Targeted Sampling of Elevation Data Based on Spatial Uncertainty of Prior Measurements

Samsuzana Abd Aziz
Iowa State University

Brian L. Steward
Iowa State University, bsteward@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/abe_eng_conf

Part of the Bioresource and Agricultural Engineering Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/abe_eng_conf/41. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Targeted Sampling of Elevation Data Based on Spatial Uncertainty of Prior Measurements

Abstract
An efficient sampling strategy should address knowledge gaps, rather than exhaustively collect redundant data. In this study, spatial uncertainty in DEM estimates was used to locate targeted sampling areas in the field. An agricultural vehicle equipped with RTK-DGPS was driven across a 2.3 ha field area to measure the field elevation. Data were collected at 3.05 m (10 ft) intervals in a continuous fashion at a speed of 9.6 mph. A geostatistical simulation technique was used to simulate field DEMs with different measurement pass intervals and to quantitatively assess the spatial uncertainty of the DEM estimates. The high uncertainty areas for each DEMs were classified using image segmentation methods and targeted sampling was performed on those areas. The resulting DEMs were compared with each other to evaluate the effect of including targeted measurement on DEM accuracy. 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 results show that spatial uncertainty models are useful to design targeted sampling for field mapping. The method is not limited to map elevation data but can be extended for mapping other spatial data.

Keywords
Digital elevation model, spatial uncertainty, targeted sampling, sequential Gaussian simulation

Disciplines
Bioresource and Agricultural Engineering

Comments
ASABE Paper No. 084826
Targeted Sampling of Elevation Data Based on Spatial Uncertainty of Prior Measurements

Samsuzana Abd Aziz
Iowa State University, 139 Davidson, Ames IA 50011, USA, suzana@iastate.edu.

Brian L. Steward
Iowa State University, 214D Davidson, Ames IA 50011, USA, bsteward@iastate.edu.

Written for presentation at
2008 ASABE Annual International Meeting
Sponsored by ASABE
Rhode Island Convention Center
Providence, Rhode Island
June 29 – July 2, 2008

Abstract. An efficient sampling strategy should address knowledge gaps, rather than exhaustively collect redundant data. In this study, spatial uncertainty in DEM estimates was used to locate targeted sampling areas in the field. An agricultural vehicle equipped with RTK-DGPS was driven across a 2.3 ha field area to measure the field elevation. Data were collected at 3.05 m (10 ft) intervals in a continuous fashion at a speed of 9.6 mph. A geostatistical simulation technique was used to simulate field DEMs with different measurement pass intervals and to quantitatively assess the spatial uncertainty of the DEM estimates. The high uncertainty areas for each DEMs were classified using image segmentation methods and targeted sampling was performed on those areas. The resulting DEMs were compared with each other to evaluate the effect of including targeted measurement on DEM accuracy. 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 resampling. The results show that spatial uncertainty models are useful to design targeted sampling for field mapping. The method is not limited to map elevation data but can be extended for mapping other spatial data.

Keywords. Digital elevation model, spatial uncertainty, targeted sampling, sequential Gaussian simulation.
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 farmers 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 patterns within the field over time. It may involve the use of models linking the various data sets within a geographic information system (GIS), which can provide information on the site-specific variability of the attribute in question. 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. However, the information may contain deviations from the truth or errors that constitute uncertainty.

In this study, the information about spatial uncertainty of elevation estimates from prior measurements was used as a rational basis for a future sampling plan to improve the accuracy of field DEMs. A geostatistical simulation technique was used to assess the accuracy and spatial uncertainty of elevation estimates. The simulation process produced multiple estimates (realizations) for a particular location and provided a range within which the true estimate lies (Wechsler, 2007). Information about spatial uncertainty was used to delineate the areas 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.

**Methods**

**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, Ill) operating at 1 Hz with a vertical static root-mean-squared error (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.

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 noise 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 were 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 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 advantages of using this technique are it preserves the flavor of real world variability and spatial correlation in the estimates; and also honors the observed data exactly 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 its extremely congenial properties of multi-Gaussian assumption. Using this technique, the spatial uncertainty of the elevation was modeled by generating multiple realizations of the elevation estimate in 1 m gridded DEMs. Each realization was randomly drawn from the probability distribution function (pdf) derived based on the first and second order statistics of the kriging estimate in each grid. 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 also known as E-type estimate of the grid. 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 was used as the uncertainty estimate of the grid estimate to characterize the areas that need to be re-sampled. In Matlab (The Mathworks, Natick, Mass.), an image segmentation algorithm was performed using a simple thresholding technique, where in our study; we chose the estimation variance threshold to be 0.04 m. Region classification was performed 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 area. Zeros binary value were assigned to the area that has value less than the threshold value. The process essentially transformed the DEM into a 1-bit binary image by allocating every grid in the DEM 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 which would not be considered as an area of interest. The Matlab morphological operations function bwmorph was used to perform a 'cleaning' operation, followed by 'filling' and 'removing' operations.

To mimic the real application of targeted sampling, where new samples should be taken only in the areas of interest; unused measurement passes that fell in the areas 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 area 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 pass were used to uniformly simulate the effect of adding new non-targeted sampled data within the division of data sub-groups.

**Data Analysis**

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).

Each generated DEM from each sampling types and calibration sub-groups were compared to the validation dataset from the validation group which had not been used to simulate the surface. Root mean squared error (RMSE), a 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.

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

\[ p = \frac{\partial z}{\partial x}, \]

\[ q = \frac{\partial z}{\partial y}. \]

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 \( w \).

\[ p \approx \frac{(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)}{8w} \]

\[ q \approx \frac{(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)}{8w} \]

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

\[ S = \sqrt{p^2 + q^2} \]

Based on this formulation, uncertainties in the slope, \( \Delta S \) were calculated using the sensitivity coefficients with respect to the nine estimates of the elevation, \( z_i \), each with their own uncertainty, \( \Delta z_i \), which obtained from the conditional simulation method:
\[ \Delta S = \sqrt{\left(\frac{\partial S}{\partial p}\right)^2 + \left(\frac{\partial S}{\partial q}\right)^2} \]

where \( \Delta p = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial p}{\partial z_i}\right)^2 (\Delta z_i)^2} \),

and \( \Delta q = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial q}{\partial z_i}\right)^2 (\Delta z_i)^2} \).

The uncertainty of the derived slope for each DEM was visually assessed using contour plot. 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.

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 value ranged around 0.1 to 0.16 m at this area. The uncertainty is small in the south and northwest of the study area where elevation is flatter (plain region). The variance ranged around 0 to 0.04 m 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 passes 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 have multiple variance modes 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 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 measurement intervals has variances distribution in the smallest range, from 0 to 0.19 m, 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 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.](image)

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 1-bit 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 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 m². With interval widths larger than 10 m, the high uncertainty regions ranged from 8,700 to 12,400 m².

Figure 4: Targeted areas 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 areas 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.](image)

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). 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.

![Graph showing RMSE of slope derived from DEMs](image)

**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 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.](image)

**Conclusion**

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 areas 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


