Spatial Variability of Kiwifruit Quality in Orchards and Its Implications for Sampling and Mapping

James A. Taylor¹

Australian Center for Precision Agriculture, McMillan Building, A03, The University of Sydney, NSW, Australia, 2006

John-Paul Praat and A. Frank Bollen

Supply Chain Systems Research Group, Lincoln Ventures Ltd, Private Bag 3062, East Street, Ruakura Research Centre, Hamilton, New Zealand

Additional index words. average variograms, sampling scheme, *Actinidia deliciosa*

Abstract. Published statistics on the spatial variation of fruit quality observed in orchards has been rudimentary to date. Dry matter and fruit weight data were collected spatially within 11 kiwifruit (*Actinidia deliciosa* var. *deliciosa* ‘Hayward’) orchards in the Bay of Plenty, New Zealand, to characterize the variability in fruit quality in terms of nonspatial and spatial statistics. Fruit weight was statistically more variable and exhibited a stronger spatial structure than the dry matter data. Individual variograms were derived for each orchard and then all the data were collated into average variograms for both quality attributes. The average variogram parameters were used to determine the optimum spacing for grid sampling to achieve a desired level of confidence when interpolating the data. A grid spacing of 28 m appears suitable for mapping fruit quality, provided sufficient area exists to collect enough points to perform block kriging. Plots of individual orchard and average variograms, and a table of nonspatial and variogram statistics are presented as a reference for future work in this area.

Harvester-mounted yield sensors have prompted many investigations into the amount of spatial variability present in grain, cotton, viticulture, and some mechanically harvested horticultural crops (usually vegetable crops). Investigations into the spatial variability of hand-harvested crops, such as kiwifruit (*Actinidia deliciosa* var. *deliciosa* ‘Hayward’), are much less common because of the difficulty in gathering detailed site-specific harvest data (Schueller et al., 1999). In the kiwifruit literature, published information on within-orchard variability of kiwifruit quantity and quality has been limited to basic examples, nonspatial statistics, and maps (Gillgren, 2001; Praat and Bollen, 2007; Praat et al., 2003). In general, mean orchard (or orchard block) statistics are all the data that are presented to the grower (Praat et al., 2003). Information on the spatial variation of quality should be of interest to growers because it forms the basis of price premiums. In 2001, a harvesting protocol for kiwifruit was developed to semi-automate the collection of bin-specific quality data (Gillgren, 2001). The protocol uses barcodes to track

fruit from locales within the orchard through the supply chain and is summarized in “Materials and Methods.” In 2003 and 2004, the harvesting protocol was used to collect and assign quality data spatially to several orchards. This produced a more complete set of data on spatial variation that can be visually represented as maps (Fig. 1).

In addition to providing a visual representation of the variation, these spatial data can be modeled using average variograms (McBratney and Pringle, 1999) to provide some quantitative geostatistics. McBratney and Pringle (1999) have shown that average variograms provide several benefits. First, they provide a benchmark of the values for variogram parameters that are to be expected for a given property. Second, given no a priori information on variability in a production system, they provide a basis for determining sampling density, particularly for grid-based sampling.

An understanding of how crop production varies spatially within a production system has become increasingly significant with the advent of precision agriculture technologies and methodologies. Although these new technologies and methodologies have predominantly been applied to broad-acre cropping systems, experiences in viticulture (Proffitt et al., 2006) indicate that they are relevant to horticulture. Precision agriculture is a management philosophy that aims to make decisions on input or operational procedures at a scale below the field scale. It advocates a shift from uniform management across individual fields according to the field average to

differential management of each site within a field according to the site requirements.

When considering whether site-specific management decisions are feasible, then both the magnitude of variation (i.e., the amount of variation) and the spatial structure of variation (i.e., how the variation is arranged in space) are important. When the magnitude of variation is small, then uniform management is likely to be preferable. If the variation is randomly distributed in space, then it is difficult to adjust management given current technology. However, if the variation has a spatial structure, then management can be differentially controlled at a subfield level. The spatial structure of variation also provides information that can be used to determine the density of sampling required to achieve a predetermined level of confidence in interpolation for map generation.

The objective of this report is to provide some indication of the magnitude of variability and the spatial structure associated with the variability of two fruit quality parameters, dry matter (DM, measured as a percentage) and fruit weight (Ω , measured in grams) in kiwifruit orchards located around the Bay of Plenty, New Zealand. These two parameters, DM and Ω , were selected because they are important quality determinants for market segregation and both have good calibrations with the near-infrared (NIR) grading system used to measure quality. The soluble solids concentration (or sugar concentration, measured as a percentage) of the fruit is also an important quality attribute. However, the current NIR calibration does not produce reliable results, and the soluble solids concentration data were not considered in this study.

Materials and Methods

Eleven commercial orchards near Te Puke, Bay of Plenty, New Zealand, were used as study sites. Data were collected during the kiwifruit harvest from May to June. The majority of data were collected during the 2003 harvest; however, one orchard was monitored in 2002 and another was monitored in both 2003 and 2004. Each orchard is given an alpha indicator, with each year considered to be an independent observation; thus, there were 12 orchard “years” (orchards A–L). Some orchards consisted of a single block whereas others had multiple blocks under the same management regime that were aggregated to form a single orchard unit.

Each orchard block is divided into bays, which are the basic vine trellis units. Bays are typically 5 × 5 m. The center of each bay was geographically referenced and the bays labeled with a unique barcode. Harvest bins were also labeled with a unique barcode that was different to the bay barcodes. During harvest each bin was referenced to the bays from which fruit was picked into the bin using the barcodes. Typically, four to five bays are picked into a single harvest bin.

At the pack house the bin barcodes were referenced to the time that the fruit was put

Received for publication 31 Aug. 2006. Accepted for publication 9 Dec. 2006.

Supported by Zespri Innovation Company Ltd. We thank the management and staff at ELEOS pack house, and the local growers around Te Puke, New Zealand, who kindly provided access to their orchards and fruit. The assistance of Darryl Gillgren in data preparation is also acknowledged.

¹To whom reprint requests should be addressed; e-mail james.taylor@usyd.edu.au

Fig. 1. Example of a bay-based representation of dry matter production (measured as a percentage) across a kiwifruit orchard near Te Puke, New Zealand. Data were derived from analysis of the mean fruit quality within harvest bins and then transferred back to the bays from which the bin was harvested. Courtesy of Darryl Gillgren, Lincoln Ventures Ltd., New Zealand.

through the kiwifruit quality grader. The inline NIR Compac Kiwifruit Grader (Compac Sorting Equipment Ltd., Onehunga, NZ) measures the DM and Ω of each fruit. Individual kiwifruit data were aggregated to produce mean bin DM and Ω . The mean bin results were referenced back to the bays in the orchard from where the kiwifruit was picked using the barcode information. The centroid of the assigned bays was used to reference the data geographically. A full description of this approach is given in Gillgren (2001).

Nonspatial statistics. The orchard name and orchard level nonspatial descriptive statistics—mean (μ), variance (σ^2), coefficient of variation (CV), skewness (κ), and number of bins (n)—for each quality attribute in each orchard were calculated.

Spatial statistics. Individual variogram clouds for each orchard were plotted in Vesper (Minasny et al., 2005). Preliminary geostatistical analysis showed that the variogram cloud for the majority of data was best fit by a spherical model (Eq. [1]) according to the Akaike Information Criteria (Webster and McBratney, 1989).

The spherical model

$$\gamma(h) = \begin{cases} c_0 + c_1 \left\{ \left(\frac{3h}{2a} \right) - 0.5 \left(\frac{h}{a} \right)^3 \right\} & \text{for } h \leq a, \\ c_0 + c_1 & \text{for } h > a \end{cases} \quad [1]$$

where c_0 is the nugget variance, $c_0 + c_1$ is the sill, and a is the range.

Standardizing the variogram model makes comparison between orchards easier. Thus, for each orchard data the theoretical spherical variograms were calculated using the global variogram function in Vesper (Minasny et al., 2005) and the variogram parameters [nugget variance (c_0), sill ($c_0 + c_1$), and the range (a)] recorded with the nonspatial statistics.

The c_0 value estimates the amount of variance at a lag distance of 0 m and is a function of stochastic effects and measurement error. The c_1 value estimates the amount of autocorrelated variance in these data and contributes with c_0 to define the sill ($c_0 + c_1$) or the total amount of variance in these data. The range defines the distance over which data are autocorrelated (i.e., the distance at which the sill is reached).

The lags for each individual orchard variogram cloud were recorded, then were fourth root transformed to normalize the response (McBratney and Pringle, 1999). For each lag, the mean was taken across the orchards for each variable before the mean values were raised to the fourth power to put them back on scale. Although a spherical function best represented the raw lags from the individual orchards, an exponential variogram (Eq. [2]) was the best fit to the “average” lags. The “average” variogram parameters were recorded and used to calculate the Cambardella index (Cambardella et al., 1994) and mean correlation distance (MCD) (Han et al., 1994) according to Eq. [3] and Eq. [4].

The exponential model

$$\gamma(h) = c_0 + c_1 \left\{ 1 - \exp\left(-\frac{h}{r}\right) \right\} \quad [2]$$

In this model the range is denoted by r , and its relationship to a in the spherical model is defined as $a = 3r$.

The Cambardella index

$$\text{Cambardella Index} = \frac{c_0}{c_1 + c_0} \bullet 100 \quad [3]$$

where c_0 is the nugget, $c_0 + c_1$ is the sill, and less than 25 indicates a strong spatial dependency, 25 to 75 indicates a moderate spatial dependency, and more than 75 indicates a weak spatial dependency.

The mean correlation distance

$$\text{MCD (m)} = \frac{3}{8} \bullet \frac{c_1}{c_0 + c_1} a \quad [4]$$

where c_0 is the nugget, $c_0 + c_1$ is the sill, a is the range.

Both the Cambardella index and MCD provide some indication of the spatial structure in these data. The Cambardella index is a ratio between the nugget (c_0) and the sill ($c_0 + c_1$), and thus measures the amount of variance in these data that is autocorrelated and potentially manageable. Although no account is taken in the index of the range parameter, smaller values are indicative of a stronger spatial structure (Han et al., 1994). The MCD is an empirical index, calculated in meters, that was originally derived for soil properties. The MCD includes the range of the data, as well as the ratio between the nugget and sill, to provide an estimate of the distance over which these data are autocorrelated. The greater the MCD, the greater the spatial structure.

Optimizing the grid-sampling size. With a priori knowledge, randomized site-directed sampling schemes have been shown to be preferable to grid sampling (Pocknee et al., 1996). However, when a priori knowledge does not exist, grid sampling can be effective, especially when it is optimized using block-kriging and block-kriging parameters

Table 1. Nonspatial statistics and variogram parameters for dry matter and fruit weight for each orchard in the study.

Orchard	No. of bins (n)	Mean (μ)	Variance (σ^2)	Skewness	CV	Nugget variance (c_0)	Sill variance ($c_0 + c_1$)	Range (a) (m)
Dry matter (%)								
A	534	15.20	0.19	0.11	2.92	0.264	4.655	163.7
B	181	16.12	0.06	-0.02	1.56	0.045	0.083	183.5
C	91	15.93	0.06	-0.07	1.52	0.006	0.017	66.0
D	403	15.35	0.35	-0.47	3.86	0.242	1.915	237.4
E	116	13.85	0.15	-0.91	2.83	0.275	0.452	67.7
F	276	15.86	0.16	0.14	2.54	0.128	0.172	116.5
G	80	16.07	0.10	-0.53	2.00	0.035	0.091	46.2
H	247	16.09	0.11	-0.35	2.07	0.029	0.060	166.4
I	169	16.11	0.05	-0.30	1.46	0.022	0.047	84.4
J	639	16.45	0.10	0.22	1.87	0.013	0.064	192.1
K	155	14.58	0.25	-0.14	3.40	0.511	4.745	127.5
L	139	15.90	0.07	0.49	1.65	0.093	0.115	103.9
Fruit wt (g)								
A		116.42	35.28	-0.45	5.10	7.496	46.526	179.9
B		113.68	25.50	0.33	4.44	13.720	30.360	207.1
C		110.29	21.62	-0.12	4.21	3.935	26.905	81.9
D		112.84	24.40	-0.07	4.38	10.260	30.470	146.7
E		110.25	19.62	0.34	4.02	16.200	21.341	62.6
F		105.23	9.30	0.10	2.90	5.349	12.028	250.0
G		106.45	19.10	0.02	4.10	4.782	24.782	134.8
H		111.01	25.81	-0.05	4.58	6.154	39.714	250.0
I		112.94	50.69	-0.67	6.30	20.630	67.630	124.9
J		110.66	25.20	0.29	4.54	8.508	30.128	199.7
K		110.90	22.47	-0.92	4.27	12.710	25.750	18.0
L		96.95	16.89	0.46	4.24	2.704	31.704	250.0

CV, coefficient of variation.

(Burrough, 1991). Given a certain kriging block size and sampling grid size, McBratney and Pringle (1999) have shown that estimates of kriging variance (σ_k^2) and the $\approx 95\%$ confidence interval (CI) of kriging prediction (where $\approx 95\% \text{ CI} = 2 \cdot \sqrt{\sigma_k^2}$) can be made using the equations presented in Webster and Oliver (1990).

An average kriging variance has been used in this calculation per the method of McBratney and Pringle (1999) (i.e., an average kriging variance for the block in relation to the grid point was taken for 25 points around the node). The output was plotted as contour maps to display the interaction between varying block length and grid size on the average kriging variance. Only square blocks have been used in this analysis; thus, block size has been specified by block length (in meters).

Results and Discussion

Nonspatial statistics. For the two quality parameters, DM and Ω , the nonspatial descriptive statistics for each orchard are given in Table 1. The number of bins (n) recorded in each orchard gives an indication of the relative size of the area sampled. On average a bin covers $\approx 112 \text{ m}^2$ or 0.011 ha . The DM μ and σ^2 are fairly constant across the orchards. Harvest time is determined by DM, so it would be expected that there is little variation between orchards. The range in mean Ω across the orchards was large (19.5 g); however, the σ^2 and CV were fairly constant across the orchards, indicating that within each orchard a similar level of variability exists. No measurement of skewness fell outside the range $-1 < \kappa < 1$, and the majority of measurements (21 of 26) were in the range $-0.5 < \kappa < 0.5$, indicating normal distributions.

Average variograms. The variograms presented here (Fig. 2) are done so to provide an indication of the spatial structure of kiwifruit quality parameters and as a reference for any future spatial analyses.

The parameters for the average variograms, the average Cambardella index values and the average MCD are given in Table 2. Fruit weight exhibited the larger average range (238.4 m) and MCD (68.4 m), whereas DM had a smaller average range (166.5 m) and MCD (40.4 m). The Cambardella index confirms this, with Ω having a smaller score than DM. These results indicate that there is a stronger opportunity to manage Ω spatially within orchards. Again, there is a potential management bias in this result because of the use of the DM values in determining harvest date.

Determining an optimal grid-sampling scheme. Isoline plots of $\approx 95\%$ CIs for kriging predictions of DM and Ω are presented in Fig. 3. These were derived using the average variogram parameters presented in Table 2. The isoline plots allow practitioners to select a suitable kriging block length and grid size to achieve a desired level of confidence in any

Fig. 2. (A, B) Individual orchard and average variograms for dry matter (A) and fruit weight (B; measured in grams). Average variograms are in bold; individual orchards are labeled with their alpha indicator (Table 1).

Table 2. Average variogram parameters and spatial indices calculated from the average variogram parameters for dry matter and fruit weight.

	Dry matter (%)	Fruit wt (g)
c_0 (nugget) (units ²)	0.06	7.45
$c_0 + c_1$ (sill) (units ²)	0.17	32.08
r (m)	55.49	79.47
Effective range ² (m)	166.50	238.40
Cambardella index	35.30	23.20
Mean correlation Distance (m)	40.40	68.60

²Exponential models were fitted to the average variograms then the exponential range (r) was adjusted to the same scale as the spherical models ($a = 3r$).

interpolation and mapping of the grid survey data. Although encouraging users to select their own threshold, McBratney and Pringle (1999) suggest that the $\approx 95\%$ CI should be

$\approx 10\%$ of the mean value of the desirable attribute. We have chosen to use one-third of the mean SD of each variable from these 12 data sets because the kiwifruit industry relies

Fig. 3. (A, B) Isolines plots of $\approx 95\%$ confidence intervals of kriging predictions for dry matter (A) and fruit weight (B) with varying block lengths (measured in meters) and sampling grid spacing (measured in meters). Average variogram parameters (c_0 , $c_0 + c_1$, and r) are held constant at the values given in Table 2.

on measures of σ to determine price premiums. For DM and Ω , this equates to 0.12% and 1.61 g respectively.

When designing a single sampling scheme for multiple variables, such as DM and Ω , the grid size needs to be optimized across the variables. This may be possible by using a constant survey grid size, but varying the kriging block length between variables. From Fig. 3, an $\approx 95\%$ CI of 0.12 for DM can be achieved with a grid spacing of 28 m and block length of 24 m. The $\approx 95\%$ CI for Ω grid spacing can also be achieved at a 28-m grid spacing using a block length of ≈ 32 m. Thus, by varying the block length, the sampling of different variables can be done at concurrent locations on a 28-m grid while achieving the desired level of confidence for each variable.

A 28-m grid produces ≈ 12 samples/ha. In general, ≈ 100 samples are required to produce effective variograms (Webster and Oliver, 1992). Most of the orchard blocks used in this study are less than the 8 ha required on a 28-m grid to achieve enough samples for variogram estimation. Orchards, though, may be more than one block, and sampling may be done across several contiguous blocks within a maturity area. However, only blocks under a uniform management system and crop type should be joined together into larger sampling areas. In situations when sampling areas are less than 8 ha, the addition of nested transects (Pettitt and McBratney, 1993) emanating from randomly selected grid nodes in random directions can be used to increase the number of samples and to ensure good variogram estimation.

The nugget variance (c_0) in these data is calculated from a large number of fruit samples (≈ 1200 – 1500 fruit/bin), but probably contain a significant contribution from error/noise in the NIR grader sensor. For hand-sampled data sets, there will be variation in the samples according to position in the canopy and position in the cordon (Miles et al., 1996). The implication of this is that although a sampling grid can be recommended from the isoline plots, where the fruit are picked and how many are picked will determine how accurately quality is measured at a given point. The calculation of the number of fruit required at each sample point is beyond the scope of this data set and requires knowledge of the amount of variation observed within a vine and between adjacent vines. If this variance is known, then the number of fruit required to measure quality parameters within a certain CI or threshold level, such as the average nugget variance, can be determined. Research into this area is being conducted but is yet to be published (Prof. Ray Littler, University of Waikato, New Zealand, pers. comm., May, 2006).

After samples have been taken, a variogram model can be derived from the data to estimate the nugget and sill variance. Given this information, the block kriging size can be adjusted in the interpolation process to achieve or approach an acceptable $\approx 95\%$ CI. Fig. 4 is an example of how this may be applied. It illustrates how the $\approx 95\%$ CI changes for both DM and Ω with varying block lengths and variogram sill values (given all other parameters are kept constant according to Table 2 and a survey grid spacing of 28 m).

The analysis presented here has focused on the use of average variograms to determine a regular (square) sampling grid size. However, it may be preferable to use the variogram information to design an irregular sampling grid. Recent work (Marchant and Lark, 2007) has demonstrated the use of spatial simulated annealing to design irregular sampling schemes. This process minimizes the total error incurred from both variogram estimation and the kriging process over a distribution of variograms for alternative sampling schemes. The average and individual variograms presented (Fig. 2) provide information on the plausible distribution of variograms over which such simulations should occur.

Conclusion

Kiwifruit quality is variable within orchards, with different quality variables exhibiting different levels of nonspatial and spatial variation. Fruit weight appears to exhibit greater spatial variation than DM using the MCD and Cambardella index. This indicates that there may be more advantage to manage Ω spatially than DM. A collection of nonspatial statistics, variogram parameters, and theoretical variograms plots for 11 orchards were presented here as a reference for future use.

Fig. 4. (A, B) Isoline plots of $\approx 95\%$ confidence intervals for kriging predictions of dry matter (A) and fruit weight (B) with varying block lengths (measured in meters) and sill variances (measured in square units). The c_0 and r values are held constant (Table 2), and the (square) sampling grid size is constrained to 28 m.

Average variograms for each quality variable have been calculated from the individual variograms. These have been used to optimize grid-sampling schemes for the purposes of mapping given no a priori information. Contour plots to assist researchers and growers select a grid size with an appropriate $\approx 95\%$ CI have been presented for general use. Optimal grid size will depend on the level of confidence desired; however, a 28-m grid appears to be a good compromise across the two quality attributes presented. After the grid survey, the collected data can be used to

derive variograms for each quality attribute and, given all other parameters remain constant, the block size can be refined further according to the derived sill values.

If the actual variogram parameters for a quality attribute are known, these parameters should be used, in lieu of the average variogram parameters, to optimize the sampling grid size.

Literature Cited

Burrough, P.A. 1991. Sampling designs for quantifying map unit composition, p. 89–126. In:

- M.J. Mausbach and L. Wilding (eds.). Spatial variabilities of soils and landforms. Special publ. 28. Soil Sci. Soc. of America, Madison, Wisc.
- Cambardella, C.A., T.B. Moorman, J.M. Novak, T.B. Parkin, D.L. Karlen, R.F. Turco, and A.E. Konopka. 1994. Field-scale variability of soil properties in central Iowa soils. *Soil Sci. Soc. Amer. J.* 58:1501–1511.
- Gillgren, D. 2001. Finding the fruit: A spatial model to access variability within a kiwifruit block. *Proc. SIRC 2001, 13th Annu. Colloq. of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand*, 2–5 Dec. 2001.
- Han, S., J.W. Hummel, C.E. Goering, and M.D. Cahn. 1994. Cell size selection for site-specific crop management. *Trans. Amer. Soc. Agr. Eng.* 37:19–26.
- Marchant, B. and M. Lark. 2007. Optimized sample schemes for geostatistical surveys. *Math. Geol.* (In press).
- McBratney, A.B. and M.J. Pringle. 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agr.* 1:125–152.
- Miles, D.B., G.S. Smith, and S.A. Miller. 1996. Within plant sampling procedures: Fruit variation in kiwifruit vines. *Ann. Bot. (Lond.)* 78:289–294.
- Minasny, B., A.B. McBratney, and B.M. Whelan. 2005. VESPER version 1.62. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW. 2006. <www.usyd.edu.au/su/agric/acpa>.
- Pettitt, A.N. and A.B. McBratney. 1993. Sampling designs for estimating spatial variance components. *Appl. Stat.* 42:185–209.
- Pocknee, S., B. Boydell, H.M. Green, D.J. Waters, and C.K. Kvien. 1996. Directed soil sampling, p. 159–168. In: P.C. Robert, R.H. Rust, and W.E. Larson (eds.). *Proc. 3rd Intl. Conf. Precision Agr. ASA/CSSA/SSSA*, Madison, Wisc.
- Praat, J.-P. and A.F. Bollen. 2007. Characterising spatial variation in quality. In: *Proc. 6th Intl. Kiwifruit Symp., ACTA Horticulture*. (In press).
- Praat, J.-P., F. Bollen, D. Gillgren, J. Taylor, A. Mowat, and N. Amos. 2003. Using supply chain information: Mapping pipfruit and kiwifruit quality. *Acta. Hort. (ISHS)* 604:377–385.
- Proffitt, T., R. Bramley, D. Lamb, and E. Winter. 2006. Precision viticulture: A new era in vineyard management and wine production. *Winetitles Pty. Ltd., Ashford, South Australia*.
- Schueller, J.K., J.D. Whitney, T.A. Wheaton, W.M. Miller, and A.E. Turner. 1999. Low-cost automatic yield mapping in hand-harvested citrus. *Computers Electronics Agr.* 23:145–154.
- Webster, R. and A.B. McBratney. 1989. On the Akaike Information Criterion for choosing models for variograms of soil properties. *J. Soil Sci.* 40:493–496.
- Webster, R. and M.A. Oliver. 1990. Statistical methods in soil and land resource survey. Oxford University Press, New York.
- Webster, R. and M.A. Oliver. 1992. Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* 43:177–192.