A simple index to determine if within-field spatial production variation exhibits potential management effects: application in vineyards using yield monitor data

James Taylor, Bruno Tisseyre, Corentin Leroux

A simple index to determine if within-field spatial production variation exhibits potential

management effects: Application in vineyards using yield monitor data.

James A. Taylor[1,*], Bruno Tisseyre[1], Corentin Leroux[1,2]

[1]UMR ITAP, Montpellier SupAgro, Irstea, Univ. Montpellier, Montpellier, France, 34000

[2]SMAG, 55 Rue Euclide, Montpellier, France, 34000

[*]E-mail:james.taylor@irstea.fr; ph: +33 617672155

Abstract: As agricultural data and information becomes more abundant, diagnostics are needed to quickly and efficiently interrogate these data. Indices exist to identify sensor data with structured spatial variation, conducive to site-specific management. However, these indices do not indicate if this spatial variation is driven by managerial or environmental effects. A new index is proposed to identify perennial (or ordered row) fields that are likely or highly likely to have management effects within the spatial pattern of sensor data. This is determined by investigating differences in anisotropic (directional) variograms parallel and perpendicular to the direction of management (row orientation). Small differences are indicative of isotropic (environmental-driven) variation. Large differences indicate row and management effects. The index is derived, run on a database of 1080 simulated fields and applied to yield data from 124 vineyard blocks to assess index performance and response to different levels of variation. Simulations showed that the index is non-responsive to the magnitude of variation but responds strongly to anisotropy in the data. The stochastic variance in the data was observed to have an effect on index response and may be problematic when applied to noisy data sets. The index scores for the simulated and real-world data showed a similar pattern of response and the index was able to identify vineyard blocks where differential row management had generated differing yield responses. The index scores are continuous and some general guidelines for use of the index are proposed.

Keywords: anisotropy, variograms, viticulture, horticulture

Introduction

High-resolution proximal crop data are widely available in many agricultural systems. These include data from yield monitors and canopy sensors. With increasing volumes of data being generated, end-users (growers and agronomists) need rapid and effective diagnostics to interrogate the quality of the data. This is especially needed before deciding if a data layer should be used in decision-making processes. Incorrect or poor quality data that is used in visualisation or in decision processes will generate poor decisions and undermine user confidence in the data and the process.

Some examples already exist of diagnostics that operate on the information within raw sensor data to help inform end-users. These diagnostics have been primarily used to understand the level of variation with-in fields and to quantify the opportunity for a field to be suitable for site-specific management. To date, proposed "opportunity indices" have considered both the magnitude and spatial structure of yield (Pringle et al., 2003, de Olivera et al., 2007) and canopy (Monso et al., 2013) variation and/or the technical constraints imposed by operations (Tisseyre and McBratney, 2008) and zoning (Roudier et al., 2011). Such indices consider the relationship between stochastic and structured variation, which is critical for considering the amount of production variation that is potentially manageable. They have been based on assumptions of isotropy in the data, i.e. that variation occurs equally in all directions. However, as well as understanding the potential opportunity for site-specific management, end-users also need an indication of whether the spatial variation in the data is likely to be associated with (spatial) differences in management, or if it is driven by environmental factors. If variation in a field is caused by management effects, then the first step is most likely to be to correct management before considering further opportunities for site-specific management.

Understanding the potential sources of variation is particularly important in ordered, perennial cropping systems, with vineyards being an example. In such systems, management is very focused on activities along a row. There exists the potential for adjacent rows to receive different levels of management. A very practical example of this is in the level of pruning (bud numbers left) by different individuals working along adjacent rows, which can impact yield and canopy development (Bates, 2008). Other possibilities include incorrect machinery set-up or equipment operations on one row of a multi-row machine, such as a sprayer. Such management effects could leave a systematic pattern of crop response in a vineyard, orchard or field. If this

is the case, then differences in anisotropic, or directional, analysis of crop data should provide an indication of whether there is an effect of management on production.

Figure 1 illustrates the potential problem using vineyard yield monitor data. It is clear in Figure 1 that the region denoted by the area A, has an irregular yield pattern that is likely to be following an environmental effect – and likely to be soil variation. Patterns flow across block boundaries. In contrast, the area denoted B has a yield pattern that is aligned with the rows, indicative of different management in different sections of the vineyard blocks (NB: all blocks are the same variety and same trellis system). There will of course be intermediate situations where there are some management effects overlying a spatial environmental effect.

At present, a suitable diagnostic for rapidly analysing spatial production data for potential management effects does not exist. In this paper, an index is proposed that considers how crop production, in this case yield, varies along and across trellised vineyard systems. The hypothesis is that environmental variation tends to be isotropic in nature, so similar spatial variance structures should be observed along and across rows. In contrast, management effects, if present, will be oriented along rows resulting in different spatial variance structures parallel and perpendicular to the rows.
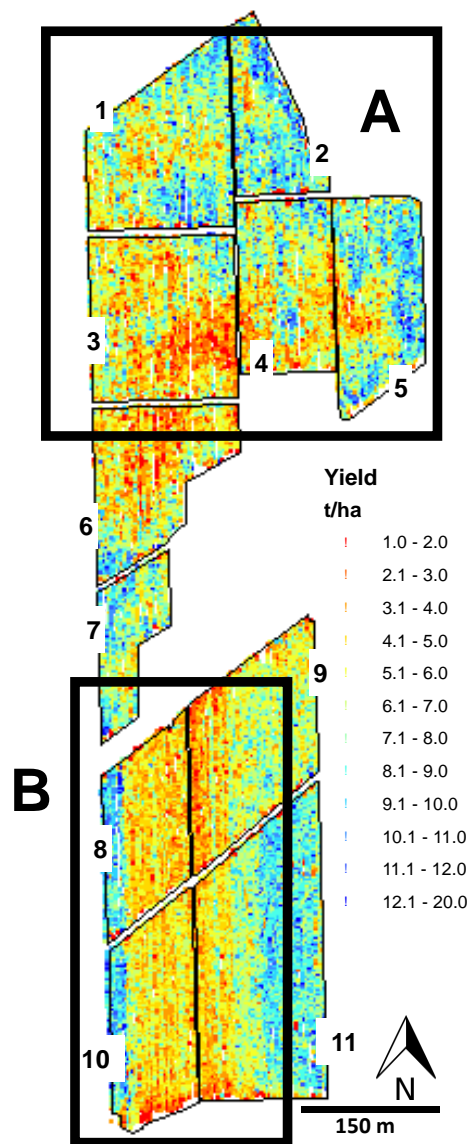
Fig. 1 A 2014 grape yield map from a Concord (*Vitis Labrusca* Bailey) vineyard in the Lake Erie Region of NY State. Boxes A and B show regions where yield relates to environmental variations (A) and where yield shows strong linear effects associated with management (B). The numbers on the graph are block identifiers (1 – 11).

Methods and Materials

The Management or Environment Index (MoE$_I$) - Derivation of MoE$_I$

Variography is a well-developed and accepted discipline in precision agriculture for interpolation (e.g. Whelan et al. 1996), sampling (e.g. Kerry and Oliver, 2003) and for diagnostic statistics (e.g. Pringle et al.

4

2003). It is possible to compute experimental variograms within a variety of different software platforms (e.g. via R packages or with Python libraries). The proposed MoE$_I$ uses this ability and is calculated by considering the difference in the area under the curve (AUC) of an anisotropic variogram generated parallel to the direction of management (rows) and the AUC of an anisotropic variogram generated perpendicular to the direction of management. The difference is then standardised against the larger of these two areas (Eq. 1). Low scores indicate a similar spatial structure in both directions, larger scores will be indicative of differences.

$$\text{MOE}_I = \frac{|AUC_\perp - AUC_=|}{\max[AUC_\perp ; AUC_=]} * 100 \qquad (1)$$

Where: $AUC_\perp$ is the area under the curve of the isotropic variogram constructed perpendicular to the direction of management (rows), $AUC_=$ is the area under the curve of the anisotropic variogram constructed parallel to (along) the direction of management.

The index will be dependent on the method used to generate the anisotropic experimental variograms and the theoretical model applied to the experimental variograms. To standardise this, there are several steps proposed to ensure that this equation can provide MOE$_I$ scores that are effective for decision-making widely within viticulture in the first instance, but also potentially in other cropping systems as well.

Derivation of the experimental variogram clouds: Every production system is potentially different, especially in the manner that the crop is established. The experimental variograms therefore need to reflect the conditions of the production system. From the production perspective, the key distances are the fixed row widths, the width of sensing operations and the typical swath width for field operations. With this in mind, the lag interval within the experimental variograms should approximate the minimum width between sensor operations (or swaths), which for vineyard yield data is a function of the row width[1], (*d*) and the maximum lag distance should be at least twice the typical field operation swath (*D*) (which may be greater than the typical sensing swath)[2]. In cases where the field operations tend to be very intense (e.g. based on 2-3 row operations), the maximum lag distance should be wider and a distance of 10 times the row width has

---

[1] For broad-acre row crops, this is likely to be a minimum sensing width, not row spacing.
[2] This is likely to be more relevant if the index is transferred to broadacre systems, where for example yield data may be collected on 6 m swaths but spraying operations may be at 36 m. In this case, the maximum lag will be 72 m to ensure two 'spray' swaths (which may be the cause of management effects) are captured in the index.

been adopted. Both conditions are likely to constrain the maximum lag to a relatively short distance (30 – 50 m). This was done deliberately to focus potential management differences on short-range variance structures and to avoid the potential effect of large spatial patterns in the data.

As the intention is to generate variograms in specific directions, a tolerance angle of 2° is specified to ensure that the anisotropic variograms generated parallel to the row orientation are only considering within-row data (Fig. 2). At this tolerance, with a row width of 3 m, data from adjacent rows should not be considered until lag distances > 80 m (assuming straight rows and fairly straight lines of georeferenced data), well beyond the maximum lag specified. Similarly, for the perpendicular anisotropic variograms, a small tolerance ensures a narrow band of data is considered. This minimises the distance between points that are located within the same row when moving sideways through the vineyard, avoiding issues with pairs within rows as well as between rows.

MoE_I Conditions:

Lag interval = row width (d)

Maximum lag = max (2x swath of widest operations (D); 10 x row width (d))

Tolerance angle = 2° (designed for vineyards at ~2-3 m row spacing[3]) but adjustable, such that the distance (Distance$_1$), where neighbouring row data (Distance$_1$) is considered for the parallel variogram is < the maximum lag distance, and the maximum distance within a row (Distance$_2$) for the perpendicular variogram is < d.

Given these constraints, anisotropic experimental variograms are generated parallel to (along) and perpendicular to the direction of management (rows) using the standard semi-variance formula (Eq. 2) constrained to an angle equal to the row orientation (θ) or perpendicular to it ($\theta + \pi/2$).

$$\gamma(h,\theta) = \left(\tfrac{1}{2}N(h,\theta)\right)\sum_{i=1}^{N(h,\theta)}[z(x_i) - z(x_i + h,\theta)]^2 \qquad (2)$$

Where θ = row orientation and the perpendicular solution utilises ($\theta + \pi/2$).

---

[3] For systems where row spacing and plant/tree/vine spacing is larger, e.g. large tree crops, a larger tolerance may be used and tolerance angle could be adapted.

These conditions and the directional (anisotropic) variogram determination are illustrated in schematic form (Fig. 2).
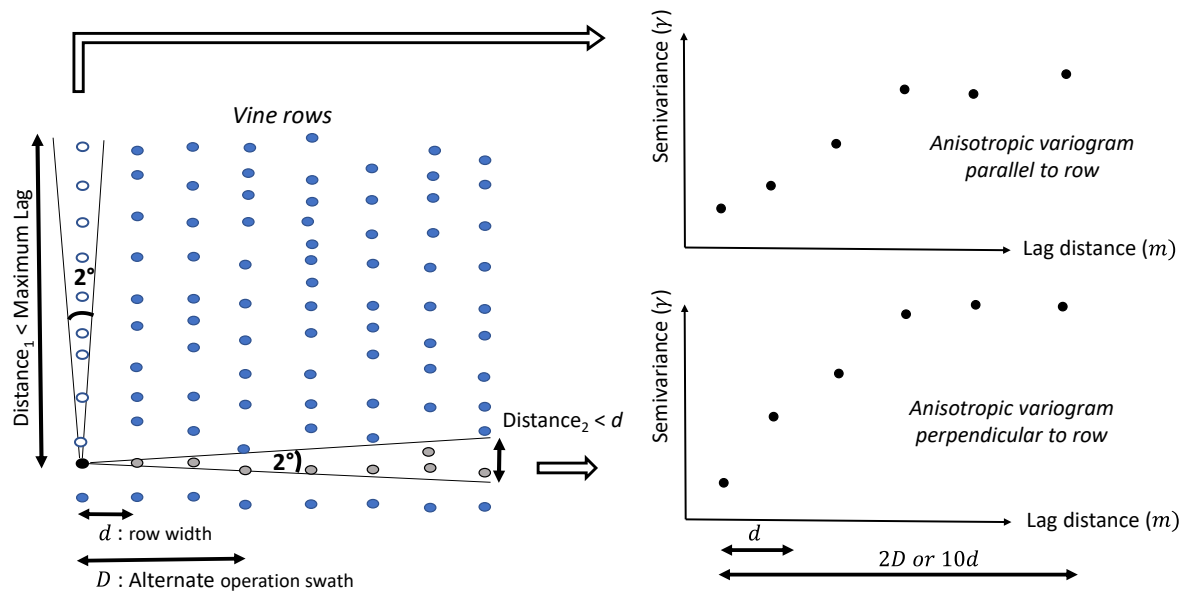

Fig. 2 Schematic indicating how the conditions specified limit the selection of data for the anisotropic experimental variograms. Image is not to scale. Points used for generating pairs with the example point (●) are indicated as open circles parallel to the row and as grey circles perpendicular. All other points are discounted from the semivariance calculations. The angle (2° here) should be chosen so that points from neighbouring rows are not considered for the parallel variogram, and points are not considered within the same row for the perpendicular variogram.

Model fitting to the experimental variogram cloud: When used for interpolation or other opportunity indices, experimental variograms are usually modelled with a standard range of models (spherical, exponential, linear with sill, gaussian, matern, etc…). However, correct automatic model fitting can be problematic using these models. The intent with an index such as this, is to ensure that it is as user-friendly and to be as robust as possible. It is not practical or desirable to have users verifying the quality of the theoretical model fit to the experimental variogram. The sole objective is to ensure a good fit to determine the area under the curve. To achieve this, a smoothing spline fit has been employed with a smoothing parameter of 0.6, which allows some flexibility in the fit. Additionally, for each lag in the experimental variogram, a weighting for the fit was assigned based on the relative number of pairs used to generate that point. This was done to ensure that

7

any individual lag with a low level of confidence (low number of pairs) does not adversely affect the fit.

Figure 3 shows an example of the experimental anisotropic variograms generated parallel and perpendicular

to the direction of harvest and the spline fit to the data for vineyard yield collected with an on-harvester yield

monitor. The high semivariance of the first lag in the anisotropic experimental variogram perpendicular to

rows (Fig. 3) is associated with a low number of pairs (n = 74) in its determination. The spline only shows a

small change associated with this value due to the weighting in the fit. Most of the other lags, in both of the

anisotropic experimental variograms, have a number of pairs > 1000.

Calculation of AUC: With the smoothing spline defined above, the numerical integral is calculated and used

for the AUC of each anisotropic variogram. These values are used in Equation 1.

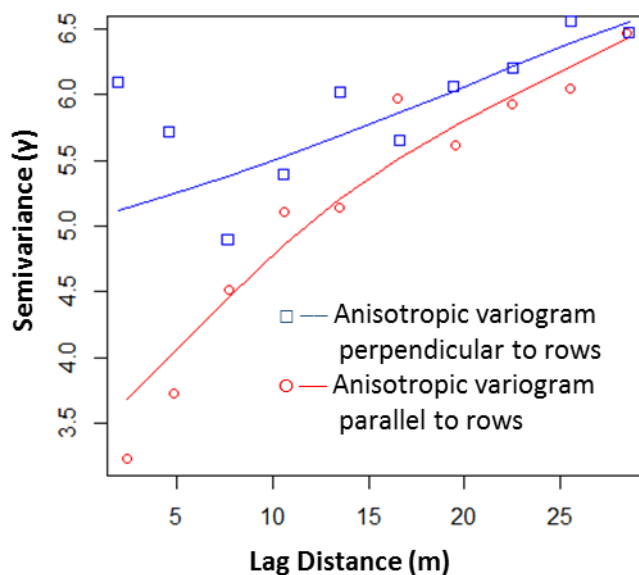**Fig. 3** An example of the anisotropic experimental variograms and weighted spline fits derived from yield

monitor data collected in a Concord vineyard block, Westfield, NY, USA.

MoE$_I$ applied to simulated yield data

To assess the performance of the MoE$_I$, simulated 1 ha 'vineyard' yield data sets were generated with a

known spatial structure (and no induced errors) using the yield data simulator of Leroux et al. (2017). The

row width was set at 5 m and the inter-vine distance within a row at 3 m. The choice of 5 m rows and 3 m vine planting is slightly larger than is typical in most systems (often approximately 2.5 m x 1.5 m), but was increased to reduce computational time for the simulations while approximating typical row to vine spacing ratio. Simulations were targeted to have an average yield of 7 – 7.5 Mg/ha with variations in:

- The **coefficient of variation** (CV) ==> from 10 % to 60 % in steps of 10% (6 levels)

- The **range of the variogram of the simulated yield** ==> from 30 m to 90 m in steps of 30 m (3 levels)

- The spatial structure in the data represented by the **nugget to sill ratio** ==> from 30 % to 70 % in steps of 10 % (5 levels), where 30% indicates that the nugget and (partial) sill comprise 30% and 70% of total semivariance respectively, and

- The **anisotropy** in the data represented by the percentage of anisotropy introduced ==> from 30 % to 90 % in steps of 20 % (4 levels), where 90 % anisotropy is generated by making the perpendicular variogram range $1/10^{th}$ the parallel variogram range.

The simulator was run 3 times to generate 3 repeats for all possible combinations of variables. In total, there were 1080 simulated fields (6 x 3 x 5 x 4 levels x 3 repeats).

These variables were selected to simulate different levels in the amount (magnitude) of variation and the spatial structure to the variation based on reported values from a global survey of vineyard yield monitor data (Taylor et al., 2005). Different levels of anisotropy were also selected. There was no reported literature on observed anisotropy in vineyard yield data so the range of anisotropy ratios selected (10 -70 %) was based on the authors' experience and expectations.

In the first instance, simulation is useful for providing an expected population distribution of MoE_I scores as a reference for any real-world applications. This was assessed using a histogram analysis and summary statistics. In the second instance, simulation can provide an indication of the sensitivity of the MoE_I scores to the different facets of yield variation changes. ANOVA was used to interrogate independently the effect of each variable (CV, variogram range, Nugget:Sill ratio and anisotropy) on the MoE_I scores and the behaviour of the index.

For each simulated field, the R program (R Core Team, 2017) was used to generate the experimental directional variogram (gstat package), fit a smoothing spline (stats package) and to determine the AUC (MESS package) for the parallel and perpendicular anisotropic variograms. The data was exported and the MoE_I calculated (Eq. 1) in JMP (v12; SAS Institute, Cary, NC, USA). Histograms and ANOVA were performed in JMP. Maps of the simulated yield data for selected blocks were generated using ArcMap (ESRI, Redlands,CA, NY, USA).

Real-world Application of MoE_I:

A Concord enterprise (~170 ha over 35 blocks), spread over several locations near Westfield NY and within the Lake Erie viticulture region, was chosen as a detailed test site. The vineyard in Figure 1 is one of these locations. All the enterprise vineyards have been yield mapped for 4 years (2014-17) using the Advanced Technology Viticulture grape yield monitor (Adelaide, South Australia, Australia). This provided a relatively temporally rich data set for viticulture (see Taylor et al., 2016 for further details on the yield monitoring process). Most of the vineyard blocks are oriented North-South (as shown in Fig. 1), although some other blocks (not shown) do have other orientations, which are accounted for in the MoE_I calculation. There were some blocks where yield data were not collected or only partially collected in some years. Blocks with no data or < 500 data points were omitted and, after removal, a total of 124 vineyard block-years (over the 4 years) were available for the analysis. The vineyard management team is progressive and often performs on-farm trials to assess new approaches. Some of these have been highlighted in Figure 1 (Area B) for one year, however experimentation is only limited to a small area in any given year.

The yield data for each year was cleaned before analysis to remove extreme values (<0.25 and >30 Mg/ha). It was then trimmed to within 3 standard deviations of the mean. Row width in these vineyards is approximately 2.7 m. Operations are done on 3-row sections (~8 m), so a max lag distance of 27 m (10 x row width) was used. The MoE_I was calculated as described above using R scripts and JMP.

Histograms of scores from all blocks are presented as well as detailed scores and anisotropic variograms for the blocks presented in Figure 1. Maps of the trimmed yield data were generated in ArcMap. Variogram plots were produced in R.

Results and Discussion:

Simulation

The simulated outputs generated a range of MoE_I scores from 0.1 to 60.6. The distribution was positively

skewed with a peak at very low values (<2), a second peak at 12, a spike about 22 and then a long tail. This

indicates that there may be overlapping distributions in the total population. The median MoE_I score was

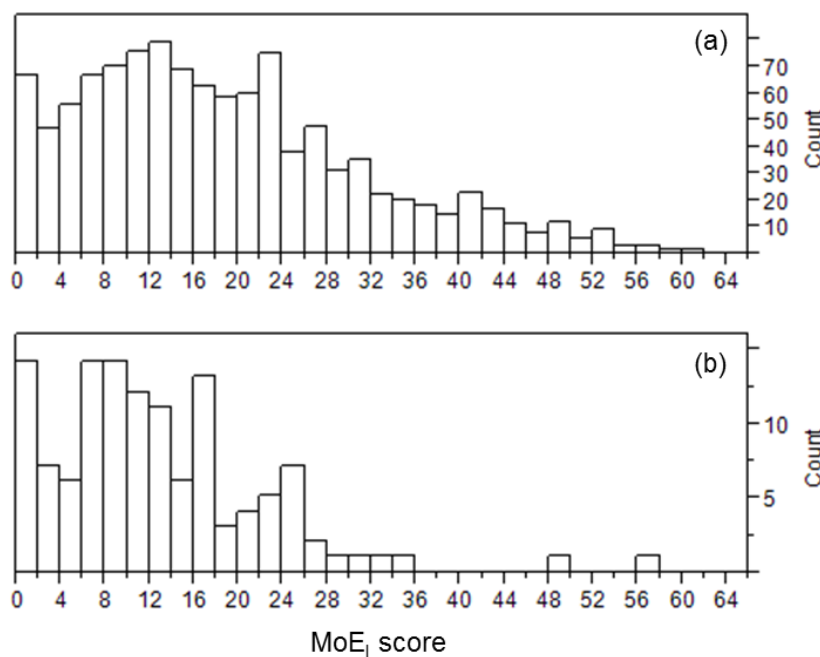16.5 and mean score was 18.5 in the simulated data.

Fig. 4 Histograms of the MoE_I scores generated (a) across the simulated data sets (n=1080) and (b) using

yield monitor data from a Concord enterprise (2014-17; n=124).

The ANOVA results (Table 1) showed that there was no effect of CV on the MoE_I scores. Since the MoE_I

is a ratio between the anisotropic variograms, it was not expected that the magnitude of variation in the

system would affect the score and the simulation results showed this. The variogram range affected MoE_I

values (lower scores) at shorter ranges (30 m). This range approximates the max lag distance defined for the

index in this system. At longer ranges, there was no statistical effect on the scores. The mean MoE_I scores

were statistically significantly different between all Nugget:Sill ratio levels and the level of anisotropy

applied also generated statistical differences in mean MoE_I scores for all levels. Increased differences in the level of anisotropy parallel and perpendicular to rows was expected (and the intent of the index). Increasing the Nugget:Sill ratio reduces the expected spatial patterning in the data and the level of autocorrelation in the data. Less autocorrelation resulted in less potential to differentiate between anisotropic variograms. This indicates that 'noisy' data sets, with a high level of stochastic variance, may not be suitable for analysis with this index and care should be taken in conditions where the nugget variance represents a large proportion of the total variance. Nugget:Sill ratios > 50% in vineyard yield monitor data are common in viticulture, particularly in Australian systems (see summary data on the Cambardella Index (Cambardella et al., 1994) presented in Taylor et al. (2005), so it is an expected effect).

Table 1 Mean responses for each adjusted variable. For a given variable, all permutations of the other variables were included in determining the mean value. Significant differences were determined using Tukey's Means comparison following ANOVA. Different letters indicate significant difference within a variable ($p<0.05$) (but not between variables).

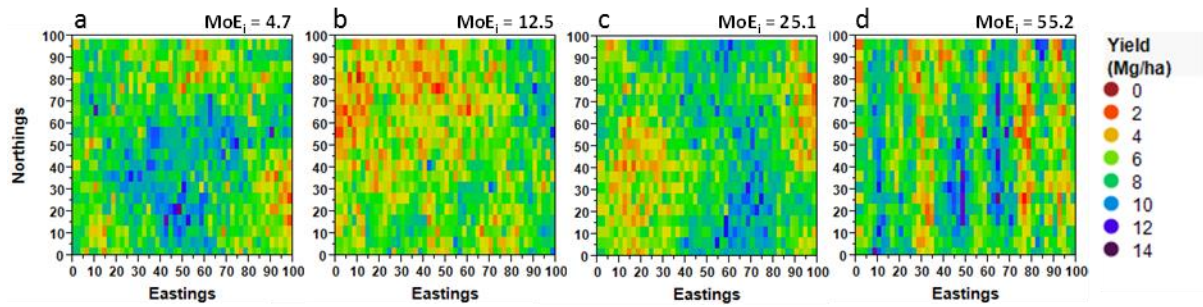| Variable	Level	Mean MoE_I
CV	10%	18.95[a]
CV	20%	19.01[a]
CV	30%	18.31[a]
CV	40%	18.11[a]
CV	50%	18.56[a]
CV	60%	18.37[a]
Range	30m	13.96[ba]
Range	60m	20.38[b]
Range	90m	21.32[b]
Nugget:Sill	30%	26.27[a]
Nugget:Sill	40%	22.59[b]
Nugget:Sill	50%	18.86[c]
Nugget:Sill	60%	13.95[d]
Nugget:Sill	70%	11.10[e]
Anisotropy	30%	8.78[d]
Anisotropy	50%	14.96[c]
Anisotropy	70%	22.38[b]
Anisotropy	90%	28.09[a]

Fig. 5 Simulated yield data sets selected randomly from within populations of MoE_I scores of a) 0-10, b) 10-20, c) 30-40 and d) >40. The actual MoE_I score for each field is indicated on each plot. Simulated yield maps show the shift from large, irregular spatial patterns (left - a), likely associated with natural environmental factors, to 'striped' patterns associated with management effects (right – d).

Figure 5 shows four examples of simulated yield data selected randomly from different parts of the MoE_I population. The progression from having large coherent, non-vertical variation (left, a - b) to very 'striped' vertical effects in the data (right, d) is clearly evident. Figure 5c shows a simulated dataset with an above median MoEI score (25.1) and presents an interesting case. The trend East-West of low-high-low yield is evident and approximates a vertical structure but it is not clear if this is following management or aligned environmental variation. It illustrates that this is an indicative tool and should help to identify blocks (fields) of interest, and not be used without verification.

Real World Application

The distribution of MoE_I scores acquired from the Concord enterprise yield monitor data (2014-17) is shown in Figure 4b (on the same x-axis as the simulated data histogram). It has the same approximate shape and positive skew as the simulated data (Fig. 4a), with a peak at low values and secondary peaks at 8 and 16, shifted slightly from those in the simulated data, followed by a long tail and a range from 0.03 to 56.43. The median (11.1) and mean (12.8) scores were slightly lower than the simulated data. The simulated and real-world data are behaving in a similar way, providing a validation for the simulated outcomes. Again, it appears that there may be several populations in the real yield data.

To further interrogate the MoE_I performance, the MoE_I scores for 11 blocks (shown in Fig. 1) across the four years of measurement are given in Table 2. Scores tended to range from 5 to the mid-high 20s, with two outliers with scores > 40 recorded. There is no clear stable temporal pattern to the block scores, with most blocks recording both higher and lower scores in different years. The two high outliers in 2014, Blocks 8 and 10 (Fig. 1), have a yield response that is strongly affected by management effects, in particular differences in pruning strategies (*pers. comm.* Mr Thom Betts, Betts Vineyards Ltd, Westfield NY, USA). In subsequent years, this management effect was removed (by using a uniform canopy management strategy). In the southern section, Blocks 8 and 10, that exhibited high MoE_I scores in 2014, showed a decreasing pattern in MoE_I scores over subsequent years when more 'normal conditions' prevailed. Block 11 was an exception, exhibiting similar scores in all four years and the same East-West yield trend. However, while known, the reason for this is unclear to the vineyard managers.

All blocks where the MoE_I scores were > 25 (n=12) were discussed with the vineyard manager. In all cases, the manager could identify varying management effects, often without viewing the yield map. These were usually either a change in pruning strategy or a difference in crop thinning operations that produced a 'management' effect on the spatial yield pattern. Blocks 4 and 5 in 2014 (Fig. 1) were 2 of the blocks discussed and present interesting cases. On first viewing, they do not seem to have strong management effects (part of Zone A in Fig. 1). However, Block 4 was used for cover crop trials in 2014 and a higher yielding middle section resulted. In Block 5, the eastern-most part of the block missed a pruning treatment resulting in higher yields. These trends are evident with a closer inspection of the raw yield data (Fig. 1).

Table 2 Annual MoE_I scores for each block within a vineyard, Westfield, NY, USA derived from yield monitor data collected over 4 seasons (2014-17). Missing data indicates that insufficient yield monitor data were collected in that block in a particular year to perform the analysis.

Block ID	2014	2015	2016	2017
1	6.67	10.32	11.18	9.09
2	20.86		7.84	14.81
3	12.19	6.48	1.44	9.58
4	27.10		8.76	10.68
5	25.85		2.76	21.39
6	9.94	15.17	12.62	16.82
7	9.14	12.67	16.88	2.85
8	31.95	7.38	0.16	13.75
9	16.07	9.23	8.85	12.03
10	33.55	6.40	6.12	9.95
11	17.06	18.60	16.15	17.72

To further illustrate the behaviour of the index, some examples of the anisotropic variograms and the raw yield maps are shown in Figure 6. The two example blocks, Blocks 3 and 8, are again derived from the vineyard in Figure 1. Block 3 has been taken from the northern section, that exhibited more natural, irregular yield patterns in 2014, and Block 8 from the southern section that had management effects in 2014 (Fig. 1). The two blocks represent high, low and intermediate MoE_I scores across the 4 years (Table 2).
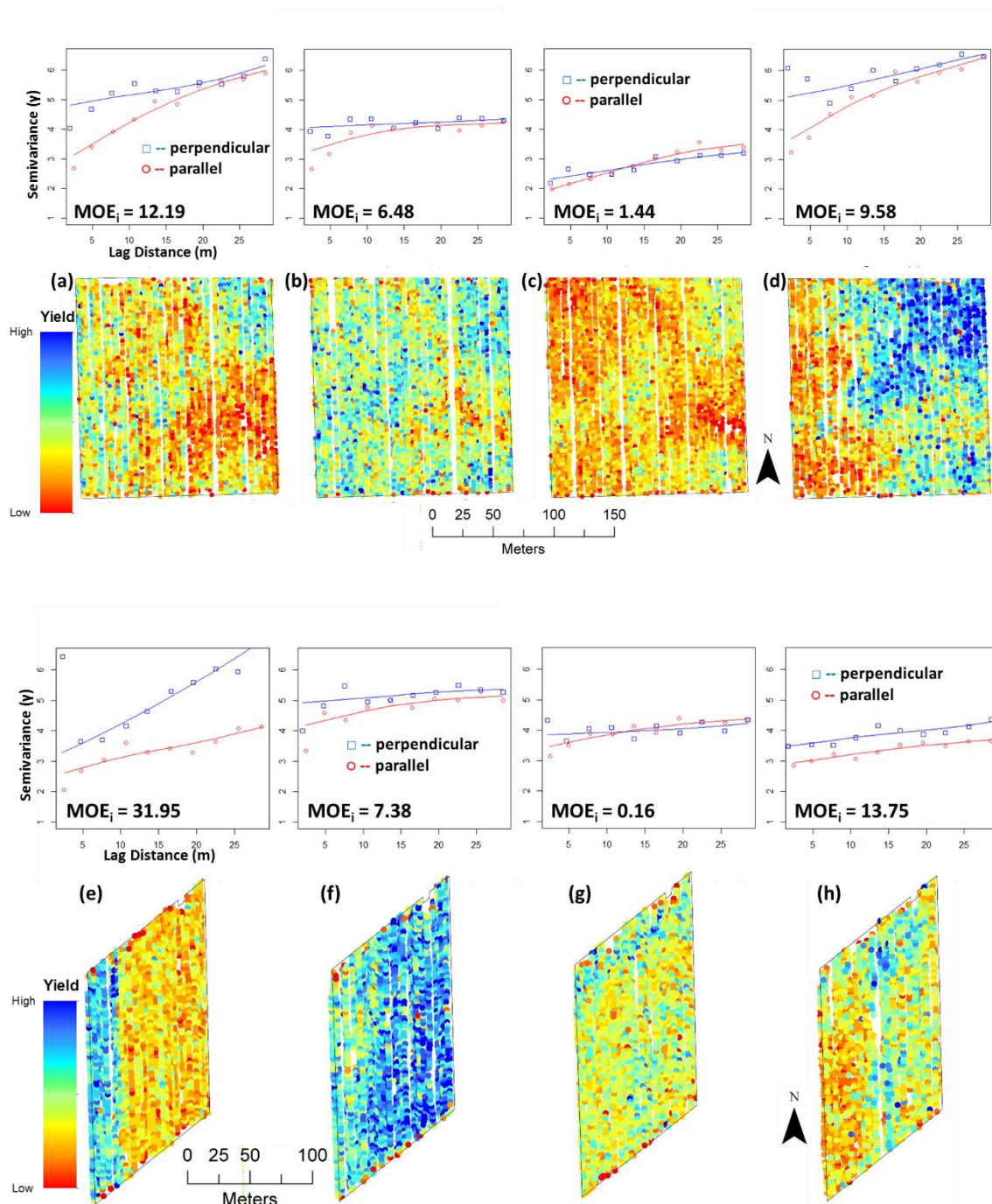
Fig. 6 Anisotropic experimental variograms and weighted spline fits parallel (○) and perpendicular (□) to management, with the corresponding raw yield maps from two blocks (Block 3 top; Block 8 bottom) from 2014 (left) to 2017 (right). All the variograms are plotted on a common scale. Yield is also plotted on a common scale across the years.

The maps of Block 3 (Fig. 6a-d) show no defined 'striping' in any year. There is a strong visible trend in the block from east-west (low-high), especially in 2017. However, the MoE_I scores are low to moderate (< 13) across all years. In contrast, the management effect in Block 8 in 2014 (Fig. 1 and Fig. 6e) is clearly defined and produced a high-yielding western part in a generally low-yielding year. This produced a strong deviation in the parallel and perpendicular variograms, which resulted in the high MoE_I (31.95) (Fig. 6e). The following year, 2015 (Fig. 6f), exhibited generally high yields across the block but does show an inverse of the 2014 pattern, with the western side slightly lower yielding than the rest of the block. However, management in 2015 was uniform, and the difference is a legacy effect of the 2014 management effect. Thought this effect was visible in 2015, especially with the 2014 data supplied, there is no distinct striping pattern in the Block 8 2015 yield data. The MoE_I was low in 2015 and by 2016 was approaching 0, with yield patterns in 2016 linked to underlying soil patterns, especially in the north of the block (Fig. 6g). In 2017 (Fig. 6h), the MoE_I rises again (13.75) and the western third of the block was lower yielding, a result of different pruning management in this section (*pers. comm.* Mr Thom Betts, Betts Vineyard Ltd, Westfield, NY, USA). The more 'natural' yield pattern that exists in the 2016 map was only evident in the eastern half of the block in 2017.

General Discussion

The simulation results showed that the index, as intended, is responding to differences in anisotropy and the amount of stochastic (nugget) variance in the data. It is not responding to the total amount of (aspatial) variance in the data. It was observed in the results that the variogram range had an effect when the range was less than the maximum lag set for the index. However, at longer ranges, there was no effect of the variogram range. Reported ranges for yield variograms tend to be larger than the likely maximum lag (e.g. a median range of 89.2 m from 162 blocks reported by Taylor et al. 2005). Therefore, the variogram range is likely to have a low impact on the MoE_I. By design, the MoE_I is constrained to short-range variation (10 x row width or 2 x max swath operation) thus the choice of a variogram range of 30-90 m was considered sufficient for index evaluation.

From the real-world results presented here, the MoE_I has flagged blocks that may have or are very likely to have management effects contributing to the spatial variation in within-block yield response. It appears to

have a strong potential to be used as a diagnostic index in precision agriculture. When discussing the outcomes with the vineyard manager, the manager could identify issues with the higher scoring blocks and was very interested to understand further the effect that management was having on production. The index was able to identify patterns that were associated with management effects that were not immediately obvious to the researchers, but were confirmed in discussions with the manager and closer inspection.

The simulation and Concord vineyard results indicated that MoE_I values of <10 are typical for vineyards where variability in production is caused predominantly by underlying environmental factors. Scores > 25 were always associated with blocks where management had strongly affected the spatial pattern of production. Based on the distribution of responses in Figure 4, and the information in Table 2 and Figure 6, some preliminary rules have been proposed in Table 3. These are designed to provide a guide for users and summarises some initial potential thresholds to interpret the MoE_I scores. It must be stressed that these are derived for yield monitor data in Concord systems in the Lake Erie Region. Since the MoE_I is standardised against the maximum AUC recorded, it is likely that these thresholds, or similar values will hold true in other applications, but should be validated before use. In cases where the stochastic variance (high Nugget:Sill ratio) is known or expected to be high, these thresholds will be lower as the MoE_I is less sensitive as the Nugget:Sill ratio increases (Table 2). This may be real production variation or induced variation (measurement error), for instance via a poorly installed/calibrated sensor or incorrect operation. The index cannot differentiate the reason for the elevated stochastic variance, but thresholds can be altered if the level of stochastic variance is likely to be dampening the MoE_I scores. The assumption is that the rank of the fields would still be maintained provided there is confidence in the experimental variograms.

Table 3 Threshold values and interpretation of MoEI scores within a viticulture system (based on Concord production in the Lake Erie region).

MoE$_I$ Score	Interpretation	Suggested action
<6	Variation is isotropic and highly likely driven by environment	Calculate other opportunity indices and consider differential management if necessary.
6-12	Variation is mostly isotropic and likely driven by environment	Consider a check of the data before calculating other opportunities and consider differential management.
12-18	Variation contains some anisotropy and is possibly associated with management	Visually check map of the block/field to determine if management effects are dominating. Consider data for site-specific management.
18-30	Variation exhibits anisotropy that is likely caused by management	Visually check map of the data with the intent of understanding where management effects are, their cause and remediation. Do not considering using data for PA unless management effects are permanent.
> 30	Variation exhibits strong anisotropy that is highly likely caused by management	Visually check map of the data with the intent of understanding where management effects are, their cause and remediation. Do not consider using data for PA unless management effects are permanent.

In the real-world scenario presented here, the fields flagged with high MoE$_I$ scores were all identified as being affected by varying management operations. This may not always be the case, and it is certainly possible, though not expected, that a change/error in the sensor during operations could cause a similar effect and generate a high MoE$_I$ score. Sensor malfunction should always be considered and discounted as a first step. The quality of the GNSS could also be an issue. Very low quality receivers, with low accuracy, could generate meandering and criss-crossing tracks that would confound the analysis. The MoE$_I$ score may not be sensible if individual rows cannot be defined in the plot of the raw data. Mid to high MoE$_I$ scores should only be treated as indicative of a potential management effect on the spatial pattern. Blocks need to be investigated to confirm management effects. This is likely to need plotting of the raw or interpolated data for visualisation of the spatial patterns and discussions with the vineyard (or farm) manager.

The intent with the MoE$_I$ is to identify blocks/fields where management may confound implementation of future differential (site-specific) management. Where management practices have been identified as driving the spatial variation, care should be taken when incorporating the data into a decision process. In most cases, the advice is likely to be to remove the data unless the management effect is expected to be constant or repeated. In the latter case, it is relevant to include the effect in differential decision and management

processes. Care should also be taken when considering using the 'management-affected' data with other indices, such the original Opportunity Index (Pringle et al. 2003) or the Technical Opportunity Index (Tisseyre and McBratney, 2008).

While its application here has been targeted at vineyard yield monitor data, theoretically, it could equally be applied to other vineyard production data, including remote- or proximal-sensed canopy data or other proximal vineyard data, such as the high resolution berry count data of Mirbod et al. (2016). Similarly, as indicated in the index description, it could potentially be applied across a wide range of crop systems, provided the MoE$_I$ conditions are suitably adjusted. Alternative perennial systems, organised in straight rows, are the most obvious applications. However, annual row crop data could be analysed if operations occur in straight lines and that turns, particularly headland and border operations, are removed before analysis.

Conclusions

An index has been proposed that examines ordered production data, with yield data used as an example, for spatial patterns that may be attributed to management rather than environmental variability. The index is based on differences between variography applied along and perpendicular to the direction of operations in a production system. When applied to actual and simulated vineyard yield monitor data, the index was effective at identifying fields where management effects were likely. The simulation data showed that the index responds to changes in anisotropy and in the amount of structured variation in the data (Nugget:Sill ratio), which are of interest, but was not affected by the magnitude of variation or the range (distance) of autocorrelated variance in the data, which are not of interest in this situation.

The proposed index fills a gap for a diagnostic tool to identify data sets that may be compromised and need validation before being used in data analytics and decision systems. It is proposed to be used with, not instead of, existing diagnostic statistics and indices in precision viticulture and agriculture. While defined here and applied to a relatively large viticulture data set, more widespread application will be needed to determine its suitability in other vineyard systems, with other crop types and with other sensors and to correctly interpret the index scores.

References

Bates, T. (2008). Pruning level affects growth and yield of New York concord on two training systems. American Journal of Enology and Viticulture, 59 (3), 276-286.

Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F., et al. (1994). Field-scale variability of soil properties in central Iowa soils. Soil Science Society of America Journal, 58, 1501–1511.

de Olivera, R.P., Whelan, B., McBratney, A. & Taylor, J. (2007). Yield variability as an index supporting management decisions: $YIELD_{EX}$. In J.V. Stafford (ed.). Precision Agriculture '07. Proceedings of the 6th European Conference on Precision Agriculture, The Netherlands: Wageningen Academic Publishers.

Kerry, R. & Oliver, M.A. (2003). Variograms of Ancillary Data to Aid Sampling for Soil Surveys. Precision Agriculture, 4(3), 261-278.

Leroux, C., Jones, H., Clenet, A., Dreux, B., Becu, M., & Tisseyre, B. (2017). Simulating yield datasets: an opportunity to improve data filtering algorithms. In J. A Taylor, D. Cammarano, A. Prashar & A. Hamilton (eds). Proceedings of the 11th European Conference on Precision Agriculture, Advances in Animal Biosciences, 8, 600-605.

Mirbod, O., Yoder, L., & Nuske, S. (2016). Automated Measurement of Berry Size in Images, IFAC-PapersOnLine, 49(16), 79-84.

Monsó, A., Arńo, J., & Martínez-Casasnovas, J.A. (2013). A simplified index to assess the opportunity for selective wine grape harvesting from vigour maps. In J.V. Stafford (ed). Precision Agriculture '13. Proceedings of the 9th European Conference on Precision Agriculture, The Netherlands: Wageningen Academic Publishers.

Pringle, M.J., McBratney, A.B., Whelan, B.M. & Taylor, J.A. (2003). A preliminary approach to assessing the opportunity for site-specific crop management in a field, using a yield monitor. Agricultural Systems, 76, 273-292.

R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Roudier, P., Tisseyre, B., Poilvé, H. & Roger, J.-M. (2011). A technical opportunity index adapted to zone-specific management. Precision Agriculture, 12(1), 130-145.

Taylor, J.A., Tisseyre, B., Bramley, R.G.V. & Reid, A. (2005). A comparison of the spatial variability of vineyard yield in European and Australian production systems. In J.V. Stafford (ed). Precision Agriculture '05. Proceedings of 5th European Conference on Precision Agriculture, The Netherlands: Wageningen Academic Publishers.

Taylor, J.A., Sanchez, L., Sams, B., Haggerty, L., Jakubowski, R., Djafour, S., et al., (2016). Evaluation of a grape yield monitor for use mid-season and at harvest. Journal International des Sciences de la Vigne et du Vin, 50(2), 57-63.

Tisseyre, B. & McBratney (2008). A technical opportunity index based on mathematical morphology for site-specific management: An application to viticulture. Precision Agriculture, 9(1-2), 101–113.

Whelan, B., McBratney, A. & Viscarra Rossel, R. (1996). Spatial Prediction for Precision Agriculture. In P.C. Robert, R.H. Rust & W.E. Larson (eds). Proceedings of the 3rd International Conference on Precision Agriculture., Madison, WI, USA: CSSA/ASA/SSSA. pp. 331-342. 10.2134/1996.precisionagproc3.c37