

Developing a GIS-based Spatial Decision Support System for Automated Tree Crop Management to Optimize Irrigation Inputs

Aviva Peeters^{1,2}, Alon Ben-Gal², Amots Hetzroni³ and Manuela Zude⁴

¹*The Faculty of Food, Agriculture and Environmental Sciences, The Hebrew University of Jerusalem, Israel*

²*Institute of Soil, Water and Environmental Sciences, Agricultural Research Organization, Gilat Research Center, Israel*

³*Institute of Agricultural Engineering, Agricultural Research Organization, Volcani Center, Israel*

⁴*Leibniz Institute for Agricultural Engineering, Potsdam-Bornim, Germany*
apecters@bgu.ac.il

Abstract: In recent years there has been a growing interest in the application of geographic information systems (GIS) for the development of spatial decision support systems (SDSS). The current paper presents the development of a GIS-based SDSS for precision management of tree crops. The SDSS is being developed within the framework of the '3D-Mosaic' project which aims at optimizing water consumption in orchards by employing precision agriculture methods through the application of information and communication technology (ICT). Orchards in Turkey and in Germany are monitored for plant and fruit growth under different environmental conditions and irrigation regimes. An autonomous platform is employed for the acquisition of plant data using vision systems and laser scanners for yield and leaf area prediction, and stationary crop quality sensors with wireless data transmission to the platform. Data collected by these systems on fruit development, fruit quality and tree condition, combined with acquisition of abiotic data regarding soil, terrain and climatic conditions, are used to construct a 3D GIS database. The database is currently analyzed using spatial statistical methods to assess the spatial variability of the orchard and of its environmental conditions, to determine correlations between the two and to develop management zones. These will be used for building an SDSS for site-specific orchard irrigation to support farmers in efficient and sustainable management of orchards.

Keywords: GIS; Management Zones; Precision Agriculture; SDSS; Spatial Statistics.

1 INTRODUCTION

Ensuring food security within a changing global climate together with the growing concern in reducing the environmental footprint of farming while increasing the economic viability of agricultural practices has resulted, in the last few decades, in the development of precision agriculture. Research and practice in precision agriculture aim at sustainably optimizing the management of agricultural fields by addressing the spatial variability in plant and environment, such as soil properties. Remote sensing, information technology (IT) and geospatial methods are used to quantify spatial variability in agricultural fields (Corwin and Plant 2005a). This variability is used as the basis for developing management zones that will optimize inputs, such as irrigation and fertilization, while improving the quantity and quality of yields and assuring sustainable agricultural practices. Spatial analysis is considered the foundation of geographic information systems (GIS) and consists of the

methods and techniques for analyzing data in its spatial context (Longley et al. 2005). It provides also the framework for applying spatial statistics in which location, in addition to the attribute values, plays a significant role in the statistical analysis. Spatial statistical methods can be used to recognize and quantify patterns and relations which vary in space, i.e. quantify the statistical significance of a recognized pattern (Mitchell 2005). Spatial statistics are therefore suitable for analyzing and quantifying the spatial variability of various variables related to crops and to their environment. In addition, the strength of the relationships between independent (explanatory) variables and plant-related variables can be weighted. This is valuable for precision agriculture as it allows identifying site-specific management zones. Maps of these zones can be used as the basis for developing a spatial decision support system (SDSS) for sustainable orchard management.

An SDSS is a computer-based system designed to aid in solving complex semi-structured spatial problems of multiple criteria and typically consists of a GIS+DSS (Densham 1991). Using spatial and non-spatial data stored in geodatabases, an SDSS builds models to simulate dynamic spatial processes and evaluate the effect of different future scenarios.

An SDSS typically includes the following main components: (a) a database management system such as a GIS, (b) a set of potential analytical models used to simulate future scenarios and (c) a graphical user interface (GUI) which provides the user with a decision making environment to interact with the computer. In addition, an SDSS should support the input of spatial data, the representation of spatial relations, the application of spatial and statistical analysis and a variety of outputs such as maps and reports. The following presents the first steps in the development of a conceptual model for coping with spatial variability in agricultural tree crops through the use of spatial statistical methods.

A prototype of a GIS-integrated model to generate management zones and an SDSS for the use of stakeholders in optimizing tree crop management is under development. All analysis was performed using the ArcGIS[®] Desktop software package (ESRI 1984).

2 METHODOLOGY

The proposed model consists of: (a) identifying spatial field variability, (b) quantifying its significance, (c) recognizing the dominant independent variables which influence yield, (d) constructing zones of common yield-affecting properties and (e) developing an SDSS for producing site-specific management maps. The current paper presents the development of steps a-c. The remaining steps are still under development. In addition, a case study is presented to demonstrate preliminary results in mapping the spatial variability, which will form the basis of the SDSS.

2.1 Recognizing Spatial Variability and Quantifying its Significance

Spatial statistical analysis allows quantifying the distribution of features or of their attributes in space by evaluating whether or not and to what extent a pattern is clustered or dispersed. (Mitchell 2005). The spatial statistical test which was adopted in the current research for identifying the type of exhibited pattern compares, in general, the observed distribution to a hypothetical random distribution using the same data within the same study area. The extent of deviation from a random distribution indicates whether or not and to what extent the exhibited pattern is clustered or dispersed. The focus of the statistical analysis is on the spatial pattern formed by the attribute values of variables associated with the trees, the fruits or the environmental parameters. The objective of the statistical analysis is to quantify the probability that the distribution of values associated with the features is due to random chance. The *Getis-Ord General G Statistic* (Getis and Ord 1992) identifies whether concentrations of high or low values exist over an area i.e. whether hot-spots (clusters of high values) or cold-spots (clusters of low values) can be identified. The output of the *General G statistic* returns, in addition to the

observed General G, the Variance, the z-score and the p-value. The z-score and p-value are measures of statistical significance, and help to determine whether an observed pattern of clusters is statistically significant with regard to a confidence level. Once it is established that significant patterns exist in the data, the locations in the study area where those patterns are located are defined. The *Getis-Ord Gi** Statistic (Getis and Ord 1995), known also as hot-spot analysis, identifies where high values or low values are spatially clustered by comparing the values of neighboring features to each feature's value within a specified distance. The statistic indicates the degree to which each feature is surrounded by features with similarly high or low values. This statistic allows recognition of the location of the variability in space of clusters in the data values.

2.2 Developing Management Zones

Management zones allow site-specific agronomic strategies and practices within an agricultural field (Khosla et al. 2010). Recognizing relevant sub-field areas is challenging because of the variety of interacting variables that influence crop yields and the complexity in defining the most important variables. Various methods exist for dividing fields into potential management zones such as the map overlay approach, or different unsupervised clustering algorithms (Fraisie et al. 2001) for example the fuzzy clustering algorithm used by Fu et al. (2010). The main objective in developing management zones is to recognize natural clusters in the data that can define a unique zone. To recognize the most dominant variables which influence the yields, in the current research, spatial statistical methods are employed to identify significant correlations between plant-related variables and environmental variables and quantify the variability of these correlations in space. This variability will be used for defining the management zones.

For now two main questions have been addressed in the analysis: are there any statistically significant clusters in the data and where are those clusters located? Once the spatial variability has been quantitatively established, the next question is what is causing these clusters? To analyze the driving mechanisms behind the recognized clusters two statistical methods will be applied in which plant related data such as tree trunk size, fruit size and yields act as the dependent variables. Different combinations of multiple independent (explanatory) variables will be used to model and explore the spatial relationships between plant and abiotic (environmental) conditions. Using statistics to measure relationships allows to examine to what extent the value of the dependent variable changes when the value of the explanatory variable changes. As a starting point the *Ordinary Least Squares* (OLS) regression technique is used. OLS provides a global model using a single regression equation to model the process in the whole tested area. In addition to the simple linear regression method, the statistic outputs a set of diagnostic results which can indicate if a non-stationary process i.e. regional variation is taking place. If this is the case a *Geographically Weighted Regression* (GWR) should be applied. GWR (Fotheringham et al. 2002) as opposed to OLS provides a local model which fits a regression equation to every feature in the data and outputs a set of regression coefficients for each location (Lloyd 2010) and is therefore a truly spatial regression method. This method provides a powerful statistical analysis for modeling linear spatial relations which vary in space. In addition, GWR does not only analyze the correlation but also demonstrates the degree (strength) of correlation for each independent variable at each location and how this correlation changes in space. It is thereby possible to identify the most important variables which define the observed clusters and subsequently develop specific management zones.

2.3 Developing a GIS-based SDSS

To support the decision maker a formal model that can generate different scenarios is usually required. Various models have been designed for the development of SDSS, and consequently for the development of the SDSS, some of which have

been modified from DSS to account for the spatial domain (Sprague 1980). While DSS have been applied in agriculture for solving problems related to risk reduction and farm management, application of SDSS in agriculture is a rather recent development. SDSS are especially suitable for addressing agricultural problems as agriculture is inherently a spatial phenomenon in which location plays an important role in defining the spatial and temporal variability of the field. Lal et al. (1993), for example, used site-specific crop simulation models to predict yield in different soil and climatic conditions by incorporating the models in a GIS using a software interface. The current research proposes to employ the GWR method as the basis for the decision model. The method of GWR allows, in addition to modeling relationships, to predict the impact of future values under different scenarios and to build prediction models for developing the SDSS. Prediction explanatory variables such as irrigation inputs can be used, for example, to output different scenarios that can be used to develop site-specific maps of irrigation inputs. Figure 1 illustrates the schematic model of the proposed GIS-based SDSS. The GIS framework provides, in addition to the statistical analysis capabilities, the link to the geodatabase (input) consisting of data sources such as tables, maps and remotely-sensed data, and the link to a comprehensive presentation (output) of the information for decision makers. The entire process can be combined in a GIS-based web application for easy access of decision makers.

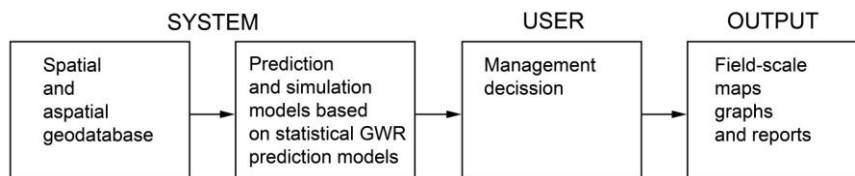


Figure 1. Schematic model of proposed GIS-based SDSS

3 APPLICATION OF MODEL AND PRELIMINARY RESULTS

3.1 Case Study

Two test sites - one in Turkey and one in Germany - are used as field trials for collecting data and for testing the adapted autonomous platform equipped with the plant sensors. The sites differ in terms of their climatic conditions (Mediterranean vs. North European), topography, soil and crop (grapefruit orchard in Turkey vs. plum orchard in Germany). The current paper presents only the test site in Turkey, for which more data, has presently been collected and processed. Part of the data, such as fruit yield still needs to be collected while other data such as the automated fruit recognition is still under processing. The test site in Germany is in full progress and will be used, once all data is collected, for verifying the model. The test site in Turkey is situated in the experimental farms of the University of Cukurova, in the vicinity of the city of Adana (Figure 2) and consists of a grapefruit orchard of approximately 4 ha with 621 trees (red polygon) from which a section of 9 rows of 23 trees has been monitored. Plant-related and abiotic data was collected by the various 3D-Mosaic teams in different time periods and consists of: tree location, trunk and fruit circumference, laser scanning (LiDAR) of tree canopies and ground surface, automated fruit recognition by the use of photogrammetric methods, fruit maturity assessment through non-invasive fruit sensors, leaf area index (LAI) and leaf water potential (LWP) measurements, soil water content, irrigation scheme, meteorological data and apparent soil electrical conductivity (ECa) measurements. The data was organized in a single GIS geodatabase using the ArcGIS® Desktop software package (ESRI 1984) to ensure topological relations, maintain the integrity of the data and assist in locating errors and inconsistencies among the layers.

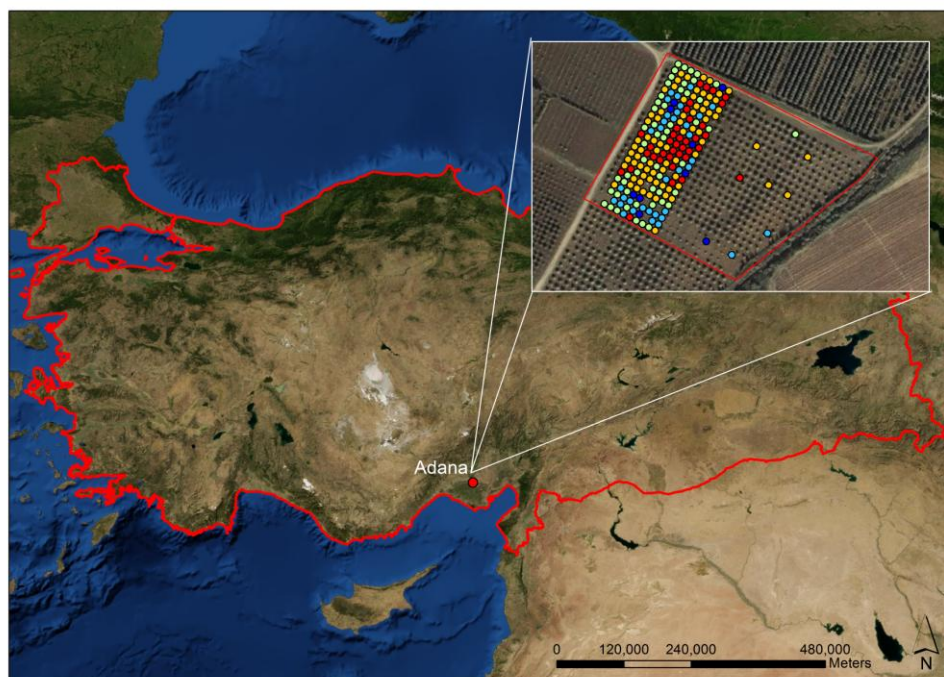


Figure 2. Location of test site in Adana. The inserted rectangle shows the orchard's borders highlighting the monitored trees which are color coded according to their trunk circumference size. Base image from ArcGIS® basemap imagery

3.2 Preliminary Results of the Spatial Statistic Analysis

The data was first analyzed to recognize global trends and patterns and then to identify correlations between abiotic (independent) and plant-related (dependent) variables using the spatial and non-spatial statistic methods described in section 3. The following presents the results of the tree-trunk size analysis in relation to the ECa. ECa is an accepted measurement for assessing the spatial characteristics of the soil such as its water content, salinity and texture. Existing researches correlating yield data and ECa have found mixed results and inconsistent correlation between crop yield and ECa. ECa point data was interpolated using an *Inverse Distance Weighted* (IDW) interpolator. In previous studies which compared IDW with various kriging (a geostatistical interpolator) approaches for interpolating soil properties such as ECa found that in some cases IDW performed better while in others kriging is better (Corwin and Lesch 2005b). The current research found that IDW performed best for this type of data as data sampling was dense with regard to local variation and evenly distributed. Validation results using the method of cross-validation resulted in a smaller prediction error and root mean square error for the IDW method.

While non-spatial methods indicated weak correlations between tree trunks and ECa (Figure 3), i.e. no significant global trends, spatial statistical methods suggested a statistically significant variability among the tree trunks. The z-score of 2.94 and p-value of 0.0032 indicate that the observed pattern of clusters of high values (large circumference values) is statistically significant with a 99% confidence level. Figure 4 illustrates the output of the *Getis-Ord Gi** statistic delineating the locations of clusters.

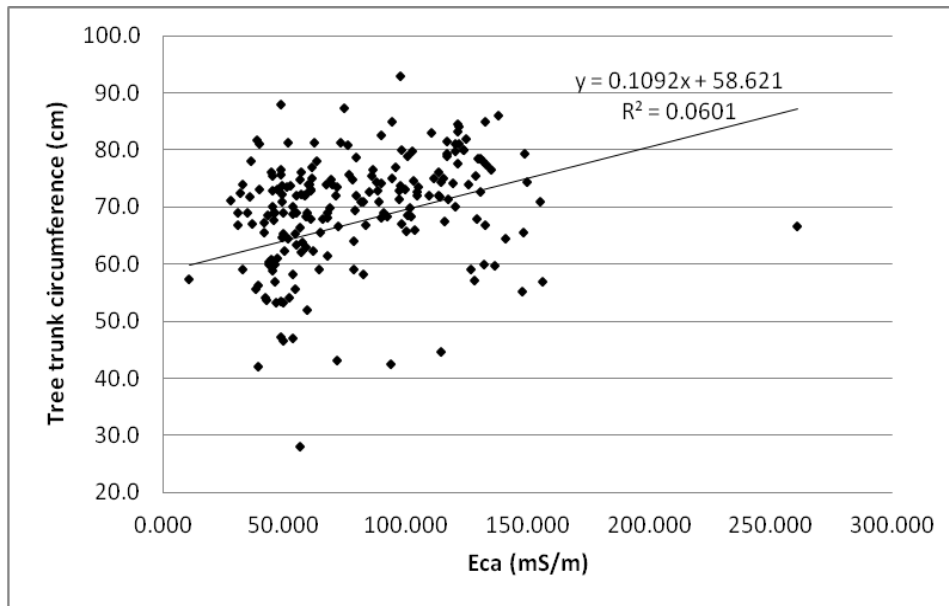


Figure 3. ECa values plotted against trunk circumference values. Linear regression results in very small R^2 values indicating that the variables, when using global regression, are only weakly related.

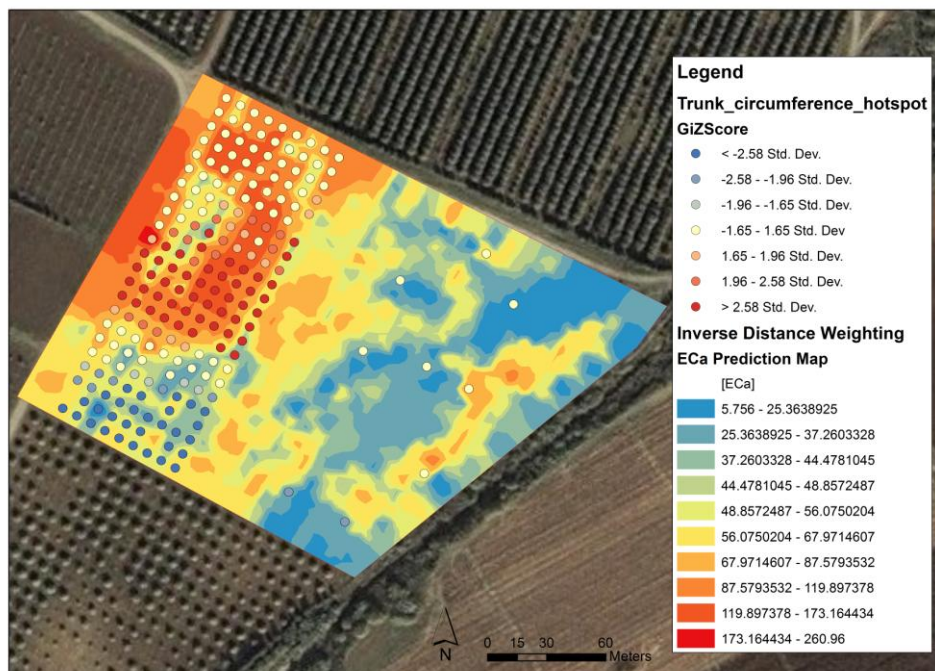


Figure 4. Identifying the location of clusters of high and low values using the Hot-Spot analysis (*Getis-Ord G_i^* statistic*) overlaid over an interpolated ECa surface. Dark red dots indicate statistically significant clusters of high values (large tree trunk circumference) while dark blue dots indicate statistically significant clusters of low values (small tree trunk circumference). Base image from ArcGIS® basemap imagery

Both OLS and GWR were then applied using tree trunk circumference as the dependent variable and ECa as the independent variable in order to analyze whether or not ECa is an important variable in defining the size of tree trunks and how this correlation varies in space. The results of the GWR indicate that there is

much variation in the strength of the relationship between ECa and trunk circumference. To evaluate the goodness of fit of the model locally the model outputs local R^2 values. Figure 5 demonstrates this spatial variation as local R^2 values vary from 0.00002 to 0.27. This indicates that by using ECa as an explanatory variable the model explains, at its best, only 27% of the tree trunk distribution. Therefore, additional explanatory variables should be combined to improve the model.

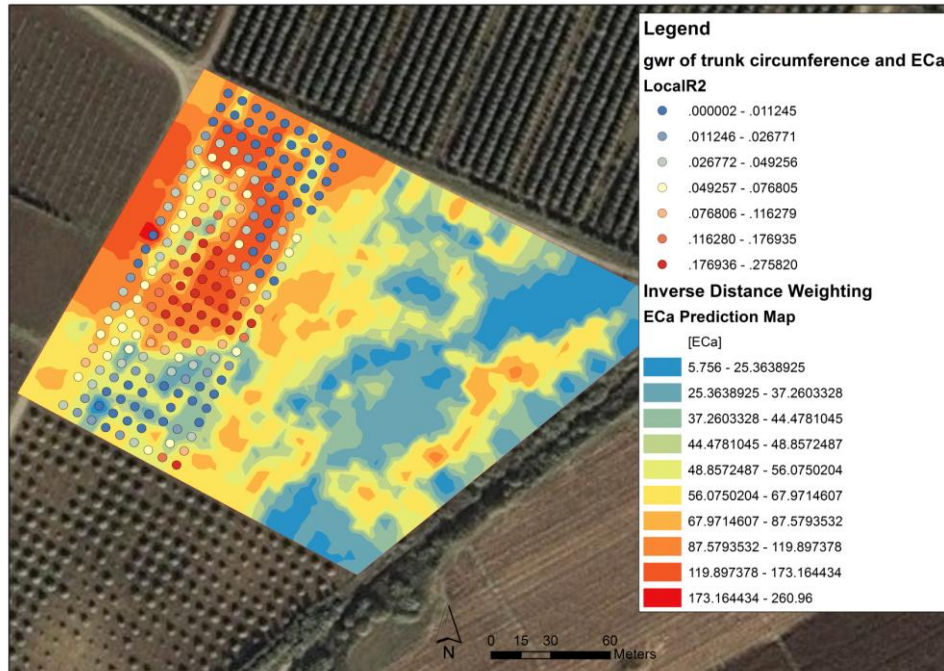


Figure 5. GWR applied to tree trunk circumference as the dependent variable and ECa as the explanatory variable. Base image from ArcGIS® basemap imagery

Once all variables; tree-related and abiotic variables, are analyzed and after recognized correlations are quantified, the most important variables can be identified based on the strength of the correlation and management zones can be defined.

4 SUMMARY

Efficient management of tree crops is essential for realizing sustainable agricultural production while maintaining food security and profitability. Precision agriculture enhances sustainable agricultural practices through site-specific crop management. However, recognizing the variables that are most important for defining the quality and quantity of the yield is complex due to the high spatial and temporal variability of agricultural fields. Although the research is only at the onset of its potential, preliminary results demonstrate the capabilities of a spatial statistical approach for quantifying plant-environment relationships that can be used for sustainable agricultural practices. For the first time, a wide range of tree-related and environmental variables will be considered in the definition of management zones. Developing one common GIS-based platform for both the zones definition and the development of the SDSS allows the entire process to be automated. In addition, a web-based GIS interface can be developed to be used as the decision making tool. To conclude, the proposed GIS-integrated model for defining management zones is unique for its spatial statistical approach which allows evaluating the influence and the spatial variability of a complex set of variables on yields, simulating management scenarios within the SDSS and consequently developing informed decisions that can be applied by decision makers.

ACKNOWLEDGMENTS

The presented research is part of the 3D-Mosaic funded by the European Commission's ERA-Net ICT-Agri project. Grateful thanks to Riza Kanber, Mustafa Unlu, Jana Beerbaum and Robin Gebbers for experimental management and data collection.

REFERENCES

- Corwin, D.L. and R.E. Plant, Applications of apparent soil electrical conductivity in precision agriculture, *Computers and Electronics in Agriculture*, 46(1-3), 1–10, 2005a.
- Corwin, D.L. and S.M. Lesch, Characterizing soil spatial variability with apparent soil electrical conductivity, I. Survey protocols, *Computers and Electronics in Agriculture*, 46(1-3), 103–133, 2005b.
- Densham, P.J. Spatial Decision Support Systems, In: Maguire, D.J., Goodchild, M.S. and D.W. Rhind (eds.), *Geographical Information Systems: Principles and Applications*, Longman, 403-412, London, 1991.
- ESRI, *ArcGIS Desktop 10.0*, Environmental Systems Research Institute, Inc., Redlands, 1984.
- Fotheringham, A.S., Brunsdon, C. and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons, Chichester, 2002.
- Fraisse, C.W., Sudduth, K.A. and N.R. Kitchen, Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity, *Trans. ASAE* 44(1), 155-166, 2001.
- Fu, Q., Wang, Z. and Q. Jiang, Delineating soil nutrient management zones based on fuzzy clustering optimized by PSO, *Mathematical and Computer Modelling*, 51(11-12), 1299–1305, 2010.
- Getis, A. and J.K. Ord, The analysis of spatial association by use of distance statistics, *Geographical Analysis* 24(3), 189-206, 1992.
- Getis, A. and J.K. Ord, Local spatial autocorrelation statistics: Distributional issues and an application, *Geographical Analysis* 27(4), 287-306, 1995.
- Khosla, R., Westfall, D.G., Reich, R.M., Mahal, J.S. and W.J. Gangloff, Spatial variation and site-specific management zones. In: Oliver, M.A. (ed.), *Geostatistical Applications for Precision Agriculture*, Springer, 195-219, Berlin, 2010.
- Lal, H., Hoogenboom, G., Calixte, J.P., Jones, J.W. and G.H. Beinroch, Using crop models and GIS for regional production analysis, *Trans. ASAE* 36(1), 177-184, 1993.
- Lloyd, C.D., *Spatial Data Analysis*, Oxford University Press, Oxford, 2010.
- Longley, P.A., Goodchild, M.F., Maguire, D.J. and D.W. Rhind, *Geographic Information Systems and Science*, John Wiley & Sons, West Sussex, 2005.
- Mitchell, A., *The ESRI Guide to GIS Analysis, Vol. 2: Spatial Measurements & Statistics*, ESRI Press, 238 pp., Redlands, 2005.
- Sprague, R.H., *A framework for the development of decision support systems*, Management Information Sciences 4, 1-26, 1980.