Comparing high density LIDAR and medium resolution GPS generated elevation data for predicting yield stability

J.M. McKinion*, J.L. Willers, J.N. Jenkins

USDA-Agricultural Research Service, Genetics and Precision Agriculture Research Unit, P.O. Box 5367 Mississippi State, MS 39762, United States

Abstract

High density light detection and ranging (LIDAR) imaging has been shown to be able to define yield stability areas of a field for multi-cropping. Since LIDAR imaging is expensive and not widely available, it was hypothesized that medium resolution GPS elevation data which is commonly collected with variable rate technology (VRT) controllers and crop yield monitors could be used in lieu of LIDAR imaging. If proven, growers would be able to construct yield stability maps of their fields without the expense of obtaining LIDAR imaging. After substituting medium resolution GPS elevation data derived from the crop yield monitors, the procedure developed for developing a crop yield stability map was invoked and tested. The hypothesis that medium resolution GPS data could be used in lieu of LIDAR data was found to be invalid as the map generated incorrectly identified both high yield and medium yield areas of the field as low yielding areas as well as the inverse. While disappointing, high resolution GPS data from real-time kinetics (RTK) systems is yet to be tested and may offer an additional avenue to developing crop yield stability maps.

1. Introduction

Crop stability patterns for a field have been shown to hold for multi-crops grown over several years using high density LIDAR imaging as the base map. The hypothesis tested in this paper is whether medium density elevation data from GPS data collected from yield monitors is sufficient to generate topographical maps for comparable use. This is the third paper in a series exploring the use of digital elevation maps (DEMs) in concert with crop yield maps to describe areas of yield stability in a commercial field. In the first paper (McKinion et al., 2010), an analytical procedure using ESRI ARCGIS® (2009), ERDAS Imagine® (2009) and the SAS® (2008) statistical software packages to develop a map representing areas of yield stability using five years of crop yield data of cotton and corn and a high density DEM generated from LIDAR flown on a fixed wing aircraft (Huising and Pereira, 1998) was described. Krigged, normalized crop yield maps were used in this process. In later work (McKinion, personal communication), the analytical process was shown to be improved via the addition of Krigged Veris® apparent electrical conductivity data with both shallow and deep apparent conductivities used. However, the Veris® data alone did not produce acceptable stability patterns. The final map produced was a three-color map using manually supervised classification selection by the five-year crop yield average. While LIDAR generated high density DEMs are becoming more available, they are not ubiquitously available. Growers, however, who are using precision agriculture do have access to GPS generated DEMs. This manuscript focuses on whether GPS generated DEMs can be used in lieu of LIDAR generated maps to produce corresponding yield stability maps of a field.

2. Literature review

This is the third paper in a series which explores the hypothesis that yield stability patterns exist in fields which, once identified via the application of GIS, spatial analysis, and statistical methodologies, can be used by growers to optimize crop planning and production on a management zone basis (McKinion et al., 2010). Ping and Dobermann (2005) report that yield mapping is one of the most widely used precision farming technologies. However, as more and more yield monitors are used and multiple-years of data are accumulated, there is an increasing need for robust data processing and interpretation techniques. As pointed out by Kaspar et al. (2003), the use of yield maps in decision making for the next season is difficult due to problems of interpretation. Permanent spatial factors that affect yield either directly or indirectly are landscape position, terrain attributes, erosion class and soil properties (Spomer and Piest, 1982; Stone et al., 1985; Jones et
3. Materials and methods

3.1. Synopsis of procedures

A synopsis of the methodology developed for generating multi-crop yield stability maps from McKinion et al. (2010) follows. The Statistical Analysis System (SAS®) from the SAS Institute (2009), ARCGIS® from ESRI Corp. (2009) and Imagine from ERDAS, Inc (2009) were used in this work. Fig. 1 is a flowchart of the procedures used to help the reader to readily see the analysis process.

A LiDAR DEM map of the Paul Good Farm (Headquarters 03° 07′ 28.80″ N 088° 29′ 46.58″ W) was obtained. ARCGIS® was used to extract Field 160 for this study. The Field 160 map was subsequently used as the extraction layer to remove any areas outside of the feature of this map. This step ensured that each and every map added to the geodatabase had exactly the same area and that each point or pixel would correspond exactly to the same point or pixels develop in other maps. Field 160 map was then Krigged to generate an elevation surface map. The next step was to extract the Krigged map using the original Field 160 map. Maps of aspect, curvature and slope were generated from the extracted, Kriged elevation map of Field 160. These four maps will be subsequently called Elevation, Aspect, Curvature and Slope.

Crop yield monitor maps for the years 2001 through 2005 were collected on the Good Farm. In 2001, 2003 and 2004, cotton was grown in Field 160. In 2002 and 2005, corn was grown. The yield maps were first processed using ARCGIS® whereby Krigging was applied to generate yield grid maps. The five maps were next extracted using the Field 160 elevation map (small e). All nine of the maps (five yield grid maps and maps of Elevation, Aspect, Curvature and Slope) were exported from ARCGIS® and imported into Imagine. Imagine was used to generate a data stack which consisted of columns of easting, northing, five crop year yields, elevation, aspect, curvature, and slope. Thus, a dataset was created in which each point, or pixel, in the maps were uniquely identified by an easting and northing location point. The coordinate system used was a projected coordinate system in UTM in NAD1983 in Zone 16N.

This dataset was then exported as a .dbf file for importation into SAS® for statistical analysis procedures. The first step was to convert the yield data for cotton and corn into a normalized yield from 0 to 255 to prevent statistical abnormalities due to differing yield ranges and to remove weather effects. The SAS® Fast-Cluster procedure was invoked with the five normalized yield columns and the Elevation, Aspect, Curvature and Slope columns as the analysis variables. Twenty clusters were specified to obtain a good range to delineate transitions in the field and this column was added to the dataset. Regression on the 20 clusters yielded a statistically significant $R^2$ with a $P > F$ for $<0.0001$ in each case.

The new dataset, including the cluster data, was exported from SAS® and imported into ARCGIS® where a new map was created which showed Field 160, as represented by the cluster data. By referring to the yield data as summarized by the SAS® Means Proc by cluster, an interpretation could then be made.

The Veris® data were then added to the SAS® dataset after Krigging and extracting map operations were performed in ARCGIS® as outlined above. These maps were then exported to Imagine to add the Veris® shallow and deep conductivity columns to the dataset. After exporting to SAS®, a new FastCluster analysis was performed with the same variables as before, but with the Veris® information also included. Detailed analysis showed that overall variability explained by Proc GLM regression by cluster improved by over 30%. Proc Means was again invoked by cluster and this time the number of clusters was reduced manually from 20 clusters to three clusters after examining yield breaks for the 20 cluster table. The dataset which now included the Veris® data and the 20 cluster numbers was then recoded down to 3 cluster numbers manually using Microsoft Excel. After exporting to ARCGIS® a final three-color map showing yield stability by high, medium and low yielding areas was produced (Fig. 2). For those who need a more detailed explanation of the procedures and the methodology developed, see McKinion et al. (2010) (Fig. 3).
Fig. 1. Flowchart of the analysis process to produce a yield stability map from five years of normalized crop yield data. Three different software systems were used in the analysis. The boxes in yellow represent procedures executed in the ESRI ARCGIS system. The box in green represents processes executed in the ERDAS Imagine system. The boxes in orange represent processes executed in the Statistical Analysis Software Institute SAS system. In the last two yellow boxes yield stability maps with 20 and 3 yield areas were generated. The 20-color yield map allows the observer to see fine details of transition versus the topology of the field. The three-color map represents a more practical application map for the grower to use. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.2. Application of GPS elevation data

Five crop years of yield data and one year of Veris® data were collected at the Paul Good Farm. Each of these datasets possessed an elevation component along with yield and easting and northing.
Fig. 4. Digital elevation map of Field 160 from a LIDAR image which has been degraded to a 3 m by 3 m pixel resolution for use as the base map for topology.

Fig. 5. Difference map created by subtracting the GPS map in Fig. 2 from the LIDAR map in Fig. 3.

4. Analysis of results

High density LIDAR data was used to produce the map in Fig. 2. The pixel representation is actually a degradation of the 10 cm accuracy of the original DEM aggregated up to a 3 m by 3 m pixel to correspond to the yield maps generated by corrected GPS-based yield monitors. The three-color map representing yield stability areas which are high, medium and low production areas is well-defined and statistically valid.

The elevation data which was used to generate Fig. 8 was developed from corrected GPS-based yield monitors and Veris® soil conductivity equipment. These data were also corrected to produce the elevation data at the surface of the field. The six datasets were averaged to produce a single elevation map that attempted to reduce the effects of yaw, pitch and roll of the GPS antenna as the harvest equipment moved across the field and GPS errors (as described by Yao and Clark (2000) and Kumar and Moore (2002))...
due to atmospheric aberrations and the dilution of precision and geometric dilution of precision of the GPS signals).

The difference maps shown in Figs. 5 and 7 indicate that the height correction and the Omnistar correction for the averaged dataset prove to be fairly accurate, but the appearance of the striations in the maps is problematic. Striation could originate from the north–south direction of travel of the harvest equipment in the field, particularly in the areas of the field where two harvesters were used in parallel, one with yield monitor and the other without.

Comparing the highly organized elevation map produced by LIDAR data shown in Fig. 4 with the GPS\textsubscript{Elev} map shown in Fig. 5 again illustrates the striation in Fig. 6 versus Fig. 4. Even so, there was still sufficient organization in the GPS\textsubscript{Elev} surface map such that the procedure could potentially be used with corrected GPS elevation data for favorable deployment.

Comparing the stability map (generated by LIDAR data combined with several years of yield data and Veris\textsuperscript{®} conductivity information shown in Fig. 2) with the map generated from GPS elevation data, it was concluded that the GPS elevation data were not of sufficient quality nor dense enough to produce a usable yield stability map. The path of the harvest equipment with the yield monitor in Field 160 was predominantly in the north–south directions. This pattern could have contributed to striations in the topology maps. The map in Fig. 8, identifies some areas which are the same as the map in Fig. 2; however, it misidentifies many areas of the field in all three categories of high, medium, and low yield and is very different from the yield stability map developed using the high density, more accurate LIDAR DEM. The map produced using GPS data from the harvester is not of sufficient quality to be used for making management decisions.

### Table 1

Basis for reducing 20 yield cluster groups down to 3 cluster groups. The upper group is the high yielding areas. The middle group separated by the heavy horizontal lines is the medium yielding areas, and the lower group is the low yielding areas of Field 160. Each observation point (or pixel) represents a 3 m by 3 m ground area. The yields are normalized over the range of 0–255, so that cotton and corn yields could be statistically analyzed together. The presence of a significant break in yield average was used as the determining factor for the three classes of yield.

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Number of observations</th>
<th>5 year average yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2186</td>
<td>166.2</td>
</tr>
<tr>
<td>6</td>
<td>1009</td>
<td>163.3</td>
</tr>
<tr>
<td>11</td>
<td>1993</td>
<td>159.1</td>
</tr>
<tr>
<td>2</td>
<td>777</td>
<td>146.3</td>
</tr>
<tr>
<td>13</td>
<td>1140</td>
<td>144.3</td>
</tr>
<tr>
<td>12</td>
<td>536</td>
<td>143.2</td>
</tr>
<tr>
<td>5</td>
<td>5790</td>
<td>142.1</td>
</tr>
<tr>
<td>8</td>
<td>25,053</td>
<td>141.1</td>
</tr>
<tr>
<td>1</td>
<td>847</td>
<td>136.4</td>
</tr>
<tr>
<td>19</td>
<td>635</td>
<td>136.3</td>
</tr>
<tr>
<td>3</td>
<td>6412</td>
<td>128.7</td>
</tr>
<tr>
<td>14</td>
<td>1299</td>
<td>126.9</td>
</tr>
<tr>
<td>7</td>
<td>2238</td>
<td>106.4</td>
</tr>
<tr>
<td>10</td>
<td>257</td>
<td>105.3</td>
</tr>
<tr>
<td>15</td>
<td>5302</td>
<td>104.3</td>
</tr>
<tr>
<td>17</td>
<td>2540</td>
<td>101.6</td>
</tr>
<tr>
<td>18</td>
<td>741</td>
<td>100.4</td>
</tr>
<tr>
<td>16</td>
<td>705</td>
<td>99.6</td>
</tr>
<tr>
<td>20</td>
<td>495</td>
<td>97.7</td>
</tr>
<tr>
<td>9</td>
<td>956</td>
<td>88.5</td>
</tr>
</tbody>
</table>
four rows contiguously. As the yield monitor and the Veris machine traversed the field, the antenna was subject to yaw, pitch and roll movements which were recorded in the elevation data. We first krigged the elevation data on a year by year basis and also krigged the Veris elevation data to create surface maps. These surface maps were then corrected for the antenna evaluation to arrive at the soil elevation. By averaging the five years of krigged, corrected elevation data with the single year of the Veris krigged, corrected elevation data as shown in Fig. 6, an attempt was made to not only compensate for yaw, pitch and roll errors in the antenna movement but also to allow for non-contiguous yield monitor and Veris machine paths in parts of Field 160. At this point the procedure developed for LIDAR elevation data was invoked and a crop yield stability map was produced as shown in Fig. 8. Even though the statistics showed that the clustered yield data points were significant, the resulting elevation map was very noisy and showed striations in each part of Field 160 which were not present in the elevation map produced by LIDAR measurements. Averaging should have reduced some of the noise in the elevation signals, but the elevation map produced by GPS does not appear to be of sufficient quality to generate a reliable crop stability map. While parts of Field 160 were contiguously sampled with the yield monitor and with the Veris machine, misidentification of high, medium and low yield areas appeared not only in the non-contiguous sample areas but also in the contiguous areas, indicating that the density of sampling paths was not the principle fault for misidentification. The conclusion then must be that the medium resolution GPS elevation data were not of sufficient resolution and accuracy needed to produce the crop yield stability map.

6. Conclusions

Methodology was developed using GIS and image analysis software in conjunction with statistical analysis software to identify areas of yield stability in a commercial field growing multi-crops over time. This methodology was based on the availability of a high density LIDAR digital elevation map of the field. The hypothesis explored in this manuscript evaluated whether corrected GPS elevation data collected by harvesters could substituted for the LIDAR data in the procedure and be able to produce similar, usable stability maps. The same crop yield data and the same Veris shallow and deep apparent conductivities were used in this analysis. The same procedure was applied to create elevation, aspect, curvature and slope topology maps from the GPS elevation data used in the analysis. The yield stability map produced from GPS data was very different and conflicted with the LIDAR based analysis. It was concluded that harvester GPS data is not suitable for producing high quality yield stability maps using this methodology. An open question is whether the more accurate real time kinetics (RTK) elevation data could be used with this methodology to develop usable yield stability maps.

References


