

## Spatial variation in tree characteristics and yield in a pear orchard

Eileen M. Perry · Raymond J. Dezzani · Clark F. Seavert · Francis J. Pierce

Published online: 21 March 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** We examined the spatial structure of fruit yield, tree size, vigor, and soil properties for an established pear orchard using Moran's  $I$ , geographically weighted regression (GWR) and variogram analysis to determine potential scales of the factors affecting spatial variation. The spatial structure differed somewhat between the tree-based measurements (yield, size and vigor) and the soil properties. Yield, trunk cross-sectional area (TCSA) and normalized difference vegetation index (NDVI, used as a surrogate for vigor) were strongly spatially clustered as indicated by the global Moran's  $I$  for these measurements. The autocorrelation between trees (determined by applying a localized Moran's  $I$ ) was greater in some areas than others, suggesting possible management by zones. The variogram ranges for TCSA and yield were 30–45 m, respectively, but large nugget variances indicated considerable variability from tree to tree. The variogram ranges of NDVI varied from about 14–27 m. The soil properties copper, iron, organic matter and total exchange capacity (TEC) were spatially structured, with longer variogram ranges than those of the tree characteristics: 31–95 m. Boron, pH and zinc were not spatially correlated. The GWR analyses supported the results from the other analyses indicating that assumptions of strict stationarity might be violated, so regression models fitted to the entire dataset might not be fitted optimally to spatial clusters of the data.

**Keywords** Tree fruit · Moran's  $I$  · Geographically weighted regression · Spatial variability · Management zones

---

E. M. Perry (✉) · F. J. Pierce  
Department of Primary Industries, PB260, Horsham, VIC 3402, Australia  
e-mail: eileen.perry@dpi.vic.gov.au

R. J. Dezzani  
Geography Department, University of Idaho, Moscow, ID, USA

C. F. Seavert  
Oregon State University Mid-Columbia Agricultural Research and Extension Center,  
Hood River, OR, USA

## Introduction

Site-specific management (SSM) has been developed and applied to annual cropping systems since the late 1980s, and there is an abundance of literature about the impact of soil, microclimate and other landscape-scale influences on crop production. In a review of SSM and related technologies, Plant (2001) stated that “Three criteria that must be satisfied in order for SSM to be justified are, (1) that significant within-field spatial variability exists in factors that influence crop yield, (2) that, causes of this variability can be identified and measured, and (3) that, the information from these measurements can be used to modify crop management practices to increase profit or decrease environmental impact.” In relation to the first two criteria, much of the published work to date has focused on annual cropping systems with mechanical harvesting technology to provide spatial data on yield. The research on perennial crops generally lacks discussion about the effect of variation in the individual plant, on yield or on the delineation of management zones; the latter are based on soil properties and landscape scale factors.

The research that has been done does indicate that perennial crop yields vary spatially. Whelan and McBratney (2001) included a comprehensive review of the available literature (Table 3), highlighting the spatial structure of the observed variability. The variogram analyses cited had ranges from 22 m (for soil moisture) to 180 m (for soil P). Pozdnyakova et al. (2005) measured and analyzed yields in cranberry bogs at three scales representing within- and between-field variability. The results of their variogram analysis indicated that at both spatial scales a spherical model provided a good fit with ranges of about 3.5 m within fields and 2300 m between fields. Zaman and Schumann (2006) showed that the normalized difference vegetation index (NDVI) and soil organic matter (SOM) could be used to delineate low, medium and high yielding management zones for variable-rate applications of soil amendments within a citrus grove. There was a strong soil influence on productivity as excessive Cu in the low SOM areas of the grove induced Fe deficiency, which was visible as chlorotic disorders of the foliage and stunted tree growth. They fitted spherical variogram models to NDVI, soil and leaf nutrient data with ranges of 234–255 m (the extent of the citrus grove was about 800 m × 800 m). The soil and plant variables recorded were significantly different in each of the delineated management zones. Similarly, Castrignanò et al. (2008) generated maps of the risk of soil salinization for a citrus growing area. The maps categorized the risk as high, medium and low salinity, based on soil properties and distance to the sea. Multi-collocated indicator cokriging and factorial kriging were used; the variogram models had ranges of 750–2000 m. Other recent research involving perennial crops certainly highlights the effect of spatial influences at landscape scales, for example in grapes (e.g. Lamb et al. 2004; Bramley and Hamilton 2004; Johnson et al. 2003) and citrus crops (Tumbo et al. 2002; Whitney et al. 2001a, b).

Tree fruit differs from most annual and some perennial crops in that most of the plants grown within an orchard block are cloned, suggesting that there should be minimal biological variability between them compared with annual crops. However, orchard management and site-specific soil and climate effects are superimposed on this homogeneity. Some management practices such as pruning, thinning and harvesting are done on a per tree basis, whereas other management practices, such as irrigation and fertilizer application are applied uniformly over larger areas, up to entire orchard blocks. Therefore we would expect to see both tree to tree variability, as well as landscape scale effects based on the soil, microclimate and management. Wang et al. (2006) recorded

yield, cost and profit data by tree over a 0.87 ha block of 272 pear trees. They used data envelope analysis, a non-parametric linear modeling method used in economics, to determine the revenue capacity efficiency for each tree, and to identify the limiting inputs (soil properties, micro- and macro-nutrients, or other environmental effects). Much of the soil data was aggregated, and spatial differences were compared by dividing the field into 14 blocks. However, even with this coarse blocking factor, spatial effects were evident.

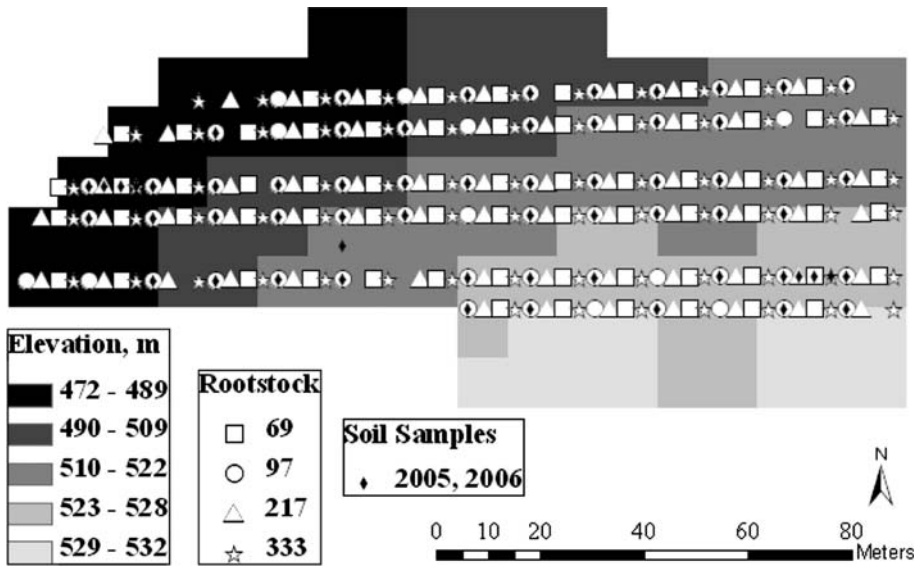
In this paper we used this same unique data set of individual tree yield and trunk size data as Wang et al. (2006), but also included higher spatial resolution soil data and remotely sensed data to address unanswered questions related to management zones for these tree fruit. We also applied spatial statistics that have not been used for precision agriculture, and that might be useful for justifying management zones. Specifically, we wanted to determine whether yield (per tree) and tree sizes were spatially correlated or unrelated over the field, and if correlated to determine the spatial structure. We used the autocorrelation statistic, Moran's  $I$  (described later), which was applied both globally and locally. We also examined whether spatial stationarity (Brunsdon et al. 1998) holds, i.e., whether linear regression models described for the entire orchard block hold equally well for any subset of the block. This has important implications for management zones: if these relationships do not hold, at what spatial scales do the models differ? To address these questions, we investigated spatial stationarity by applying geographically weighted regression (described later).

### Description of the orchard site

The research orchard used for this study is described in Wang et al. (2006). It is located near Hood River, Oregon, approximately longitude  $121^{\circ}41'$  west and latitude  $45^{\circ}41'$  north. Soil samples from the site indicate that the soil texture is a sandy loam with an average of  $16 \text{ cmol kg}^{-1}$  organic matter. The mean growing season (March–August) precipitation measured at the site weather station from 1971 to 2004 was 0.208 m (Oregon State University 2005), therefore, the tree fruit crops depend on additional water from irrigation. The research orchard is a pear (*Pyrus communis*) rootstock trial of d'Anjou variety planted on four different rootstocks (Fig. 1). The orchard is approximately 170 m west to east and 60 m north to south in extent. The tree spacing is approximately 3 m along the rows, which run west to east, and 6 m between rows.

### Tree and soil measurements

We focused on data that were recorded and related to individual trees. Trunk cross-sectional area (TCSA) was determined from circumference measurements made in 2003 just above the graft line of each tree. Yield data were collected in 2002 and 2003, expressed as total fruit picked (kg) per tree. Remote sensing was used to access crop vigor during 2002 and 2003. Space Imaging IKONOS imagery ([www.spaceimaging.com](http://www.spaceimaging.com)) were acquired in June, July and August of 2002, and Digital Globe Quickbird imagery ([www.digitalglobe.com](http://www.digitalglobe.com)) was acquired in June of 2003. The imagery was converted to



**Fig. 1** Layout of research orchard with soil sampling locations

reflectance values based on within scene targets and the normalized difference vegetation index (NDVI; Rouse et al. 1974) was determined as:

$$NDVI = \frac{R_{nIR} - R_{red}}{R_{nIR} + R_{red}}, \tag{1}$$

where  $R_{nIR}$  is the near infrared band and  $R_{red}$  is the red band of the imagery. The 0.6 m panchromatic band was registered to an orthorectified aerial image, and the tree crowns were delineated. The corresponding pixel for each tree was selected from the co-registered NDVI image. Soil samples were taken in 2005 and 2006 near a subset of the trees in the orchard block (Fig. 1). The soil was sampled at a depth of 0–0.015 m from 4 to 5 sites within a 1 m radius of each tree and mixed together to form a composite sample. Soil analysis of these samples was done by a commercial laboratory using standard techniques to determine TEC, pH, organic matter and crop nutrients. A subset of the soil samples was analyzed for soil texture (sand, silt and clay).

**Statistical methods**

For the analysis of spatial heterogeneity and stationarity, we used Moran’s *I* (Moran 1948), localized application of Moran’s *I* (Anselin 1995), geographically weighted regression (Fotheringham et al. 2002) and variogram analysis. We briefly introduce and describe our application of these statistical tools and then provide some explanation of our determination and relevance of spatial stationarity.

Moran’s *I* is an index of the degree of spatial autocorrelation, as described by Cliff and Ord (1981):

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n \mathbf{W}_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n \mathbf{W}_{ij}(x_i - \bar{x})^2}. \quad (2)$$

The global autocorrelation ( $I$ ) is computed for the variable of interest ( $X$ ) based on the variable mean ( $\bar{x}$ ) and contiguity matrix  $\mathbf{W}_{ij}$  for all  $n$  spatial units indexed by  $ij$ . The expected mean and variance for random distributions is determined as

$$E(I) = \frac{1}{(n-1)}. \quad (3)$$

$$E_N(I) = \frac{n^2 s_1 - n s_2 + 3 s_0^2}{s_0^2 (n^2 - 1)}, \quad (4)$$

where

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n \mathbf{W}_{ij}, \quad (5)$$

$$S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{W}_{ij} + \mathbf{W}_{ji})^2, \quad (6)$$

and

$$S_2 = \sum_{i=1}^n \left( \sum_{j=1}^n \mathbf{W}_{ij} + \sum_{j=1}^n \mathbf{W}_{ji} \right)^2. \quad (7)$$

Then the test statistic under a null hypothesis of complete spatial randomness is:

$$z = \frac{I - E(I)}{\sigma^2(I)}. \quad (8)$$

Positive values of  $I$  indicate spatial dependence among values. If  $Z \geq 1.96$  and  $I \approx 1.0$  then  $X$  is strongly spatially structured at the  $\alpha = 0.025$  level of confidence; if  $Z \leq -1.96$  and  $I \approx -1.0$  then  $X$  is uniform at  $\alpha = 0.025$ ; if  $-1.96 < Z < 1.96$  and  $I \approx 0$  then  $X$  is uncorrelated.

Anselin (1995) showed that Moran's  $I$  can be applied locally to evaluate the degree of autocorrelation for a given location, where the summation for  $ij$  is over a local neighborhood, resulting in a value  $I_i$  computed for each sample. The localized Moran's  $I$  provides a test of locational relevance of the global Moran's  $I$  and thus serves to provide locational information on deviations from an expectation of strict stationarity (Anselin 1995). As with the global Moran's  $I$ , large positive values indicate strong local autocorrelation. In addition, the expected value, variance,  $Z$  statistic and probability of  $Z$ ,  $p(Z)$ , were computed for each sample. In this paper, we have used the term spatial clustering to indicate neighborhoods of trees or soil samples displaying a high degree of spatial autocorrelation. The size of these neighborhoods was determined based on the dimensions of the kernel used, and the degree of clustering was measured by  $p(Z)$ . Both the global and localized Moran's  $I$  were computed using ArcGIS software (ESRI, Redlands, CA, USA).

Geographically weighted regression (GWR) provides a tool for identifying when the assumptions of strict stationarity do not hold, as well as a means of mitigating any violations using localized model fitting. For our applications, a local linear regression was

fitted to each tree or soil sample. Model parameters were determined based on the samples within the band width (kernel size) surrounding that tree, allowing the relationships being measured to vary over space:

$$y(\mathbf{g}) = \beta_0(\mathbf{g}) + \beta_1(\mathbf{g})x_1 + \beta_2(\mathbf{g})x_2 + \dots + \beta_n(\mathbf{g})x_n + \varepsilon(\mathbf{g}), \quad (9)$$

where  $\mathbf{g}$  refers to a location at which estimates of the GWR parameters are obtained and the  $\beta$  are the regression coefficients. The parameters are determined as:

$$\beta' = (X^T \mathbf{W}(\mathbf{g})X)^{-1} X^T \mathbf{W}(\mathbf{g})Y, \quad (10)$$

where  $\mathbf{W}(\mathbf{g})$  is a matrix of weights specific to location  $\mathbf{g}$  such that observations nearer to  $\mathbf{g}$  are given greater weight than observations further away. This local linear regression generates a coefficient of determination,  $R^2$ , and model parameters at each point (tree). Details of the method are described in Fotheringham et al. (2000, 2002). The GWR analysis was done with the GWR3 software (University of Newcastle upon Tyne, Newcastle upon Tyne, England). To determine the optimum band width for localized fitting, we minimized Akaike's information criterion (AIC; Akaike 1974), following Hurvitch et al. (1998) and Fotheringham et al. (2002):

$$AIC = 2k + n \left[ \ln \left( \frac{2\pi RSS}{n} \right) + 1 \right], \quad (11)$$

where  $k$  is the number of parameters (one for this research),  $n$  is the number of samples, and  $RSS$  is the residual sum of squares.

Fixed band widths starting with this optimum size were used and then increased until the results were not significant. Adaptive band widths were also used to determine the number of samples to include in a kernel, as an alternative to a fixed distance. We computed the significance for the spatial variation in the estimates of the model parameters using a Monte Carlo significance method (Hope 1968; Fotheringham et al. 2000). This significance tests the null hypothesis of spatial independence, that is, of the likelihood of the spatial model parameters given that the global regression model actually holds.

Variogram analysis (Deutsch and Journel 1998) was used to quantify the spatial scales of yield, TCSA and NDVI. GenStat (Payne et al. 2008) was used to compute the experimental variograms and to fit several models. Variograms of the soil properties were estimated by the residual maximum likelihood (REML) method using the MLREML program of Pardo-Igúzquiza (1997) as this method may require fewer data (e.g., Pardo-Igúzquiza 1998; Kerry and Oliver 2007) than the more usual method of moments variogram estimator. In this approach, linear combinations of the data (generalized increments) are used rather than the original data. The model parameters are calculated directly from the generalized increments of a covariance matrix of the full data. Note that for the REML method there is no experimental variogram, so only the fitted model and parameters are shown.

## Stationarity and isotropy

Several definitions of stationarity are used in this study and are either implemented as assumptions or tested for validity. Stationarity implies that a generating process is consistent over the study area for which observations are obtained, and it is essential for the creation of consistent point and interval estimators resulting from spatial processes.

Usually, a process is assumed to be stationary, but new methods have been developed to assess whether this is the case. The assumptions of stationarity and isotropy required for the statistical tools used in our study are summarized in Table 1.

Strict stationarity assumes that the distribution of the generating process is consistent over the study area and is invariant to changes in location (Schabenberger and Gotway 2005a). Strict stationarity implies that a single generating process and finite-dimensional single and joint distributions apply to all the data in the set of observations. The assumption can be evaluated by calculating Moran’s *I* function. In the computation of Moran’s *I*, only a single sample mean and variance are used as the observations are all assumed to be derived from the same generating distribution regardless of location. If strict stationarity is violated by a spatially dependent process that cannot be observed directly, then estimates based on the assumption might not be consistent or efficient estimators of the parameters of the generating distribution. For an estimator that is computed from spatially generated data, if the variable response is not stationary, the estimator cannot be guaranteed to be consistent or unbiased. Several recent studies highlight this issue for regression analysis in particular (Anselin and Kelejian 1997; Beran and Hall 1992; Lahiri et al. 2002).

The residual maximum likelihood (REML) variogram estimator requires a joint normal distribution and depends on the weaker assumptions of second-order stationarity where the mean,  $\mu = E[X(\mathbf{s})]$ , is constant for all  $\mathbf{s}$ , and the a priori variance of the process,  $\sigma^2 = E[\{X(\mathbf{s}) - \mu\}^2]$  is assumed to be finite and, as for the mean, the same everywhere. When two points  $\mathbf{s}_i$  and  $\mathbf{s}_j$  do not coincide, their covariance,  $C$ , depends on their separation and not on their absolute positions, and this applies to any pair of points separated by the lag  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$  (a vector in both distance and direction), so that

$$\begin{aligned}
 C(\mathbf{s}_i, \mathbf{s}_j) &= E[\{X(\mathbf{s}_i) - \mu\}\{X(\mathbf{s}_j) - \mu\}] \\
 &= E[\{X(\mathbf{s})\}\{X(\mathbf{s} + \mathbf{h})\} - \mu^2] \\
 &= C(\mathbf{h}),
 \end{aligned}
 \tag{12}$$

which is also constant for a given  $\mathbf{h}$ . This constancy of the first and second moments of the process constitutes second-order or weak stationarity.

Sometimes the variance appears to increase indefinitely as the extent of the area increases. The covariance cannot be defined then because we cannot insert a value for  $\mu$

**Table 1** Stationarity and isotropy assumptions for statistical methods used

Assumptions	Moran’s <i>I</i> (global spatial autocorrelation)	LISA and kriging (local spatial autocorrelation)	Residual maximum likelihood (REML) variogram estimator	Variogram (geostatistics)	GWR (local regression estimates)
Stationarity	Strict	Local	Second-order	Intrinsic and second-order	Local
Isotropy vs. Anisotropy	Isotropy	Isotropy; Anisotropy can be managed through the variogram model in kriging		Isotropy; anisotropy managed by subsetting data according to direction between pairs of observations	Depends on kernel range and weights

into Eq. 12. This is a departure from second-order stationarity. Matheron's (1963) solution to this is the weaker intrinsic hypothesis of geostatistics (Cressie 1993). Although the general mean might not be constant, it would be for small lag distances and so the expected differences would be zero as follows:

$$E[X(\mathbf{s}) - X(\mathbf{s} + \mathbf{h})] = 0, \quad (13)$$

and the expected squared differences for those lags define their variances

$$E = \left[ \{X(\mathbf{s}) - X(\mathbf{s} + \mathbf{h})\}^2 \right] = \text{var}[X(\mathbf{s}) - X(\mathbf{s} + \mathbf{h})] = 2\gamma(\mathbf{h}), \quad (14)$$

where  $\gamma(\mathbf{h})$  is the semivariance at lag  $\mathbf{h}$ . As for the covariance, the semivariance depends only on the lag and not on the absolute positions of the data. If the process  $X(\mathbf{s})$  is second-order stationary, the semivariance and covariance are equivalent. However, if the process is intrinsic only there is no equivalence because the stationary covariance function (Eq. 12) does not exist. The following equation gives Matheron's (1965) variogram estimator which is the method commonly used

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2(m(\mathbf{h}))} \sum_{i=1}^{m(\mathbf{h})} \left[ x(\mathbf{s}_i) - \left( x(\mathbf{s}_i + \mathbf{h}) \right)^2 \right], \quad (15)$$

where  $m(\mathbf{h})$  is the number of paired comparisons.

If a generating process is stationary for small subregions or geographical subsets and the generating process can be disaggregated into simpler stationary processes where the mean is consistent over short distances or for smaller subsets, then the process is said to exhibit local stationarity (Schabenberger and Gotway 2005b). The local indicators of spatial association (LISA) statistics assume local stationarity and estimate local differences using proximal dyadic pairs of observations. The LISA (e.g. localized Moran's  $I$ ) is a local disaggregation of the global Moran statistic that enables the local assessment of spatial dependence and the evaluation of globally or strictly non-stationary dependence. The GWR estimates also assume local stationarity as defined by weighted locally stationary processes using a variety of sampling kernels. Within an estimation kernel, observations are assumed to be generated from a locally stationary process. The methods developed by Fotheringham et al. (2000) can determine statistically significant kernel ranges or sizes within which stationarity is locally consistent. For kriging, local stationarity is also sufficient because it is a local estimator.

## Results and discussion

To address our first objective, that is, determining whether yield and tree size (TCSA) are spatially clustered or autocorrelated, we computed Moran's  $I$  for the entire orchard. Table 2 gives Moran's  $I$  for the yields of 2002 and 2003 both by rootstock, and based on all of the rootstocks combined. Note that the Moran's  $I$  values are positive, with large positive  $Z$  values that are significant ( $p < 0.01$ ). For comparison, Moran's  $I$  for rootstock (which we would expect to be non-clustered) is  $-0.047$  with a  $Z$  value of  $-7.18$ . This indicates that the yields for 2002 and 2003 are spatially correlated. Likewise, the Moran's  $I$  values (Table 2) for the tree trunk TCSA measurements also suggest spatial clustering of the tree characteristics, regardless of whether the trees were analyzed according to rootstock or when combined. In addition to TCSA, we analyzed NDVI as a surrogate measure of tree vigor. As with the yields



**Table 2** Global Moran's  $I$  for TCSA and yield by rootstock

$n$	278 TCSA (All)	69 TCSA (69)	66 TCSA (97)	68 TCSA (217)	72 TCSA (333)
Moran's $I$	0.04	0.04	0.09	0.10	0.05
$Z$	7.05	3.18	5.77	6.11	4.02
$p(z)$	<0.001	0.0025	<0.001	<0.001	<0.001
	Yield 2002 (All)	Yield 02 (69)	Yield 02 (97)	Yield 02 (217)	Yield 02 (333)
Moran's $I$	0.09	0.12	0.10	0.04	0.08
$Z$	14.90	7.52	6.67	3.20	5.33
$p(z)$	<0.001	<0.001	<0.001	0.0024	<0.001
	Yield 2003 (All)	Yield 03 (69)	Yield 03 (97)	Yield 03 (217)	Yield 03 (333)
Moran's $I$	0.09	0.07	0.11	0.07	0.10
$Z$	15.90	4.54	6.71	4.64	6.59
$p(z)$	<0.001	<0.001	<0.001	<0.001	<0.001

TCSA Trunk cross-sectional area

and TCSA, global Moran's  $I$  was computed for the available image dates when the trees were in full canopy cover in 2002 and 2003. The Moran's  $I$  and corresponding  $p$  ( $Z$ ) given in Table 3 indicate that NDVI is also spatially clustered or autocorrelated.

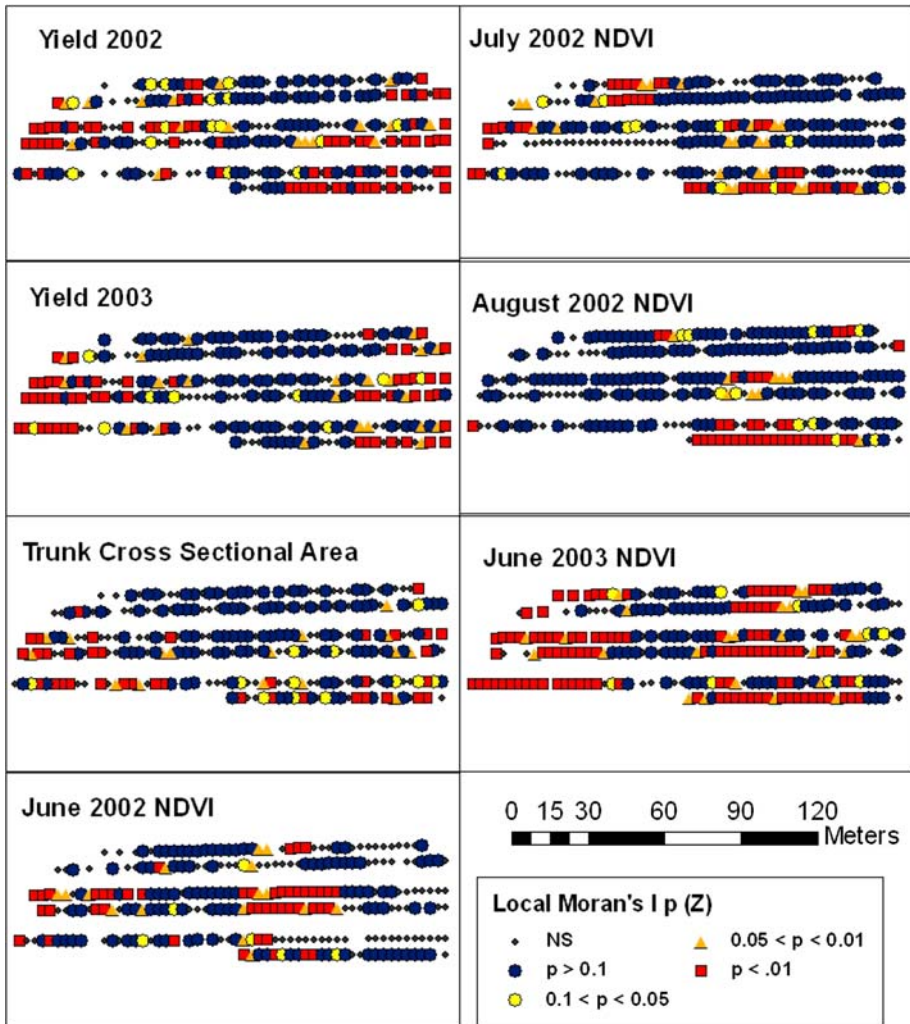
If these variables are spatially correlated; what are the scales of the spatial structures? We applied the localized Moran's  $I$  to assess the degree of autocorrelation for each tree relative to its neighbors. For this, the orchard was not subdivided on the basis of rootstock and a 30 m kernel was used, in order to include trees both within the row and across rows. The resulting maps of local Moran's  $I$  for yield and TCSA are shown in Fig. 2. Negative values of Moran's  $I$  are plotted as 'NS'. Positive values are plotted according to the probability of the corresponding  $Z$  value. The maps of yield and TCSA show similar patterns for the resulting probabilities, with more significant values of  $Z$  (larger positive values of Moran's  $I$ ) at both the eastern and western ends of the orchard block, indicating stronger similarities among adjacent trees in these areas. Likewise, localized Moran's  $I$  was computed for NDVI for 2002 and 2003 (Fig. 2). The resulting maps differ somewhat among image dates, but are similar to TCSA, especially for the June 2003 NDVI data. The combined results suggest that there are distinct zones that could be differentiated for management.

Given the evidence of spatial correlation and differences in neighborhood correlations, we wanted to address the question of strict spatial stationarity: do regression models described for the entire block hold for every subset of the block equally? We evaluated this

**Table 3** Global Moran's  $I$  for NDVI for all rootstocks

	June-02	July-02	August-02	June-03
$n$	278	278	278	278
Moran's $I$	0.04	0.09	0.10	0.04
$Z$	7.06	5.78	6.11	3.18
$p(z)$	<0.001	<0.001	<0.001	0.003

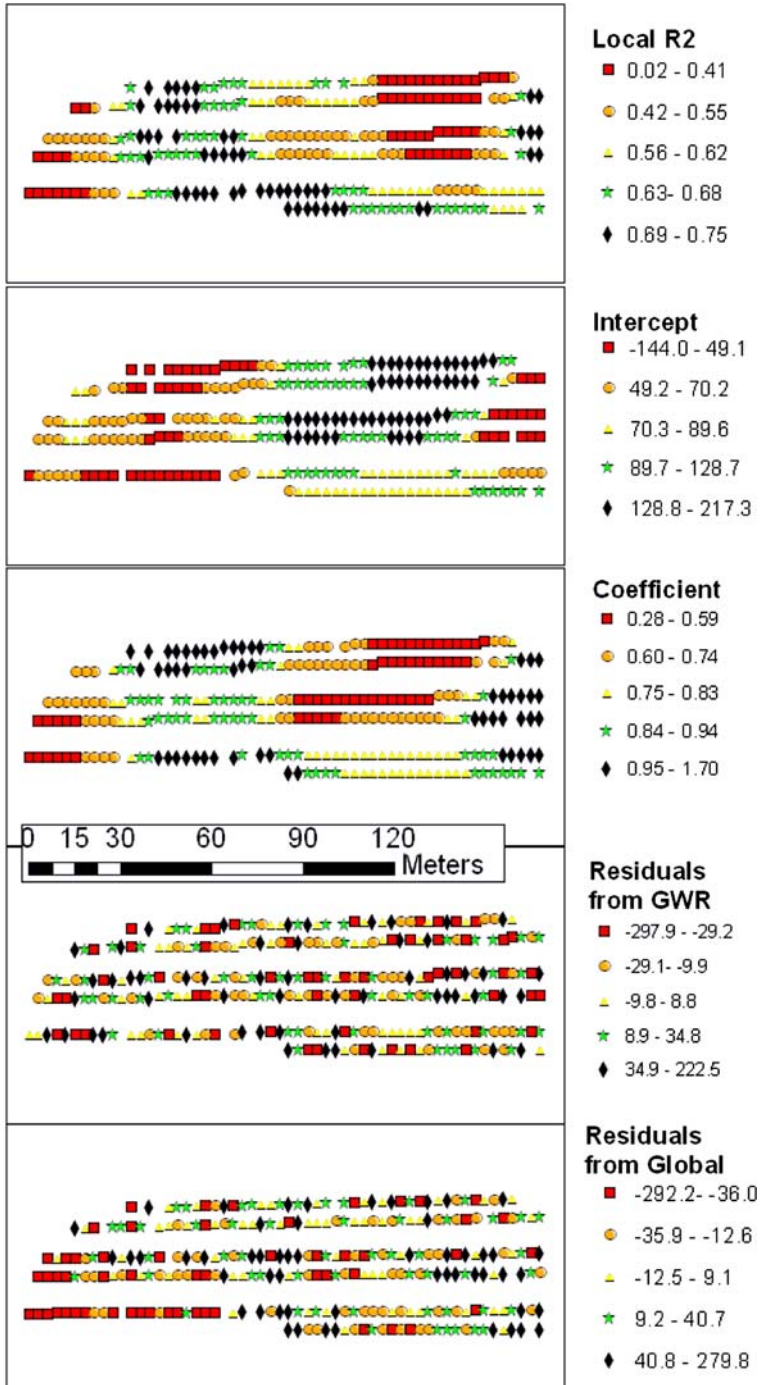
NDVI Normalized difference vegetation index



**Fig. 2** Local Moran's *I*, computed with a 30 m kernel (window) and using all rootstocks

using GWR, and selected models based on measurements that showed large correlation coefficients (not shown). These were the yields for 2003 and 2002, and also the yield for each year (as dependent variables) and TCSA. The GWR was also computed for several image dates for NDVI on TCSA, and for June 2003 NDVI on June 2002 NDVI.

The results of the GWR based on yield and TCSA were somewhat contradictory. The localized  $R^2$ , intercept and coefficients show obvious patterns (Fig. 3). Likewise, the local  $R^2$  values and decrease in AIC generally show an improvement over the global values (Tables 4, 5). The residuals from the GWR appeared random (Moran's *I* resulted in a Z score of  $-1.8$ , not shown) and slightly smaller than when using global regression as seen in Fig. 3d, which again indicates an improvement in the fit of the model. However, the spatial significance tests failed for the models at some kernel sizes greater than the initial one selected by the minimization of the AIC. For the fixed bandwidths, the tests resulted in



**Fig. 3** Local geographically weighted regression applied to the relationship of 2003 yield on 2002 yield

**Table 4** Comparison of global regression with GWR results

	<i>n</i>	Globally adjusted $R^2$	Local $R^2$		AIC	Bandwidth (m)	Monte Carlo spatial significance test, <i>p</i> value	
			Min, Max	% $R^2 >$ Global $R^2$			Intercept	Coefficient
<i>Yield 2003 on Yield 2002</i>								
Global parameters	278	0.47			3065.5			
GWR			0.02, 0.75	73	3013.3	9.5*	0.03	0.02
			0.47, 0.79	100	3023.0	20.0	0.03	0.14
			0.58, 0.76	100	3032.4	30.0	0.06	0.32
<i>Yield 2002 on TCSA</i>								
Global parameters	278	0.26			3024.2			
GWR			0.40, 0.95	100	2958.9	6.4*	0.17	0.25
<i>Yield 2003 on TCSA</i>								
Global parameters	278	0.41			3094.8			
GWR			0.50, 0.86	100	3048.0	10.0*	0.03	0.01
			0.52, 0.72	100	3052.9	20.0	0.15	0.11
			0.57, 0.71	100	3056.5	30.0	0.43	0.18

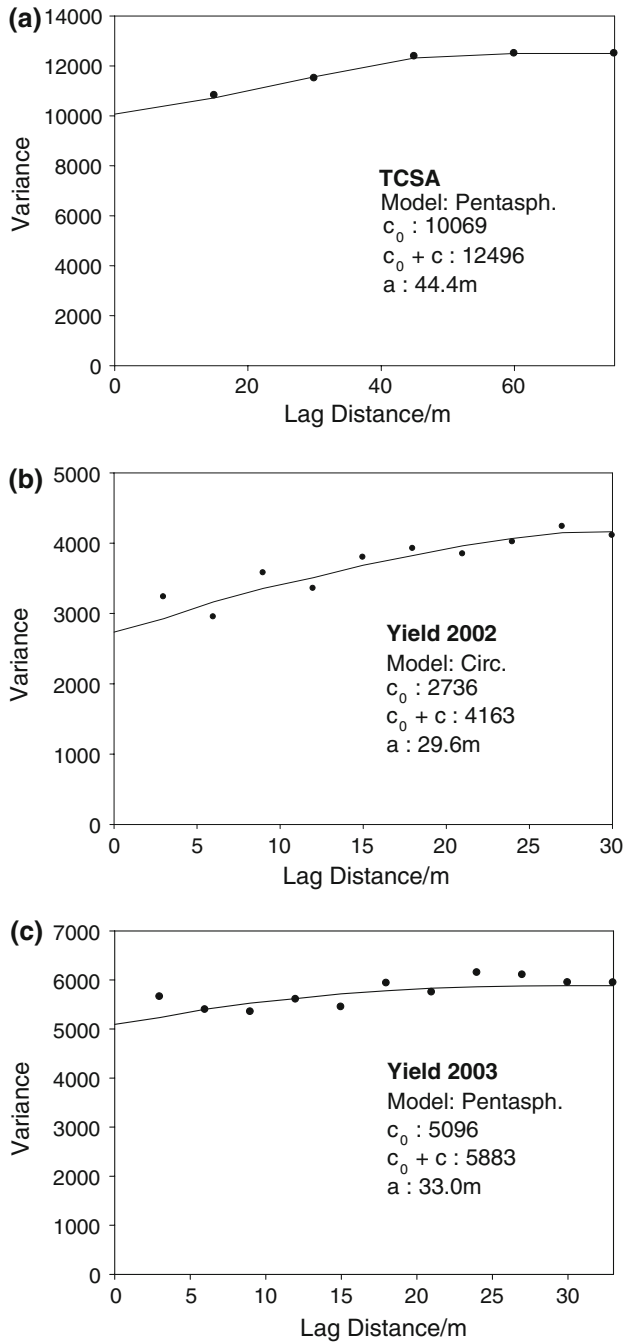
GWR is geographically weighted regression

\* Band width selected by AIC minimization

**Table 5** Comparison of global regression with GWR results using an adaptive kernel

	$n$	Global adjusted $R^2$	Local $R^2$		AIC	Bandwidth ( $n$ Samples)	Monte Carlo spatial significance test, $p$ value	
			Min, Max	% $n >$ Global $R^2$			Intercept	Coefficient
<i>Yield 2003 on Yield 2002</i>								
Global parameters	278	0.47			3065.5			
GWR			0.33, 0.84	94	3020.9	62*	0.07	0.06
<i>Yield 2002 on TCSA</i>								
Global parameters	278	0.25			3030.3			
GWR			0.33, 0.93	100	2965.4	26*	0.22	0.48
<i>Yield 2003 on TCSA</i>								
Global parameters	278	0.40			3097.0			
GWR			0.57, 0.73	100	3064.1	139*	0.16	0.26

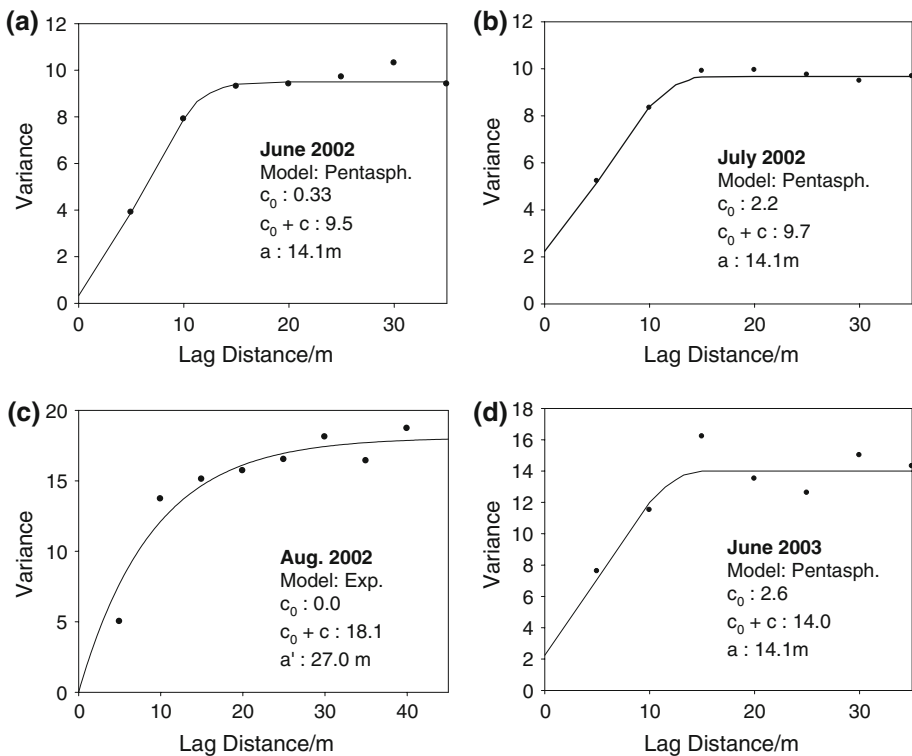
\* Band width selected by AIC minimization



**Fig. 4** Variograms of: **a** tree cross-sectional area (TCSA), **b** 2002 yield and **c** yield 2003. Variograms were computed for all trees ( $n = 278$ ) and all directions. TCSA was computed with a lag value of 15 m, whereas the yields for 2002 and 2003 were computed with lags of 3 m. The model functions are: Circ. is circular and Pentasph. is pentaspherical, and the parameters are:  $c_0$  is the nugget variance,  $c + c_0$  is the sill variance and  $a$  is the range of spatial dependence

$p > 0.1$  for the coefficient of 2002 yield at bandwidths of 20 m and 30 m, and for most of the 2002 and 2003 yield results on TCSA (Table 3). When using the bandwidth based on the number of samples (adaptive bandwidth), the spatial test was significant at  $p < 0.1$  only for the 2003 yield on 2002 yield, and  $p$  was  $>0.1$  for the other two models (Table 4). All of the regression models for GWR on NDVI failed the spatial significance test. Since the poor model fitting overall suggested that the models were not entirely explaining the variance in the dependent variables, other models were fitted with additional variables, such as soil characteristics and elevation, without any noticeable improvements (not shown).

The variogram analysis results for TCSA and yield for 2002 and 2003 are shown in Fig. 4. Variograms were computed based on individual rootstocks (not shown) as well as for all rootstocks combined, but were found to be very similar. Likewise, the variograms computed in all directions, across rows, along rows and in the direction of elevation change (130 degrees from grid north) were similar, therefore only the omni-directional results are shown (Fig. 4). After experimenting with different lag intervals for the variogram of TCSA, an interval of 15 m was used which corresponds with the distance between the same rootstock values. The lag value used for the 2002 and 2003 yields was 3 m, i.e., the distance between the trees. The variograms for TCSA and yield have ranges of 30–45 m and large nugget variances (Fig. 4). Variograms determined for NDVI recorded on several



**Fig. 5** Variograms of NDVI by date. Variograms were computed for all trees ( $N = 278$ ) over all directions, and using a lag value of 5 m. The model functions are: Pentasph. is pentaspherical and Exp. is exponential, and the parameters are:  $c_0$  is the nugget variance,  $c + c_0$  is the sill variance,  $a$  is the range of spatial dependence for the pentaspherical function and  $a'$  is an approximate range for the exponential function based on three times the model's distance parameter

dates showed little difference in structure by direction or rootstock (not shown). Figure 5 shows the variogram for each NDVI image date over all directions and all rootstocks. The variogram models for June and July of 2002, and June of 2003 indicate ranges of approximately 14 m, which represents a local neighborhood given the tree spacing (3 m between trees and 6 m between rows). The variogram for NDVI August 2002 has a range that is similar to that for yield 2002.

The TCSA, yield and NDVI are directly related to tree by tree variability, but how does their variation relate to the underlying spatial structure of the orchard's soil? As for the tree properties, we computed Global Moran's  $I$  and variograms for the soil properties. The results of these analyses varied according to the soil property or nutrient. Organic matter, TEC, Zn, Cu and Fe have significant values for the global Moran's  $I$  (Table 6). Variograms of the soil properties were estimated by the residual maximum likelihood (REML) method (Pardo-Igúzquiza 1997, 1998) as there were fewer soil data than tree and NVDVI data. Variograms were estimated successfully by REML for Fe, OM, Cu and TEC (Fig. 6); these have longer ranges (31–95 m) than those for the tree properties (14–44 m). Variograms could not be estimated for Zn, B and pH. These results suggest that Cu, Fe, Om and TEC are spatially structured.

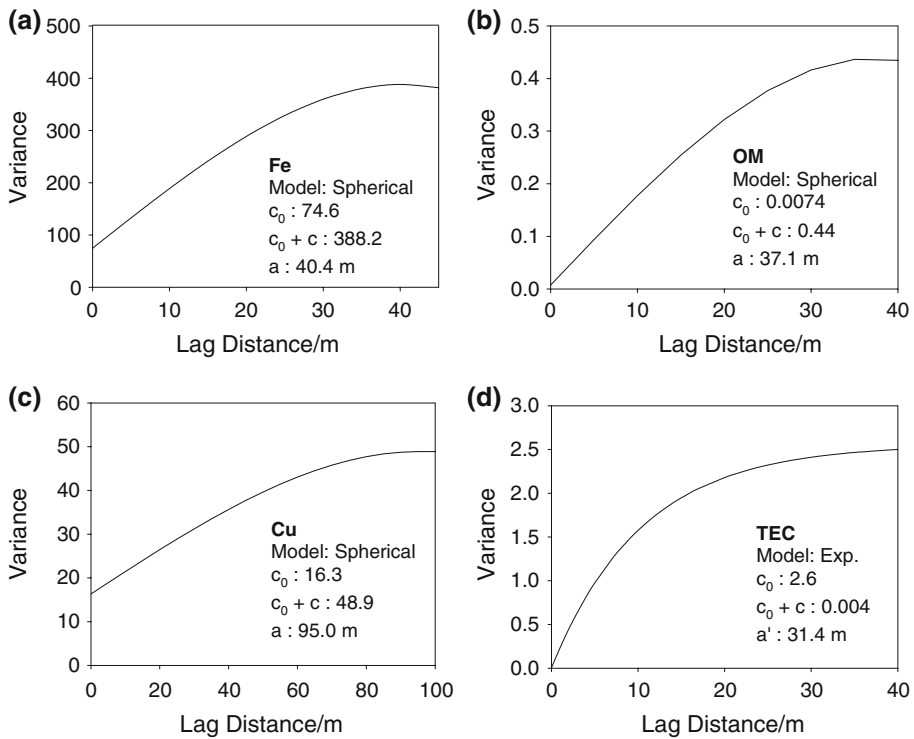
As an additional test of spatial stationarity, we applied GWR to model TEC on OM. Although we realized that this is not of particular interest as a model for prediction, the relationship has a large global  $R^2$ , and we used it to determine whether a localized model improved the fit, and at what scale. The soil data were interpolated using an inverse distance weighting quadratic function, which is localized and makes no assumptions of stationarity. The results are given in Table 7. Based on minimization of the AIC, a kernel of 13 samples was used. The localized  $R^2$  values did show improvement for most of the orchard block compared with the global regression, and the AIC has reduced significantly

**Table 6** Global Moran's  $I$  for soil

	TEC	pH	Organic matter	Ca	Mg
$n$	61	61	61	61	61
Moran's $I^a$	1.02	0.01	0.29	0.25	0.21
$Z$	20.08	0.47	5.84	5.01	4.42
$p(z)$	<0.001	0.36	<0.001	<0.001	<0.001
	Mn	Cu	Fe	P	B
$n$	61	61	61	61	61
Moran's $I^a$	0.26	0.61	0.17	0.45	0.05
$Z$	5.56	12.26	3.63	9.14	1.47
$p(z)$	<0.001	<0.001	<0.001	<0.001	0.14
	Zn	Sand	Silt	Clay	
$n$	61	61	61	61	
Moran's $I^a$	0.18	-0.04	0.10	0.05	
$Z$	3.97	0.12	2.60	1.74	
$p(z)$	<0.001	0.40	0.01	0.09	

<sup>a</sup> Note: these were computed with 30 m distance, inverse distance weighted, except 60 m fixed kernel used for soil texture





**Fig. 6** Variograms of soil properties estimated by residual maximum likelihood (REML), where Exp. is exponential, and the parameters are:  $c_0$  is the nugget variance,  $c + c_0$  is the sill variance and  $a'$  is an approximate range for the exponential function based on three times the model's distance parameter

**Table 7** Comparison of global regression with GWR results for total exchange capacity on organic matter

	Global $R^2$	Local $R^2$		$AIC$	Bandwidth ( $n$ samples)	Monte Carlo spatial significance test, $p$ value	
		Min, Max	% $n >$ Global			Intercept	Coefficient
Global parameters	0.71			524.8			
GWR		0.52, 0.99	98	100.3	13*	>0.001	>0.001

\* Band width selected by  $AIC$  minimization

from the global value. The spatial tests (Table 7) indicate significance at  $p < 0.0001$ . Overall, these results again suggest that the soil data are non-stationary in the strict sense.

The scales of the spatial structure differ somewhat between the tree-based measurements (yield, size and NDVI) and the soil properties. Variograms of yield and TCSEA have ranges of approximately 30–45 m. The large nugget variances for yield and tree size indicate considerable variability from tree to tree. Variograms of NDVI for several image acquisition dates indicate a short range of variation of approximately 14–27 m. Variograms of OM, Fe, TEC and Zn have a longer range of spatial dependence, from 31 m to 95 m.

In summary, analyses using both global and localized Moran's  $I$  suggest that yield, TCSA and NDVI are strongly spatially clustered or autocorrelated. The autocorrelation between trees (determined by a localized Moran's  $I$ ) is greater in some areas than others. Most soil variables also show significant spatial correlation in the global Moran's  $I$ ; the exceptions are pH, soil texture and B. The GWR analyses support the conclusions of non-stationarity, in the strict sense, of the data, but the GWR spatial significance tests failed at some scales, so we were unable to use these analyses to determine the spatial scales.

## Conclusions

So what do these results suggest for management of the orchard block? The spatial structure of clustering of tree TCSA, yield and NDVI, as shown by the Moran's  $I$  values, would suggest management zones on a scale of about 30 m. However, the large nugget variances and short ranges of variograms for many of the crop variables measured (e.g. 14–27 m for NDVI) favor management by tree, in spite of the longer scales of variation for some of the soil variables as determined by variogram analysis. The spatial non-stationarity, in the strict sense, in these data has implications in addition to the impacts on management zones. Regression models fitted to the entire dataset might be not be fitted optimally if there are spatial clusters in the data. Future assessment of perennial crop datasets such as this might include techniques such as autoregression to build spatial effects directly into the models. In addition, Markov random fields would enable further examination of specific spatial effects at different scales within a model.

**Acknowledgments** Funding for this research came from the USDA CSREES Initiative for Future Agriculture and Food Systems (IFAFS) Grant No. 2001-52103-11323. The satellite imagery utilized in this work was acquired through the NASA Scientific Data Purchase by staff at the NASA John C. Stennis Space Center Earth Science Applications Directorate. The authors are grateful to Dr. Frank Yin of the University of Tennessee, Department of Plant Sciences for the use of soils data from the research site. We are also indebted to Professor Margaret Oliver for the generous assistance in variogram analysis, and all of the reviewers who provided their perspectives on the work.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*, 93–115.
- Anselin, L., & Kelejian, H. H. (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review*, *20*, 153–182.
- Beran, R., & Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions. *Annals of Statistics*, *20*, 1970–1984.
- Bramley, R. G. V., & Hamilton, R. P. (2004). Understanding variability in winegrape production systems. 1. Within vineyard variation in yield over several vintages. *Australian Journal of Grape and Wine Research*, *10*, 32–45.
- Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression-modelling spatial non-stationarity. *The Statistician*, *47*, 431–443.
- Castrignanò, A., Buttafuoco, G., & Puddu, R. (2008). Multi-scale assessment of the risk of soil salinization in an area of south-eastern Sardinia (Italy). *Precision Agriculture*, *9*, 17–31.
- Cliff, A. D., & Ord, J. K. (1981). *Spatial properties: Models and applications* (pp. 19–21). London, England: Pion.
- Cressie, N. A. C. (1993). *Statistics for spatial data, revised edition*. New York: Wiley.
- Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide*. New York, USA: Oxford University Press.

- Fotheringham, S. A., Brunson, C., & Charlton, M. (2000). *Quantitative geography, perspectives on spatial data analysis*. London, England: Sage Publications.
- Fotheringham, S. A., Brunson, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, England: Wiley.
- Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society Series B*, 3, 582–598.
- Hurvitch, C. M., Siminoff, J. S., & Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B*, 60, 271–293.
- Johnson, L. F., Roczen, D. E., Youkhana, S. K., Nemani, R. R., & Bosch, D. F. (2003). Mapping vineyard leaf area with multispectral satellite imagery. *Computers and Electronics in Agriculture*, 38, 33–44.
- Kerry, R., & Oliver, M. A. (2007). Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*, 140, 383–396.
- Lahiri, S. N., Lee, Y., & Cressie, N. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference*, 103, 65–85.
- Lamb, D. W., Weedon, M. M., & Bramley, R. G. V. (2004). Using remote sensing to predict grape phenolics and colour at harvest in a Cabernet Sauvignon vineyard: Timing observations against vine phenology and optimising image resolution. *Australian Journal of Grape and Wine Research*, 10, 46–54.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58, 1246–1266.
- Matheron, G. (1965). *Les variables régionalisées et leur estimation*. Paris: Masson et Cie.
- Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society Series B*, 10, 243–251.
- Oregon State University. (2005) Archived climate data from Hood River Exp'l Station. <http://www.ocs.oregonstate.edu/index.html>
- Pardo-Igúzquiza, E. (1997). MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. *Computers and Geosciences*, 23, 153–162.
- Pardo-Igúzquiza, E. (1998). Inference of spatial indicator covariance parameters by maximum likelihood using MLREML. *Computers and Geosciences*, 24, 453–464.
- Payne, R. W., Murray, D. A., Harding, S. A., Baird, D. B., & Soutar, D. M. (2008). *GenStat for windows (11th Edition) introduction*. Hemel Hempstead, England: VSN International.
- Plant, R. E. (2001). Site-specific management: The application of information technology to crop production. *Computers and Electronics in Agriculture*, 30, 9–29.
- Pozdnyakova, L., Gimenez, D., & Oudemans, P. V. (2005). Spatial analysis of cranberry yield at three scales. *Agronomy Journal*, 97, 49–57.
- Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., & Harlan, J. C. (1974). Monitoring the vernal advancements and retrogradation of natural vegetation. Greenbelt, MD: NASA/GSFC final report, 371 pp.
- Schabenberger, O., & Gotway, C. A. (2005a). *Statistical methods for spatial data analysis. Chapter 2: Some theory on random fields* (p. 43). Boca Raton, FL: Chapman and Hall.
- Schabenberger, O., & Gotway, C. A. (2005b). *Statistical methods for spatial data analysis. Chapter 8: Non-stationary covariance* (pp. 425–426). Boca Raton, FL: Chapman and Hall.
- Tumbo, S. D., Whitney, J. D., Miller, W. M., & Wheaton, T. A. (2002). Development and testing of a citrus yield monitor. *Applied Engineering in Agriculture*, 18, 399–403.
- Wang, C., Färe, R., & Seavert, C. F. (2006). Revenue capacity efficiency of pear trees and its decomposition. *Journal of the American Society for Horticultural Science*, 131, 32–40.
- Whelan, B. M., & McBratney, A. B. (2001). The “null hypothesis” of precision agriculture management. *Precision Agriculture*, 2, 265–279.
- Whitney, J. D., Ling, Q., Wheaton, T. A., & Miller, W. M. (2001a). A DGPS yield monitoring system for Florida citrus. *Applied Engineering in Agriculture*, 17, 115–119.
- Whitney, J. D., Ling, Q., Wheaton, T. A., & Miller, W. M. (2001b). A citrus harvesting labor tracking and yield mapping system. *Applied Engineering in Agriculture*, 17, 121–125.
- Zaman, Q.-U., & Schumann, A. W. (2006). Nutrient management zones for citrus based on variation in soil properties and tree performance. *Precision Agriculture*, 7, 45–63.