detection. Overall, two years of GPS surveys of elevation data from field operations could improve the accuracy of the field DEM by 50% relative to the first DEM.
- With minimum control of errors in elevation measurement surveys, the effect of these GPS errors can be reduced with appropriate data processing to reduce the effect of discontinuities and combine multiple survey data using methods that take into account the confidence in estimates based on their measurement support.
- With a large number of measurement surveys, fuzzy logic and weighted averaging methods had about the same performance; however, DEM error associated with the weighted averaging method decreased more consistently as more measurement surveys were used.

Acknowledgement

This research of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 3612, was supported by Hatch Act and State of Iowa funds and Malaysian Academic Staff Training (SLAB) scholarship from Malaysian Ministry of Higher Education.

References

Ackermann, F. 1999. Airborne laser scanning -- Present status and future expectations. ISPRS J. *Photogram. and Remote Sensing* 54: 64-67.

ASAE Standards. 2005. S390.4: Definition and Classifications of Agricultural Field Equipment. St. Joseph, Mich.: ASAE.

Berry, J. K., J. A. Delgado, R. Khosla, and F. J. Pierce. 2003. Precision conservation for environmental sustainability. *J. Soil and Water Cons:* 58(6): 332-339.

Bishop, T. F. A., and A. B. McBratney. 2002. Creating field−extent digital elevation models for precision agriculture. *Precision Agric*. 3(1): 37−46.

Brown, C. 2008. Plan a switch to less tillage. Wallace Farmer: Iowa Learning Farm, IA. Available at www.wallacesfarmer.com. Assessed September 10, 2008.

Buick, R. 2006. RTK base station networks driving adoption of GPS +/- 1 inch automated steering among crop growers. RTK Network White Paper: Trimble Agricultural Division, Westminster, CO. Available at: www.trimble.com. Accessed March, 31, 2007.

Clark, R. L., and R. Lee. 1998. Development of topographic maps for precision farming with kinematic GPS. *Trans. ASAE* 41(4): 909−916.

de Jong, S. M., M. L. Paracchini, F. Bertolo, S. Folving, J. Megier, and A. P. J. de Roo. 1999. Regional assessment of soil erosion using the distributed model SEMMED and remotely sensed data. *Catena* 37: 291-308.

Evans, D. L., and J. Apel. 1995. Spaceborne synthetic aperture radar: Current status and future directions. A report to the Committee on Earth Sciences, Space Studies Board, National Research Council. NASA Technical Memorandum 4679. Pasadena, Cal.: NASA Jet Propulsion Laboratory. Available at: http://southport.jpl.nasa.gov/nrc/nrcT_of_C.html Accessed 14 July 2003.

Farrell, J., and M. Barth, 1999. *The Global Positioning System and Inertial Navigation*. New York, N.Y.: MacGraw-Hill.

GeoCommunity, 2007. USGS DEM Resources: 1995-2007. GIS Data Depot, Niceville, FL.: MindSites Group, LLC Available at: http://data.geocomm.com/. Accessed 23 April 2006.

Ghidey, E., E. E. Alberts and L. A. Kramer. 2001. Comparison of measured and WEPP predicted runoff and sediment loss from deep Loess soil watershed. *Proc. Int. Conf. in Soil Erosion Research.* 131-134. J. C. Ascough II and D. C. Flanagan eds. St. Joseph, Michigan: ASAE.

Isaaks, E. H., and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. New York, N.Y.: Oxford University Press.

James, R., 1994. GPS and Differential GPS: An error model for sensor simulation. *Position Location and Navigation Symposium* 260-266, IEEE, Piscataway, NJ.

Lin, C-Y., and W-T. Lin. 2001. An automated delineation of slope length for watershed upland erosion and sediment yield estimation. *Proc. Int. Conf. in Soil Erosion Research.* 423-426. J. C. Ascough II and D. C. Flanagan eds. St. Joseph, Michigan: ASAE.

Maidment, D. R. 1996. GIS and hydrologic modeling - an assessment of progress. Proceedings of Int. Conf. of Intergrating GIS and Env. Modeling. Available at: www.ce.utexas.edu/prof/maidment/gishydro/meetings/santafe/santafe.htm. Assessed September 10, 2008.

Oost, K. V., G. Govers, and P. Desmet. 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *Landscape Ecology* 15(6): 577-589.

Ouyang, D., J. Bartholic and J. Selegan. 2005. Assessing sediment loading from agricultural croplands in Great Lake basin. *The Journal of American Sci*. 1(2): 14-21.

Patterson, T., 2001. DEM manipulation and 3-D Terrain Visualization: Techniques Used by the U.S. National Park Service. In *Cartographica*, 89-101. L. Hurni, K. Kriz, T. Patterson and R. Wheate, eds. Toronto, ON: University of Toronto Press Inc.

Renschler, C. S., D. C. Flanagan, B. A. Engel, L. A. Kramer, and K. A. Sudduth. 2002. Site−specific decision−making based on RTK GPS survey and six alternative elevation data sources: Watershed topography and delineation. *Trans. ASAE* 45(6): 1883−1895.

Ritsema, C. J., K. O. Trouwborst, V. Jetten, S. Ledin, L. Rui, L. Baoyuan, and F. Boije. An interdisciplinary approach for soil and water conservation planning to improve the sustainability of land-use on the Loess Plateau in China. *Proc. Int. Conf. in Soil Erosion Research.* 431-434. J. C. Ascough II and D. C. Flanagan eds. St. Joseph, Michigan: ASAE.

Sarangi, A., C. A. Cox, C. A. Madramootoo. 2007. Evaluation of the AnnAGNPS model for prediction of runoff and sediment yields in St Lucia watersheds. *Biosystems Engineering* 97: 241-256.

Scherzinger, B., J. Hutton, and M. Mostafa. 2007. Chapter 10: Enabling technologies. In *Digital elevation model technologies and applications*: *The DEM Users Manual, 2nd Edition*, 354-355. D.F. Maune, ed. Bethesda, Maryland: ASPRS.

Schmidt, F., and A. Persson. 2003. Comparison of DEM data capture and Topographic Wetness Indices. *Precision Agriculture* 4: 179-192.

Sidler, R., 2003. Kriging and conditional geostatistical simulation based on scaled-invariant covariance models. Diploma Thesis. Institute of Geophysics, Department of Earth Science, Swiss Federal institute of Technology Zurich.

Trangmar, B. B., R. S. Yost, and G. Uehara. 1985. Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy* 38: 45−94.

Vaze, J. and J. Teng. 2007. High resolution LiDAR DEM – How good is it? *Proc. International Congress on Modelling and Simulation* 692-698. L. Oxley, and D. Kulasiri eds. Caberra, Australia: MODSIM.

von Ka´rman, T. V. 1948. Progress in the statistical theory of turbulence. *Proc. Natl. Acad. Sci* 34: 530–539.

Westphalen, M. L., B. L. Steward, S. Han. 2004. Topographic mapping through measurement of vehicle attitude and elevation. *Trans. ASAE* 47(5): 1841−1849.

Wilson, J. P. and J. C. Gallant. 2000. *Terrain Analysis: Principles and Applications*. New York: John Wiley & Sons, Inc.

Wilson, R. C., R. S. Freeland, J. B. Wilkerson, and W. E. Hart. 2005. Interpolation and data collection error sources for electromagnetic induction-soil electrical conductivity mapping. *Applied Engineering in Agric*. 21(2): 277−283.

Wise, S. M. 1998. The effect of GIS interpolation errors on the use of digital elevation models in geomorphology. In *Landform Modelling and Analysis*, 139−164. S. N. Lane, K. S. Richards, and J. H. Chandler, eds. Chichester, U.K.: John Wiley and Sons.

Zhang, Q., and S. Han. 2002. An Information table for Yield Data Analysis and Management. *Precision Agriculture* 83(3):299-306.

CHAPTER 4. TARGETED SAMPLING OF ELEVATION DATA BASED ON SPATIAL UNCERTAINTY OF PRIOR MEASUREMENTS

Presented at ASABE Annual International Meeting, 2008. ASABE paper no: 084826

S. Abd Aziz, B. L. Steward, M. Karkee

Abstract

Field sampling can be a major expense for planning within-field management in precision agriculture. An efficient sampling strategy should address knowledge gaps, rather than exhaustively collect redundant data. Modification of existing schemes is possible by incorporating prior knowledge of spatial patterns within the field. In this study, spatial uncertainty of prior DEM estimates was used to locate targeted sampling regions in the field. An agricultural vehicle equipped with RTK-DGPS was driven across a 2.3 ha field area to measure the field elevation in a continuous fashion. A geostatistical simulation technique was used to simulate field DEMs using measurements with different pass intervals and to quantitatively assess the spatial uncertainty of the DEM estimates. The high uncertainty regions for each DEMs were classified using image segmentation methods and targeted sampling was performed on those regions. The addition of targeted measurements significantly reduced the time dedicated for the re-sampling effort and resulted in DEMs with lower RMSE. For the widest interval between sampling passes, the RMSE of 0.46 m of the DEM was reduced to 0.25 m after adding the targeted measurements which was close to the 0.22 m RMSE of DEM with whole field re-sampling. The estimated sampling time for targeted sampling was more than 50% lower than re-sampling the whole field. The results

show that spatial uncertainty models are useful to design targeted sampling to help reduce the cost while maintaining the accuracy of the measurements. The method is not limited to map elevation data but can be extended for mapping other spatial data.

Introduction

Precision agriculture is a farming system aims on improving yields and product quality while reducing input cost and minimizing environmental impact. The important key to efficient and effective precision agriculture is to match resource inputs to the spatial and temporal variability of attributes within farm fields through site-specific management. In the past, managers used estimates of average conditions of farm attributes for the whole field and treated farm fields uniformly as single units. Site-specific management, however, requires an understanding of spatial variability within the field, and hence sampling is needed to estimate attributes at a finer than whole-field scale.

Field sampling can be a major expense for planning within-field management in precision agriculture. Locating the samples inappropriately or taking more samples than are needed can result to extra expense. Taking too few samples on the other hand, may not help understanding the variability within the field. Conventionally, grid sampling was used in gathering field attributes. Sample points were located at the nodes or centers of square, rectangular or other regular shaped grids on the field, where the locations can be established and maintained using GPS. Gridded schemes are convenient to locate and analyze, but, like traditional simple random sampling schemes, may be inefficient to precisely capture the spatial variability of the attributes and somewhat ignores actual local variability.

Recently, continuous vehicle-based sampling has been widely investigated due to proliferation of automatic guidance systems on agricultural vehicles with high-accuracy GPS capability and advance sensor technology. It requires less labor and offers a rapid and relatively easy way for producers to obtain field data. Example includes vehicle-mounted GPS systems to collect elevation data (Clark and Lee, 1998; Westphalen et al., 2004; Schmidt et al., 2003), continuous soil sampling systems to sample soil attributes on-the-go (Kataoka et al., 2004), an autonomous underground *Soil Scout* for monitoring soil properties like moisture, temperature, nutrient level and pH (Tiusanen, 2006) and electrical conductivity (EC) mobile sensors to measure soil EC continuously in the field (Grisso et al., 2007; Ehsani and Sullivan, 2006). Vehicle-based sampling is characterized as highly dense data along the travel path and no samples between the paths. Again, like grid sampling, the question comes back to where exactly to sample to efficiently capture the variability in the field.

An efficient sampling strategy should address knowledge gaps rather than exhaustively collect redundant data. Hence, a “smart sampling” plan should be conducted for efficient data collection and improve estimates of the variability. Modification of existing schemes is possible by incorporating prior knowledge of spatial variability within the field. Field elevation in the form of digital elevation models (DEMs) is among the most important attributes that can provide information relating the spatial variability in the field (Kravchenko, A. N. and Bullock, 2000; Kaspar et al., 2003; Rampant and Abuzar, 2004; Jiang and Thelen, 2004). This paper reports on research to investigate a method to efficiently implement vehicle-based sampling to collect elevation data in the farm fields.

Many studies have sought to quantify DEM accuracy and compare the accuracy of DEMs produced using different data sources and production methods. By comparing with

higher accuracy data sources, measures such as standard deviation or root mean square error (RMSE) are typically used to represent the DEM quality. Such non-spatial statistics are global measures and specifically do not provide an accurate assessment of how precise each grid in a DEM represents a true elevation (Wise, 1998; Wechsler, 2007). Moreover, in the absence of higher quality data, it is impossible to quantify such measures. A number of authors recognized that spatial simulation methods can be used for uncertainty assessment of DEM quality (Hunter et al., 1995; Holmes et al., 2000; Carlisle, 2005; Wechsler and Kroll, 2006). The simulation process accounts for spatial correlation in the data to produced multiple estimates (realizations) for each particular location in the DEM. These realizations provide a range within which the true estimate lies (Wechsler, 2007). The variance of the elevation realizations for each DEM grids can be used as an uncertainty measure of the estimate in the grids.

In this paper, the uncertainty and spatial distribution of elevation estimates from prior measurements was used as a rational basis for a future sampling plan to improve the accuracy of field DEMs. The uncertainty of elevation estimates across the DEMs was assessed using geostatistical simulation technique to delineate the regions in the field that needed to be re-sampled. Additional samples can be targeted and obtained from specified locations rather than re-sampling the whole field. The objective of this study is to develop a targeted sampling method based on spatial uncertainty of prior measurements for topographic mapping.

Methodology

Field Study and Data Preparation

Data were collected from a portion of 6.5 ha (16-acre) field that had been chisel-plowed after the previous corn crop had been harvested. Elevation data were collected using a self-propelled agricultural sprayer (model 4710, Deere & Co., Moline, Ill.) equipped with real-time kinematic differential GPS (RTK DGPS) receivers (StarFire RTK, Deere & Co., Moline, III) operating at 1 Hz with a vertical static RMSE of less than 1.5 cm. The GPS receiver was mounted at a height of 3.81m above the field surface. The vehicle was driven over a 2.3 ha (5.7-acre; 247.55 m wide by 294.96 m long) area of the field at a speed between 6.4 to 9.7 km/h (4 to 6 mph) along northwest-southeast in a headland pattern with opposite travel directions on adjacent paths (Westphalen et al., 2004). The passes were 3.05 m (10 ft) apart.

Since the raw data is in the format of a geographic coordinate system consisting of longitude, latitude, and altitude, data projection was done to convert the raw data set into a projected coordinate system. Projection was required for spatial data analysis so that analysis proceeded using units of length in the horizontal plane. The standard USGS Universal Transverse Mercator (UTM) format was used with UTM grid zone of 15N for the coordinate projection with all units for easting, northing and elevation in meters.

Vehicle-based RTK-DGPS accuracy relies on the continued availability of differential corrections broadcast from dedicated base station receivers. Loss or interruption of the DGPS correction signal will affect the GPS positioning and attitude measurement,

which introduces errors in the range of centimeters. Errors may also occur when satellites appear or leave the field of view during the GPS data collection.

An algorithm was developed to detect measurement discontinuities for data correction. Discontinuity correction in the horizontal plane was accomplished by shifting sequential measurements to minimize discontinuities along the vehicle path. The discontinuities in elevation measurements were corrected by re-estimating the value using the mean of the nearest high accuracy neighboring points.

Every other measurement point along the travel passes was sub-sampled and used as the calibration group. The remaining measurements were used as validation group to measure the quality of the simulated elevation. To simulate the calibration data group was jackknifed into seven separate sub-groups by skipping data along passes at a regular interval. It started with skipping every one pass to produce measurement consisted of every second pass of vehicle measurements. Consequently the number of passes skipped was increased until the widest interval of every eighth pass (seven passes skipped). These subgroups corresponded to intervals of 6.10 m, 9.15 m, 12.20 m, 15.25 m, 18.30 m, 21.35 m and 24.40 m between passes, respectively. These seven datasets became the initial-sampling data from which the field DEMs were simulated to assess the uncertainty in the elevation estimates.

Uncertainty Assessment

The spatial uncertainty of the elevation is modeled using a conditional geostatistical simulation method. The simulations model the uncertainty in the elevation spatial distribution based on the data sources available near each point of the DEM grids. Likely realizations of elevation estimates were randomly drawn from the possible distribution of the elevation

provided by the local conditional cumulative distribution function (ccdf) for each grid. The advantage of using this technique is that it preserves the nature of real world variability and spatial correlation in the estimates without the smoothing of the interpolated estimates which usually occur in kriging (Goovaerts, 1997). Among many other conditional simulations techniques, sequential Gaussian simulation (SGS) is by far the most widely used to estimate continuous variable like elevation; because of the inherent structure of the Gaussian model makes determining local ccdf fairly straight forward. Using this technique, simple kriging is used within the simulation routine to establish uncertainty models of elevation estimates at every DEM grid location. Multiple realizations of elevation estimates were randomly drawn from the ccdf derived based on the kriging estimate and its associated variance. The sample variogram of the data was fit with a linear variogram model with a 20 m lag distance and zero nugget effect. The search radius of the kriging estimator was set to the range of the variogram and a minimum of 16 data points. A total of 100 simulations were run resulting in 100 realizations in each DEM grid. The average of the realizations in each grid was calculated to produce the mean estimate which is also known as E-type estimate of the grid. The E-type estimate across the DEM was used to produce the map of DEM estimates. The variance of the realizations also known as conditional variance was used to quantify the uncertainty of the DEM estimates. Detailed descriptions of SGS algorithm can be found in Goovaerts (1997). The *gstat* program in R statistical software (Free Software Foundation, Inc., Boston, MA) was use to perform the SGS.

Targeted Sampling

The conditional variance quantified in each grid from the simulation process was used as the uncertainty estimate of the DEM. This produced conditional variance maps which were used to characterize the regions that need to be re-sampled. In Matlab (The Mathworks, Natick, Mass.), an image segmentation algorithm was performed using a simple thresholding technique, for our study we chose the estimation variance threshold to be 0.04 m^2 because the histogram of the variances showed a distinguished separation at 0.04 m^2. Region classification was performed on the conditional variance maps to classify the regions that exceed the threshold value. This is done by allocating a binary value equal to 1 in every grid in that region. Zeros binary value were assigned to the region that has value less than the threshold value. The process essentially transformed the DEM into a binary image by allocating every grid cell in the DEM as either black or white, depending on their value. The algorithm proceeded with morphological operations to filter segmentation noise and scattered unconnected pixels. Scattered unconnected pixels may correspond to random noise introduced from SGS and should not be considered a valid region of interest. The Matlab morphological operations function *bwmorph* was used to perform a 'cleaning' operation, followed by 'filling' and 'removing' operations to remove the noise.

In the application of targeted sampling, new samples should be taken only in the regions of interest. Hence, unused measurement passes that fell in the regions that exceed the estimation variance threshold were added to each initial-sampling sub-group. Only one unused pass in between initial measurement passes in the delineated region were used to uniformly simulate the effect of adding new targeted sample data within the division of data

sub-groups. Then SGS was performed on the new sampling sets to produce an improve DEM estimates, as well as its associated uncertainty.

For comparison purposes, non-targeted sampling was conducted for each data sub-groups. Unused passes were added in between initial measurement passes across the whole study area. Again, only one unused pass in between initial measurement passes was used to uniformly simulate the effect of adding new non-targeted sampled data within the division of data subgroups.

Data Analysis

Sampling Time

The amount of time spent to collect data for each sub−group within each sampling type was estimated based on the travel distance and the vehicle speed used for travelling along the passes as well as making turns. As the speed when traveling along the passes was in the range of 6.4 to 9.7 km/h (4 to 6 mph), the minimum speed, 6.4 km/h was used to estimate the travel time. The speed which making turns between passes was slower and estimated around 3.2 km/h (2 mph).

DEM error Estimation

Each generated DEM from each sampling types and calibration sub-groups was compared to the validation dataset from the validation group which had not been used to simulate the surface. RMSE, the typical measure of DEM error (Wise, 1998), was calculated by subtracting the elevation of the nearest estimated point from that of each validation point.

The DEMs produced from initial-sampling were used as the control to evaluate the effect of adding new targeted and non-targeted sample data in mapping the field elevation.

Slope Estimation

One of the common needs in quantitative DEM interpretation is to determine the slope which is the rate of elevation change in the direction of the steepest descent. DEM slope is frequently used to determine water flow direction in hydraulic analysis or surface erosion and environmental impact in agricultural and environmental studies. To study the effects of sampling procedures on slope prediction, the slope derivatives from each generated DEMs were calculated using ArcGIS (Version 9.2, ESRI, Redlands, Cal.). The DEMs were imported into ArcGIS and a slope calculation extension was used in the ArcMap Spatial Analyst to automatically calculate the slope. The accuracy of the slope was quantified by comparing the estimated value with the slope derived from the DEM developed using validation data. The RMSE was calculated by subtracting the estimated slope in each grid from the slope value in the corresponding grid of the validation DEM.

Slope Uncertainty

Calculation of slope in ArcGIS is based on the first partial derivatives of elevation, z (Burrough and McDonell, 1998):

$$p = \frac{\partial z}{\partial x}, \tag{1}$$

$$q = \frac{\partial z}{\partial y}. \tag{2}$$

where p is the change of height in the directions of x (easting) and q is the change of height in the direction of y (northing) axes. The values of the partial derivatives were in 3 x 3 neighborhoods of elevation points approximated using equations 3 and 4. The top row of the 3 x 3 neighborhood points are represented by z_1, z_2, z_3, the middle row by z_4, z_5, z_6 and the bottom row by z_7, z_8, z_9. The distance between adjacent points or the grid size is denoted by w.

$$p \approx \frac{(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)}{8w} \tag{3}$$

$$q \approx \frac{(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)}{8w} \tag{4}$$

The slope, S of a grid was calculated as a change of height within the distance unit shown in equation 5.

$$S = \sqrt{p^2 + q^2} \tag{5}$$

Based on this formulation, the uncertainties in the slope, ΔS were calculated using the sensitivity coefficients with respect to the nine neighboring estimates, z_i, each with their own uncertainty, Δz_i (obtained from the conditional simulation method):

$$\Delta S = \sqrt{\left(\frac{\partial S}{\partial p}\right)^2 (\Delta p)^2 + \left(\frac{\partial S}{\partial q}\right)^2 (\Delta q)^2}, \tag{6}$$

where

$$\Delta p = \sqrt{\sum \left(\frac{\partial p}{\partial z_i}\right)^2 (\Delta z_i)^2}, \tag{7}$$

and

$$\Delta q = \sqrt{\sum \left(\frac{\partial q}{\partial z_i}\right)^2 (\Delta z_i)^2}. \tag{8}$$

The uncertainty of the derived slope for each DEM was visually assessed using a contour plot of the mean elevation and the variance.

Results

The conditional variance maps produced using SGS reveal clear correlation of the uncertainties in DEM with the slope of the land surface (Figure 1). A visual inspection of the maps shows conditional variance is larger at the steepest area (northeast) of the fields where elevation values change the most. The variance values ranged from 0.1 to 0.16 m^2 in this area. The uncertainty is small in the south and northwest of the study area where elevation is flatter (plain region). The variance ranged from 0 to 0.04 m^2 in this area.

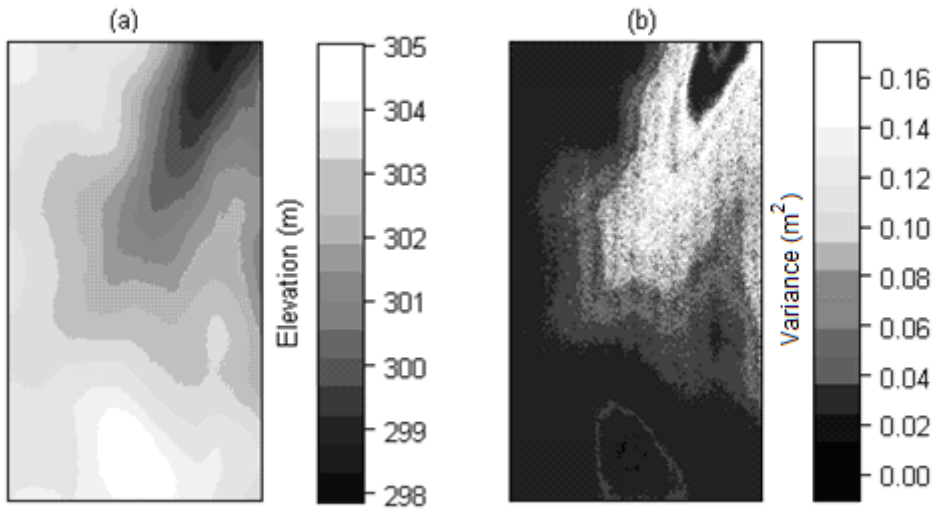

Figure 1: (a) The E-type estimates map and its corresponding (b) conditional variance map of DEM generated from SGS using measurements with 6.05 m pass intervals.

The histograms of the values of the grids in the conditional variance maps across measurements subgroup were plotted to verify the appropriateness of the chosen threshold value (Figure 2). The histograms show a strong multimodality because the simulation process relies not only on the variability of the elevation values but also on the distance to the sampling measurements. As the measurements were collected systematically along parallel

passes, the simulation process seemed to capture the pattern. In all cases, the mode with the highest frequency had values ranging from 0 to 0.04 m^2 and was clearly separated from the other modes. This distribution corresponded to grids that have little change in elevation and were situated closer to sampling measurements. The sub-group with 6.05 m^2 measurement intervals has variances distribution in the smallest range, from 0 to 0.19 m^2, relative to other measurement sub groups. As the measurement interval increased, the distribution of the variances spread to larger ranges. Thus the 0.04 m^2 variance threshold was adequate to classify the variance estimates into high and low uncertainty classes.

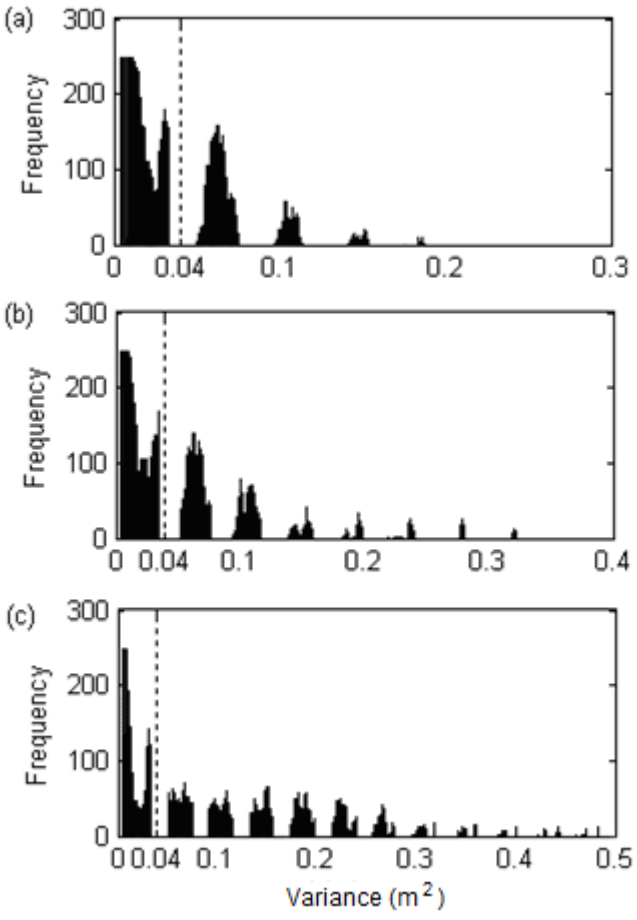

Figure 2: Histograms of SGS variance estimates using measurements passes at (a) 6.05 m, (b) 15.25 m and (c) 24.40 m intervals.

After segmentation and morphological operations, the field was classified into high uncertainty and low uncertainty regions. Intuitively, the sparser the measurement passes, the more uncertain the estimated values were. Visually, the DEM developed using measurements with passes interval of 24.4 m has the largest high uncertainty region (Figure 3 (c)). The high uncertainty region for DEM developed using measurements with passes interval of 6.10 m was smaller and located at region where elevation values change the most (Figure 3(a)). In this case, the SGS captured the actual elevation variability.

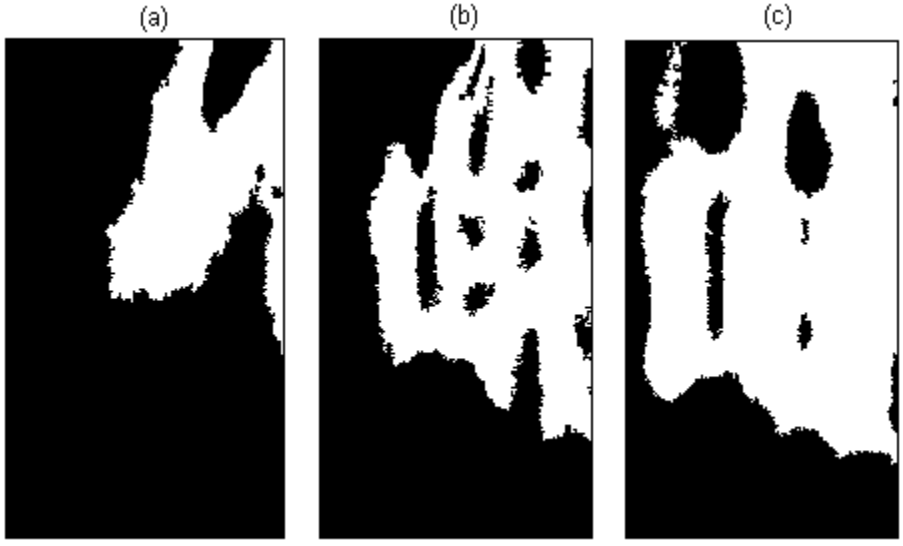

Figure 3: The conditional variance maps for measurements at (a) 6.05 m, (b) 15.25 m and (c) 24.40 m passes intervals were transformed into binary images. Targeted regions (white) were classified using the image thresholding technique in Matlab followed by the Matlab morphological operations functions such as 'cleaning', 'filling' and 'removing'.

The size of the high uncertainty regions decreased as the interval width of measurements passes used in data sampling decreased (Figure 4). This result shows that besides elevation variability, the uncertainty also depends on the distance between the estimates and the sampling locations. For this study field, the sampling measurements with

interval width less than 10 m adequately captured the spatial variability in the elevation and have uncertain regions of about 3,500 m^2. Although the number of measurement passes for 9.15 m was substantially lesser than 6.10 m sub-group, the high uncertainty region area was about the same for both interval sub-groups (Table 1). This result shows that given this information, one might want to sample at 9.15 m interval rather than 6.10 m interval because both would capture similar variability of the field. With interval widths larger than 10 m, the high uncertainty regions ranged from 8,700 to 12,400 m^2. The substantial jump of uncertain region area between 9.15 m and 12.2 m in the graph was due to the effect of skipping the measurement passes. Skipping three or more measurement passes resulting to substantially lesser measurement passes for analysis (Table 1).

Table 1: Number of measurement passes used for each measurement interval sub-group.

Interval Subgroup	Number of passes skipped	Number of measurement passes used for analysis
6.10 m	1	13
9.15 m	2	9
12.20 m	3	6
15.25 m	4	6
18.3 m	5	5
21.35 m	6	4
24.4 m	7	3

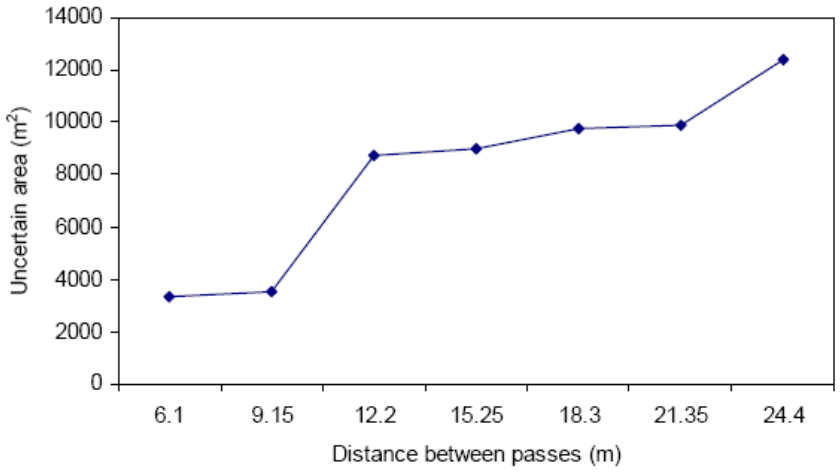

Figure 4: Targeted regions characterized based on conditional variance of DEMs increases as the interval width between passes increases.

For each measurement subgroup, targeted sampling was located in the high uncertainty regions by adding a measurement pass in between the initial (first) measurement passes (Figure 5).

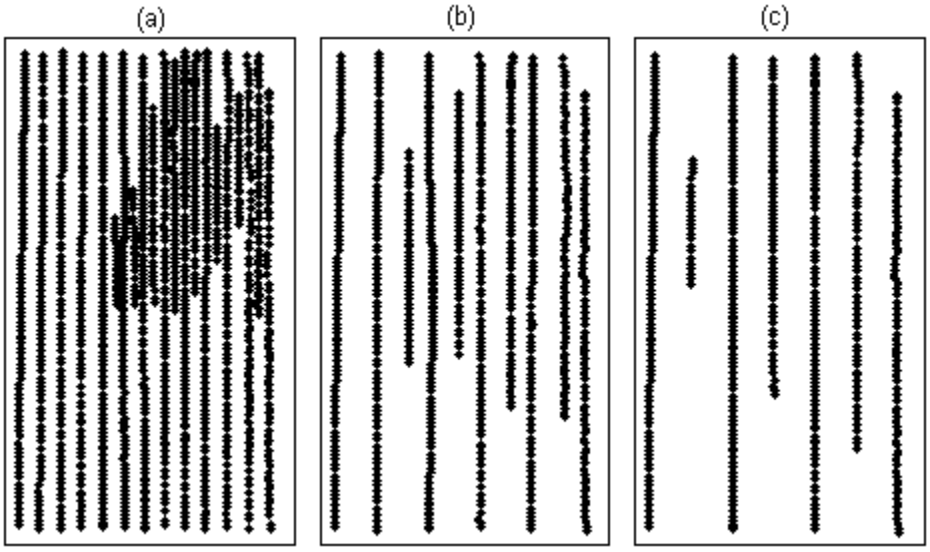

Figure 5: Measurements passes at (a) 6.05 m, (b) 15.25 m and (c) 24.40 m intervals with additional targeted measurements between the passes.

The collection time to additionally targeted and non-targeted sample was estimated across measurement subgroups. For both cases, the estimated time decreased as the distance between passes increased (Figure 6). The estimated time ranged from around 16 minutes to an hour for measurements with additional non-targeted sampling and around 11 to 35 minutes for measurements with additional targeted sampling. Targeted sampling significantly reduced the time for re-sampling. This reduction is important in minimizing the cost of data.

The RMSE of DEMs developed using measurements subgroups and with additional targeted and non-targeted measurements between the passes increased as the distance between passes increased (Figure 7). Additional targeted and non-targeted sampling significantly reduced the RMSE of the DEMs developed using the initial (first) measurements. For the smallest measurement interval of 6.10 m, the RMSE of the DEM was 0.08 m and decreased to 0.07 and 0.05 m with additional targeted and non-targeted measurements respectively. For the widest measurement interval of 24.40 m, the RMSE of the DEM was 0.45 m and decreased to 0.25 and 0.22 m with additional targeted and non-targeted measurements respectively. Although the RMSEs of DEM developed with additional targeted measurements are slightly higher than with the additional non-targeted measurement, the estimated time spent for targeted sampling was substantially lower than non-targeted sampling. For distance between passes of 15.25 m, the RMSE for sampling with additional targeted and non-targeted measurements were not much different from each other (0.14 and 0.13 m respectively). However, the estimated sampling time was more than 50% lower than for non-targeted sampling. The targeted sampling method could help reduced the data collection time which may result in lower cost while maintaining the accuracy of the measurements.

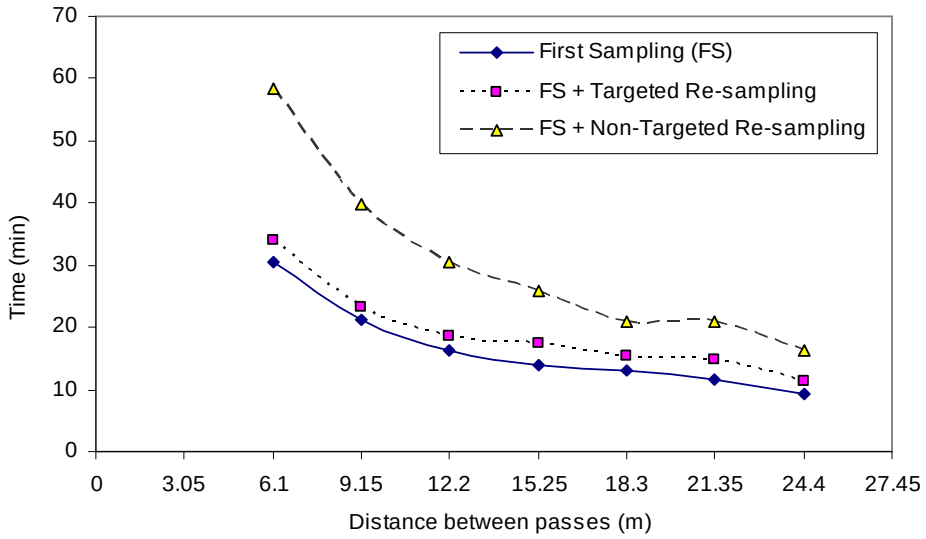

Figure 6: Estimated time to collect the elevation data with additional targeted and non-targeted measurements across distance between passes.

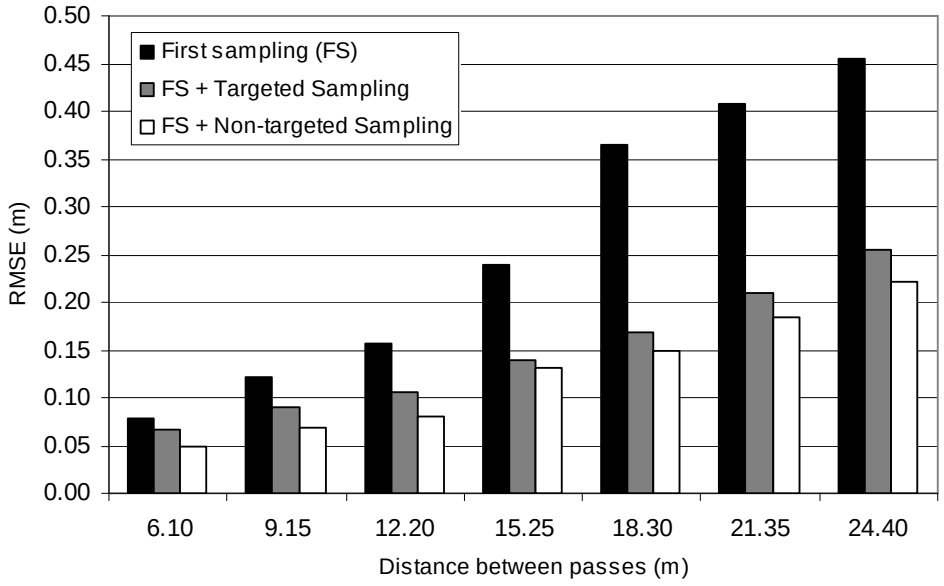

Figure 7: RMSE of DEMs developed using measurements across different passes intervals and with additional targeted and non-targeted measurements between the passes.

The RMSE of the slope estimates increased as the interval distance between measurements passes increased (Figure 8). Additional measurements slightly improved the slope estimation for smaller measurement intervals, and more significant improvement was observed for larger measurement intervals. For the smallest measurement interval of 6.10 m, the RMSE of the slope derived from the DEM was 1.6% and decreased to 1.5 and 1.4% with additional targeted and non-targeted measurements respectively. For the widest measurement interval of 24.4 m, the RMSE of the slope derived from the DEM was 2.8% and decreased to about 2.2% with additional targeted or non-targeted measurements. The difference of slope RMSE between DEMs with additional targeted and non-targeted measurement was very small, hence the targeted sampling which requires less time for data collection is preferable.

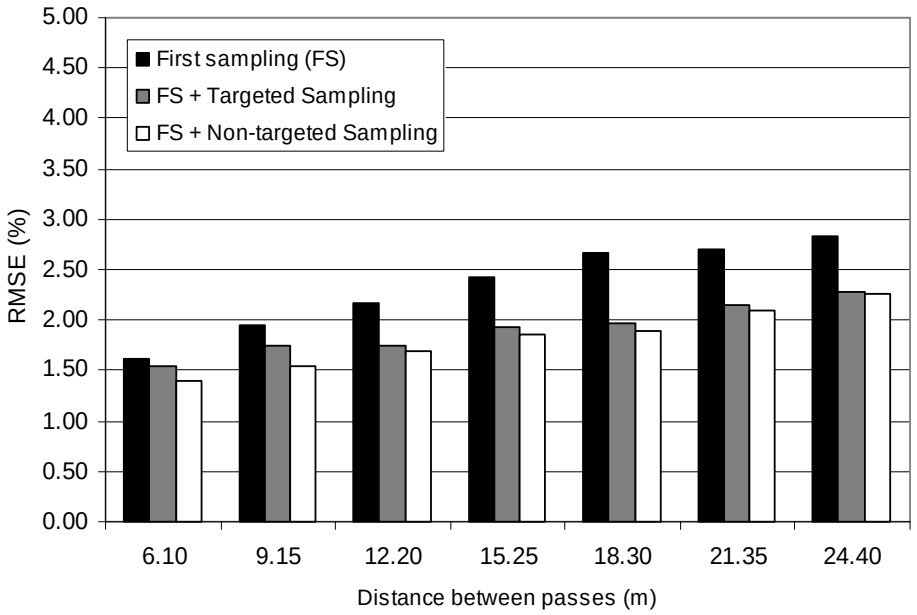

Figure 8: RMSE of slope derived from DEMs developed using measurements across different passes interval and with additional targeted and non-targeted measurements between the passes.

Generally, the additional re-sampled measurements led to better estimation of the field DEM and its derived slope parameter. The quantitative results were confirmed by visual inspection of contour plots and generated from the DEMs at different passes intervals (Figure 9). The addition of measurements either through the targeted or non-targeted sampling led higher spatial frequency content in the contour lines. This higher frequency content may indicate that these DEMs are resolving on real topographic features as confirmed by the statistical error measures in some cases. For the DEMs developed using measurement passes at 24.4 m interval, the sparcity of data led to substantial distortion in the DEM interpolated from the first sampling. The distortion was reduced with the addition of measurements either through the targeted or non-targeted sampling.

The calculated slope ranged from 0 to about 13% (Figure 10). The maps of the slope show clear pattern of surface changes related to the DEMs. The pattern of the slope changes was visibly more related to the DEM as the additional targeted or non-targeted measurements were added. The estimated uncertainty of the slope derivation exhibits a pattern similar to the estimated conditional variance of the DEM. For measurement passes at 15.5 m intervals, the slope uncertainty ranged to around 0.05%. The addition of measurements passes either through the targeted or non-targeted sampling substantially reduced the uncertainty of the derived slope (Figure 10). The information about uncertainty in the slope derivatives may be useful to study the propagation of error induced from deriving the parameter from the simulated elevation estimates.

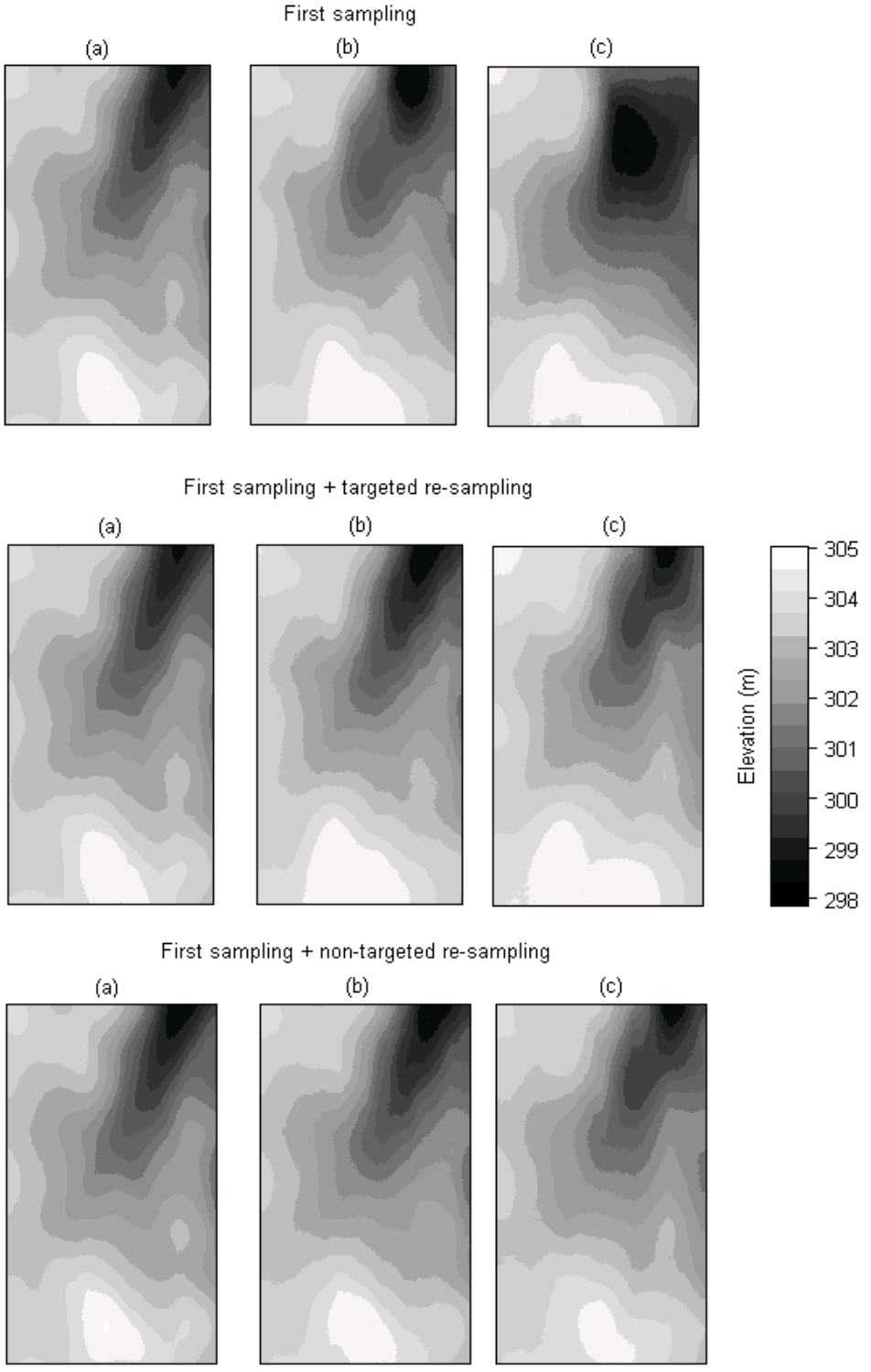

Figure 9: Plots of 0.35 m contours of DEMs using measurements at (a) 6.05 m, (b) 15.25 m and (c) 24.40 m passes intervals.

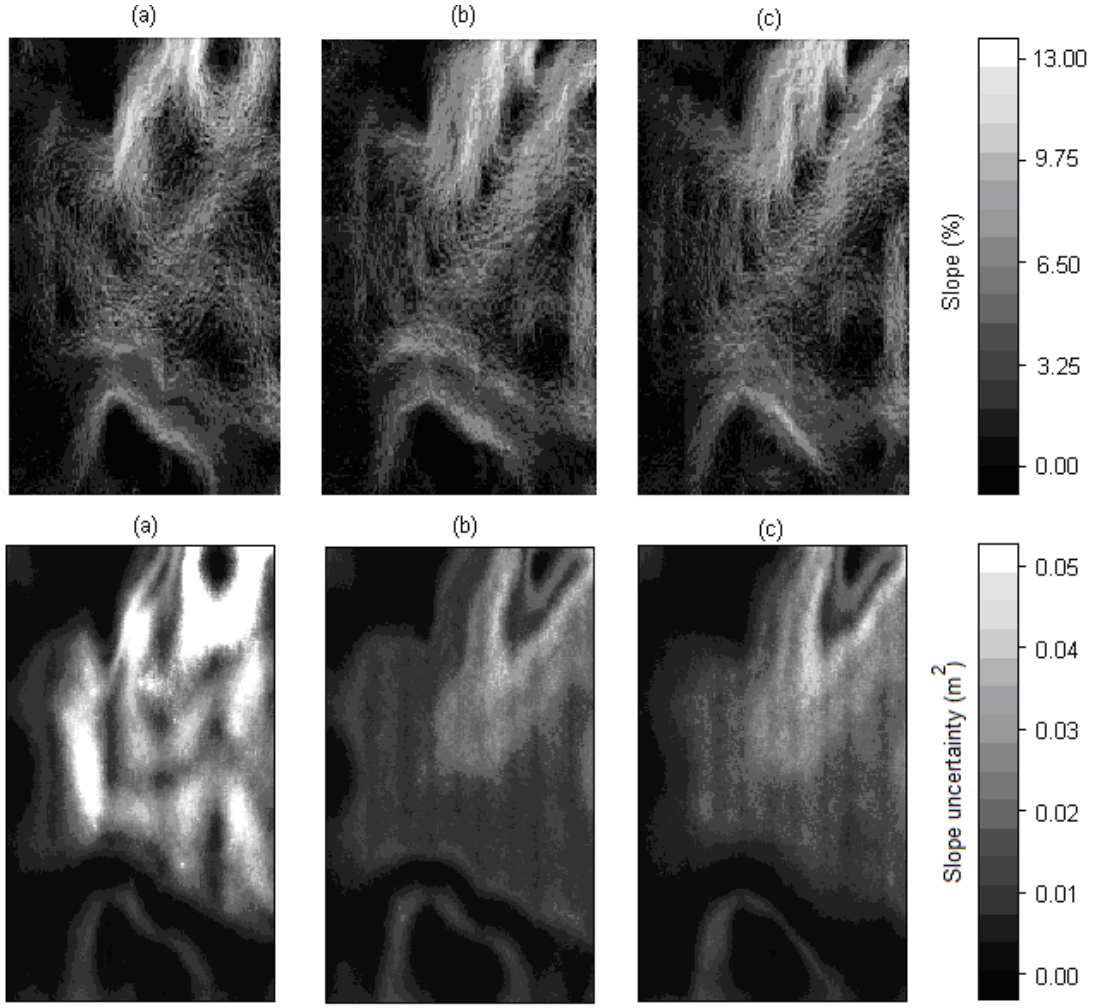

Figure 10: (a) Plots of slope using measurements at 15.25 m passes intervals (top) and its associated uncertainty maps (bottom). Additional targeted and non-targeted measurements were added to generate maps in (b) and (c) respectively.

Conclusions

From this study, a few conclusions can be drawn:

- Uncertainty assessment using SGS quantified the variability of attributes in the field based on available sampled data. The information may aid producers in designing a more efficient sampling strategy by targeting only regions of interest

in the field for re-sampling consideration. Over all interval widths of the measurement passes, the introduction of targeted measurements reduced the time required for data collection and resulted in DEMs with relatively low RMSE values. Use of targeted sampling procedure may efficiently aid farm attribute estimation for site specific management practice.

- The addition of targeted measurements significantly reduced the RMSE of slopes derived from DEMs generated using measurement passes at different interval widths. The information about uncertainty in spatial attribute estimation is useful to study error propagation induced from deriving parameters of interest related to the attributes.

References

Burrough, P. A., and R. McDonnell. 1998. *Principles of geographical information systems.* Oxford Uni. Press, New York.

Carlisle, B. H. 2005. Modelling the spatial distribution of DEM error. *Transactions in GIS* 9(4):521-540.

Clark, R. L., and R. Lee. 1998. Development of topographic maps for precision farming with kinematic GPS. *Trans. ASAE* 41(4): 909−916.

Eshani, R., and M. Sullivan. Soil Electrical Conductivity (EC) Sensors. Extension Factsheet, AEX- 565-02: Ohio State University Extension. Available at http://ohioline.osu.edu/aexfact/pdf/0565.pdf. Accessed March 3, 2006.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation.* Oxford Uni. Press, New York.

Grisso, R., W. G. Wysor, D. Holshouser, and W. Thomson. 2007. Precision farming tools: Soil electrical conductivity. Virginia Cooperative Extension Publication 442-508: College of Agriculture and Life Sciences, Virginia Polytechnic Institute and State University. Available at: http://www.ext.vt.edu/pubs/bse/442-508/442-508.pdf. Accessed March 7, 2008.

Holmes, K. W., O. A. Chadwick, and P. C. Kyriakidis. 2000. Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology* 233: 154-173.

Hunter, G. J., M. Caetano, and M. F. Goodchild. 1995. A methodology for reporting uncertainty in spatial database products. *URISA Journal* 7(2): 11-21.

Jiang, P., and K. D. Thelen, 2004. Effect of soil and topographic properties on crop yield in a north-central corn-soybean cropping system. *Agronomy Journal* 96: 252-258.

Kaspar, T. C., T. S. Colvin, D. B. Jaynes, D. L. Karlen, D. E. James, D. W. Meek, D. Pulido, and H. Butler. 2003. Relationship between six years of corn yields and terrain attributes. *Precision Agric*. 4(1): 87−101.

Kataoka, T., M. Saito, H. Okamoto, S. Hata, and M. R. Ehsani. Development of Continuous Soil Sampling Machine System. ASAE Paper No. 041047. St. Joseph, Mich.: ASAE.

Kravchenko, A. N. and D. G. Bullock. 2000. Correlation of corn and soybean grain yield with topography and soil properties. *Agron. J.* 92(1): 75-83.

Rampant, P. and M. Abuzar, 2004. Geophysical tools and Digital Elevation Models: tools for understanding crop yield and soil variability. Supersoil 2004: Proceedings of the 3rd Australian New Zealand Soils Conference. Available at: http://www.regional.org.au/au/asssi/. Accessed March, 9,2007.

Schmidt, J. P., R. K. Taylor, and R. J. Gehl. 2003. Developing topographic maps using a sub−meter accuracy global positioning receiver. *Applied Eng. In Agric.* 19(3): 291−300.

Tiusanen, J. 2007. Validation and results of the Soil Scout radio signal attenuation model. *Biosystems Engineering* 97(1): 11-17.

Wechsler, S. P. 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrol. *Earth Syst. Sci*. 11: 1481-1500.

Wechsler, S. P., and C. N. Kroll. 2006. Quantifying DEM uncertainty and its effect on topographic parameters. *Photogrammetric Engineering & Remote Sensing* 72(9): 1081-1090.

Westphalen, M. L., B. L. Steward, and S. Han. 2004. Topographic mapping through measurement of vehicle attitude and elevation. *Trans. ASAE* 47(5): 1841−1849.

Wise, S. M. 1998. The effect of GIS interpolation errors on the use of digital elevation models in geomorphology. In *Landform Modelling and Analysis,* 139−164. S. N. Lane, K. S. Richards, and J. H. Chandler, eds. Chichester, U.K.: John Wiley and Sons.

CHAPTER 5. ASSESSING THE EFFECTS OF DEM UNCERTAINTY ON SOIL LOSS ESTIMATION IN AGRICULTURAL FIELD

A paper to be submitted to Transactions of the ASAE.

S. Abd Aziz, B. L. Steward, M. Karkee

Abstract

The slope length and steepness (LS) factor is one of the factors in the Revised Universal Soil Loss Equation (RUSLE) needed to predict average annual soil loss. The LS factor is often derived from digital elevation models (DEM). DEM errors and uncertainty could affect LS factor estimation and consequently soil loss prediction. However, DEM uncertainties were not always accounted for and the effects were not always evaluated in soil loss prediction. This study compared the soil loss prediction of a 62.81 ha agricultural crop area using a 7.5-minute US Geological Survey (USGS) DEM and DEMs developed using RTK-DGPS and dual frequency (DF)-DGPS field surveys. Spatial prediction and uncertainty analysis was carried out using sequential Gaussian simulation (SGS). A total of 50 equiprobable DEM realizations were produced using SGS to assess DEM uncertainty and quantify its effect in the soil loss prediction. DEM uncertainty substantially affects the resulting soil loss prediction. The uncertainty of the annual soil loss estimates across the study field was represented as the 95% confidence interval (CI). For DF- DGPS DEM and USGS DEM, the percentage of the field that have soil loss CI value greater than two tons/acre/year were 20% and 30% which were substantially larger than the percentage area in RTK DEM (0.41%). Average annual soil loss map showed that USGS DEM contained

artifacts and underestimated the soil loss prediction in many areas of the field. The results suggested that higher accuracy DEMs such as generated using RTK-DGPS measurements are more appropriate for soil loss prediction in an agricultural field. Quantification of the DEMs uncertainty and its effect on the soil loss prediction was useful to better judge the reliability of the result.

Introduction

Erosion is one of the most important agricultural management problems. Soil erosion due to water occurs through detachment and transport of soil from land by water, including runoff from melted snow and ice. Topography is a major factor affecting soil erosion by water (Fangmeier et al., 2006). Naturally, the steeper the slope of a field, the greater amount of soil loss due to erosion by water. Soil erosion by water also increases as the slope length increases due to the greater accumulation runoff.

In 1960s, the Universal Soil Loss Equation (USLE) was developed to predict soil erosion by water primarily for croplands. The USLE is based on simple empirical relationships, implemented through the use of tables, figures, and homographs of data collected over years from 1940s to the 1970s (Wischmeier and Smith, 1965). Later, the Revised Universal Soil Loss Equation (RUSLE1), a software version of a greatly improved USLE for any land uses was released in the early 1990s (Renard et al., 1997) followed by RUSLE2 in 2003 (USDA, 2008). Generally, RUSLE predicts longtime average annual soil loss, based on six factors including rainfall erosivity, soil erodibility, slope length and steepness, cover management, and support practice. RUSLE accounts for topographic effects through the product of the slope length, L, and steepness, S, sub-factors, which when

combined are called the topographic factor or the LS factor. The LS factor represents the ratio of soil loss on a given slope length and steepness to soil loss from a slope that has a length of 22.13 m, and a uniform steepness of 9% where all other factors are the same (McCool et al., 1971).

An important data layer for estimating soil erosion is topography which is often extracted from digital elevation models (DEM). Many environmental studies used DEMs to derive LS factor in erosion risk prediction. Lu et al. (2004) for example, mapped soil erosion risk in a large area of the Brazilian Amazonia forest using RUSLE with a 30-m DEM digitized from a 1:100 000 topographic map. They found that the majority of the study area had LS values less than 2.5 and most of the forest area had low erosion risk. Hoyos (2005) created a 25-m resolution DEM from a contour map of a 52 km^2 coffee-growing region in Columbia to calculate the LS factor for soil erosion prediction in that area. They found that the relationship between LS factor and soil erosion potential had a correlation coefficient (Spearman r) ranging from 0.57 to 0.59; indicating an evidence of topography influence on soil erosion potential in that area. Lee and Lee (2006) generated a 20-m resolution DEM of 274 km^2 Bosung basin, Korea by digitizing and interpolation contour lines in a 1:5000 scale topographic map. They used the DEM as a parameter input to RUSLE. Their study implied that the topographic LS factor, which is directly derived from the DEM, is sensitive to grid size, and the optimal resolution to quantify soil loss in the RUSLE model for the study site was 125 m.

These studies demonstrated the use of DEMs for soil erosion prediction in environmental studies over large scale (watershed-scale) areas. For a relatively small scale (field-scale), a reliable field DEM is vital because estimation of LS from an unreliable DEM

would propagate error into soil loss predictions, which could potentially lead to a poor conservation practices in the agricultural field. A field DEM can be generated using existing publically available US Geological Surveys (USGS) DEMs or more accurate measurements collected from GPS-aided farm operations (Renschler and Flanagan, 2008). Depending on data sources, methods and procedures to generate the field DEMs, the DEM estimates contain errors (Holmes et al., 2000; Wechsler and Kroll, 2006; Weschler, 2007). DEMs errors affect LS factor prediction (Renschler et al., 2001) or any other DEM derived parameters (Weschler, 2007). Though this is well known, the DEMs are often used as the true field surface, and the topographical uncertainty is not always accounted for in applications.

The Root Mean Square Error (RMSE), the typical global measure of DEM accuracy does not provide an accurate assessment of how precise *each grid* in a DEM represents topographical features (Wise, 1998; Wechsler, 2007). Hence a number of researchers have investigated spatial simulation methods to assess the uncertainty of elevation estimates in each DEM grid (Hunter et al., 1995; Holmes et al., 2000; Carlisle, 2005; Wechsler and Kroll, 2006). The simulation process accounts for spatial correlation in the data to produce equiprobable estimates (realizations) for each particular grid in the DEM. These realizations provide a range within which the true estimate lie and can be used to quantify the uncertainty at each particular DEM grid (Wechsler, 2007).

It is important to assess the uncertainty associated with DEM elevation estimates, so that the propagation of these errors can be accounted for in the other derived parameters or models. In this study, a 7.5-minute USGS DEM and GPS field measurements were used to develop DEM from which LS factors were calculated and average annual soil loss was

predicted for an agricultural field. The objectives of this study were to 1) assess the uncertainty in field DEMs elevation estimates and their effect on soil loss prediction; and 2) compare soil loss prediction uncertainty calculated using a USGS DEM and DEMs developed using GPS field measurements.

Materials and Method

Elevation Data

The study field was located in Boxholm, Iowa and covered an 62.81 ha (795 m wide by 790 m long) agricultural crop area. The field had an elevation range of 1140 m to 1162 m (22 m elevation difference) and a 0 to 16.74 degree slope. Elevation data was collected during a seeding operation using an agricultural implement equipped with a real-time kinematic differential GPS (RTK-DGPS) receiver (StarFire RTK, Deere & Co., Moline, III) with a vertical static root-mean-squared error (RMSE) of less than 0.025 m. Another set of elevation measurements was collected using a dual frequency (DF) DGPS receiver (StarFire SF2, Deere & Co., Moline, III) mounted on a John Deere harvester during a harvesting operation. The DF-DGPS receiver has vertical static RMSE of around 0.1 m. For both field operations, the vehicle traveled along 10 m swaths in the West-East direction.

A 7.5-minute USGS DEM of Boxholm located in Boone County, Iowa was acquired from an online GIS data provider (GeoCommunity, 2007). This DEM covered an area of 147.11 km^2, has a 30 m cell resolution and generated by contour digitization, either photogrametrically or from existing maps with vertical root mean square error (RMSE) of 7.0 m to 15.0 m. At least 28 test points (20 interior points, 8 along the edges) located on bench marks, spots elevation, or points on contour from existing source maps were used by USGS

to calculate the RMSE (US Geological Survey, 1987). The DEM has an absolute vertical elevation error tolerance of 50 m for any grid node when compared to the test points. USGS also has set a standard that any array of points in the DEM cannot encompass more than 49 contagious elevations to have error greater than 21 m. The elevation data in the USGS DEM within the boundary of the study field (0.63 km^2) was used in this study.

Exploratory data analysis

The GPS measurements histograms were slightly skewed to the left though the distribution is generally normal, indicating that a small percentage of measurements have high elevations (Fig. 1b, d). The RTK-DGPS measurements (7097 points) ranged from 1139.74 to 1162.01 m with a mean of 1149.62 m and standard deviation of 4.36 m. The DF-DGPS measurements (6874 points) had very similar elevation values, ranged from 1139.65 to 1161.88 m with a mean of 1149.24 m and standard deviation of 4.32 m. The USGS dataset histogram revealed a strongly multimodal distribution as a result of sparse pattern of elevation data points which failed to capture the continuity and surficial detail in the field topography. There were low elevation patterns in the middle and southwest of the field study that were smoothed in the USGS dataset (Fig. 1e). The USGS dataset also underestimated the elevation, as the elevation values ranged from 1121 m to 1136 m; about 20 m lower than GPS measurements. This may due to systematic errors as the result of the procedures used in the USGS DEM generation process that cause bias in the elevation. The USGS DEM was co-registered with the RTK-DGPS elevation measurements. From this point forward, USGS co-registered elevations were used for further analysis.

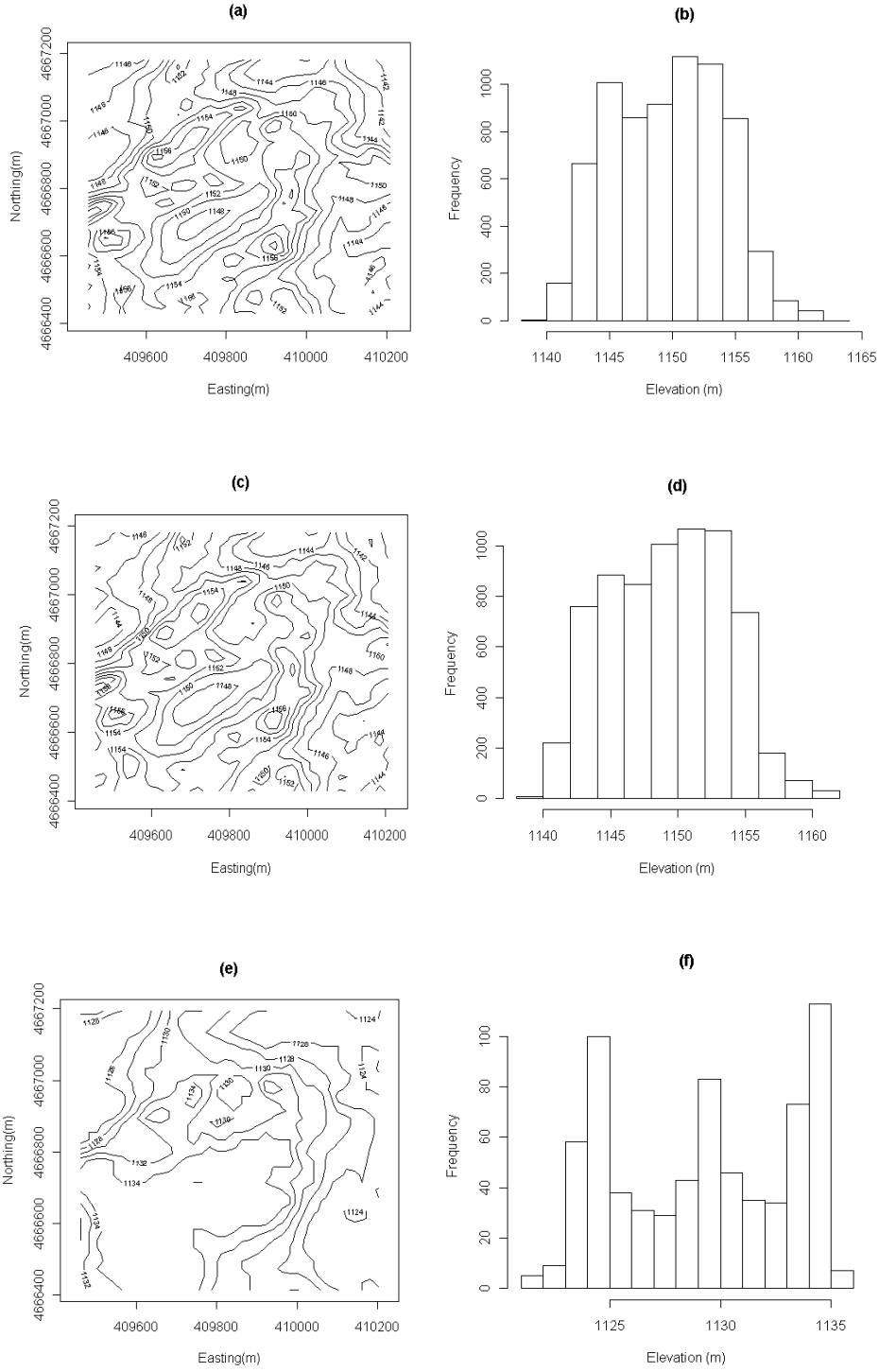

Figure 1: Contour maps and histograms of the (a, b) 7097 points RTK-DGPS, (c, d) 6874 points DF-DGPS measurements and (e, f) 704 points USGS DEM dataset.

DEM Development

Using each elevation dataset, the field-level DEMs were developed. This process included interpolating the GPS measurements and the 30-m USGS DEM into 10-m gridded DEMs. The grid size of 10-m was chosen as a trade-off between sufficient resolution and required computation time. For synchronization, the locations of the DEM grids were pre-defined so that each DEM developed using different datasets would use the same grid locations. This was done to ensure that the raster grids for various DEMs overlay precisely on each other and that the values can be compared to each other in the later analysis.

Elevation data were interpolated using ordinary kriging to generate the DEMs of the field. Ordinary kriging was chosen because it is a commonly used unbiased estimator that seeks to minimize error variance (Isaaks and Srivastava, 1989) which provides the best estimate of the value based on the available data. In addition, visual inspection of the data indicated no large trends, and ordinary kriging is known to be quite robust (Trangmar et al., 1985). The *gstat* program in R statistical software (Free Software Foundation, Inc., Boston, MA) was use to perform the interpolation.

First, the semivariograms of the elevation data were constructed. Usually, data that varies smoothly like elevation are generally expected to present a high spatial dependence with semivariograms that have a shallow slope near zero distance (high data similarity at short distances) (Burrough , 1987; and Valeriano et al., 2006). Topographic data also typically have an increase in semivariance with increasing distances (Fig. 2) because elevation has the potential to become more and more variable over larger distances due to surficial process (i.e. the semivariogram does not reach a sill; Holmes et al., 2000). The semivariograms from the GPS datasets showed a slightly larger increase in variance with lag

distance (steeper slope) because the large number of point measurements gave higher variance and details in the elevation within a small lag. A slightly slower increase (shallow slope) in semivariance for USGS dataset was due to the underestimation of elevation and smoothed elevation variation at the middle and southwest of the field. The USGS dataset was also more sparse, thus fewer data points were used to calculate the semivariance at each lag distance.

The sample semivariogram of each dataset was fit with a Gaussian semivariogram model because a Gaussian model presents a region of low slope near the zero distance, which is suitable for data that varies smoothly like elevation. The semivariogram models for RTK and DF-DGPS measurements were similar with a small nugget value of 0.1 m^2. For the USGS DEM data, the semivariogram had a nugget value of 0.6 m^2 (Fig. 2). The nugget values provide an indication of the amount of local variation in the dataset or an indication of the micro-spatial variability at a scale below the sampling resolution.

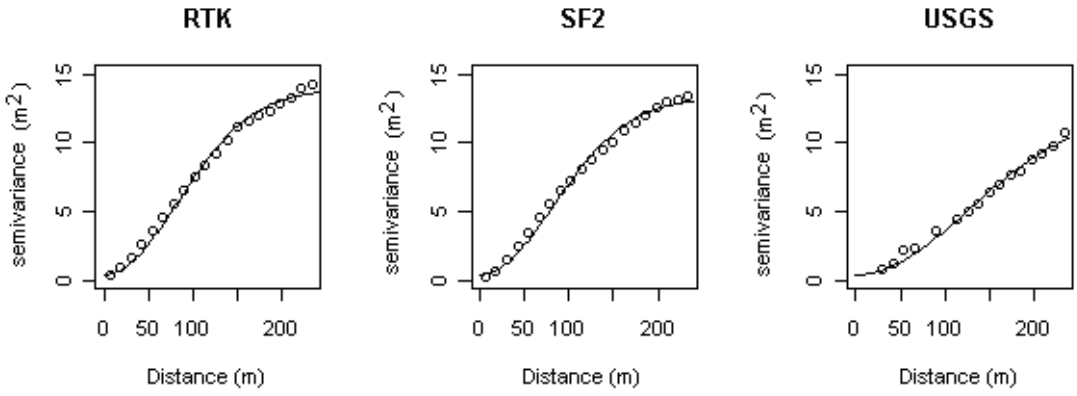

Figure 2: Semivariograms of RTK-DGPS and DF-DGPS measurements and USGS DEM data. Solid line on each semivariogram is the semivariogram model generated using the *gstat* program in R statistical software.

Using the semivariogram models, elevation data from each dataset were interpolated to 10 m common grids. A fixed radius of 60 m and a minimum of 30 data points were used to ensure enough interpolation support within an applicable computation time.

Error Simulation in Lower Accuracy DEMs

Errors were added to each of the interpolated DEMs because although kriging gave estimates with minimized error variance, it contains errors and under uncertain conditions is assumed to be one of an infinite number of equiprobable elevation realizations. Error simulation enables quantification of uncertainty associated with elevation estimates and derived parameter in each DEM grid. In this section, the procedure to assess elevation estimates uncertainty is discussed.

Researchers have used error measured at discrete points (such as from GPS surveys or data of higher resolution) to estimate and investigate DEM error and the spatial structure of the DEM error (Holmes et al., 2000; Carlisle, 2005). In this study, DF-DGPS measurements and USGS datasets had lower accuracy relative to the RTK-DGPS measurements. Therefore, RTK-DGPS elevation measurements were used as reference measurements to calculate errors contained in lower accuracy datasets. The error was calculated by subtracting the interpolated elevation at each DF-DGPS and USGS DEM grid from the nearest neighbor RTK elevation measurement.

The DF-DGPS errors had no visible spatial patterns (Fig. 3a). The histogram of the error values from DF-DGPS dataset had a roughly normal distribution, with a mean of 0.44 m, a median of 0.42 m and a standard deviation of 0.69 m, indicating that on average over the study area the DF-DGPS DEM underestimated the elevation by 44 cm (Fig. 3b). The

maximum error value of 3.14 m and minimum of -4.76 m however, showed there are large differences from RTK measurements in some areas. The semivariogram of the DF-DGPS error data showed spatial correlation, and there was a slight trend of increasing variance with distance to a 200 m range (Fig. 3c).

The USGS DEM error exhibited spatial patterns; particularly visible was the large region with negative error values to around -8 m showing a topographic depression (Fig. 3d). There were also several small regions in which positive error was observed. These regions were typically larger than the underlying 30 m spatial resolution of the original DEM. The USGS co-registered DEM error histogram was slightly skewed to the right indicating that a high percentage of grids underestimated the elevation value. The across the field, the mean error was 0.11 m, the median was 0.25 m and standard deviation was 2.28 m. The maximum (7.04 m) and the minimum (-8.45 m) error values show that there are significant differences in some areas. The semivariogram of the USGS DEM error data shows spatial correlation with a substantial increase in variance with increasing distance to a 200 m range (Fig. 3f).

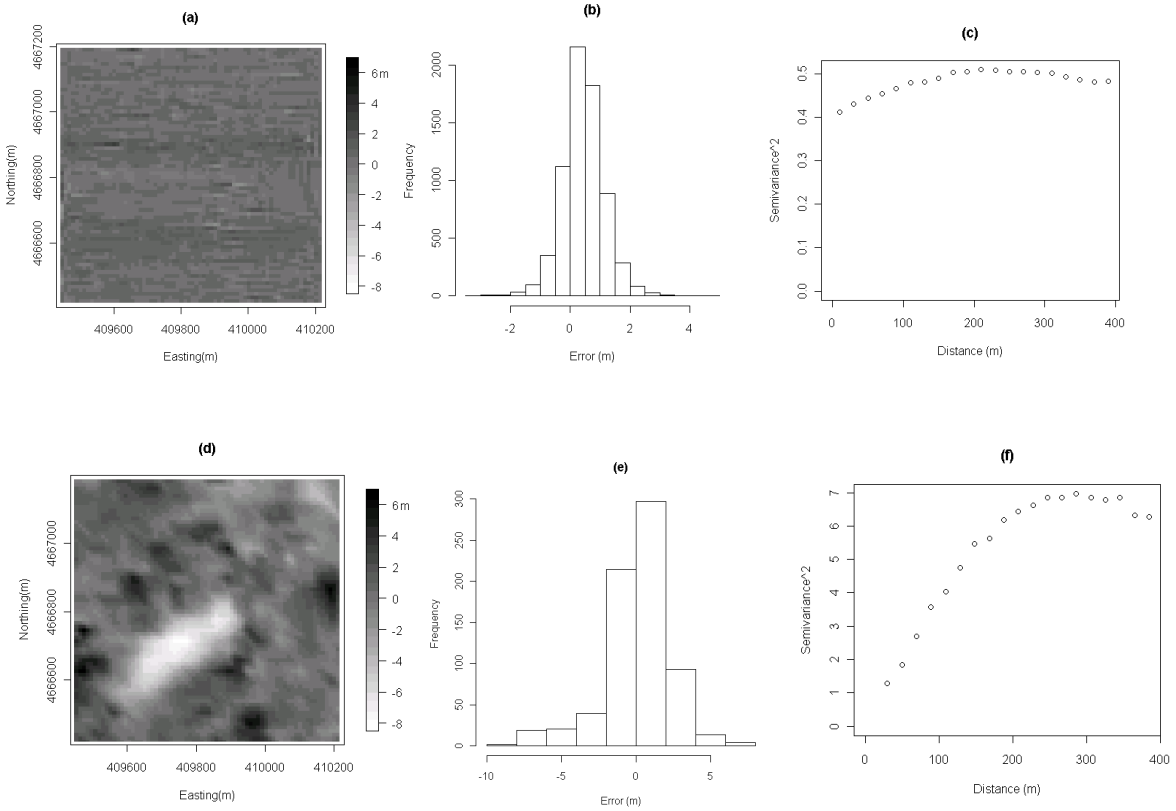

Figure 3: Plots and histograms of (a, d) 6874 points DF-DGPS measurement error and (b, e) 704 points USGS DEM error with their corresponding (c, f) semivariogram.

Many studies showed that DEM error is spatially variable (Ehlschlaeger and Shortridge, 1997; Hunter and Goodchild, 1997; Carlisle, 2005) and spatially correlated. Therefore a model of DEM error should not be random, but spatially dependent. In this study, the magnitude and spatial distribution of error in the DEMs was evaluated using a geostatistical method which recognized to be a realistic approach for DEM error modeling because it provides alternative plausible representations of possible spatial distribution of errors in a DEM (Holmes et al., 2000).

Sequential Gaussian simulation (SGS) was used to produce multiple realizations of the error value at each DEM grid based on available error data and spatial distribution of the

error. Detailed descriptions of SGS algorithm can be found in Goovaerts (1997). The SGS was implemented using *gstat* program in R (Free Software Foundation, Inc., Boston, MA). Prior to the simulation process, the normal scores transform, a non-linear transform that remaps any distribution to a normal distribution (Goovaerts, 1997), was applied to the error datasets to map the error distribution into a standard normal distribution. This was done to meet the format requirement of Gaussian simulation, which is that the univariate distribution of the error data be standard normal (Fig. 4a, and c). The semivariogram of the normal score-transformed error data was modeled for simple kriging estimation used in the simulation routine. Using *gstat*, the DF-DGPS normal score transformed error data was fit with a spherical semivariogram model with a nugget effect of 0.8 m^2, lag distance of 180 m, and sill of 1.0 m^2 (Fig. 4b). The USGS normal score transformed error data was fit with a spherical semivariogram model with a nugget of 0.3 m^2, lag distance of 200 m, and sill of 1.2 m^2 (Fig. 4d).

SGS models the uncertainty in the error data based on the normal score transformed data available near each point of the DEM grids. The simple kriging estimates (kriging prediction and its associated kriging variance) were used to establish the local conditional cumulative distribution function (ccdf) of the error estimates at every DEM grid location. Within the simulations, multiple realizations of error predictions were randomly drawn from the ccdf. Once the simulations of the normal score values have been produced, each realization must be back-transformed to the original error distribution. This process essentially consists of taking the inverse of the normal scores transform to remap the normal score distribution to the original error histogram.

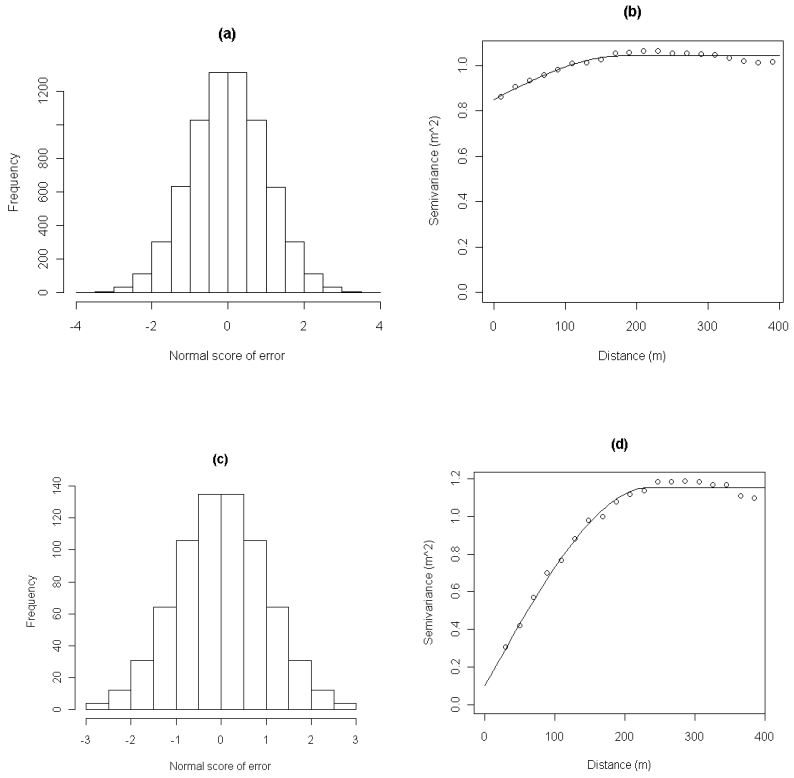

Figure 4: Histograms of the normal score-transformed of (a) 6874 points DF-DGPS measurements error and (c) 704 points USGS DEM error with their corresponding (b, d) semivariograms. Solid line on each semivariogram is the semivariogram model.

A total of 50 simulations were run resulting in 50 realizations of each DEM error map. The minimum number of needed simulations was determined when the percent difference in standard deviation of simulated errors between subsequent simulations was below 1 % and reached a steady state. The final step of the uncertainty analysis was to add the simulated error realizations of each dataset to its original kriged DEM. This created 50 equiprobable DEM realizations for each dataset to be used for soil loss prediction. These realizations provide a range within which the true estimate lies and can be used to quantify elevation uncertainty associated with each DEM and its effect on soil loss prediction.

Errors Simulation in RTK-DGPS DEM

For RTK-DGPS measurements, the vertical RMSE of the receiver was stated to be less than 0.025 m by the GPS receiver manufacturer. Typically, in the absence of higher accuracy data to calculate error values, a global accuracy measure such as RMSE is the only statistic available. When RMSE is the only information available, DEM error are often modeled based on a random process of error values with standard deviation equal to the RMSE value (Weschler, 2006), which means the DEM error is assumed to be spatially uncorrelated. This assumption is generally not appropriate for modeling error in DEMs because higher error is expected in areas of more rugged terrain (Hunter and Goodchild, 1997). Indeed a number of authors reported that the DEM errors could be larger on steep slopes (Hunter and Goodchild, 1997; Carrara et al., 1997), lower in less complex terrain (Gao, 1997), correlated with terrain ruggedness (Kyriakidis et al., 1999) and gradient and could be related to other elevation features (Ehlschlaeger and Shortridge, 1997). As RTK-DGPS measurements were collected using a moving vehicle, the assumption that the elevation errors were related to terrain variability seems appropriate because measurement errors due to vehicle dynamics resulted from the vehicle interaction with field topography or variability in the field surface.

Hence, the methodology presented here is intended to model the RTK-DGPS DEM errors based on the known RMSE and relate it to the elevation variability. The spatial distribution of RTK-DGPS DEM error related to the elevation variability was assessed using SGS. The normal score transform of the RTK-DGPS elevation measurements were used within the simulation to produce 50 realizations of elevation values at each DEM grid. First, the spatial correlation in the normal score transformed data were modeled using a Gaussian

semivariogram model with a nugget value of 0.1 m^2, similar to that of the original elevation. After simulation, the elevation realizations were back-transformed to the original elevation data distribution. This produced 50 DEM realizations which were then subtracted from the mean realizations to produce 50 realizations of error maps. Each error map provides a plausible representation of possible spatial distribution of errors in the DEM which the spatial structure related to elevation variability was accounted within the SGS routine (Fig. 5). As the RMSE of the RTK DGPS measurements was 0.025 m, the error maps were rescaled to have mean value equal to zero and standard deviation equal to the RMSE of RTK-DGPS measurements of 0.025 m. The rescaled error maps were added to the previously kriged RTK DEM to produce 50 equiprobable RTK DEM for soil loss prediction. These realizations provide a range within which the true estimate lies and can be used to quantify the RTK-DEM uncertainty and its effect on soil loss prediction.

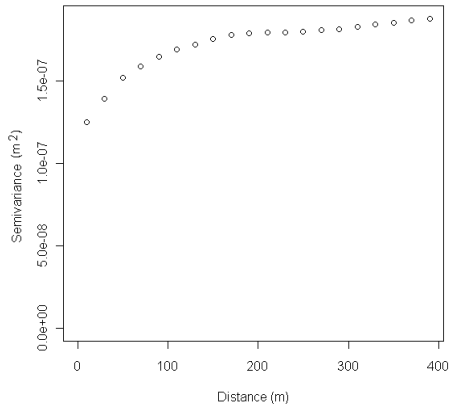

Figure 5: Semivariogram showing the spatial correlation of RTK-DGPS DEM errors produced using sequential Gaussian simulation.

Soil Loss Prediction Using RUSLE

LS topographical factor estimates were derived from the 50 realizations of DEMs from each dataset using ArcView (Version 3.3, ESRI, Redlands, Cal.) within the ArcView™ Spatial Analyst extension. The calculation was done using an Avenue script of *RUSLE3D*, an improved method for RUSLE calculation within GIS (Mitasova et al., 2001). The computation of LS factor at a point **r** = (x, y) is given by:

$$LS(\mathbf{r}) = 1.8 * [A(\mathbf{r}) / 22.13]^{0.4} [\sin b(\mathbf{r}) / 0.09]^{1.4} \quad (1)$$

where $A(\mathbf{r})$ is the upslope contributing area per unit contour width (m^2m^{-1}) and b is the slope in degrees. In this equation L and S factors were combined by substituting the parameter of slope length with upslope contributing area. The LS factor was calculated in *vector-grid* approach within Spatial Analyst operations.

This resulted to 50 LS factor maps for each dataset which were used to produce 50 equiprobable maps of the average annual soil loss of the field. The annual soil loss maps were generated by multiplying the LS factor maps with other RUSLE factors as followed:

$$E=R*K*LS*C*P \quad (2)$$

where;

E = the estimation of average annual soil loss in tons per acre per year caused by sheet and rill erosion,

R = the rainfall erosivity factor of 160 (hundreds of foot-tonf-inch per acre- year) for Boone County, IA,

K = the soil erodibility factor of 0.032 (tons-acre-hour per hundreds of foot-tonf-inch) for Clarion loam soil type,

LS = the slope length and steepness factor maps calculated in ArcView™,

C = the cover and management factor of 0.24 for spring tillage, corn-soybean rotation,

P = the support practice factor of 1 indicating no soil conservation practice.

DEM and Soil Loss Uncertainty Estimation

Using 50 elevation and 50 soil loss realizations from the three DEM sources, the uncertainty of the estimates in each grid was quantified. The elevation and soil loss uncertainty in each grid was based on the dispersion of the estimates from their mean. The dispersion was estimated by calculating the 95% confidence interval (CI) of the estimate in each grid; indicating 0.95 probability of the estimates fall in that interval. The 95% CI was calculated as the standard deviation multiplied by the critical two-tailed value of 1.96 for a standard normal distribution (Sheskin, 2004):

$$z_i = \pm 1.96 \times \sigma_i \tag{3}$$

where z_i refers to the lower and upper 95% CI of the estimates in the i^{th} grid with *i* as the indexing number of the grid across the map and σ_i refers to the standard deviation of the estimates in i^{th} grid. The 95% CI estimator provides an indication of statistical dispersion of the predicted parameters which can be used to quantify the uncertainty of the predicted parameter at each grid location. An estimate with a small CI value is more reliable than the estimate that has a high CI value. By calculating the CI value in each grid estimate across the map, the estimate spatial uncertainty at any particular location can be observed and studied.

Results and Discussions

Field DEMs

Contour plots of the DEMs developed using each elevation dataset showed that generally the field has lower elevations at the edge of southeast, northeast, and northwest of the field (Fig. 6). There were some common patterns with dense contour lines in several spots indicating high elevation gradients (Fig. 6a and b). These patterns appeared differently in the 10 m DEM developed with the USGS data (Fig. 6c). The low resolution USGS DEM data missed many topographical details in the field. Errors as the result of the procedures used in the USGS DEM generation process created artifacts specifically in the middle and southwest of the field.

In the simulation process, 50 simulated error realizations were added to the kriged DEMs to created 50 equiprobable realizations of the DEMs. The average elevations across the field study were calculated by taking the average of all the grid mean elevations across the DEM. Overall for each of the DEMs, the average estimated elevation was very similar, around 1149.74 m to 1149.95 m (Table 1). The elevation range across the field was 21.84 m for USGS DEM, 21.57 m for DF-DGPS DEM and 21.63 m for RTK DEM. The differences in elevation ranges indicate the differences in predicted field DEM elevations from each of the dataset.

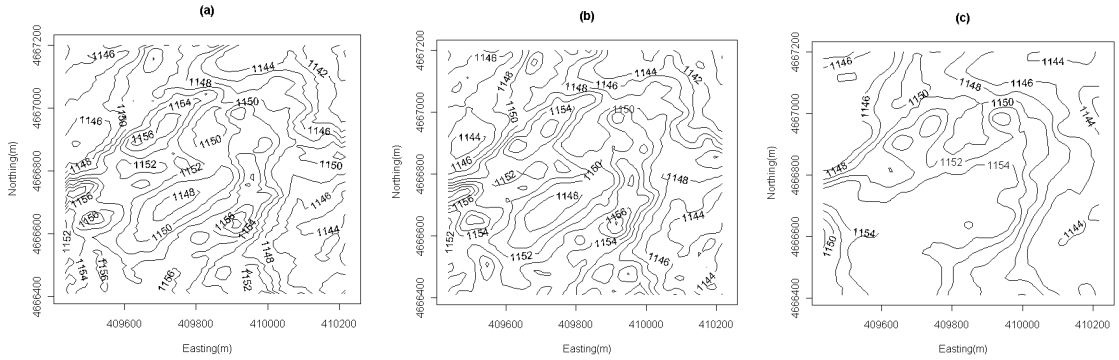

Figure 6: Contour map of the kriged DEM developed using (a) RTK-DGPS measurements, (b) DF-DGPS measurements and (c) USGS DEM dataset.

Uncertainty Estimates of DEM

Grayscale maps of the 95% CI of the elevation estimates in each grid were constructed to describe the uncertainty in the estimated elevation of the DEMs (Fig. 7). Darker color indicates higher CI values which signify higher uncertainty in the estimates. For the DEMs simulated from RTK-DGPS measurements, the 95% CI values of the grid elevations were very small, with a maximum of 0.15 m (Fig. 7a). The small uncertainty was mainly related to the variability in the RTK-DGPS elevation measurements around the grid which the magnitude was defined by the accuracy of the RTK-DGPS receiver. The uncertainty can also be related to vehicle dynamics resulting from the vehicle interaction with the micro-scale variability in the field surface. For the DEMs simulated from DF-DGPS measurements, the uncertainty of the simulated elevations in each grid was relatively higher than the RTK DEM with 95% CI values up to 3.12 m (Fig. 7b). A few dark patches in some spots of the 95% CI map show that there was a substantial deviation of the DF-DGPS elevations from the RTK DEM which indicated high degree uncertainty in the that area. The uncertainty in the elevation estimates was more clearly distinguished in the 95% CI map of

the DEMs simulated from the USGS dataset (Fig. 7c). The high value of 95% CI in grid elevation estimates was clearly observed in the area that has large elevation error as computed in the error analysis. These areas were characterized in northeast, middle and southwest regions of the study field with 95% CI value ranged to 4.33 m. Small predicted elevation CI in RTK DEM grids show that the degree of certainty in the grid elevation was higher using the RTK measurements compared to the DF-DGPS measurements and USGS dataset. The uncertainty estimates could give some insight to the modeler as to whether they might improve the elevation dataset based on the uncertainty of the grid elevations, and thus be better suited to understand the uncertainty in the output of their models.

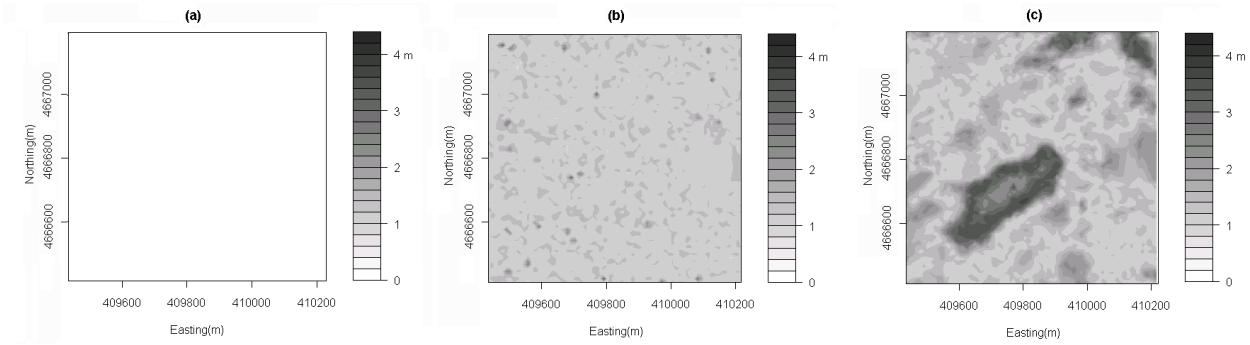

Figure 7: Contour map of 95 % CI for 50 equiprobable DEMs elevations using (a) RTK-DGPS measurements, (b) DF-DGPS measurements and (c) USGS DEM dataset.

Overall across the study field, the average uncertainty of the grid elevation estimate was higher for USGS DEM (Table 1). The average 95% CI estimates of the simulated elevation values in each grid across the field study for RTK DEM, DF-DGPS DEM and USGS DEM were 0.08 m, 1.35 m and 1.72 m respectively. The range of CI values across the field was 0.14 m for RTK DEM, 2.93 m for DF-DGPS DEM and 3.48 m USGS DEM. Although the difference in elevation between SF2 DEM and RTK DEM was smaller

compared to the difference in elevation between USGS DEM and RTK DEM (Fig. 3), the range of 95% CI across SF2 DEM was relatively high indicating there was high uncertainty in some areas.

It is the responsibility of the modeler to determine whether the uncertainty in these DEMs will affect the results of their applications that utilize the parameters derived from the DEMs. The effect of the DEM uncertainty on average annual soil loss prediction of the field was discussed in the following section.

Predicted Average Annual Soil Loss

The predicted average annual soil loss for kriged DEMs using GPS measurements were similar to each other (Fig. 8a, b). Generally, most of the sloped areas in the field show denser contour lines which signify high values of estimated soil loss. The high soil loss values appeared in a few spots with the highest peak at the west of the field. The contour pattern of soil loss estimates derived from the kridged DF-DGPS DEM (Fig. 8b) shows similarity with the soil loss pattern from RTK DEM (Fig. 8a), though most of the high soil loss areas were underestimated. For the USGS DEM, a contour map of the soil loss estimates lost a lot of detail in the patterns compared to soil loss from RTK DEM (Fig. 8c). The areas that have high uncertainty of elevation values characterized in the previous section (in northeast, middle and southwest regions) were the areas that have the most differences in soil loss prediction relative to the soil loss estimates derived from RTK DEM. The contour map of soil loss for simulated USGS DEM shows sparser contours which indicate that the soil loss values were underestimated in many areas of the study field compared to RTK DEM.

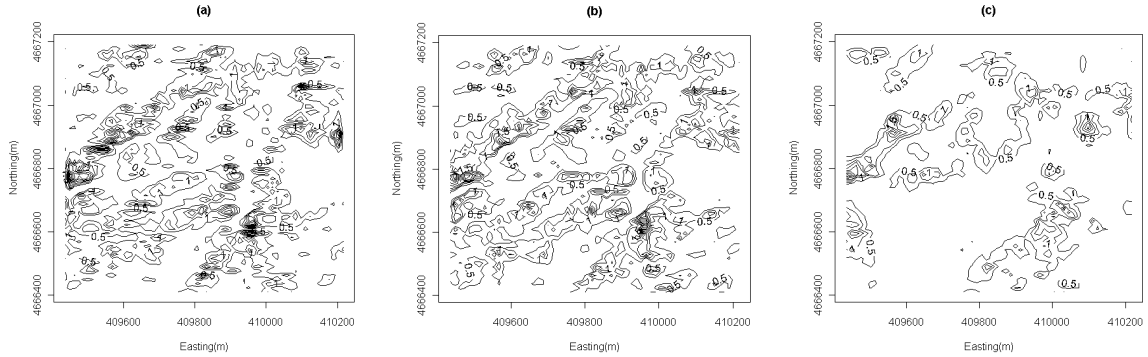

Figure 8: Contour map of soil loss calculated from kriged DEMs using (a) RTK-DGPS measurements, (b) DF-DGPS measurements and (c) USGS DEM dataset.

The mean of simulated average annual soil loss across the field study was calculated by averaging all the mean soil loss values across the entire DEMs. For each of the DEM, the average annual soil loss for the whole field study was very similar, ranged from 0.60 to 0.62 tons/acre/year (Table 1). However, the range of average annual soil loss across the field study for RTK DEM was 8.29 tons/acre/year, higher than 5.31 tons/acre/year for DF-DGPS DEM and 5.76 tons/acre/year for USGS DEM. For the RTK-DGPS DEM, 95% of the grid soil loss values were smaller than 2.11 tons/acre/year. For DF-DGPS DEM and USGS DEM, 95% of the grid soil loss values were smaller than 1.52 tons/acre/year. Smaller range of soil loss estimates in DF-DGPS and USGS DEM signify that DF-DGPS DEM and USGS DEM underestimated the soil loss value at some areas in the field compared to RTK DEM.

The estimate of soil loss errors relative to soil loss predicted from the kriged (undisturbed) RTK DEM were calculated by subtracting the soil loss values for each simulated DEM from the soil loss value of the undisturbed RTK DEM . The percentage of the areas that have mean error value greater than 2 tons/acre/year were estimated (Figure 9). About 3%, 5% and 7% of the study field area has soil loss error greater than 2 tons/acre/year

for RTK, DF-DGPS and USGS DEM respectively. Mostly, the areas that have error greater than 2 tons/acre/year were located in the sloping areas. The error estimates from DF-DGPS and USGS DEM were higher than RTK DEM, though the differences were not substantial because when adding the simulated elevation errors to each of the DEM, the process essentially corrected the DEM to be similar to kriged RTK DEM. However, the corrected errors have uncertainties associated with them which were not usually accounted for in applications.

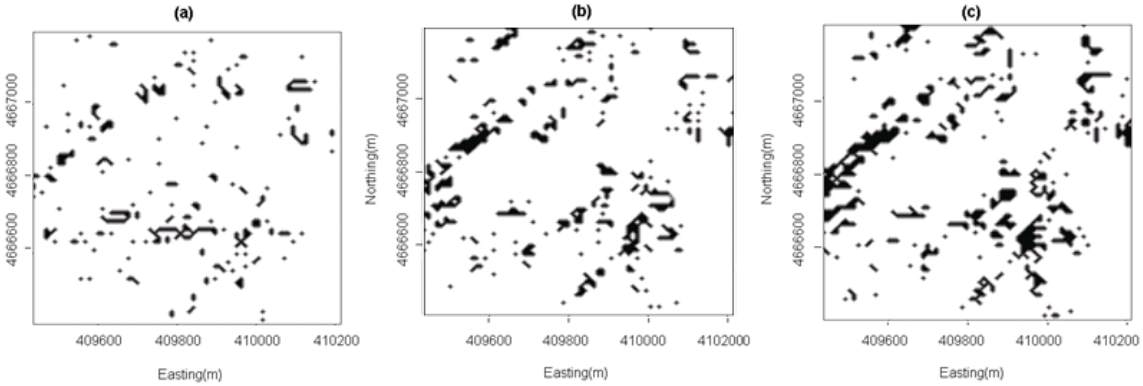

Figure 9: The soil loss error estimates greater than 2 tons/acre/year calculated by subtracting soil loss estimate from simulated (a) RTK DEM, (b) DF-DGPS DEM and (c) USGS DEM from soil loss predicted from kriged (undisturbed) RTK DEM.

Uncertainty Estimates of Predicted Soil Loss

The grayscale map of the 95% CI of the 50 equiprobable soil loss value in each grid was constructed to describe the uncertainty in the soil loss prediction (Fig. 10). For the DEMs simulated from RTK-DGPS measurements, the 95% CI values of the grid soil loss were generally small across the field, with only a few spots that have high CI value ranged to 3.40 tons/acre/year (Fig. 10a). For the DF-DGPS DEMs, the uncertainty of the predicted soil loss in each grid was relatively higher across the whole field with 95% CI values up to 4.98

tons/acre/year in (Fig. 10b). Some dark patches in many spots of the 95% CI map indicated high degree uncertainty in the estimates. The uncertainty in the soil loss estimates was more clearly distinguished in the 95% CI map of the predicted soil loss from the USGS DEM (Fig. 10c). The high value of 95% CI in soil loss estimates was clearly spotted in the area that has large elevation error as computed in the error analysis. These areas were characterized in some northeast, middle and southwest regions of the field study with 95% CI value up to 7.53 tons/acre/year. Low 95% CI of soil loss estimates from RTK DEM show that the degree of certainty in the soil loss estimates was higher using the RTK-DGPD DEM compared to the DF-DGPS DEM and USGS DEM.

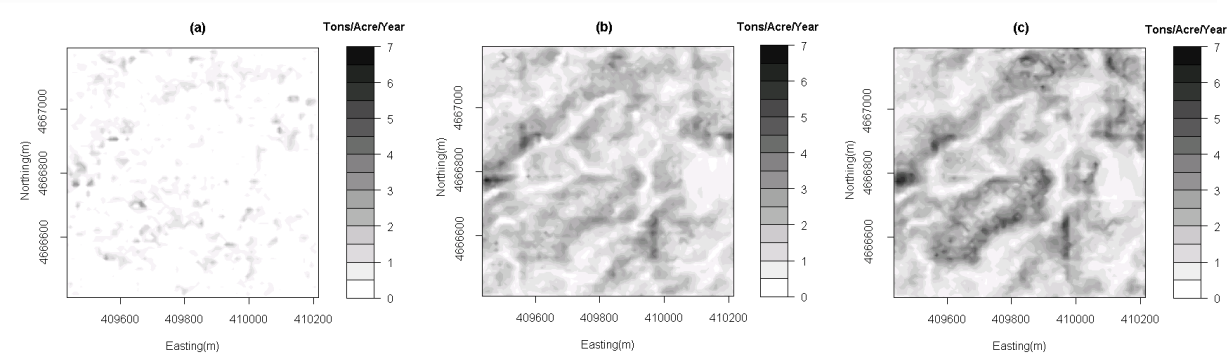

Figure 10: Contour map 95% CI of soil loss estimates calculated using 50 simulated DEMs using (a) RTK-DGPS measurements, (b) DF-DGPS measurements and (c) USGS DEM dataset.

Similar to the average uncertainty of the elevation estimates, the average uncertainty of the soil loss factor estimate across the study field was higher for USGS DEM (Table 1). Quantitatively, the average 95% CI estimate of the soil loss values across the field study for RTK DEM, DF-DGPS DEM and USGS DEM were 0.23, 1.51 and 1.65 tons/acre/year respectively. Although the errors in DF-DGPS DEM were smaller than the errors in USGS

DEM, the impact of the errors on soil loss prediction for DF-DGPS DEM was substantially high as of in the USGS DEM. This shows that even a small amount of elevation error in DEMs greatly affected the result of the soil loss estimation.

Table 1. Summary statistics of 50 equiprobable elevations and soil loss across the study field

Statistics of 50 estimates	Dataset	Elevation across field (m)				Soil loss across field (Tons/acre/year)			
		Avg. (μ)	Range	5^{th}%	95^{th}%	Avg. (μ)	Range	5^{th}%	95^{th}%
Mean	RTK	1149.95	21.63	1143.05	1156.23	0.60	8.29	0.00	2.11
	DF -DGPS	1149.93	21.57	1143.01	1156.22	0.59	5.31	0.03	1.52
	USGS	1149.74	21.84	1142.69	1156.07	0.62	5.76	0.05	1.52
95% CI	RTK	0.08	0.14	0.04	0.11	0.23	3.40	0.00	0.89
	DF-DGPS	1.35	2.93	1.05	1.62	1.51	4.98	0.21	2.98
	USGS	1.72	3.40	1.16	3.30	1.65	7.53	0.27	3.74

To compare the uncertainty of average annual soil loss estimates from different DEMs, the percentage of the areas that have 95%CI value greater that 2 tons/acre/year were estimated. Only 0.41% of the area in the soil loss map predicted from RTK DEM has 95% CI values more than two tons/acre/year (Fig. 11a). The percentages of the area that has soil loss 95% CI value greater than two tons/acre/year for DF-DGPS DEM and USGS DEM were substantially higher, 20% and 31% respectively (Fig. 11b,c). These areas were generally located on the slopping area of the field. Overall, the errors in USGS DEM greatly affect the soil loss prediction derived from the DEM. The uncertainty estimators such as the 95% CI value of the soil loss estimates which were calculated on grid by grid basis enable quantification and visualization of the impacts of the DEM errors in the soil loss prediction.

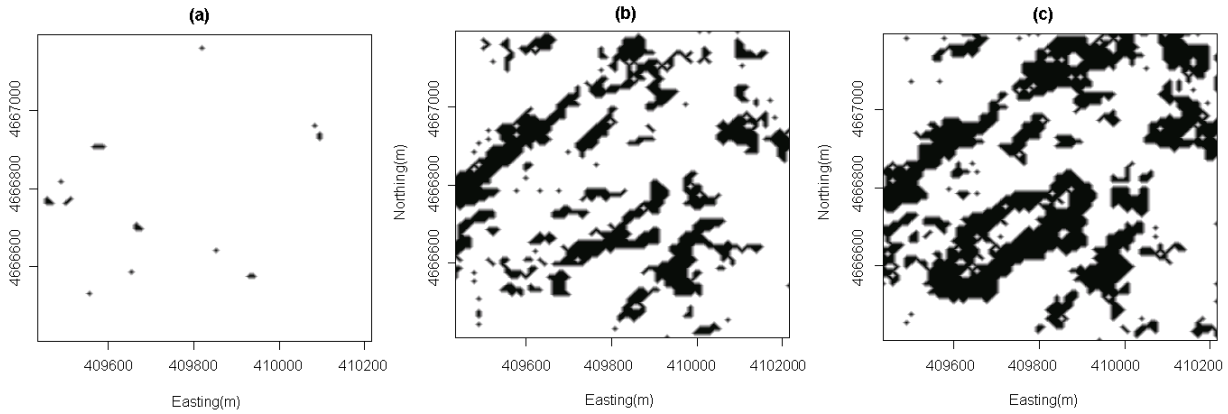

Figure 11: Shown in black are the areas that have 95% CI estimates greater than two tons/acre/year for soil loss calculated using (a) RTK-DGPS DEMs, (b) DF-DGPS DEMs and (c) USGS DEMs.

Conclusion

The uncertainty in the estimated DEMs affected the result of soil loss prediction in this study. From this study, several conclusions can be drawn:

1. The average uncertainty value for DF-DGPS DEM (95% CI = 1.35 m) and USGS DEM (95% CI = 1.72 m) were significantly higher than the uncertainty of RTK DEM (95% CI = 0.08 m). Although the errors in DF-DGPS measurements were smaller (ranged from -4.76 m to 3.14 m) than the errors in USGS DEM dataset (ranged from -6.88 m to 8.78 m), the impacts of DEM errors on the DEM elevation estimates were substantial.
2. Even small errors in the DEM elevation produced large deterioration of the annual soil loss prediction in the study field. The percentage of the areas that have 95% CI of the annual soil loss estimates greater than two tons/acre/year were 20% and 30% for

DF-DGPS and USGS DEM which were significantly higher than the percentage area of 0.41% for RTK DEM.

3. The quantification of uncertainty estimator such as 95% CI values in each grid enables a thorough assessment of the prediction uncertainty. The uncertainty estimators which were calculated on grid by grid basis provide visualization of the impacts of the DEM errors in the soil loss prediction.

Thorough evaluation of the uncertainty in the elevation is needed for appropriate conclusion on the impact of the DEM errors on soil loss prediction. The quantification of the uncertainty estimators in grid by grid basis enables more precise assessment of reliability of the predicted estimates across the study area. The impact of the DEM uncertainty on soil loss prediction can also be discerned through visualization of the estimator grids like 95% CI maps as shown in this study.

The need for use of accurate DEMs in soil loss estimation is evident. For accurate LS factor estimation to be used in RUSLE equation for soil erosion prediction, high accuracy DEM such as using RTK-DGPS measurements is required. Otherwise, one would like to take into account the prediction uncertainty such as using the quantified uncertainty estimators to classify the areas that has unreliable prediction, which provide guidelines for error reduction and management planning.

References

Burrough, P.A., 1995. Spatial aspects of ecological data. In *Data Analysis in Community and Landscape Ecology,* 213–251. R. H., Jongman, C. J. F. ter Braak, O. F. R. Van Tongeren, eds. Wagenigen, The Netherlands: Cambridge University Press.

Carrara, A., G. Bitelli and R. Carla. 1997. Comparison of techniques for generating digital terrain models from contour lines. *Int. Jounal of Geographical Info. Science* 11:451-473.

Carlisle, B. H. 2005. Modeling the spatial distribution of DEM error. *Trans. of GIS* 9(4): 521-540.

Ehlschlaeger, C. R., and A. Shortridge. 1997. Modelling elevation uncertainty in geographical analyses. *In Proc. of the 7th Int. Symposium of Spatial Data Handlin*g, 585-595. M. J. Kraak and M. Molenaar, eds. London: Taylor and Francis.

Fangmeier, D. D., W. J. Elliot, S. R. Workman, R. L. Huffman and G. O. Schwab. 2006. *Soil and water conservation engineering*, 5th Edition. Thomson Delmar Learning, New York.

Gao, J. 1997. Resolution and accuracy of terrain representation by grid DEMs at micro-scale. *Int. Jounal of Geographical Info. Science* 11:199-212.

GeoCommunity, 2007. USGS DEM Resources: 1995-2007. GIS Data Depot, Niceville, FL.: MindSites Group, LLC Available at: http://data.geocomm.com/. Accessed 23 April 2006.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*. Oxford Uni. Press, New York.

Holmes, K. W., O. A. Chadwick, and P. C. Kyriakidis. 2000. Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology* 233: 154-173.

Hoyos, N. 2005. Spatial modeling of soil erosion potential in a tropical watershed of the Colombian Andes. *Catena* 63: 85-108.

Hunter, G. J., M. Caetano, and M. F. Goodchild. 1995. A methodology for reporting uncertainty in spatial database products. *URISA Journal* 7(2): 11-21.

Hunter, G. J., and M. F. Goodchild. 1997. Modeling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis* 29: 35:49.

Isaaks, E. H., and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. New York, N.Y.: Oxford University Press.

Kyriakidis, P.C., A. M. Shortridge and M. F. Goodchild. 1999. Geostatistics for conflation and accuracy assessment of digital elevation models. *Int. Jounal of Geographical Info. Science* 13: 677-707.

Lee, G. S. and K. H. Lee. 2006. Scaling effect for estimating soil loss in the RUSLE model using remotely sensed geospatial data in Korea. *Hydrol. Earth Syst. Sci.* 3:135-157.

Lu, D., G. Li, G. S. Valladares and M. Batistella. 2004. Maping soil erosion risk in Rondonia, Brazilian Amazonia: Using RUSLE, remote sensing and GIS. *Land Degrad. & Develop*. 15:499-512.

McCool, D. K, G. R. Foster and G. A. Weesies. 1997. Slope length and steepness factors (LS). In *Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE), Agriculture Handbook No. 703*, 101-251. Renard, K. G., G. R. Foster, G. A. Weesies, D. K. McCool and D. C. Yoder, eds. Washington, D. C: U. S. Gov. Printing Office.

Mitasova, H., W. M. Brown, M. Hohmann and S. Warren. 2001. Using Soil Erosion Modeling for Improved Conservation Planning: A GIS-based Tutorial, Urbana-Champaign, IL.: Geographic Modeling Syst. Lab, UIUC. Available at: http://skagit.meas.ncsu.edu/~helena/gmslab/reports/CerlErosionTutorial/denix/default.htm. Accessed 23 July 2008.

Renard, K. G., G. R. Foster, G. A. Weesies, D. K. McCool and D. C. Yoder. 1997. Predicting soil erosion by water. *Agriculture Handbook No. 703.* Washington, D. C: U. S. Gov. Printing Office.

Renschler, C. S., and D. C. Flanagan. 2008. Site-specific decision-making based on RTK GPS survey and six alternative elevation data sources: Soil erosion predictions. *Trans. ASABE* 51(2): 413-424.

Renschler, C. S., D. C. Flanagan and B. A. Engel. 2001. Data accuracy issues in spatially distributed soil erosion modeling: What does decision-making gain? ASAE Paper No. 701P0007. St Joseph, Mich.: ASAE.

Sheskin, D. 2004. *Handbook of Parametric and Nonparametric Statistical Procedures. 3rd Edition*. Chapman & Hall/CRC. FL.

Trangmar, B. B., R. S. Yost, and G. Uehara. 1985. Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy* 38: 45−94.

USDA. 2008. User's Reference Guide: Revised Soil Loss Equation Version 2 (RUSLE2). Agricultural Research Service. Washington, D.C.: USDA National Agricultural Research Service. Available at: www.ars.usda.gov. Accessed 14 October 2008.

USGS. 1987. Standard for Digital Elevation Models, Part 2: Specifications. U.S. Department of the Interior, Washington, DC: USGS. Available at http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dem/2DEM0198.PDF. Assessed 30 October 2008.

Valeriano, M. M, T. M. Kuplich, M. Storino, B. D. Amaral, J. N. Mendes Jr. and D. J. Lima. 2005. Modeling small watersheds in Brazilian Amazonia with shuttle radar topographic mission- 90 m data. *Computer and Geosciences* 32: 1169-1181.

Wischmeier, W. H. and D. D. Smith. 1965. Predicting rainfall-erosion losses from cropland east of the Rocky Mountains. *Agiricultural Handbook No. 282.* Washington, D. C: U. S. Gov. Printing Office.

Wise, S. M. 1998. The effect of GIS interpolation errors on the use of digital elevation models in geomorphology. In *Landform Modelling and Analysis,* 139−164. S. N. Lane, K. S. Richards, and J. H. Chandler, eds. Chichester, U.K.: John Wiley and Sons.

Wechsler, S. P. 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrol. *Earth Syst. Sci*. 11: 1481-1500.

Wechsler, S. P., and C. N. Kroll. 2006. Quantifying DEM uncertainty and its effect on topographic parameters. *Photogrammetric Engineering & Remote Sensing* 72(9): 1081-1090.

CHAPTER 6. GENERAL CONCLUSSIONS

This research provides the basis of field DEMs development and the assessment of their uncertainty. One of the objectives was to develop a methodology for utilizing repeated elevation field surveys during typical farming operations for the development of field DEMs as a by-product of GPS-aided farm operations. Repeated GPS surveys of elevation data improved the DEM accuracy over time. This research presented two methods for combining the measurements of multiple surveys, one was using fuzzy logic and another was using weighted averaging technique. These two methods were more robust to measurement errors and resulted in improved performance over the grid-wise averaging method. This study demonstrates the importance of passing along a measure of estimate confidence in the process as measurements from new surveys are added.

The second objective was to develop a targeted sampling method based on spatial uncertainty of prior measurements for topographic mapping. Uncertainty assessment using SGS quantified the variability of attributes in the field based on available sampled data. The information may aid producers in designing more efficient sampling strategies by targeting only regions of interest in the field for re-sampling consideration. The introduction of targeted measurements reduced the time required for data collection and resulted in DEMs with relatively low RMSE. Use of targeted sampling procedure may efficiently aid farm attribute estimation for site specific management practice.

The information about uncertainty in spatial attribute estimation is useful in understanding error propagation induced from deriving parameters of interest related to the attributes. The third objective of the study was to assess the uncertainty in field DEMs

elevation estimates and their effect on soil loss prediction. Using SGS, the quantification of the uncertainty estimators on a grid by grid basis enables more precise assessment of reliability of the predicted soil loss estimates across the study area. The impact of the DEM uncertainty on soil loss prediction can also be discerned through visualization of the estimator grids like 95% CI maps as shown in this study.

The need for accurate DEMs in soil loss estimation is evident. For accurate LS factor estimation to be used in RUSLE equation for soil erosion prediction, high accuracy DEM such as using RTK-DGPS measurements is required. Otherwise, one should take into account the prediction uncertainty such as using the quantified uncertainty estimators to classify the areas that has unreliable prediction, which provide guidelines for error reduction and management planning.

Summary

GPS-equipped farm vehicles enable landowners to utilize elevation data during normal field operations for the development of agricultural field DEM. Generation of DEMs from measurements acquired with such systems provided users with additional benefits from the original capital investment in the equipment. This research provided extensive but useful guidance on appropriate procedures involved in the development of field DEMs for land users to take full advantage of the existing technology. The DEMs can be further integrated into existing operational environment such as yield mapping and auto-guidance systems where topographic information can be a great support.

Digital elevation models, like other maps, are models that deviate from reality. Depending on process, methods and procedures to generate the DEMs, the topographic

parameters derived from a DEM contain uncertainties. In this study, the uncertainty of DEM estimates was assessed and found to be useful to enhance the sampling strategy in improving the accuracy of the DEMs. The effect of DEM uncertainty on topographic parameters was investigated and found that DEM uncertainty has a substantial impact on soil erosion prediction which may affect the consequence management decisions such as the decision on how much biomass need to be removed from the field for conservation practice. Many users particularly farmers, may not be knowledgeable about the theory, so they will appreciate the guidance about appropriate analysis that helps them make good choices for their data and applications.

Suggestions for Future Work

This research has a great potential to be expanded to make the DEMs application feasible and more profitable. The resulting field DEMs developed using repeated GPS surveys over many years may change if the topography or elevation changes due to excessive soil erosion or other disturbance. The analysis and method to detect the effect of topography or elevation changes due to excessive soil erosion or other disturbance on DEM development is an interesting point of future work. This work could incorporate external data support such as rainfall intensity and soil erosivity maps to produce more accurate prediction of DEMs elevation.

The procedure for designing targeted sampling based on spatial uncertainty of prior measurements can be extended to applications over larger and more complex agricultural area. The result of the analysis on a larger and more varying terrain fields may be more interesting to be investigated to study the effect of terrain complexity and the robustness of

the method. The application of the methodology on other spatial sampling applications such as soil and yield sampling could also be a valuable future work.

The method of investigating DEM uncertainty propagation through the intermediate estimates of slope and upslope contributing area need to be developed for better understanding of the impact on soil loss prediction within GIS. With better understanding of the error propagation, the resulting soil loss estimation from a field DEM can be tested and integrated into application such as automated systems for biomass harvest equipment for conservation practices. A thorough procedure and methodology is needed to study the effect of the estimation uncertainty on such agricultural application for better farm management decisions.

Based on the experience gained from this work, following future work is recommended:

1. Make the DEM development algorithm more robust by taking into account of more exceptional cases such as elevation changes or disturbance.
2. Apply the procedure for designing targeted sampling over larger and more complex agricultural area.
3. Test the procedure for designing targeted sampling for soil and yield mapping.
4. Run extended statistical analysis investigating DEM uncertainty propagation through slope and upslope contributing area for soil loss prediction within GIS.
5. Incorporate the field DEMs on automated systems in biomass harvest equipment for conservation practice.

References

Agarwal, P. K., L. Arge and A. Danner. 2006. From point cloud to grid DEM: A scalable approach. In Progress in Spatial Data Handling. 12th International Symposium on Spatial Data Handling 771-788. A. Riedl, W. Kainz, and G. Elmes, eds., Springer Berlin Heidelberg.

Billingsley, J. 2000. Automatic guidance of agricultural mobiles at the NCEA. *Industrial Robot: An International Journa*: 27(6): 449-457.

Bingner, R.L., and F. D. Theurer. 2001. Topographic factors for RUSLE in the continuous-simulation, watershed model for predicting agricultural, non-point sources pollutants (AnnAGNPS). ASAE Paper No: 701P0007. ASAE, St.Joseph, MI.

Burrough, P.A., 1995. Spatial aspects of ecological data. In *Data Analysis in Community and Landscape Ecology*, 213–251. R. H., Jongman, C. J. F. ter Braak, O. F. R. Van Tongeren, eds. Wagenigen, The Netherlands: Cambridge University Press.

Clark, R. L., and R. Lee. 1998. Development of topographic maps for precision farming with kinematic GPS. *Trans. ASAE* 41(4): 909−916.

Cressie, N., 1989. Geostatistics. *The American Statistician* 43(4): 197-202.

Deutsch, V. D., and A. G. Journel. 1998. GSLIB geostatistical software library and user's guide. Oxford, London: Oxford University Press.

Dosskey, M.G, D.E. Eisenhauer, and M.J. Helmers. 2005. Establishing conservation buffers using precision information. *J. Soil and Water Cons.* 60(6): 349-354.

Endreny, T. A., and E. F. Wood. 2001. Representing elevation uncertainty in runoff modeling and flowpath mapping. *Hydrol. Process*. 15: 2223-2236.

Franklin J., P. McCullough, and C. Gray, 2000. Terrain Variables Used for Predictive Mapping of Vegetation Communities in Southern California. In *Terrain Analysis: Principles and Applications* 331-53. J. P. Wilson and J. C. Gallant, eds., New York, John Wiley & Sons

Fried, J.S., D.G. Brown, M.O. Zweifler, and M.A. Gold. 2000. Mapping contributing areas for stormwater discharge to streams using terrain analysis. In *Terrain Analysis: Principles and Applications* 183-203. J. P. Wilson and J. C. Gallant, eds., New York, John Wiley & Sons.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*. Oxford Uni. Press, New York.

Holmes, K. W., O. A. Chadwick, and P. C. Kyriakidis. 2000. Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology* 233: 154-173.

Hunter, G. J., M. Caetano, and M. F. Goodchild. 1995. A methodology for reporting uncertainty in spatial database products. *URISA Journal* 7(2): 11-21.

Hunter, G. J., and M. F. Goodchild. 1995. Dealing with error in spatial databases: a simple case study. *Photogrammetric Eng. & Remote Sensing* 61(5): 529-537.

Isaaks, E. H., and R. M. Srivastava. 1989. An Introduction to Applied Geostatistics. New York, N.Y.: Oxford University Press.

Journel, A. G., and CH. J. Huijbregts. 1978. Mining Geostatistics. New York, N.Y.: Academic Press.

Kim, H., J. R. Arrowsmith, C. J Crosby, E. Jaeger-Frank, V. Nandigam, A. Memon, J. Conner, S. B. Badden and C. Baru. 2006. An Efficient Implementation of a Local Binning

Algorithm for Digital Elevation Model Generation of LiDAR/ALSM Dataset. Lidar/ASLM Knowledge Base, Tempe, AZ.; Active Tectonics Research Group, Arizona State Univ. Avaiable at: http://lidar.asu.edu/knowledgebase.html. Assessed 28 October, 2008.

Mackey, B.G, I. C. Mullen, K. A. Baldwin, J. C. Gallant, R. A. Sims, and D. W. Mckenney. 2000. Toward a Spatial Model of Boreal Forest Ecosystems: The Role of Digital Terrain Analysis. In *Terrain Analysis: Principles and Applications* 391-423. J. P. Wilson and J. C. Gallant, eds., New York, John Wiley & Sons.

Matheron, G. 1973. The intrinsic random functions and their applications, *Adv. In Appl. Probability.* 5(439-468).

Maune, D. F. 2007. Digital elevation model technologies and applications: The DEM Users Manual, 2nd Edition. Bethesda, Maryland: ASPRS.

Murthy, A. R., I. V. Murali Krishna and M. S. R. Murthy. 2004. Precision Conservation of Natural Resources for Sustainable Development. Proceedings of Geosciences and Remote Sensing Symposium 2004: 4605-4608. IEEE, Piscataway, NJ.

Oksanen, J. and T. Sarjakoski. 2005. Error propagation of DEM-based surface derivatives. *Computer & Geosciences* 31(8): 1015-1027.

Olea, R. A. 1999. Geostatistics for Engineers and Earth Scientists. Norwell, MA: Kluwer Academic Publishers.

Pelletier, G., and S. K. Upadhyaya. 1999. Development of a tomato load/yield monitor. *Computers and Electronics in Agriculture* 23(2): 103-117.

Reid, J. F., Q. Zhang, N. Noguchi, and M Dickson. 2000. Agricultural Automatic Guidance Research in North America. *Computers and Electronics in Agriculture* 25: 155-167.

Renschler, C. S., and D. C. Flanagan. 2008. Site-specific decision-making based on RTK GPS survey and six alternative elevation data sources: Soil erosion predictions. *Trans. ASABE* 51(2): 413-424.

Renschler, C. S., D. C. Flanagan, B. A. Engel, L. A. Kramer, and K. A. Sudduth. 2002. Site−specific decision−making based on RTK GPS survey and six alternative elevation data sources: Watershed topography and delineation. *Trans. ASAE* 45(6): 1883−1895.

Valeriano, M. M, T. M. Kuplich, M. Storino, B. D. Amaral, J. N. Mendes Jr. and D. J. Lima. 2005. Modeling small watersheds in Brazilian Amazonia with shuttle radar topographic mission- 90 m data. *Computer and Geosciences* 32: 1169-1181.

Vellidis, G., C. D. Perry, J. S. Durrence, D. L. Thomas, R. W. Hill, C. K. Kvien, T. K. Hamrita, and G. Rains. 2001. The peanut yield monitoring system. *Tran. ASAE* 44(4): 775–785.

Wang, P., J. Du, X. Feng, and S. Hu. 2006. Effect of DEM uncertainty on distributed hydrological model TOPMODEL. Proceedings of Geosciences and Remote Sensing Symposium 2006: 1074-1077. IEEE, Piscataway, NJ.

Wechsler, S. P. 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrol. *Earth Syst. Sci*. 11: 1481-1500.

Wechsler, S. P., and C. N. Kroll. 2006. Quantifying DEM uncertainty and its effect on topographic parameters. *Photogrammetric Engineering & Remote Sensing* 72(9): 1081-1090.

Westphalen, M. L., B. S. Steward, and S. Han. 2004 Topographic mapping through measurement of vehicle attitude and elevation. *Trans. ASAE 47(5):* 1841−1849.

Will, J., Stombaugh. T., Benson, E., Noguchi, N., Reid. J.F., 1998. Development of a flexible platform for agricultural automatic guidance research. ASAE Paper No: 983202. ASAE, St.Joseph, MI.

Wise, S. M. 1998. The effect of GIS interpolation errors on the use of digital elevation models in geomorphology. In Landform Modelling and Analysis, 139−164. S. N. Lane, K. S. Richards, and J. H. Chandler, eds. Chichester, U.K.: John Wiley and Sons.

Whitney, J. D., J. D., Q. Ling, W. M. Miller, and T. A. Wheaton. 2001. A DGPS yield monitoring systems for Florida citrus. *Applied Engineering in Agriculture:* 17(2): 115–119.