Precision Agriculture, 6, 341–357, 2005

Spatial Influence of Topographical Factors on Yield of Potato (*Solanum tuberosum* L.) in Central Sweden

ANDREAS PERSSON andreas.persson@jti.slu.se
Swedish Institute of Agricultural Engineering (JTI), P.O. Box 7033, SE-750 07, Uppsala, Sweden

PETTER PILESJÖ petter.pilesjo@giscentrum.lu.se
GIS-centre, Lund University, Sölvegatan 12, SE-223 62, Lund, Sweden

LARS EKLUNDH lars.eklundh@nateko.lu.se
Department of Physical Geography and Ecosystems Analysis, Lund University, Sölvegatan 12, SE-223 62, Lund, Sweden

Abstract. This study has evaluated the sampling density for creation of high-resolution digital elevation models (DEMs) for precision agriculture purposes. The relationships between yield and topographical factors were investigated in a study area located in the central Sweden province of Dalarna. The DEM data sampling was carried out with a RTK-GPS system. A dense sampling scheme was employed and data was divided into two for both interpolation and validation. Kriging interpolation was used for DEM generation. From the DEM, topographical parameters were extracted and topographical indices were estimated. The indices were calculated with slope length and its vertical and horizontal components. The drainage area for a point of interest and the relationship of this area to the total drainage area were also estimated. The relationship of yield and the topographical parameters and indices was investigated using both circular and spatial statistics. A spatial regression was used to calculate a model for the relationship. Up to 20% of the yield could be explained in the final model for one of the fields.

Keywords: Yield, Topographical indices, Potato, DEM, Drainage area, Spatial regression

Introduction

Yield variation and its relation to topography

The aim of site-specific agriculture is to optimise the use of spatial and temporal management strategies. Such optimisation can improve crop yield and quality and reduce the risks for nutrient and pesticide leakage. During the past decade, several projects have focused on quantifying and characterising variation in factors such as crop yield, soil properties and precipitation and their interrelationships.

Generally, yield variation has been expected to be related to variation in properties of the underlying soil. Research has focused on topsoil depth, soil organic matter content, clay content, pH, phosphorus content, and, of course, soil nitrogen content. For instance, a negative correlation between crop yield and phosphorus content was reported by Webster (Lake *et al.*, 1997). Electrical conductivity and topographic attributes were used for management zone delineation in Missouri, USA. The

'goodness' of management zones was tested against grain yield data (Fraisse *et al*., 2001). However, not all of the variation in yield can be explained by these variables. Fiez *et al*. (1994) found a positive correlation between yield variation and soil water content. Schneider *et al*. (1997) found a positive correlation between elevation and potato yield in an irrigated field. In dry-land farming systems, the opposite relationships can be expected. Landscape position and plant available water were positively correlated with corn silage yield in a study by Afyuni *et al*. (1993). It has also been noted that soil moisture affects yield in different ways depending on other factors (e.g. temperature and radiation).

Digital elevation models (DEM) can be used to describe the surface elevation of the landscape. DEMs are interpolated from known elevation points. Data sources for the interpolation of topography include topographical maps, aerial photographs and GPS-collected data (Burrough and McDonnell, 1998). In interpolation, the value of an unknown point is estimated using the values of surrounding known points. In most interpolation algorithms, a measured point close to the unknown point has a larger influence than a point farther away from the unknown one. One of the most frequently used interpolation algorithms is kriging (Burrough and McDonnell, 1998). Kriging is a geo-statistical method that models continuous data as the sum of three components: a structural component that shows the mean trend in the spatial dataset, a spatially dependent stochastic component, and a spatially uncorrelated random component that describes the noise (Cressie, 1991; Burrough and McDonnell, 1998). A semivariogram is used to determine the input parameters for the interpolation. Kriging has been used for interpolating surface elevations by e.g. Gao (1997) and Blomgren (1999). In precision agriculture, kriging has been used for several purposes, e.g. soil analysis by Moore *et al*. (1993) and for spatial crop modelling by Sudduth (1997).

Apart from the interpolation algorithm, the result of a spatial interpolation also depends on the spatial distribution of the input data points. It has been shown that data sampled along isolines (lines with equal elevation steps in between) often generate terraces or striping in DEMs due to the nature of most local interpolation algorithms. A convenient way to handle this problem is to ensure that the density of the data points is equal in all directions (Eklundh and Mårtensson, 1995).

Topographical parameters, such as slope, aspect, and drainage area, are relatively easy to derive from a gridded digital elevation model. The grid has the form of a matrix with equally sized cells in rows and columns. Each cell is assigned a value for the characteristic that it represents. In DEMs, a cell value represents the elevation of the midpoint of the corresponding area of ground. The cell size and the characteristic that it describes are homogeneous within a DEM (Pilesjö, 1992).

Topographical indices are sometimes useful when explaining the topographical influence on different phenomena. Simple indices explaining topographical position have been used for decades (e.g. Skidmore, 1989), while more complex indices are frequently used in hydrological modelling for instance. An example of the latter type is the frequently used wetness index based on slope and drainage area (Moore *et al*., 1991). This index has been tested for different drainage area algorithms against field measurements by Schmidt and Persson (2003).

DEMs with a high accuracy are a necessity in precision agriculture and hydrological modelling. Although commercial maps with a scale of 1:20,000 are available

in Sweden, this resolution in contour interval (5 m) is not sufficient for fields that are no larger than a couple of hectares. The resolution cannot be used to explain the soil moisture movements needed to estimate plant available water. Instead, an accuracy on decimetre or centimetre level is required.

Past research has mainly focused on the topography being divided into different slope classes, and has not treated the slope as a continuous surface. Instead, classes such as top slope, mid slope and foot slope have been used (Changere and Lal, 1990; Manning *et al.*, 2001).

Objectives

The objectives of this study were:

- to evaluate the field sampling distribution and density for creation of accurate DEMs in cultivated landscapes and
- to investigate the relationship between yield variation of potato and a number of topographical parameters and indices.

To achieve these objectives, a number of steps were undertaken. First, elevation data were collected. This was followed by interpolation of the data into gridded DEMs. From these DEMs, the topographical parameters and indices were extracted and estimated and then comparisons between yield and topography were made.

Materials and methods

Study area

Two fields in the province of Dalarna, central Sweden, served as the study sites (latitude 60°24′35″, longitude 15°47′10″). The fields have homogeneous sandy, silty soils. They are situated on a flood plain formed by the movements of a river. A hillside towards the river is the dominating characteristic of the two fields. Undulating features within the fields are superimposed on the slope and are more pronounced for the larger of these fields (the fields are called 'larger' and 'smaller' below). The maximum and minimum elevations differ by 15.1 m. The maximum gradient is 12.1°. The mean gradient of the fields is 5.6° with a standard deviation of 1.8°.

Generation of digital elevation models

Field data collection. All elevation data were registered by the NAVSTAR Global Positioning System (GPS). A RTK (Real Time Kinematic) system, consisting of two Trimble Site Surveyor 4400 units, recorded positions both vertically and horizontally with a precision of millimetres, using both carrier wave measurements and code measurements (Hofmann-Wellenhof *et al.*, 1997).

The mobile GPS unit was carried by an all terrain vehicle (ATV). The GPS antenna was mounted above one of the front wheels of the ATV to assure sampling in

the corners and along the edges of ditches. Sampling across the fields was then conducted along transects approximately 6 m apart (Figure 1). The sampling density was 1 point per 1.5 m along the transects, determined by the GPS receiver frequency and the velocity of the ATV (approximately 3 km/h). An example of the GPS sample geometry is shown in Figure 1a.

After data collection, the dataset was split into two. Every second sampled point was used for interpolation while the remaining points were used for validation. This yielded one interpolation dataset (IDS) and one evaluation dataset (EDS) for the area to be interpolated. For both fields together, the full IDS as well as the EDS consisted of 13,369 points each.

Interpolation. After data collection, the IDS was interpolated with ordinary kriging to create continuous surfaces representing the topography. Following an analysis of the semi-variogram, the range was set to 55 m, the sill to 0.5 and the nugget to zero. The whole area, the two fields and several surrounding fields, was interpolated since data collection was not restricted to the field boundaries. The surface was then interpolated with a cell resolution of 1 m. Since the accuracy of the RTK was better than 0.001 m, the number of decimals in the interpolated cells were set to 3 (i.e. mm accuracy). After this, the individual fields were extracted from the 'main' DEM data.

The interpolated surfaces were evaluated against the EDS. Elevation values were extracted in the interpolated surfaces corresponding to the positions of the evaluation data points. The elevation values of these cells were then compared to the elevation data points measured with the RTK in the field (i.e. the EDS). Descriptive statistics for the datasets were calculated. Since errors may be positive or negative, the absolute values of error were analysed. In addition, the root mean square error (RMSE) was calculated for the interpolated surfaces.

In order to test the effect of different data sampling densities, two additional datasets were created. From the full dataset (IDS) of 13,369 points, every second point along the sampled transects was removed, creating the second interpolation

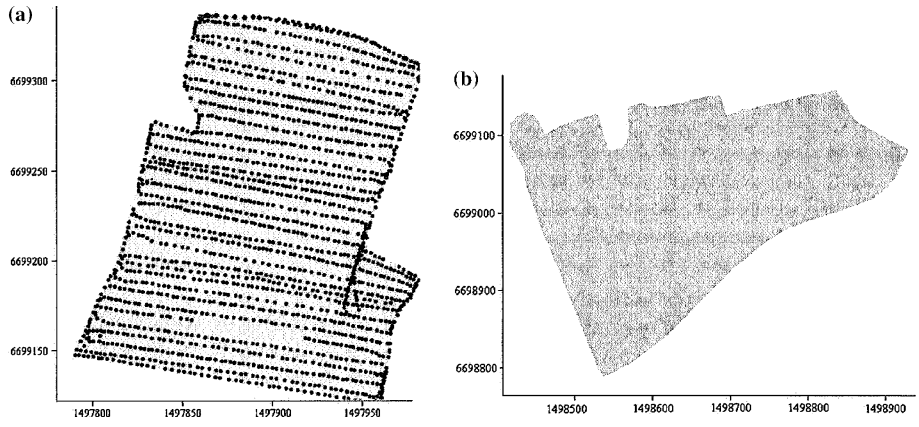

Figure 1. Boundaries of the smaller (a) and larger field (b). Data collected with an ATV driven across the field along transects 6 m apart as seen in (a). A position with x, y and z values was collected every 1.5 m. Coordinates in Swedish National Grid (RT90), meters (m).

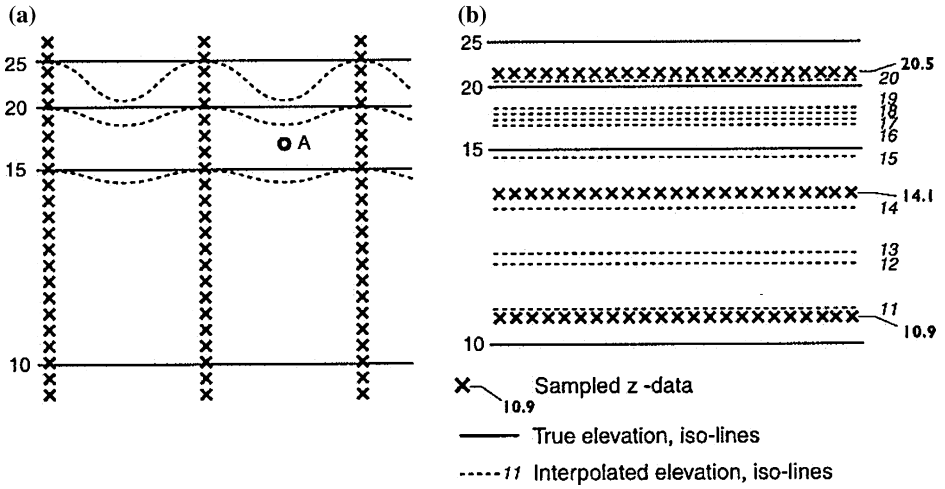

Figure 2 (a–b). Schematic graph of interpolation bias due to elevation sampling in transects perpendicular (a) and parallel (b) to the 'true' isolines; see the text for explanation.

dataset (IDS2, $n = 6684$). For the third interpolation dataset (IDS3), two out of every three points were removed ($n = 4456$).

Striping and/or terracing was revealed by visualising the interpolated surfaces, and by studying the potential water flow (often referred to as 'flow accumulation') over the interpolated surfaces. The latter was simulated with a flow accumulation algorithm presented in Pilesjö *et al.* (1998). Since the potential water flow also concentrates in smaller depressions and valleys in the terrain, the simulation easily detects these sorts of landforms. This way of revealing illogical disturbances on the surface due to interpolation errors was used to optimise the input data in order to create a DEM representing the true surface.

Figure 2 illustrates a hypothetical lack of fit between the possible isolines in the field and the interpolated isolines extracted from a 'striped' DEM. The lack of fit may occur when sampling is performed perpendicular to the isolines (i.e. when data is sampled 'up-down', as well as parallel to the possible isolines. In both cases, the anomalies depend on the distance weighting of the known input data points.

Point A in Figure 2a is situated in a concave slope between two sample transects. The up-slope points are weighted equally with the down-slope points in the interpolation. Since A is situated on a concave slope, the differences in elevation are greater up-slope than down-slope (within the same distance). This yields an overestimation of A, and the isolines will bend downwards. The more concave the slope, the greater the striping effect. If the slope were convex, the result would be an underestimation, and the isolines would bend upwards. In Figure 2b, the data points are sampled parallel to the isolines. The striping in this example can be explained by an over-representation of data points with one and the same elevation value. Close to the known points, interpolated values are more or less equal to the known values.

In the centre, between the transects, there is a higher density of interpolated isolines. These factors result in terraces/striping.

Estimation of topographical parameters. The basic characteristic of a DEM is the elevation dataset, which in itself constitutes a topographical parameter. With this parameter and the knowledge of the raster geometry, other parameters may be obtained. It was decided to estimate slope (gradient) and aspect (slope direction), as well as slope length and drainage area.

Second-order trend surfaces were calculated within a 3 by 3 cells window moving over the DEM (see Eq. (1)). The slope and aspect at the mid-point in these trend surfaces were then used to represent the window (Pilesjö *et al.*, 1998). The parameters were estimated according to Eqs. (2) and (3), following Pilesjö *et al.* (1998):

$$z(x, y) = b_0 + b_1 x + b_2 y + b_3 xy + b_4 x^2 + b_5 y^2 \tag{1}$$

$$\text{slope} = \arctan \text{sqrt}(b_1^2 + b_2^2) \tag{2}$$

$$\text{aspect} = 180 - \arctan(b_2/b_1) + 90(b_1/|b_1|) \tag{3}$$

where b_i are coefficients, z is the elevation value on the second order trend surface and x, y are the horizontal position coordinates.

Slope length (L) and drainage area up to the water divide (D) were estimated according to Pilesjö *et al.* (1998).

Generation of topographical indices. The parameters used to estimate the six topographical indices generated are presented in Figure 3. Four indices are based on slope length (Figure 3a), and two are based on drainage area (Figure 3b).

Slope length indices. Depending on how distances between different points in the DEM are defined, a number of different slope length indices can be designed and estimated. When generating the two slope length indices (SLI1 and SLI2), the dis-

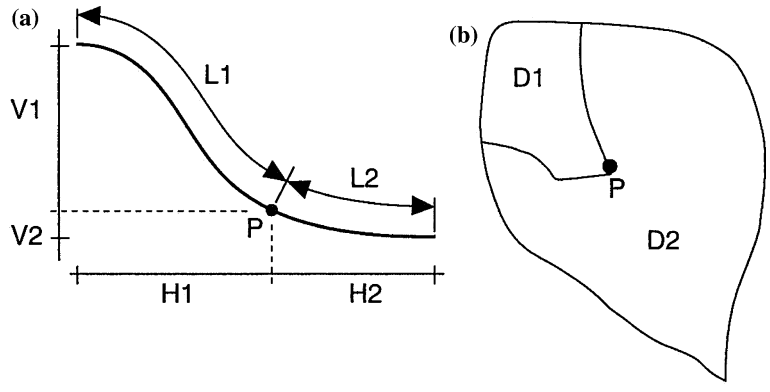

Figure 3. Topographical indices based on (a) the components of slope length and (b) drainage area. Point of interest is marked P.

tance (L_1) between the water divide and the point of interest (P) was defined and estimated as following the topography (see Figure 3a). That was also the case when estimating the distance (L_2) between the point of interest and the lowest point in the drainage basin, i.e. outlet :

$$L_1 \qquad \text{(SLI1)}$$

$$L_1/(L_1 + L_2) \qquad \text{(SLI2)}$$

If the horizontal or vertical distance along the flow path is presumed to be significantly influential on yield, it is relevant to generate slope length indices based on these distances. To test the influence of horizontal and vertical distances, a horizontal index (HI) and a vertical index (VI) were created (see Figure 3a):

$$H_1/(H_1 + H_2) \qquad \text{(HI)}$$

$$V_1/(V_1 + V_2) \qquad \text{(VI)}$$

Drainage area index. The drainage area indices (DAI1 and DAI2) were defined and estimated as the ratio between the area (D_1) above the point of interest (P), and the total drainage area of the drainage basin, D_2 (see Figure 3b):

$$D_1 \qquad \text{(DAI1)}$$

$$D_1/D_2 \qquad \text{(DAI2)}$$

The drainage areas were estimated according to Pilesjö *et al*. (1998).

Estimation of potato yield. The potato yields were estimated using a mapping technique that is under development in Sweden. The yield mapping system is based on an optical sensor that uses image analysis to count and gauge the size of the tubers (Persson, 1998). The system has been tested for 5 years in the field and in a laboratory test, which showed a deviation from the measured values of less than 2% for tuber sizes (Persson *et al*. 2004). The data used in this study was collected during the harvest period (September through October) 2002. Together with the size and amount measurements, the positions were logged with a GPS system every second.

The sensor calibration was based on a regression model with an $R^2 = 0.98$. Since the sensor readings (cross-sectional area) were correlated to weight, the relationships should not be linear. The model equation, $y = 0.010x^{1.26}$, fitted to the data in the calibration was used to calculate the yield biased on the field data.

The delay for a tuber passing through the harvester was 32 s. The band conveyors carrying the tubers were all set to a constant speed during data collection. The delay was corrected for in the post-processing of the harvest data and positions were calculated.

Yield variation and its relation to topography. In order to investigate the spatial influence of topographical factors on yield of potato, the estimated parameters and indices were statistically tested versus yield data. The relationships between the topographical data and yield were tested by the use of Pearson's correlation coefficient. C-association (Fisher, 1993) was used for testing the relationship between aspect and yield. Spatial regression and stepwise multiple regression were used in order to explain the variation in yield.

The C-association is a way of correlating a linear variable (X) to a circular variable (Θ), e.g. yield and aspect. As Θ performs a cycle from 0 to 360°, the expectance $E(X|\Theta = \theta)$ performs a sine-curve oscillation over the same range. The C-association method ranks the variables X and Θ from a sample and the ranking order numbers are then named: $s_1 \ldots s_n$ and $r_1 \ldots r_n$, respectively. A value between 0 and 1 is obtained for D_n (Eq. (4)), where a value close to unity suggests a high correlation. A significance test for the C-association was also used, called the U_n-statistic (Mardia, 1972; Fisher, 1993).

$$D_n = a_n(T_c^2 + T_s^2), \tag{4}$$

where

$$T_c = \sum_{i=1}^{n} s_i \cos r_i, \quad T_s = \sum_{i=1}^{n} s_i \sin r_i$$

and

$$a_n = \begin{cases} 1/[1 + 5\cot^2(\pi/n) + 4\cot^4(\pi/n)] & n \text{ even} \\ 2\sin^4(\pi/n)/[1 + \cos(\pi/n)]^3 & n \text{ odd} \end{cases}$$

(equations from Fisher, 1993).

The Spatial regression. Since distributed data are often autocorrelated, the coefficients for a regression may vary spatially. The method of spatial regression provides a tool to take into account the residuals for the neighbouring, autocorrelated values (Rogerson, 2001). For the response variable (y) and for each of the predictors (x) in the regression, new values (y^* and x^*) were calculated by:

$$y_i^* = y_i - \rho \sum_{j=1}^{n} w_{ij} y_j, \tag{5}$$

$$x_i^* = x_i - \rho \sum_{j=1}^{n} w_{ij} x_j, \tag{6}$$

where ρ is a measure of the correlation strength and w_{ij} is the weight of influence from autocorrelated surrounding values (equations from Rogerson, 2001). When ρ is zero, the spatial regression reverts back to a normal linear regression model.

The w-values included in the calculations of x^* and y^* may be based on a distance weight, which determines the spatial influence. In our case, a semi-variogram was used to derive the maximum distance of spatial influence. This distance was set to 30 m and the weight (w) was set to 1/distance^2. The regressions were then tested for a

number of ρ. The ρ that generated the lowest standard error of the estimate was used for further analysis.

The new set of response and predicting variables was used in a stepwise multiple regression. The threshold for including and excluding a variable was set to a 0.15. Only models where all parameters were significant ($P = 0.05$) were kept.

Results

Field sampling and evaluation of the interpolation

Visual examination revealed that striping occurred in some of the interpolated surfaces (Figure 4). A comparison with the elevation data points sampled in the field 17 (Figure 1) indicated that the striping was concentrated parallel to the sampling transects. When studying interpolations based on the first of the reduced dataset (IDS2), the negative influence of the dense sampling scheme was still apparent, though to a lesser extent. Figure 5 shows this surface where water from a flow

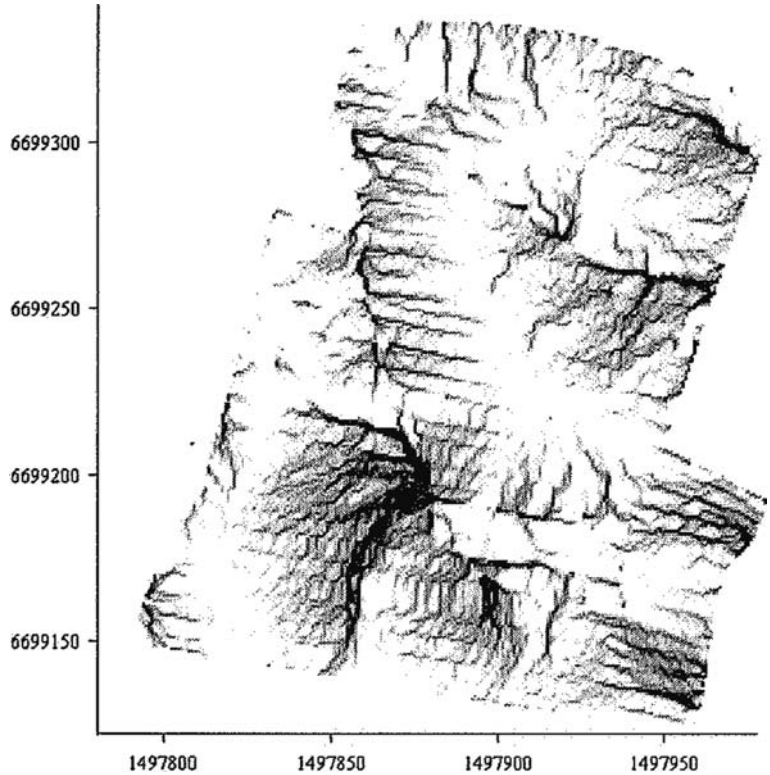

Figure 4. The flow pattern from the full dataset. The striping is obvious over the whole field. The darker the map more accumulated the flow. Coordinates in Swedish National Grid (RT90), meters (m).

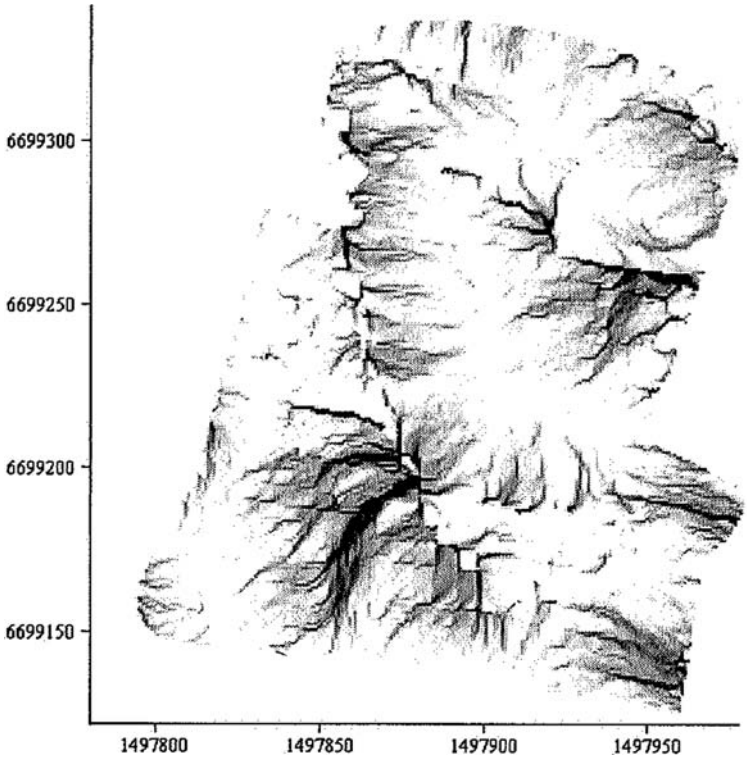

Figure 5. A dataset reduced by removal of every second point and the kriging algorithm disguise the pattern where flow accumulation is low. The striping is revealed where flow accumulation becomes higher. The darker the map, the more accumulated the flow. Coordinates in Swedish National Grid (RT90), meters (m).

accumulation calculation is forced along the stripes. The pattern was not visible for IDS3 (Figure 6), which was reduced by two out of every three points.

The results of the evaluation for each dataset are presented in Tables 1, 2, and 3, respectively. Mean values of the interpolated surfaces were within centimetres of the mean value of the sampled data. The maximum error was greater, in the order of a couple of decimetres. However, for one of the reduced datasets (IDS2), the interpolated surface had a maximum error of about 2 dm. The root mean square error was calculated to 0.036 m for the best interpolation.

Scatter plots over the value from interpolated surfaces of IDS, IDS2 and IDS3 against the EDS is presented in Figure 7(a–c). Correlation coefficients (r^2) was 0.999, 0.996 and 0.999 for IDS, IDS2 and IDS3, respectively.

Yield variation and its relation to topography

On the larger field, the yield ranged from 9 to over 52 t/ha (Figure 8). A lower yield was harvested on the smaller field (Figure 9), a minimum of 4.8 and a maximum of 27.2 t/ha were recorded. Descriptive statistics for yield are presented in Table 4.

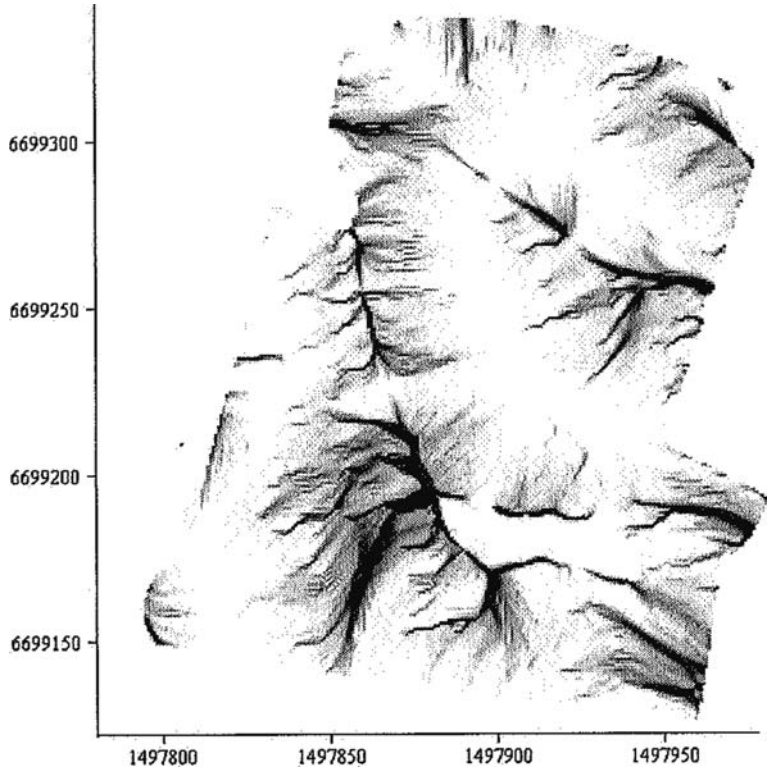

Figure 6. Flow accumulation run on the DEM from kriging interpolation with data reduced by two out of every three points gives a result that is both intuitively and physically correct. The darker the map the more accumulated the flow. Coordinates in Swedish National Grid (RT90), meters (m).

The C-association for testing the hypothesis of a relationship between aspect and yield was calculated for each field. For the smaller and larger fields, the D_n-values were 0.04 and 0.03 respectively, with a U_n-statistic of 9.1 and 6.8, respectively. The critical value for U_n is 9.1 for a significance level of 99%. This leads to an acceptance of the null-hypothesis that aspect and yields are independent. The aspect was therefore not included in the final multiple regression model for topographical influence on yield.

Table 1. Descriptive statistics of interpolation accuracy (n = 1817) on the full dataset (IDS)

	GPS data	Kriging interpolation
Mean (m)	99.999	100.008
Minimum (m)	97.031	97.071
Maximum (m)	102.921	102.877
Variance	1.033	1.027
Std. Dev.	1.016	1.013

RMSE for this dataset was 0.036.

Table 2. Interpolations on dataset reduced by every second point (IDS2) compared with ground truth data (***n* = 1810**)

	GPS data	Kriging interpolation
Mean (m)	99.993	100.001
Minimum (m)	97.031	97.519
Maximum (m)	102.916	102.709
Variance	1.026	0.901
Std. Dev.	1.013	0.949

RMSE for this dataset was 0.177.

Table 3. Interpolations on dataset reduced by 2 out of 3 points (IDS3) compared with ground truth data (***n* = 1810**)

	GPS data	Kriging interpolation
Mean (m)	99.993	100.001
Minimum (m)	97.031	97.025
Maximum (m)	102.916	102.879
Variance	1.026	1.022
Std. Dev.	1.013	1.018

RMSE for this dataset was 0.043.

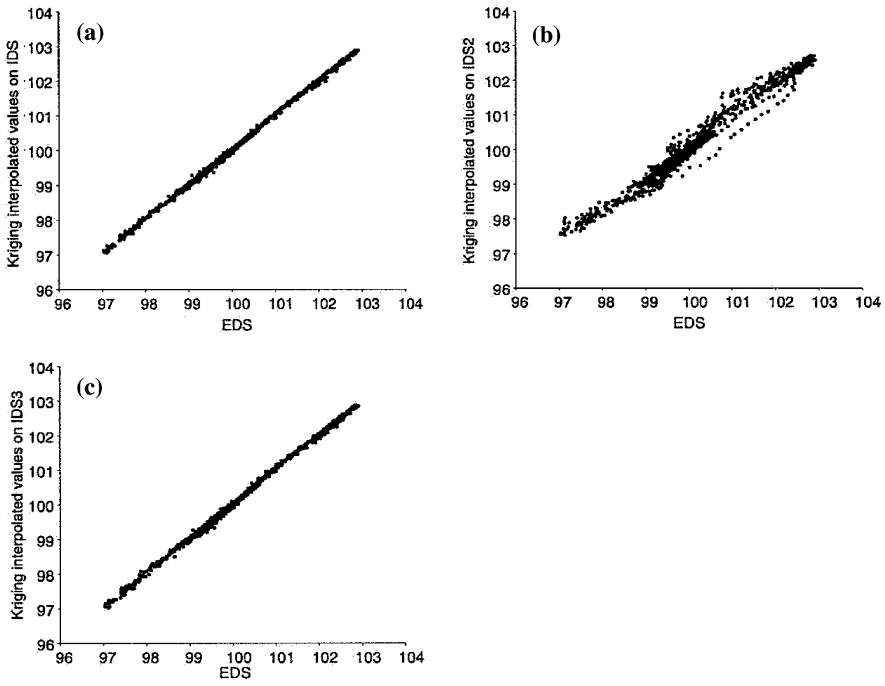

Figure 7. Scatter plots over the interpolated values from IDS (a), IDS2 (b) and IDS3 (c) against EDS. Number of evaluation points are 1817, 1810 and 1810, respectively.

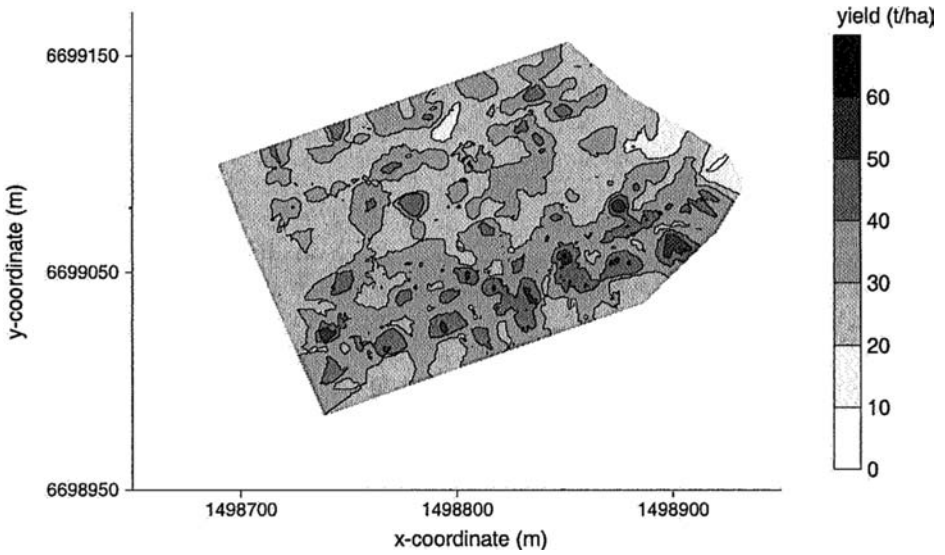

Figure 8. Yield map of the larger field. Coordinates in Swedish National Grid (RT90), meters (m).

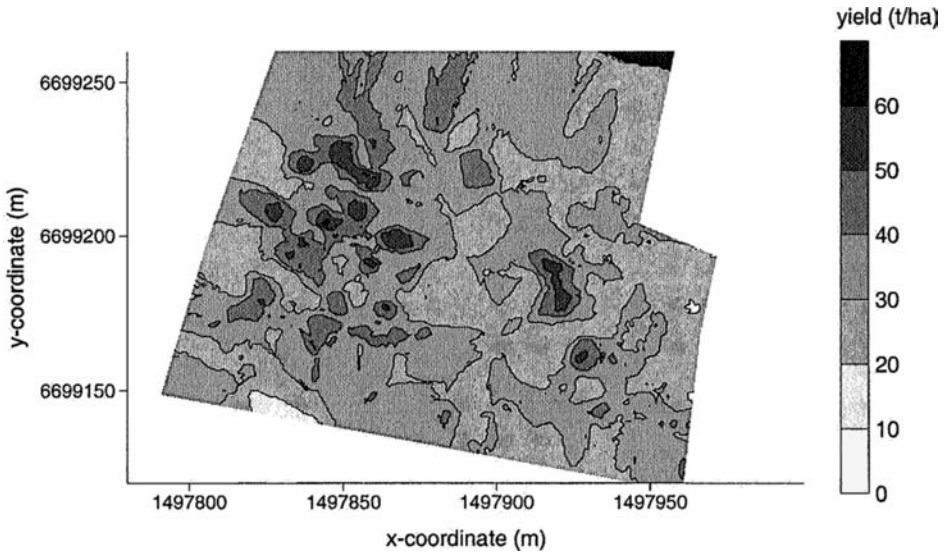

Figure 9. Yield map of the smaller field. Coordinates in Swedish National Grid (RT90), meters (m).

In the ensuing spatial and stepwise multiple regressions of yield, the predictors were elevation, gradient and the topographical indices. In the calculation of the spatial influence for the regression, the lowest standard error of estimate was found for ρ-values of 0 and 0.05 for the smaller and larger field, respectively. The data were then used in the stepwise multiple regressions for each field, in which the number of

Table 4. Descriptive statistics for the yield of the two fields

	Smaller field	Larger field
Mean (t/ha)	13.16	31.35
Minimum (t/ha)	4.86	9.56
Maximum (t/ha)	27.24	52.46
Std. Var.	4.55	6.00

Table 5. Statistics for the coefficients to the model for the smaller field

Predictor	Coeff	SE coeff	*T*	*P*
Elevation	0.6842	0.24	−2.91	0.00
Gradient	−27.57	8.48	−3.25	0.00
Horizontal index (HI)	−0.07	0.02	−4.45	0.00

$R^2 = 0.20$.

Table 6. Statistics for the coefficients to the model for the larger field

Predictor	Coeff	SE coeff	*T*	*P*
Elevation	−0.75	0.16	−4.73	0.00
Horizontal index (HI)	−0.03	0.01	−3.37	0.00
Drainage area index 2 (DAI2)	−6.92	3.63	−1.91	0.06

$R^2 = 0.18$.

predictors were reduced to elevation, gradient and the horizontal index (HI) for the smaller field. For the larger field, the predictors were elevation, the horizontal index (HI) and the drainage area index, which relates the area of a point of interest to the total drainage area for the slope (DAI2). In the multiple regression, the *T*-statistics was used to test if the coefficient was significant. The *P*-value shows the probability level for rejecting the null hypothesis that the variable does not contribute significantly to the model (given the inclusion of the other variables). The results for the smaller and larger field are presented in Tables 5 and 6, respectively.

Discussion

DEM generation

When sampling data for DEMs, a dense sampling scheme is preferred because much data may be collected to facilitate evaluation.

The patterns revealed in the first interpolations showed some systematic errors over the whole surface, caused by a dense sampling scheme. The striping, though not always obvious, greatly impacts on subsequent analyses (like drainage area) and is not revealed by descriptive statistics. With drainage pattern analysis, these errors may be detected more easily.

The descriptive statistics are used to gauge the similarity of the three different interpolations using the three interpolation datasets (i.e. IDS1, IDS2 and IDS3). The statistics in the tables (Tables 1–3) do not differ appreciably. The scatter plots (Figure 7) do not show much difference either. The correlation coefficients are close to 1 for all three datasets. However, a visual assessment of estimated drainage area clearly shows that the datasets make the flow accumulation behave quite differently, and that the flow pattern is improved as data are reduced. Taking into account that the striping only occurs on a very limited area of the DEM with a difference in elevation on the centimetre scale, the impact of the stripes is very low in the statistic.

In conclusion, we strongly recommend data to be reduced before interpolation. A rule of thumb would be to equalise inter-transect distance and the distance between points in each transect.

Topographical indices

Different topographical indices have been suggested in the literature over the years. Many indices use the length of slopes and the drainage area. In the present study, we decided against indices that are too complex since factors other than topography may influence the yield significantly. Hence, the first four indices used describe a point on the slope in relation to the slope length in horizontal, vertical and 'true' length (SLI1, SLI2, HI and VI). The fifth and sixth indices DAI1, and DAI2, are somewhat more advanced when it comes to their calculation. The drainage area above the point on the slope and in relation to the total drainage area of the slope is calculated. The form-based calculation of water movement in the algorithm (Pilesjö *et al.*, 1998) takes into account whether the water diverges or converges duo to a convex or concave slope. The indices were significantly correlated to yield.

Yield variation and its relation to topography

Topographical influence accounted for up to 20% of yield variation according to the multiple spatial regression model for one of the fields. This suggests that the composition of the landscape must not be excluded as a parameter that influences yield. Field management must take into account both the direct factors such as water distribution as well as the secondary factors such as nutrient distribution and redistribution due to topography.

The aspect of slopes in the fields investigated here was not associated with the yield according to the C-association test. Aspect might be a factor in areas with a greater gradient and perhaps in latitudes where the day length (in summer) is shorter. This is because an important yield factor is the solar radiation that the plant receives (de Temmerman *et al.*, 2002).

The fields do not have any accentuated micro-topography, which means that the elevation values and gradient values are the major parameters to take into account for the spatial regression and its limits. The distance of influence, the weight (w) in

the spatial regression, was set to 30 m, which was large enough to include the possible spatial autocorrelation. For both fields, the ρ-values were low indicating a low spatial correlation between the residuals. For the smaller field, the correlation strength (ρ) was zero, so the regression became a linear model. For the larger field, a small ρ was found and was taken into the calculations.

Multiple regression was chosen because a single topographical parameter cannot alone explain the whole variation in yield. A non-linear model might explain more of the variation using topographical parameters but, since other direct factors (e.g. soil) are not included, a linear model does not overestimate the influence in the way that a non-linear model might do (Rogerson, 2001).

In the case of the smaller field, the final model included elevation, gradient and the index that describes the slope position of the point horizontally (HI). For the larger field, the predictors included in the final model were elevation, horizontal slope position of the point (HI) and the index describing the relationship between drainage areas (DAI2). The significance levels for the parameters in this second model were greater, although the drainage area index falls just outside the 95% significance level. The drainage area and the vertical slope position both represent parameters directly affecting the water distribution in the field and, hence, the water available to the plant and nutrient dispersion.

Conclusions

The sampling scheme may contribute errors to the DEM, and thus caution is advised when collecting data for interpolation. Care should be taken with data point distribution, density and the interpolation algorithm itself.

The study clearly revealed differences between the spatial distributions of sampled data points. An even distribution of points is strongly recommended. Errors may not be obvious in the statistical evaluation, but may be revealed by the estimation of flow pattern.

Topographical factors had an influence on the yield, explaining about 20% of the yield. The impact of topography implies that the management, strategies for precision farming should include the landscape configuration.

References

Afyuni, M. M., Cassel, D. K. and Robarge, W. P. 1993. Effect of landscape position on soil water and corn silage yield. Soil Science Society of America Journal **57**(6), 1573–1580.

Blomgren, S. 1999. A digital elevation model for estimating flooding scenarios at the Falsterbo Peninsula. Environmental Modelling & Software **14**, 579–587.

Burrough, P. A. and McDonnell, R. A. 1998. *Principles of Geographical Information Systems*. (Oxford University Press, Oxford, UK) p. 333.

Changere, A. and Lal, R. 1990. Slope position and erosional effects on soil properties and corn production on a Mianian soil in central Ohio. Journal of Sustainable Agriculture **11**(1), 5–21.

Cressie, N. A. C. 1991. *Statistics for Spatial Data*. (John Wiley & Sons, New York, USA) p. 900.

Eklundh, L. and Mårtensson, U. 1995. Rapid generation of digital elevation models from topographic maps. International Journal of Geographical Information Systems **9**(3), 329–340.

Fiez, T. E., Miller, B. C. and Pan, W. L. 1994. Winter wheat yield and grain protein across varied landscape positions. Agronomy Journal **86**(6), 1026–1032.

Fisher, N. I. 1993. *Statistical Analysis of Circular Data*. (Cambridge University Press, Cambridge, UK) p. 277.

Fraisse, C. W., Sudduth, K. A. and Kitchen, N. R. 2001. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. Transactions of ASAE **44**(1), 155–166.

Gao, J. 1997. Resolution and accuracy of terrain representation by grid DEMs at a micro-scale. International Journal of Geographical Information Systems **11**(2), 199–212.

Hofmann-Wellenhof, B., Lichtennegger, H. and Collins, J. 1997. *GPS Theory and Practice*. (Springer, New York, USA) p. 389.

Lake, J. V., Bock, G. R. and Goode, J. A. (eds) 1997. General discussion I. In: *Precision Agriculture – Spatial and Temporal Variability of the Environmental Quality*. (Wiley, Chichester, UK), pp. 68–78.

Manning, G., Fuller, L. G., Eilers, R. G. and Florinsky, I. 2001. Soil moisture and nutrient variation within an undulating Manitoba landscape. Canadian Journal of Soil Science **81**(4), 449–458.

Mardia, K. V. 1972. *Statistics of Directional Data*. (Academic Press, London, UK) p. 357.

Moore, I D., Grayson, R. B. and Ladson, A. R. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrological Processes **5**(1), 3–30.

Moore, I. D., Gessler, P. E., Nielsen, G. A. and Peterson, G. A. 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal **57**(2), 443–452.

Persson, D. A., 1998. Potato yield mapping with an optical sensor. In: *Proceedings of the First International Conference on Geospatial Information in Agriculture and Forestry*, edited by B. J. Petoskey (ERIM Int., Ann Arbor, USA), pp. 618–623.

Persson, D. A., Eklundh, E. and Algerbo, P. -A. 2004. Evaluation of an optical sensor for tuber yield mapping. Transactions of ASAE **47**(5), 1851–1856.

Pilesjö, P. 1992. *GIS and Remote Sensing for Soil Erosion Studies in Semi-arid Environments*. (Lund University Press, Lund, Sweden) p. 203.

Pilesjö, P., Zhou, Q. and Harrie, L. 1998. Estimating flow distribution over digital elevation models using a form-based algorithm. Geographic Information Sciences **4**(1–2), 44–51.

Rogerson, P. 2001. *Statistical Methods for Geography*. (Sage, London, UK) p. 236.

Schmidt, F. and Persson, A. 2003. Comparison of DEM data capture and topographic wetness indices. Precision Agriculture **4**(2), 179–192.

Schneider, S. M., Boydston, R. A., Han, S., Evans, R. G. and Campbell, R. H., 1997. Mapping of potato yield and quality. In: *Precision Agriculture '97:Proceedings of the First European Conference on Precision Agriculture*, edited by J. V. Stafford (BIOS Scientific Publishers, Oxford, UK), pp. 253–261.

Skidmore, A. K. 1989. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. International Journal of Geographical Information Systems **3**(4), 323–334.

Sudduth, K. A., 1997. Spatial modeling of crop yield using soil and topographic data. In: *Precision Agriculture '97: Proceedings of the First European Conference on Precision Agriculture*, edited by J. V. Stafford (BIOS Scientific Publishers, Oxford, UK), pp. 439–447.

De Temmerman, L., Wolf, J., Colls, J., Bindi, M., Fangmeier, A., Finnan, J., Ojanpera, K. and Pleijel, H. 2002. Effect of climatic conditions on tuber yield (*Solanum tuberosum* L.) in the European 'CHIP' experiments. European Journal of Agronomy **17**(4), 243–255.