Spatial Discontinuity Analysis, a novel geostatistical algorithm for on-farm experimentation

Rudolph S.[1], Marchant B.P.[1], Gillingham V.[2], Kindred D.[3], Sylvester-Bradley R.[3]

[1] British Geological Survey, Environmental Science Centre, Keyworth, United Kingdom

[2] AgSpace Agriculture, Swindon, United Kingdom

[3] ADAS UK Ltd, Boxworth, United Kingdom

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract

Traditional agronomic experimentation is restricted to small plots. Under appropriate experimental designs the effects of uncontrolled environmental variables are minimized and the measured responses (e.g. in yields) are compared to controllable inputs (seed, tillage, fertilizer, pesticides) using well-trusted design-based statistical methods.

However, the implementation of such experiments can be complex and the application, management, and harvesting of treated areas might have to be done manually or with specialist equipment. Furthermore, these experiments only compare treatment performance over a relatively small area and the same relationships might not apply over larger management zones, or fields. In addition, the small area of the experiments might limit their precision.

These problems have motivated a number of researchers and farmers to consider field-scale experiments which tend to be based on systematic rather than randomized designs. Systematic experimental designs enable different treatments to be applied relatively simply using farm equipment. However, design-based statistical methods are not applicable for the systematic designs and correlation amongst the observed responses can lead to exaggerated estimates of the significance of any observed treatment differences. Therefore, geostatistical or model-based methods must be used to quantify and account for this correlation.

In this study the statistical analysis of such systematic experiments is considered with reference to a trial where the fertiliser nitrogen rate was varied on an arable field. The response variable was a vegetation index derived from an aerial photograph. The magnitude and significance of the average treatment effects across the experiments can be calculated if the data are represented by a linear mixed model which includes spatial correlation. If the spatial correlation is neglected then the confidence intervals for the treatment effects were erroneously small. Furthermore, a novel analysis method referred to as spatial discontinuity analysis (SDA) is proposed. SDA was used to focus on the boundaries between different treatments and to test whether there was a significant jump in the response variable where the treatment changed. In this context, the spatial correlation was advantageous since, in the absence of treatment effects, the expected differences between adjacent observations was smaller than if they had been uncorrelated. Therefore, the local treatment effects could be more easily distinguished from the underlying variation of the response variable.

Keywords:

Spatial discontinuity analysis SDA, field experimentation, geostatistics, soil heterogeneity

Introduction

Agricultural experimentation regarding the response of crops to a variety of management interventions is crucial to the improvement of farm management decisions and to progress in crop science. Traditionally, agricultural experiments have been carried out in small-scale and controlled plots. Trusted experimental designs have been employed. These designs ensure that the precision of the experimental results are maximized and the variation due to environmental factors is minimized through choice and management of the site, randomization, blocking and replication (Little and Hills, 1978). The experiments are analyzed using design-based statistical techniques such as the analysis of variance (ANOVA) that utilize the randomization to ensure that the estimates of the effects of different treatments are unbiased and largely free from assumptions. The ANOVA can be used to test the statistical significance or the probability that any observed differences in the response variable (e.g. the crop yield) could have occurred by chance rather than as a result of the treatment difference.

However, farmers sometimes question whether management recommendations inferred from plot-based experimentation are necessarily applicable to their own farms. The plot-based experiments might have been conducted in a different region, on different soil types, in different environmental conditions or using different varieties. Therefore farmers conduct on-farm experiments to verify industry claims and recommendations and to test and fine-tune new concepts, products, and systems under local environmental and management conditions. Such farmers are assisted by the increasingly availability of precision agriculture technology and the associated ease with which inputs can be varied and site-specific data can be gathered.

A wide variety of on-farm experimental designs have been implemented. These range from simple split-field designs, in which a field is divided into two halves, to complex embedded designs such as the checkerboard (Kindred et al., 2015) or sine-wave design (Bramley et al., 1999). Split-trials are relatively easy to establish and do not require high resolution electronic data recording. However, the absence of replication means that it is not possible to test formally whether the observed differences in responses are wholly the result of the different treatments or whether they are influenced by other factors (e.g. fertility, soil type, drainage or slope) that vary between two plots. In contrast, the embedded designs are highly replicated and therefore the treatment effects can be disentangled from the other sources of variation but the implementation and analysis of these experiments is time consuming and requires technical expertise, commonly more than is available on farms.

Farmers often adopt a compromise approach such as strip-trial designs where the treatments are applied in parallel strips of two or three harvester widths over the entire length of the field. The strips can be orientated perpendicular to major sources of variation (e.g. different management classes or soil zones) to explore the interaction between these factors and the treatment effects. In common with traditional plot trials the experimentalist will generally attempt to determine whether the mean yields or responses in strips under one treatment differ significantly from the mean yields in strips under a different treatment. Alternatively, Lawes and Bramley (2012) specifically focused on the difference in the response variable in adjacent strips under different treatments. Such comparisons were motivated by the ideas that sharp changes to responses (e.g. crop colour) are often observed at the boundary between treatments and that by focusing on these differences within a small area one might expect that the environmental factors which limit yield are relatively uniform and their influence relatively small.

Standard ANOVA methods require the assumption that the observations of the response variable are uncorrelated or independent. However, one would expect that the observations of the response variable across a field are likely to be spatially correlated. For example, yields recorded within a small sub-region of the field are more likely to be similar than those recorded a long distance apart because the environmental factors limiting the yield are likely

to be relatively uniform within the sub-region. In traditional plot trials, the independence assumption is appropriate because the different treatments are allocated at random. On-farm experiments are more likely to follow a systematic design where particular treatments are constrained to be situated a set distance from each other. In these trials, the independence assumption is inappropriate and the spatial autocorrelation amongst the observations must be accounted for.

Geostatistical models can be used to account for autocorrelation. The autocorrelation between observations is commonly expressed in the form of a variogram. The variogram describes how the expected squared difference between a pair of observations varies according to the lag distance separating them. The variogram for a particular set of observations can be estimated by the method of moments. First a series of point estimates of the variogram is calculated for different lag distances. Then an authorized mathematical function (Webster and Oliver, 2007) is fitted to these point estimates. Once the variogram can be expressed as a mathematical function it can be used within the kriging predictor to simulate or predict the variable at sites where it has not been observed and to calculate the uncertainty of these predictions. However, these predictions and uncertainties are only accurate if the assumptions of the geostatistical model are appropriate.

The application of geostatistical techniques to agricultural trials is far from new. McBratney (1984) was the first to map experimental treatment responses across a field by kriging and Blackmer and White (1996), Bruulsema et al. (1996), Cathcart et al. (1999), and Sadler et al. (2002) have followed suit. However, these authors did not assess whether the treatment effects were statistically significant. This shortcoming was addressed by Bishop and Lark (2006, 2007) who proposed that the yields under different treatments could be considered as coregionalized variables. Such variables might be autocorrelated and correlated to each other in a manner that varies according to the lag distance separating observations. This multivariate variation can be represented by a linear model of coregionalization (LMCR) which consists of a variogram for each variable and a cross-variogram for each pair of variables. The parameters of these variograms and cross-variograms must be constrained to ensure that negative prediction variances cannot arise. Having estimated their LMCR, Bishop and Lark (2006) used the z-statistic to test whether the yield difference was significant.

In the majority of on-farm trials, the response variable is the crop yield as measured by a yield monitor within the harvester. Lark et al. (1997) noted that the spatial resolution of such data can be limited by the mixing of grain as it travels from the header to the yield monitor and that artefacts can arise because of this grain flow delay. Additionally each observation covers an area as wide as the harvester and misleading measurements can occur if the cutter-bar is not ‘full’, the harvested rows are not parallel or if the row contains wheel tracks. These problems can be avoided if the crop response to different treatments is assessed using airborne or satellite sensors which can provide geo-referenced representation of the crop canopy. Vegetation indices have been derived from multi- and hyper-spectral images of the reflectance of the crop canopy taken from satellites or planes but the value of these images can be limited by poor resolution, insufficient availability of images and high acquisition costs. In recent years unmanned aerial vehicles (UAV) equipped with compact RGB (red, green, blue) digital cameras have become widely available and they have been suggested as a convenient and cost-effective alternative for crop monitoring. For example, the green-red vegetation index (RGVI) derived from the red and green spectral band of a standard RGB image has been used by Motohka et al. (2010) as a phenological indicator. Also, Hunt et al. (2005), Geipel et al. (2014) and Bendig et al. (2015) have demonstrated that the RGVI can be used to determine and predict biomass, nutrient status, and yield.

In this paper the geostatistical analysis of strip trials is considered with reference to a nitrogen response trial on an arable crop in southern England. The response variable is the RGVI derived from a RGB image taken from an UAV. The variation of the RGVI is represented by a linear mixed model (LMM) which separates the component of variation that

can be explained by the treatments or environmental factors from the spatially correlated residual variation. This model is used to test for significant differences in the average RGVI under each nitrogen treatment across the trial. Furthermore, a novel form of analysis is introduced which is akin to the comparisons on either side of the treatment employed by Lawes and Bramley (2012). This analysis method is referred to as Spatial Discontinuity Analysis (SDA). Whereas Lawes and Bramley (2012) used design-based t-tests to explore the significance of the observed yield differences across the line, SDA accounts for the autocorrelation amongst the data via the LMM. The effect of the autocorrelation on the precision of SDA is rather complex. If one considers a single comparison between a pair of observations situated each side of the treatment line then the autocorrelation amongst the RGVI measurements should imply that the treatment effects can be more easily distinguished from the underlying variation in RGVI. However, when the comparisons between multiple pairs of measurements are combined then the observed differences are themselves correlated and therefore the precision of the combined comparisons does not improve with the number of pairs as quickly as it would for independent observations.

Theory

Linear mixed models and hypothesis testing

The LMM divides the variation of the response variable into fixed effects and random effects. It is written:

$$\mathbf{y} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1)$$

were $\mathbf{y}$ is a vector containing n observations of the response variable, $\mathbf{M}$ is a design matrix of size $n \times q$ and the vector $\boldsymbol{\beta}$, which is of length q, contains the coefficients of the fixed effects. The residual term $\boldsymbol{\epsilon}$ is a length n vector containing random effects, which are often assumed to be a realization of a Gaussian random function with zero mean and covariance matrix $\mathbf{C}$. If the observed residuals are inconsistent with the Gaussian assumption then a transformation might be applied to the data. If the random effects are independent then $\mathbf{C}$ will be a diagonal matrix with the variance of each random effect on the main diagonal and zeros elsewhere. If the random effects are auto-correlated then $\mathbf{C}$ will include non-zero elements away from the main diagonal.

The fixed effects are the product $\mathbf{M}\boldsymbol{\beta}$. In its simplest case, when the mean of the response variable is expected to be constant across the field, $\mathbf{M}$ is a $n \times 1$ matrix and all elements of $\mathbf{M}$ are equal to one. One might hypothesise that the mean of the response will be different in a portion of the field which undergoes an experimental treatment. This assumption can be tested by adding a second column to $\mathbf{M}$ which contains one when the corresponding observation is made in the treated area and zero if it is made in the control region. Further indicator variables can be added to $\mathbf{M}$ to correspond to other treatments.

The response variable might also vary according to other environmental covariates such as soil type or elevation. It can be beneficial to include such covariates in the model to reduce the variability of the random effects and lead to more precise estimates of the treatment effects (Rudolph et al., 2016). Further columns of binary indicator covariates can be added to $\mathbf{M}$ to represent categorical covariates such as the soil type or management zone and cross terms between these factors or between these factors and the treatment effects. Linear trends in continuous covariates such as the spatial coordinates, an environmental variable such as slope or a variable derived from a proximal or remote sensor can be accommodated by adding them as additional columns to the $\mathbf{M}$ matrix. Polynomial trends can also be considered by including the square, cube and higher order products of these covariates.

As more columns are added to the $\mathbf{M}$ matrix, there is a danger that the LMM becomes

overfitted. This means that the model becomes too intricately fitted to the observed dataset but does not produce similarly reliable predictions at sites which were not used to calibrate the model. This problem can be avoided if one ensures that each additional column of $\mathbf{M}$ leads to a significant improvement to the fit of the model. Different criteria can be used to assess the improvement in fit and the Akaike Information Criterion (McBratney and Webster, 1986) is commonly used in the spatial analyses of soil properties. The LMM with the lowest AIC best manages the tradeoff between model complexity (number of parameters) and the quality of the model fit.

The entries of the random effects covariance matrix $\mathbf{C}$ can be calculated from the variogram of the residuals of the fixed effects model (i.e. $\mathbf{r} = \mathbf{y} - \mathbf{M}\boldsymbol{\beta}$). Webster and Oliver (2007) fully describe the method of moments which in this study is used to estimate the variogram from the available observations. There are a number of different authorized variogram models that can be fitted. Because of a high flexibility at small lag distances the Matérn model (Marchant and Lark, 2007) is preferred:

$$\gamma(h) = c_0 + c_1\left(1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{h}{a}\right)^{\nu} K_{\nu}\left(\frac{h}{a}\right)\right) \text{ for } h > 0 \text{ and } \gamma(0) = 0 \quad (2)$$

where h is the lag distance separating two observations, c_0 is the nugget variance and $c_0 + c_1$ the sill variance, Γ is the gamma function, K_{ν} denotes the modified Bessel function of the second kind while $\nu > 0$ and $a > 0$ are smoothness and scale parameters, respectively. Thus the variogram has four parameters c_0, c_1, a and ν which must be estimated.

Once the variogram parameters have been estimated, element i, j of the covariance matrix can be calculated using:

$$C_{ij} = c_0 + c_1 - \gamma(h_{ij}) \quad (3)$$

where h_{ij} is the lag distance and C_{ij} the covariance between points i and j. Note that, if there is no spatial correlation present in the random effects then $c_1 = 0$ and $C_{ij} = 0$ if $i \neq j$ and $C_{ij} = c_0$ if $i = j$. Thus $\mathbf{C} = c_0\mathbf{I}$ where $\mathbf{I}$ is the $n \times n$ identity matrix.

The unknown coefficients of the fixed effects $\boldsymbol{\beta}$ can be estimated by maximum likelihood:

$$\boldsymbol{\beta} = (\mathbf{M}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{M})^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}, \quad (4)$$

and the covariance matrix of these estimated coefficients is equal to:

$$\mathbf{W} = (\mathbf{M}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{M})^{-1}. \quad (5)$$

It was previously suggested that one might want to test whether the mean effect of a particular treatment is significant. This can be achieved by a z-test. If the i^{th} column of $\mathbf{M}$ indicates the presence or absence of the treatment at each site, then the z statistic is equal to:

$$z_i = \frac{\beta_i}{\sqrt{W_{ii}}}, \quad (6)$$

where β_i is the i^{th} element of $\boldsymbol{\beta}$ and W_{ii} is the i, i^{th} element of $\mathbf{W}$. The null hypothesis that the treatment has no effect on the response variable (i.e. that $\beta_i = 0$) can be rejected at the $p = 0.05$ level if $|z_i| > 1.96$. This test accounts for the spatial autocorrelation amongst the data through the covariance matrix $\mathbf{C}$. When there is no spatial correlation and $\mathbf{C} = c_0\mathbf{I}$ it reduces to a standard design-based test.

Spatial discontinuity analysis

The above test considers the average effect of the treatment on the response variable

across the experiment. This might be sufficient information for a farmer to decide whether or not to employ the treatment uniformly across the field. However, the farmer might wish to consider whether it is beneficial to vary the treatment within the field and therefore a test that can assess more localized responses is required. This idea motivated Lawes and Bramley (2012) to explore the treatment differences almost continuously along a line which divided two areas under different treatments. Their design-based test of such differences can be extended to a model-based analysis that accounts for spatial correlation and is referred to here as spatial discontinuity analysis or SDA.

SDA is based on a LMM of the response variable under the control treatment. First, a line in the field is identified where the control treatment (T_0) is applied on one side and the treatment of interest (T_1) is applied on the other side. The locations of the observations adjacent to this line where T_i is applied are referred as $\mathbf{x}_i$ and the corresponding observations of the response variable are $\mathbf{y}_i$. Then, a LMM of $\mathbf{y}_0$, the response variable under the control treatment, is estimated. This LMM and the observations of $\mathbf{y}_0$ are used to simulate 1000 realizations of $\mathbf{y}_0$ at the locations $\mathbf{x}_1$ by the LU method (Webster and Oliver, 2007). These simulations are referred as $\widetilde{\mathbf{y}_0}$. Then the difference between these simulations and the observed values of $\mathbf{y}_1$ at these sites are calculated:

$$\mathbf{d}(\mathbf{x}_1) = \widetilde{\mathbf{y}_0}(\mathbf{x}_1) - \mathbf{y}_1(\mathbf{x}_1). \quad (7)$$

Thus $\mathbf{d}(\mathbf{x}_1)$ consists of 1000 simulations of the treatment effect at each of the locations in $\mathbf{x}_1$**.** Histograms of these differences can be used to assess the significance of the treatment effect. For example, if at a single site more than 950 of the values of $\mathbf{d}$ are positive this indicates that a treatment has a positive effect on the response variable and the null hypothesis of no effect can be rejected at the p=0.05 level.

A single observation location might not be sufficiently large for a significant treatment effect to be observed. Therefore one might want to average the simulated values of $\mathbf{d}$ at multiple locations and see if these average values are significantly different to zero. By averaging across multiple locations one might expect that smaller differences in the response variable can be detected but the spatial resolution of the comparison decreases. The detectable difference at a single site and the rate of decrease of this detectable difference as it is averaged across multiple sites will depend on the amount of spatial correlation identified in the LMM.

Materials and methods

Case study

In this study a strip trial of the response of an arable crop to different nitrogen inputs is considered. The trial was located at the Geldings Ley farm (51°58'27"N 0°8'25"E), 25 km south of Cambridge, England. The field was cropped with wheat of the variety JB Diego. Nitrogen was applied at a standard rate across the field except for one 25 m wide tramline where 60 kg N ha^{-1} less was applied and one tramline where 60 kg N ha^{-1} more was applied. These 'low' and 'high' treatments were situated next to each other (see Figure 1). The tramlines were 60 m long and orientated along a gentle north-east slope.

The response variable is derived from an RGB aerial image of canopy reflectance taken from a UAV above the trial on 7th July 2015 (during grain-filling) with a Panasonic Lumix GH4 camera mounted on a DJI S900 Drone which was operating 100 m above the surface. The image covered the entirety of the two treated tramlines and one and a half tramlines of the standard or control treatment. Ground resolution of the image was 0.03 m which was geo-referenced on DGPS located wheel-ways in ArcGIS. Treatments were assigned to the image using a buffer around the center of the treatment bout of half the sprayer width.

Pre-processing

To quantify the treatment response the green-red vegetation index was calculated as:

$$\mathrm{RGVI} = \frac{\mathrm{DN}_{\mathrm{green}} - \mathrm{DN}_{\mathrm{red}}}{\mathrm{DN}_{\mathrm{green}} + \mathrm{DN}_{\mathrm{red}}} \quad (8)$$

where $\mathrm{DN}_{\mathrm{green}}$ and $\mathrm{DN}_{\mathrm{red}}$ are the digital numbers (0-255) of the green and red band of the image. The RGVI varies between -1 for green and +1 for red image pixels. High frequency noise was removed from the image by applying a Daubechies Type I two-dimensional wavelet decomposition filter which resulted in a reduction of the image resolution to 1 m. A 0.5 m wide buffer was applied along each wheel-mark and information within these areas was excluded from the statistical analyses. To simplify the analysis, the image was rotated so that the treatment-bouts ran vertically. A spatial trend in the RGVI values is apparent in the images. This trend appears to be an effect of cloud shading on the image rather than a property of the canopy.

Geostatistical analyses

A LMM was fitted to the preprocessed RGVI values. In an initial model, the standard N rate was treated as the control treatment and the 'high' and 'low' treatments were included as additional rows in the fixed effects matrix. A Matérn variogram model was used for the random effects. Then, additional spatial trend terms were added to the fixed effects matrix to explore whether they led to a reduction in the AIC. These terms were included to represent the cloud shading effect which was observed in the image. The LMM with the lowest AIC was used for the subsequent analyses.

The magnitude of the treatment effects were extracted from the estimated LMM and the significances of these effects at the field-scale were explored using the z-statistic (Equation 6). Then SDA was conducted along the lines separating the control from each of the non-standard treatments and the line separating the two non-standard treatments. Next, SDA was used to make comparisons at the scale of a single pixel and then for combined pixels to explore the rate at which the precision of the comparisons improved.

To explore the importance of the autocorrelation in the LMM, the above analyses were repeated but the parameters of the LMM were constrained such that $c_1 = 0$. Thus the RGVI values were assumed to be independent and the analyses became equivalent to a design-based approach.

Results

The high resolution RGB canopy reflectance measurements of the nitrogen fertilizer experiment are depicted in Figure 1a. An underlying trend is evident in the image that is not related to the different N treatments. The image appears to become brighter and more blurred towards the top left hand corner of the image. It is assumed that this trend is either caused by optical effects or the reflection of wind-induced motion of the crop canopy. Chlorophyll deficiency expressed as leaf yellowing is evident throughout the low level nitrogen treatment. Small and irregular symptoms of nitrogen deficiency are also evident within the control treatment. In contrast, the dark green canopy of the N-rich treatment indicates good crop performance. A sharp contrast in green intensity can be observed along the interface between the low N level and standard treatment. The clear distinction between low and high N treatment diminishes towards the northern end of the line where the image becomes brighter and more blurred.

The filtered and down-sampled RGVI image is shown in Figure 1b. The pattern of variation is similar to that described for the RGB image. The standard deviations of the high resolution RGVI image under the low, standard and high treatments were 0.035, 0.038 and 0.031 respectively (Table 1). Filtering reduced these standard deviations to 0.016, 0.024 and

0.015. The raw RGVI also had a strong negative skew which suggests that values might not be consistent with a Gaussian random effects model. However, the magnitude of this skew reduced with filtering. The LMMs described below were re-fitted after various transformations of the data but these did not lead to an improvement in the AIC.

Linear mixed models with constant, linear and quadratic spatial trends were estimated for the filtered RGVI image. The quadratic trend model had the lowest AIC and was used for subsequent analyses. The variogram of this LMM (Figure 2) possessed pronounced short scale variability and autocorrelation for distances greater than 10 m. The estimated trend surface and the residuals from this trend surface are shown in Figure 1c-d. Note that these residuals are not the same as the LMM random effects since the treatment effects have not been removed. The removal of the quadratic trend from the image further reduced the standard deviations for each treatment to 0.014, 0.011 and 0.010. Prior to the removal of the quadratic trend, the standard treatment unexpectedly had the largest average RGVI values. However when the quadratic trend was removed the RGVI increased upon moving from low to standard to high treatments.

Table 1: Descriptive statistics if the RGVI estimation derived from a) the high resolution, b) the downscaled, and c) the downscaled RGB image after removing of quadratic trend components.

a)

Treatment	Descriptive of the RGVI estimation (high resolution)							
	Mean	Median	Sd	Min	Max	Range	Skew	CV
Low	-0.048	-0.047	0.035	-1.000	0.586	1.586	-0.104	0.727
Standard	-0.084	-0.079	0.038	-1.000	0.269	1.269	-0.961	0.450
High	-0.070	-0.065	0.031	-1.000	0.256	1.256	-1.419	0.434

b)

Treatment	Descriptive of the RGVI estimation (downscaled)							
	Mean	Median	Sd	Min	Max	Range	Skew	CV
Low	0.049	0.049	0.016	0.005	0.105	0.099	0.304	0.319
Standard	0.083	0.089	0.024	0.027	0.130	0.103	-0.252	0.283
High	0.071	0.073	0.015	0.036	0.108	0.073	-0.112	0.217

c)

Treatment	Descriptive of the RGVI estimation (downscaled + trend removed)							
	Mean	Median	Sd	Min	Max	Range	Skew	CV
Low	0.048	0.048	0.014	0.003	0.093	0.091	0.084	0.286
Standard	0.074	0.075	0.011	0.009	0.106	0.096	-0.608	0.152
High	0.082	0.082	0.010	0.050	0.111	0.061	0.007	0.124

Table 2: Comparison of treatment effects at the field scale using the classical and model based approach. β is the difference between the mean response and two tested treatments. SE is the standard error of the estimated β.

Treatment comparison	Design-based		Model-based	
	β	SE	β	SE
Standard vs Low	-0.0307	0.0007	-0.0142	0.0013
Standard vs High	0.0060	0.0007	0.0052	0.0013

According to the z-tests (Equation 6) at the field-scale the low N treatment had a significant negative effect on RGVI and the high N treatment had a significant positive effect on RGVI. These results are summarized in Table 2. Note that when a design-based analysis is applied (i.e. when the random effects are assumed to be independent) that the standard errors on the treatment effects are underestimated. This illustrates how inappropriate statistical tests could potentially lead to statistical significance treatment differences being erroneously assigned to small treatment differences.

The results of the localized SDA are shown in Figure 3. When comparisons across the line are made for 1-m pixels then no significant differences are observed. However, when the pixels are merged into 10 m long blocks, significant treatment effects are evident. These results illustrate the trade-off between spatial resolution and precision for SDA comparisons. When a single pair of pixels is considered, a reasonably large difference between them might occur by chance. However, it is less unlikely that multiple pairs of pixels spanning a larger block would all have similar differences unless there was a treatment effect. In Figure 4 the improvements in precision that occur with increasing block size are illustrated. Note that if the random effects had been assumed to be independent that erroneous standard errors would have been estimated. For comparisons between a single pair of pixels the estimated standard errors would have been too large because the correlation between the adjacent RGVI values would have been ignored. However, as the block size is increased the standard errors based on the independent random effects model become too small. This reflects that the different comparisons are themselves correlated and therefore the improvement in precision from averaging over larger blocks is not as large as would be expected for randomly positioned comparisons.

Discussion

Farmers require on-farm experiments to confirm that management recommendations are suited to the local conditions found in their fields. However, it is impractical, expensive and time-consuming for farmers to establish the randomized plot experiments favored by agricultural researchers. Farmers are more likely to implement simpler experiments such as split-field comparisons or strip-trials. It is important to formally analyze the results of such trials to assess whether any potential treatment effects are statistically significant. Statistical tests are required to determine the probability that the observed results might have occurred by chance rather than because of a real effect of the treatment. Since the simple trials employed by farmers tend to be systematic rather than randomized, standard design-based tests such as ANOVA cannot be applied. Instead it is necessary to estimate a model of the underlying variation of the response variable and to explore whether the responses recorded in treated areas are significantly different. The inferences drawn from such model-based or geostatistical analyses are only valid to the extent that the model is a valid representation of the underlying variation.

In this paper two approaches for testing the significance of treatment effects in non-randomized agricultural experiments have been demonstrated. The first approach is analogous to the ANOVA in that it tests whether the mean values of the response variable are different in areas that have undergone different experimental treatments. A statistical model is used to account for the spatial correlation amongst the observations of the response variable. A farmer might use the results of such a test to decide on the optimal uniform treatment for the field. The second test which is referred as Spatial Discontinuity Analysis (SDA) looks specifically at the response variable on either side of the boundary separating different treatments. Sharp changes in the color, height or density of a crop are often evident where a management change occurs. SDA can be applied to small blocks along the boundary line to formally test whether any measured jump in the response variable is statistically significant. It is possible to use SDA to consider whether the effectiveness of the treatment varies according to the soil zone or management zone and therefore a farmer might use the results to decide upon variable treatments within the field.

The two analysis methods were applied to a nitrogen strip-trial. Fertilizer nitrogen was applied at the recommended rate in the majority of the field. The application rate of nitrogen was increased by 60 kg/ha in one strip and reduced by 60 kg/ha in another. The response variable was the RGVI derived from a RGB photograph taken from a UAV. At the field-scale both of the treatments had a significant effect on the RGVI. Also, the SDA identifies blocks of significant treatment effects along the boundaries separating the different treatments. There is a trade-off between the precision of the SDA and the spatial resolution of the comparison (i.e. the length of the block).

For both of the analyses it could be confirmed that if the data were erroneously assumed to be independent then the estimated standard errors of the treatment effects were poorly estimated. Thus, if the underlying variation is improperly modelled then false inferences about the treatment effects might be drawn. In the field-scale test the standard errors estimated from the independent model are erroneously small because the model has failed to acknowledge that the observations are correlated and therefore to some extent they replicate the same information. For the SDA on a small block the independent model over-estimates the standard errors of the treatment effects. In this circumstance the spatial correlation amongst the response variable is a benefit that means a treatment effect can be more easily identified. However, as the length of the block increases the true standard errors do not decrease at the rate that would be expected according to the independent model. This is because the series of comparisons between points on either side of the boundary are themselves correlated.

The same analysis methods can be applied to other response variables such as spectral indices derived from proximal or remote sensors or yield measurements derived from a harvester with yield monitor. The use of yield monitors leads to further challenges because factors such as the delay between the grain being cut and recorded by the sensor can lead to artefacts in the data. Such artefacts must be removed prior to statistical analyses.

The precision of these systematic experiments is largely controlled by the design of the experiment, the magnitude of the underlying variation in the field and the precision of the measurement device. The precision of the nitrogen response experiment described here might be improved by adding further strips of each non-standard treatment. It is plausible that a single strip might, by chance, coincide with a particularly fertile region of the field. In such a case a perceived treatment effect could actually arise as a result of the underlying fertility. It is less likely that multiple treated strips would all coincide with fertile regions and therefore additional strips improve the precision. The precision would improve if the length of a single strip was increased but the benefit would be limited because of the autocorrelation within the strip. The precision could also be improved by varying the treatment within the strip. These potential improvements to the precision of the experiment all increase the costs, effort and skill required to establish the experiment.

If the causes of underlying variation in the response variable are known (and if these factors are not of concern) one might use a model to explain some of the variation. For example, a smooth trend was evident in the RGVI image. The authors of this study were confident that this trend did not reflect a property of the canopy but was instead caused by blurring due to crop movement or shading by clouds. It was found that the trend could be represented by a quadratic surface and the precision of the results improved after its removal.

If a measurement device is prone to adding erroneous artefacts to the response then these might be removed by some form of pre-processing of the data. However, to a large extent an imprecise measurement device will always lessen the precision of an experiment. The spatial resolution of the experiment is also limited by the precision with which the different treatments can be applied. For instance, with sprayed applications there might be some overlap between treatments. In these circumstances it might be necessary to remove the data where the treatments overlap and therefore the precision of the experiment deteriorates.

The focus of this study has been on the use of on-farm experiments to determine the best management practices on a particular farm. However, if the results of many such experiments were combined then they could inform about the effectiveness of management practices at a much larger scale and the factors that inhibit them. The pooling and effective analysis of such results requires knowledge exchange between farmers and agricultural researchers. In the UK, the Agronōmics initiative (Kindred et al. (2016) is seeking to facilitate this knowledge exchange through the development of efficient designs for on-farm experiments, appropriate analysis methods such as those described here and the formation of knowledge exchange networks.

Figures

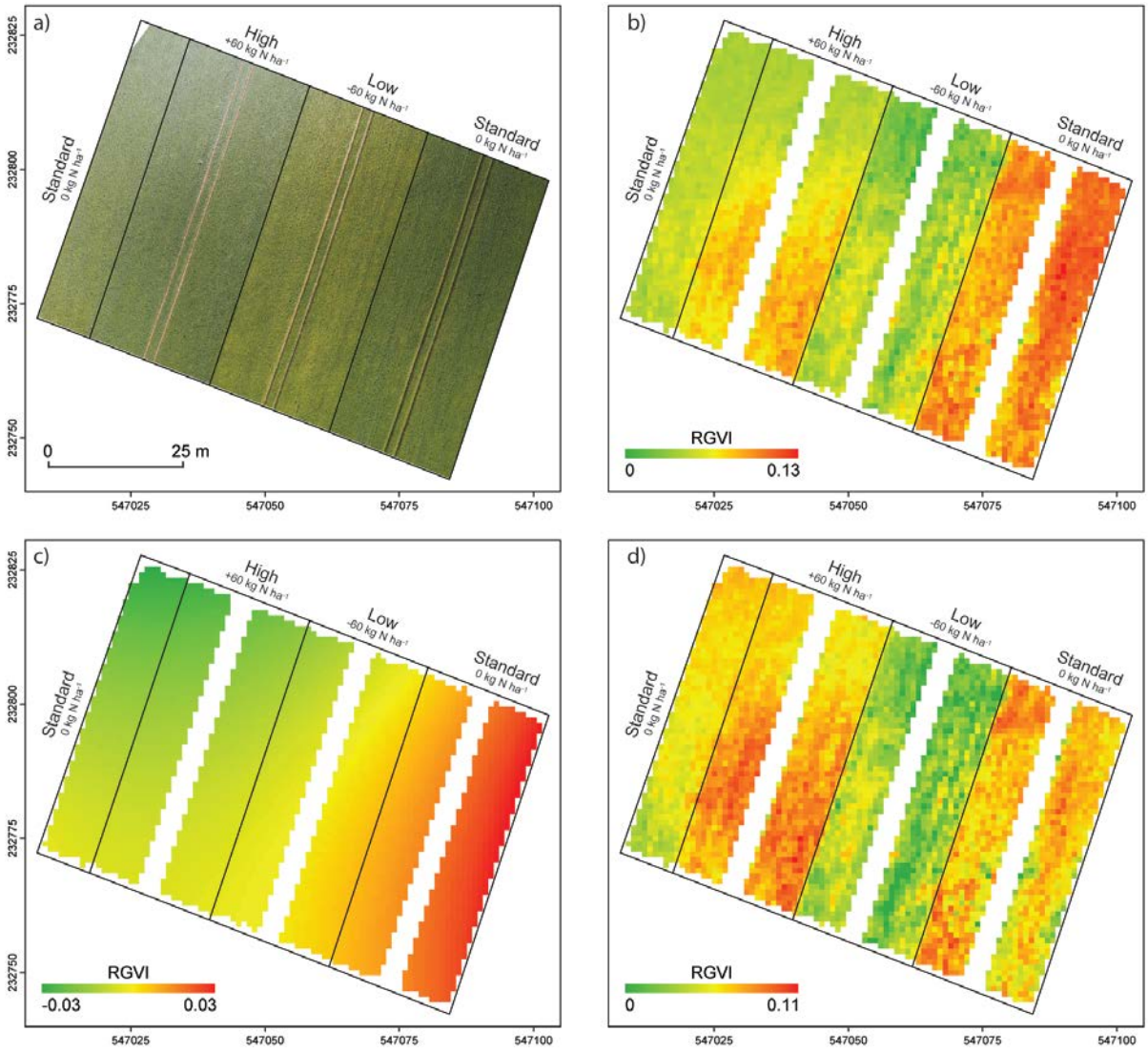

Figure 1: Strip trial design of the nitrogen response trial. a) RGB image taken by a drone at high resolution, b) the RGVI following smoothing and downsampling to 1 m resolution, c) the estimated underlying quadratic trend surface for the RGVI and d) the residuals upon subtraction of the quadratic trend surface from the RGVI image.

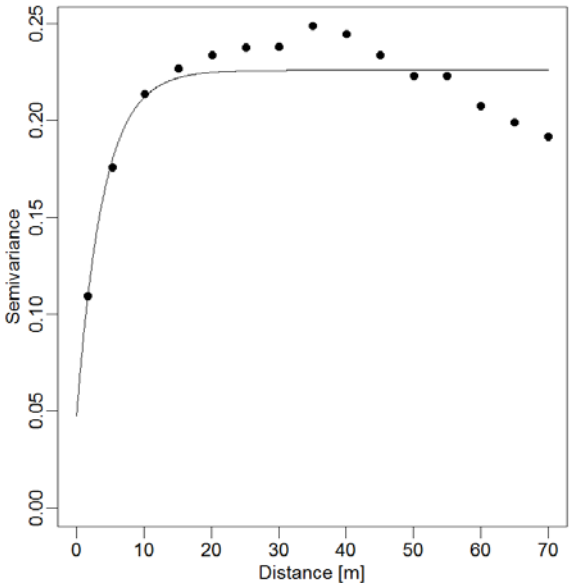

Figure 2: The estimated variogram of the LMM residuals.

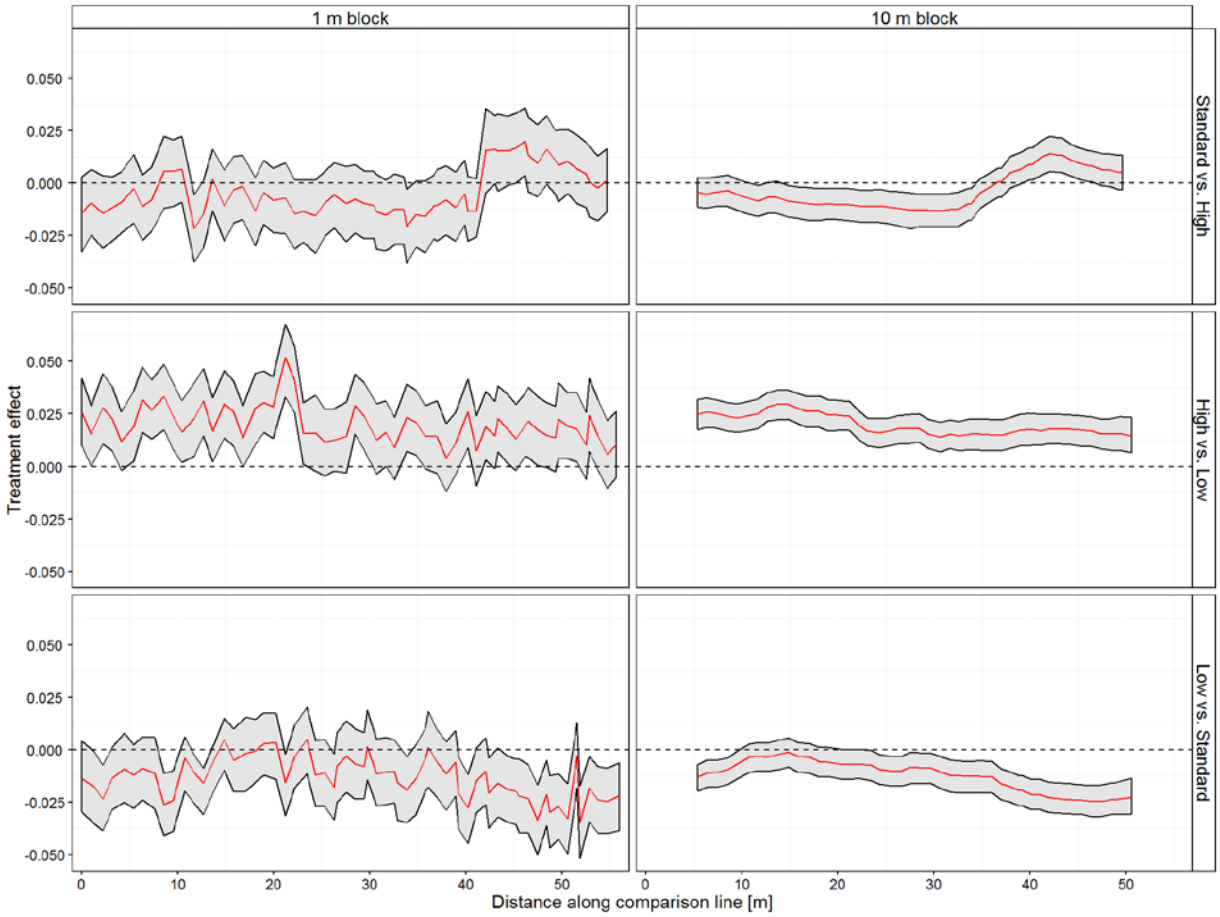

Figure 3: Results of the SDA comparing the residual RGVI on either side of the boundary between the stated treatments and for the stated block lengths. The red lines are the predicted difference in RGVI and the grey area the 95% prediction interval for this difference.

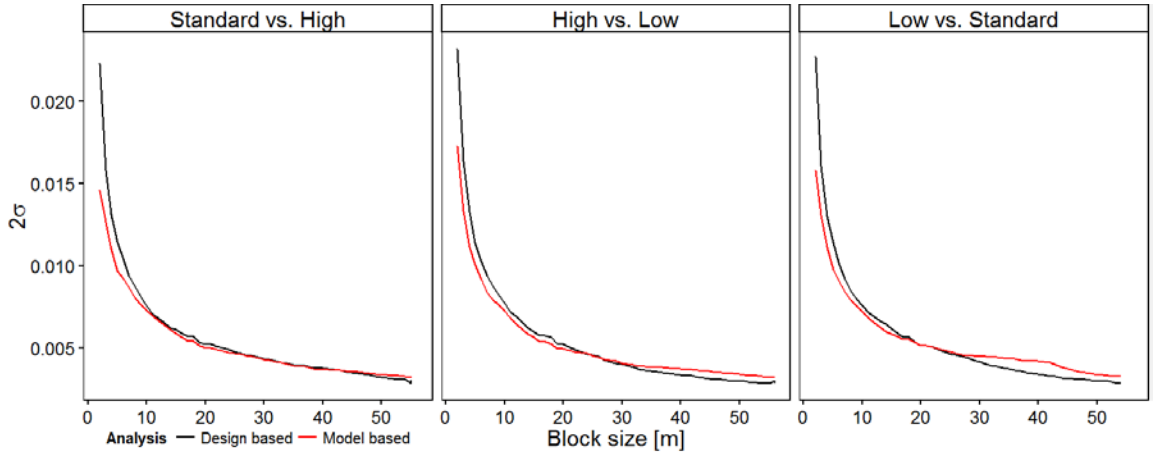

Figure 4: Illustration of the trade-off between the precision of the comparisons across the line (expressed as two times the standard error when estimating the treatment effect) for the stated treatment comparisons and the length of the block over which the comparison is made. The black lines show the estimated precision when spatial correlation is accounted for whereas the red lines show the estimated precision when it is ignored.

Acknowledgements

The methodologies described in this paper were developed as part of the Agronōmics Project which is led by ADAS UK Ltd., involves AgSpace, BASF UK Ltd, Trials Equipment Ltd., VSN International and the British Geological Survey, and is co-funded by Innovate UK. The nitrogen trial was conducted as part of the AHDB funded LearN project which is led by ADAS with partners NIAB TAG, Agrii and CF Fertilisers. The aerial photography was conducted by Paul Witney and Simon Ward of the British Geological Survey. This paper is published with the permission of the Executive Director of the British Geological Survey.

References

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M.L., Bareth, G., 2015. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. International Journal of Applied Earth Observation and Geoinformation 39, 79-87.

Bishop, T.F.A., Lark, R.M., 2006. The geostatistical analysis of experiments at the landscape-scale. Geoderma 133(1-2), 87-106.

Bishop, T.F.A., Lark, R.M., 2007. A landscape-scale experiment on the changes in available potassium over a winter wheat cropping season. Geoderma 141(3-4), 384-396.

Blackmer, A.M., White, S.E., 1996. Remote sensing to identify spatial patterns in optimal rates of nitrogen fertilization. In: P.C. Robert, R.H. Rust, W.E. Larson (Eds.), Precision Agriculture: Proceedings of the Third International Conference on Precision Agriculture. ASA-CSSA-SSSA, Madison, USA, pp. 33 – 41.

Bramley, R.G.V., Cook, S.E., Adams, M.L., Corner, R.J., 1999. Designing your own on-farm experiments: how precision agriculture can help. Grains Research & Development Corporation, Canberra, Australia.

Bruulsema, T.W., Malzer, G.L., Robert, P.C., Davis, J.G., Copeland, P.J., 1996. Spatial relationships of soil nitrogen with corn yield response to applied nitrogen. In: P.C. Robert, R.H. Rust, W.E. Larson (Eds.), Precision Agriculture: Proceedings of the Third International Conference on Precision Agriculture. ASA-CSSA-SSSA, Madison, USA, pp. 505-512.

Cathcart, R.J., O' Halloran, I.P., Bertoldi, A.P.v., 1999. Corn yield response to Delta Y predicted variable nitrogen fertilization: an Ontario field study, Proceedings of the Fourth International Conference on Precision Agriculture, St Paul, Minnesota, USA, pp. 417-427.

Geipel, J., Link, J., Claupein, W., 2014. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. Remote Sensing 6(11), 10335.

Hunt, E.R., Cavigelli, M., Daughtry, C.S.T., McMurtrey, J.E., Walthall, C.L., 2005. Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. Precision Agric 6(4), 359-378.

Kindred, D., Sylvester-Bradley, R., Clarke, S., Roques, S., Smillie, I., Berry, P., 2016. Agronōmics – an arena for synergy between the science and practice of cropping, 12th European IFSA Symposium, Harper Adams University, England.

Kindred, D.R., Milne, A.E., Webster, R., Marchant, B.P., Sylvester-Bradley, R., 2015. Exploring the spatial variation in the fertilizer-nitrogen requirement of wheat within fields. J Agr Sci-Cambridge 153(1), 25-41.

Lark, R.M., Stafford, J.V., Bolam, H.C., 1997. Limitations on the spatial resolution of yield mapping for combinable crops. J Agr Eng Res 66(3), 183-193.

Lawes, R.A., Bramley, R.G.V., 2012. A simple method for the analysis of on-farm strip trials. Agron J 104(2), 371-377.

Little, T.M., Hills, F.J., 1978. Agricultural Experimentation: Design and Analysis. John Wiley & Sons Ltd, West Sussex, England.

Marchant, B.P., Lark, R.M., 2007. The Matern variogram model: Implications for uncertainty propagation and sampling in geostatistical surveys. Geoderma 140(4), 337-345.

McBratney, A.B., 1984. The role of geostatistics in the design and analysis of field experiments with reference to the effect of soil properties on crop yield In: D.R. Nielsen, J. Bouma (Eds.), Soil Spatial Variability: Proceedings of a workshop of the ISSS and the SSSA. Wageningen, The Netherlands: Pudoc, Las Vegas.

McBratney, A.B., Webster, R., 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. J Soil Sci 37(4), 617-639.

Motohka, T., Nasahara, K.N., Oguma, H., Tsuchida, S., 2010. Applicability of green-red vegetation index for remote sensing of vegetation phenology. Remote Sensing 2(10), 2369.

Rudolph, S., Wongleecharoen, C., Lark, R.M., Marchant, B.P., Garré, S., Herbst, M., Vereecken, H., Weihermüller, L., 2016. Soil apparent conductivity measurements for planning and analysis of agricultural experiments: A case study from Western-Thailand. Geoderma 267, 220-229.

Sadler, E.J., Camp Jr., C.R., Evans, D.E., Millen, J.A., 2002. Spatial analysis of corn response to irrigation. In: P.C. Robert (Ed.), Proceedings of the 6th International Conference on Precision Agriculture and Other Precision Resources Management. ASA-CSSA-SSSA, Madison, USA, pp. 159 – 169.

Webster, R., Oliver, M.A., 2007. Geostatistics for environmental scientists. John Wiley & Sons Ltd, West Sussex, England.