Using Spatial Analysis in a GIS Environment, for the Exploration of Soil and Corresponding Leaf Chemical Components. An Application in Apple Trees of Kastoria Region, Greece

Papadopoulou M.[1], Tziachris P.[2], Lafazani P.[1], Papadopoulos F.[2], Metaxa I.[2]

1 *Laboratory of Cadastre and GIS, AUTH, Greece*
2 *Soil Science Institute of Thessaloniki, Greece.*

Abstract: One of the most noted concepts in agriculture sector currently, which gains interest worldwide, is the term "precision agriculture". Precision agriculture is based on the exact knowledge of the field's location and what its physical and chemical properties. The great number of data ("big data") that are produced by precision agriculture, need specialized tools and techniques in order to expose patterns of use and useful results. The most important tool for using and analyzing spatial data is a Geographic Information System (GIS). Along with that, specialized spatial techniques are imperative to be used in the case of spatial data and the specialized agricultural scientific research. In this paper, different GIS techniques are applied, aggregated under the term "Spatial Analysis", in a real set of data concerning apple trees, from Kastoria, Greece. As it is presented, specialized spatial analysis reveals relationships, patterns and trends of the data, that help the agronomists to reach useful conclusions and results.

1. Introduction

There are many concerns around the world and especially in Greece about environmental issues that are focusing on the impacts coming from conventional agriculture. Thus, aspects of the traditional, up till now, practice have to go through serious revision. One of the most interesting new concepts that is gaining interest worldwide is precision agriculture, since it allows site specific management practices to be applied minimizing the effect of crop growing in the environment and people. One important component of precision agriculture is soil mapping so that it is known where the field is located and what its physical and chemical properties are (Cambardella and Karlen, 1999). The great number of data ("big data") that are produced by precision agriculture, need specialized tools and techniques in order to expose patterns of use and useful results. One of the most important tool for using and analyzing spatial data is a Geographic Information Systems (GIS).

A GIS software contains many techniques that combine location with descriptive information. All these techniques aggregate under the term "Spatial Analysis" (see e.g. Mitchell ,1999, Longley, 2005).

Spatial analysis can reveal relationships, patterns and trends of the data, that is, things, which most of the times are not perceived only by mapping the data, even if advanced methods of 2-D or 3-D visualization are used.

An important category of spatial procedures concerns the analyzing of distribution patterns. These procedures discover whether the phenomena develop concentrations (clusters) within certain regions, if outliers exist among the data values, if a certain phenomenon is autocorrelated, if there are clusters of high or low values in a region etc. All these analyses are statistical methods that use space and spatial relationships directly in their mathematical computations and are gathered under the term "spatial statistics".

In general in agriculture and more specifically in soil studies, the above spatial processes interest a lot, as it appears from relevant published work in journals and in books (see e.g. Goovaerts, 1998, Lee et al., 2006, Pierce and Clay, 2007, Lakhankar et al., 2010, Lin et al., 2010, Nayanaka et al., 2010, Huo et al., 2011)

In this paper, pattern analysis is used for the exploitation of the distribution of chemical values of the soil as well as corresponding values coming from apple tree leaves, in order the behavior of the different elements, important for the trees, to be studied. For the analysis, the ArcGIS 10.3 software package was used.

With the available data, different ways of using GIS in the agricultural sector is assessed, in a real use case in Kastoria , Greece. In more detail, the behavior of samples in relation to properties like soil texture, pH, $CaCO_3$ was evaluated and their correlation to macronutrients (e.g. P, K, Mg, Ca) and micronutrients (e.g. Fe, Zn, Mn, B) in space was studied. The form of the distribution of soil and leaf data was also studied, in order, areas with concentrations of high or low values of a chemical element to be found and comparisons of leaf growth to soil components to be made.

2. Some basic elements of pattern analysis

The analysis of patterns is used for the revelation of probably underlying spatial conditions that affect the spatial objects under study.

The analysis, is made through global and/or local calculations. The global calculations are useful when the data are complex and it is necessary overall trends to be described. The local calculations give as a result the extent and location of data clustering if there is any.

The z-score and p-value indicators give the statistical significance according to which the rejection of the null hypothesis is possible. In the analysis of patterns, the null hypothesis is randomness.

There are many locations where a data value, due to various reasons, is locally different from all its closest neighbours, or globally different from all the values of the data set. A reliable analysis of patterns presupposes the recognition and removal of such different values, when they are considered as global or local outliers.

Something to note, is that the results of patterns analysis directly depends on the scale of the study. Thus, if the analysis takes place for the entire region of a distribution, the results are different from those obtained when the study is elaborated in limited neighbourhoods of the data. The distributions in the whole of the study area can appear clustered, while in a subregion can appear dispersed.

In this paper the following techniques were implemented: for the study of spatial autocorrelation the global Moran's I index was calculated. For the discovery of global and local outliers the semmivariogramme cloud and the Voronoi polygon methods were used. Last, the hot spot analysis was implemented for the study of the clustering of the data. About the pattern analysis and techniques, there is a lot of material in Mitchell, 2005, De Smith et al., 2007, Fischer and Getis, 2009, Krivoruchko, 2011.

3. Data and study area

Digitized Soil Map provides a useful electronic database for the spatial representation of the soil variation of a region, based on in situ soil sampling, laboratory analysis, GIS techniques and plant nutrition modeling, coupled with the local land cadastre (see Papadopoulos et al., 2014a). An application of this tool is conducted in the regional unit of Kastoria (Figure 1) and a part of the data are used in this research (see Papadopoulos et al., 2014b).

During this soil survey, 1260 field parcels were sampled from 2011 to 2014. From these 1260 soil survey points, 328 were chosen for the current study (Figure 1).

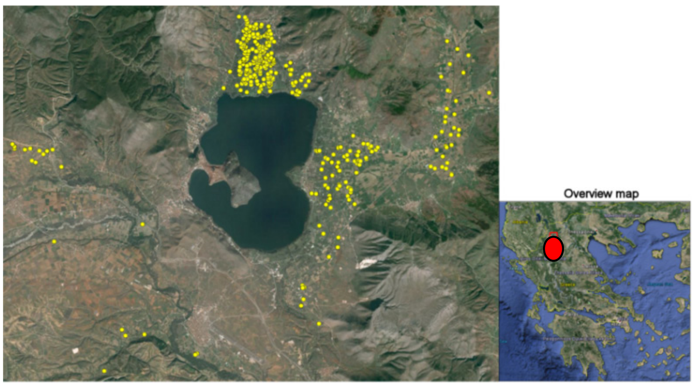

Figure 1 *The 328 survey points (in yellow colour) in Kastoria area, around Orestiada lake, in Northern Greece.*

These survey points were collected only from fields that grew apple trees and both soil and leaf data had been acquired, in order the impact of the soil parameters on the trees to be better researched and assessed.

A composite soil sample was taken up to a depth of 30 cm from each field parcel, which comprises of several sub-samples. Soil sampling took place during the period autumn-winter for three consecutive years (2010-2013). Soil samples were dried in the Soil Science Institute (SSI) laboratory and analyzed for soil texture, pH, $CaCO_3$, organic matter, EC, N, P, K, Mg, Ca, B, Mn, Fe, Cu and Zn. Apart from that, 328 leaf samples from the same fields were collected from 2011 to 2013. These leaves were also analyzed for N, P, K, Mg, Ca, B, Mn, Fe, Cu and Zn.

All the above analyses were carried out according to Methods of Soil Analysis (Sparks et al., 1996). Global Positioning System (GPS) receivers were used to identify the data locations in the National Geodetic Reference System (see Fotiou and Pikridas , 2006) at the time of sampling.

4. Spatial analysis of the soil and leaf data

4.1 Spatial autocorrelation-Moran's I

The examination of the autocorrelation of the elements took place through the the the Moran's I index. For the calculation of the index the inverse distance method was implemented. According to this, the impact of any point of the distribution to the point under investigation, was inversely proportional to its distance. The results are depicted in Table 1 and Table 2.

According to the table, C presents the largest autocorrelation globally with a z-score equal to 26.78 and zero p-value, along with the other soil texture parameters (S,SI) and pH (z-score > 13). This was something expected because these soil properties are quite stable over time and not at all or less dependent on fertilization practices.

The element P however, is the one that shows significant randomness in the distribution values as indicated by the very low z-score equal to 1.17 and the high p-value equal to 0.24083. P is one of the most cumbersome elements in the soil and consequently it would be expected to present the most possible uniformity (cluster form) in its spatial distribution. On the contrary, we see that P is presented with the higher randomness in space if compared with all the other elements, N included (z-score=3.58). A possible explanation for this fact has to do with the human intervention and specifically the way of application of P. When P is applied on the soil surface by hand, it is committed on the surface layer and does not go deeper to reach the root zone. As a result, P is not calculated in the soil analysis.

Table 1. *Moran's I spatial autocorrelation index for the soil data*

SOIL			
ELEMENT	MORANS' I	Z-SCORE	P-VALUE
S	0.16	17.70	0.000000
C	0.24	26.78	0.000000
SI	0.14	15.90	0.000000
pH	0.13	13.17	0.000000
EC	0.04	5.08	0.000000
OO	0.03	3.98	0.000069
$CaCO_3$	0.05	6.18	0.000000
N	0.03	3.58	0.000332
P	0.01	1.17	0.240830
K	0.05	6.82	0.000000
Ca	0.06	6.68	0.000000
Fe	0.05	6.06	0.000000
Zn	0.04	5.17	0.000000
Mn	0.09	10.53	0.000000
Cu	0.03	4.50	0.000000
B	0.02	2.70	0.006857
Mg	0.05	6.33	0.000000

Table 2. *The Moran's I spatial autocorrelation index for the leaf data of apple trees*

LEAVES			
ELEMENTS/PARAMETERS	MORANS' I	Z-SCORE	P-VALUE
N	0.01	1.40	0.150000
P	0.02	2.56	0.010000
K	0.06	6.83	0.000000
Ca	0.06	6.69	0.000000
Mg	0.13	15.3	0.000000
B	0.08	9.39	0.000000
Mn	0.04	4.36	0.000003
Zn	0.02	3.01	0.000001
Fe	0.04	4.32	0.000001
Cu	0.01	1.44	0.149000

In general, all Moran's I indicators are small but the null hypothesis is always rejected except for the P. The data tend to create concentrations and are spatially autocorrelated. All the rest soil elements have a cluster distribution in space with values of z-score to vary between 2.7 (for B) and 10.53 (for Mn).

The calculation of the index for the leaves gives the following Table 2. According to Table 2, two elements, Cu and N, show a randomness in the distribution values as indicated by the very low z-scores (1.44 for Cu and 1.40 for N), as well as by the high p-values (0.149 for Cu and 0.15 for N).

Cu in apple leaves shows the higher randomness in space when compared to all the other leaf elements. This is opposite to the findings in soil where the Cu showed a clustered form (z-score=4.5). The difference between soil and leaves could be the existing farmers' of making regular use of Cu formulas for the confrontation of fungal and bacterial offences, in the foliation part of trees, as well in the region of neck (Stilianidis et al., 2002).

Another element, which shows randomness in leaves is N. N is the most important nutrient for the plants and necessary in all fertilization schedules. However, an attention is needed regarding the soil fertilization units because an overdose might cause damages or degradation in cultivation. For this reason, some farmers prefer to limit the soil fertilization in winter and apply foliar sprays in summer in order the needs in N to be fullfiled. So, N is also dependent on farmers' practices as it happens with Cu.

It is also interesting that for P the null hypothesis is rejected but with a quite high p-value and low z-score. In general, the chemical elements of the leaves are spatially autocorrelated and tend to create concentrations. P in apple leaves shows a clustered distribution, although not so much statistically significant. It is possible that randomness of P in soil affects in some way the behavior of this element in the leaves, too. All the rest leaf elements have a cluster distribution in space with values of z-score to vary between 3.01 (for Zn) and 15.3 (for Mg).

In conclusion, the cluster form prevails in the spatial distribution of the elements in the soil as well as in the leaves. Randomness is rare and probably it has mainly to do with special farmers' practices.

4.2 Location of global and local outliers

4.2.1 Semivariogramme cloud

While the Moran's I index gives an overview of spatial autocorrelation in a distribution, through the semivariogramme cloud, a detailed search of autocorrelation in various locations of the distribution can be done. So it is possible to locate areas with high autocorrelation that are depicted by pairs at short distance locations with low γ value, or global outliers of the distribution values, depicted by pairs closely spaced but with high value of γ.

As an example, the semivariogram cloud of soil K is presented that shows three closely spaced clusters with high value of γ (Figure 2). These outliers signal points with major differences in values from the other points in the overall area, due to different reasons: errors in the specific samples in soil analysis, unusual agricultural intervention e.g. increased fertilization, unexpected high values that should be examined by the agronomists etc.

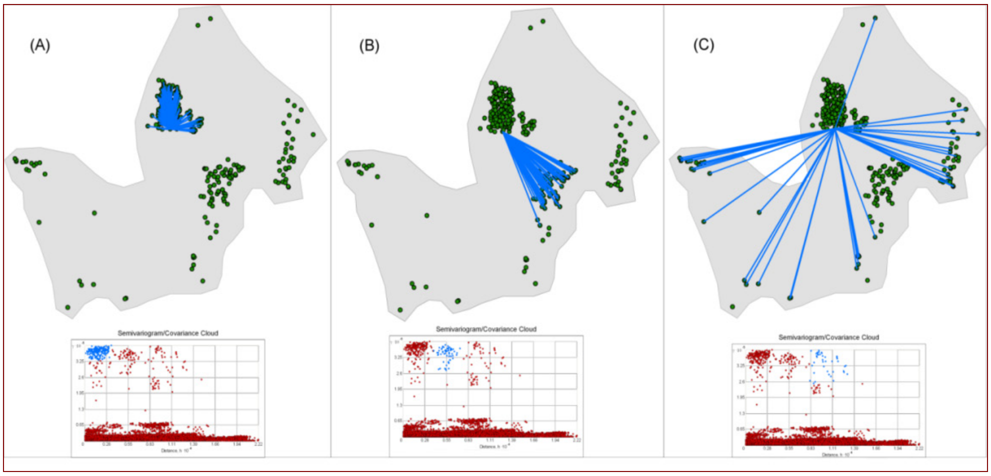

Figure 2. *(A) Global outlier that destroys autocorrelation in short distances, (B) global outlier that destroys autocorrelation in medium distances, (C) global outlies that destroys autocorrelation in longer distances. The examined parts of the semmivariogramme are in blue colour*

4.2.2 Use of Voronoi polygons

Zooming in smaller areas, the local outliers of the sample data could be further investigated, with the use of Voronoi polygons (for Voronoi polygons see Paraschakis et al, 1990). These outliers need special investigation which might need further field research, in order the agronomist to address them accordingly

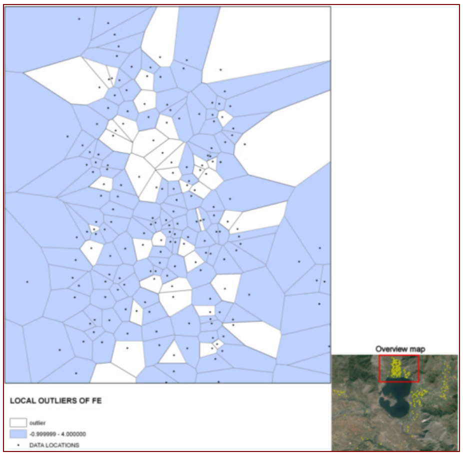

Figure 3. *Detection of local outliers of soil Fe (white areas) using Voronoi polygons in a neighbourhood of the study area.*

(e.g. to include or reject them from the samples). In Figure 3 an example of using this method for Fe is depicted.

4.3 Hot Spot analysis of the data

Searching for significant data clusters in the study area, hot spot analysis was used. All the maps below show, as an example, the most interesting results of the analysis concerning five of the chemical elements in soil and leaves.

For the calculation of a proper fixed distance band, Ripley's K function was used for ten distance bands around each point and the upper and lower limits of a confidence interval were calculated for nine permutations. The study area's borders were taken under consideration. None of bounding correction methods was used.

In the hot spot analysis the fixed distance band, which was recommended by the Ripley's K function, was used around each point in combination with the inverse distance method. In all the following Figures the red colour indicates high value clusters, the blue colour indicates low value clusters and the yellow indicates clusters that are not statistically significant.

4.3.1 Spatial clustering of K in soil and leaves

The hot spot analysis of K in the soil (Figure 4) shows three distinct areas: i) an area located to the north of Kastoria's lake with the lower concentrations of K (mean value = 69, blue points), ii) an area located to the east of Kastoria's lake, where the higher concentrations of K is found (mean value = 158, red points) and, iii) an area to the southeast of the lake without any obvious trend in concentrations of K (mean value = 90, light yellow points). One possible explanation for the lower concentrations of K in the first area could be the corresponding lower percentage of clay in the area. It is known that the heavier the soil is, the larger amounts of cations can be hold in its texture. Except of clay percentage, the farmer's practices play an important role in the distribution of K. Each farmer usually fertilizes its parcel field with the same manner over time independently of real needs. Thus, it was observed a correlation between farmers and concentration of K in the soil of their field.

The hot spot analysis of K in the leaves (Figure 5) shows two distinct areas: i) an area located to the north of Kastoria's lake with the higher concentrations of K (mean value = 0.76, red points), ii) an area located to the east and southeast of Kastoria's lake, where the lower concentrations of K are found (mean value = 0.53, blue points). It is seen that clusters of K between soil and leaves are opposite. A possible reason could be the fact, that farmers with the lower concentration of K in the soil of their fields were informed by the SSI about the poor condition of their soil and followed the proposed fertilization doses, which gave higher concentrations of K in leaves.

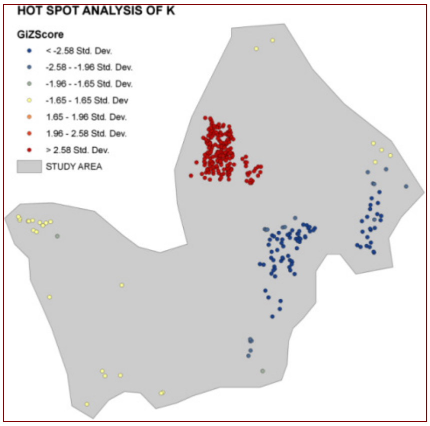

Figure 4. *Clustering of K in soil*

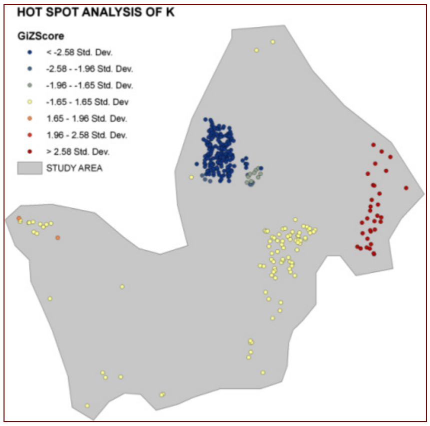

Figure 5. *Clustering of K in leaves*

4.3.2 Spatial clustering of P in soil and leaves

The hot spot analysis of P in soil (Figure 6) does not show any distinct area with lower or higher concentrations and is presented here as an example of no spatial clustering. This is in accordance with the findings of the pre-mentioned paragraph where P was found to be the element with the larger spatial randomness (z-score=1.17). Similar to soil analysis is the relative analysis of P in leaves (Figure 7), with only a minor hot area on the north (orange points).

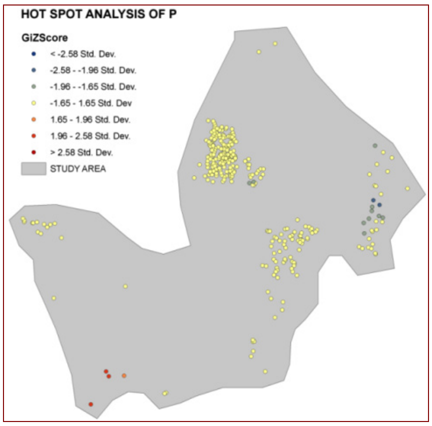

Figure 6. *Clustering of P in soil*

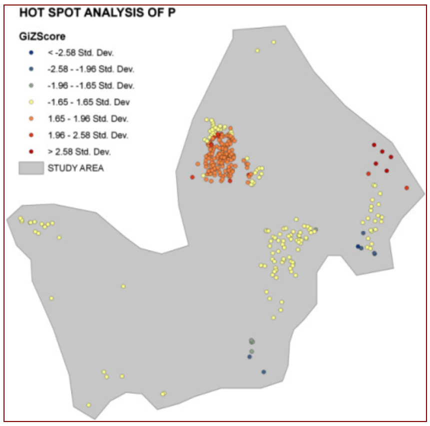

Figure 7. *Clustering of P in leaves*

4.3.3 Spatial clustering of Mg in soil and leaves

The hot spot analysis of Mg in soil (Figure 8) results in two areas: one with higher concentrations of the element (red points), located at the north to the Kastoria's lake, although with limited significance and the other in the rest of the study area without any cluster form (light yellow points).

On the contrary, the relative map of Mg in leaves (Figure 9) shows two distinct clusters, one in the north with blue points and one in the south with red points, which are located in opposite areas compared to soil clusters. This fact is difficult to be understood, however it should be noticed that the mean values of the clusters differ less than 20%. A possible explanation could the antagonism with other elements, such as potassium.

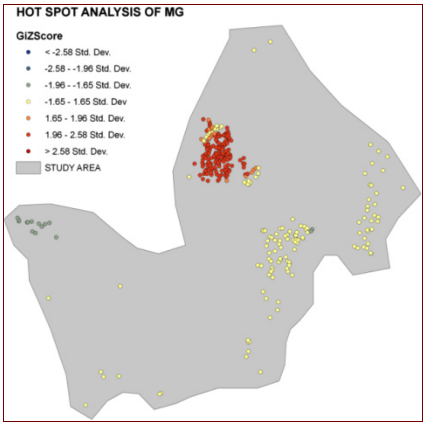

Figure 8. *Clustering of Mg in soil*

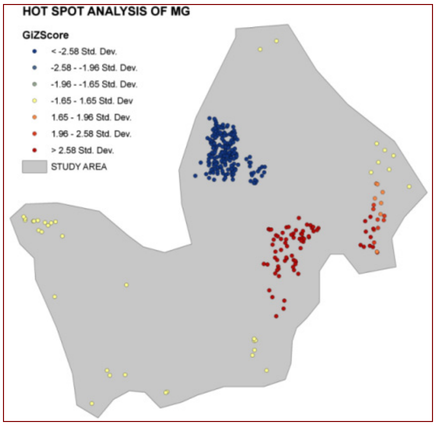

Figure 9. *Clustering of Mg in leaves*

4.3.4 Spatial clustering of Mn in soil and leaves

There are two main clusters formed by Mn in soil (Figure 10): the first one is located to the north of Kastoria's lake with the lower concentrations of Mn (mean value = 13, blue points), and the second to the southeast of the lake where the higher concentrations of Mn (mean value = 22, red points) are found.

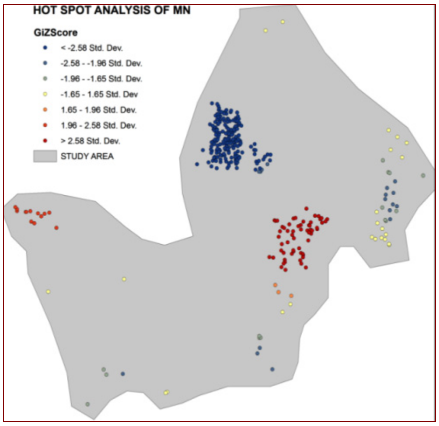

Figure 10. *Clustering of Mn in soil*

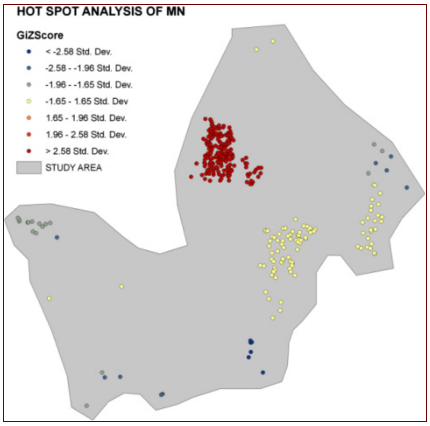

Figure 11. *Clustering of of Mn in leaves*

The corresponding clustering of Mn in leaves (Figure 11) gives one cluster with high values located to the north of Kastoria's lake (red points). The fact that the area with the lower concentrations of Mn in the soil turned into higher concentrations in the leaves can be explained by the proposed by SSI fertilization of the soil with Mn. Fertilizers passed from soil to the trees and increased the concentrations in the leaves.

4.3.5 Spatial clustering of B in soil and leaves

Two clusters are formed by B in soil (Figure 12): the first one is located to the north of Kastoria's lake with the higher values of B (mean value = 0.81), and the second to the east of the lake where the lower values of B (mean value = 0.55) are found.

The relative clustering of B in leaves (Figure 13) shows only one cluster form with the higher values located to the north of Kastoria's lake, too. Higher concentrations of B in soil and leaves were found in the same area, a fact that consists a normal situation.

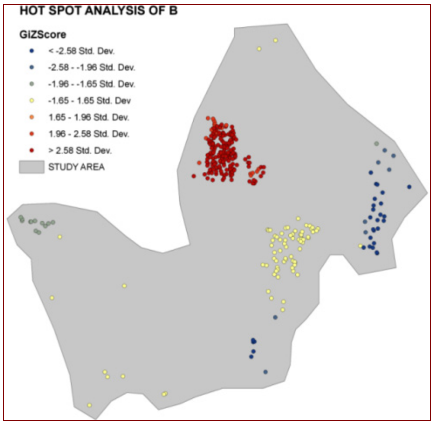

Figure 12. *Hot spot analysis of B in soil*

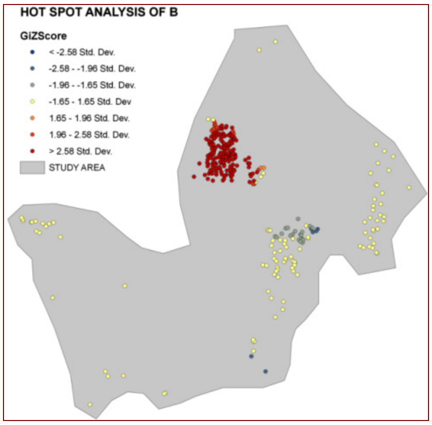

Figure 13. *Hot spot analysis of B in leaves*

4.3.6. Hot spot analysis at a large scale

As it is already mentioned, it is important to note that the results of patterns analysis directly depends on the scale of the study. The current hot spot analysis has taken place in the entire region, therefore it identifies statistically significant clusters of highs and lows for the entire study area. However focusing on a smaller area, hot spot analysis can lead to a different distribution of hot and cold spots and reveal different clusters.

As an example, focusing only on the north clustered area (red points) for the B in the leaves (Figure 13), and using again hot spot analysis only for these points, a different interesting pattern of hot and cold spots in the same area is revealed (Figure 14), which forms smaller clusters that should be further investigated by the agronomists.

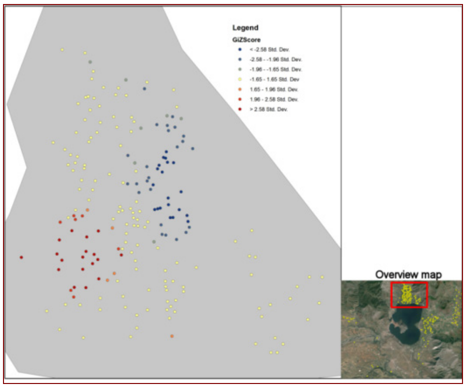

Figure 14. *Hot spot analysis of B in leaves at a larger scale for a smaller area*

5. Conclusions

As technology evolves, it becomes cheaper and easier to use. Devices and applications that collect data will multiply, along with the numbers of the data themselves, especially in the agriculture sector. Most of these data are related to a location and thereforethe agronomist needs specialized tools and techniques in order to decipher the behaviour and characteristics of the data. GISs are the specialized software that can employ the spatial parameter, and they should be combined with the right/appropriate techniques according/based to the nature of the phenomenon and the data that is to be analyzed. Some of these techniques and their results were presented in the current paper, in a real use case example regarding the agricultural sector, as guidance for future researchers, regarding spatial data in soil and leaves.

In the Kastoria area in Northern Greece, 328 soil and leaf samples were collected and analyzed in the soil laboratory of the Soil Science Institute of Thessaloniki. The data were studied with a GIS system, using different spatial analysis techniques.

In the beginning the spatial autocorrelation of the soil and leaf samples was estimated (Table 1, Table 2) as a measure of the degree to which a set of spatial objects and their corresponding values tend to create concentrations or tend to be dispersed. For this reason the global Moran's I index was used as an indicator that describe the autocorrelation throughout the study area, along with the z-score and the corresponding p-value.

Next, the examination for global and local outliers in the area was conducted with the use of the semivariogram cloud and the Voronoi polygons techniques. For example the K elements in the Kastoria area shows three closely spaced clusters with high value of γ in deferent distances, that present potential global outliers that should be further examined (Figure 2).

One interesting result for the study area was, that parameters in the soil that are quite stable over time and less depended on fertilization practices (soil texture parameters, pH), present larger autocorrelation globally. On the other hand the element P in the soil, shows significant randomness in the distribution values, with a possible explanation the human intervention and specifically the way of the application of phosphorus fertilization in the apple trees. Cu and N in the apple tree leaves present higher randomness among the different chemical parameters, That exists due to the farmers' practices to apply foliar sprays that increase these elements in the leaves.

Another spatial analysis that is presented, is the hot spot analysis that helps to identify statistically significant clusters of high or low values (Figures 4-14). The appropriate distance for the hot spot analysis was given by the Ripley's K function, which was calculated for the soil as well as the leaf samples. The analysis revealed areas of hot and cold spots that could help the agronomists to better assess the cultivation of the area. For example the clusters of the element K (a significant element for apple trees) in the soil could be the result of the differences in the soil texture of the area or/and differences in farmers' practices, something that needs further investigation. Also differences in the clusters between soil and leaves for the same area like in the elements K, Mg, Mn, or shift on the zooming scale reveals interesting patterns both for the overall study area and for smaller neighborhoods that should be further analyzed by agronomists.

Overall as it was presented in this paper, the optimum use of the spatial dimension for the agricultural data, in conjunction with the appropriate tools and techniques in a GIS environment, improves the decision support processes and enhance the quality and quantity of the agricultural production.

6. References

Cambardella, C. & Karlen, D., 1999, Spatial analysis of soil fertility parameters. *Precision Agriculture, 1*(1), 5-14.

De Smith, M. J., Goodchild, M. F., & Longley, P, 2007, *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*: Troubador Publishing Ltd.

Fischer, M. M. & Getis, A., 2009, *Handbook of applied spatial analysis: software tools, methods and applications*: Springer.

Fotiou, A. & Pikridas, C., 2006.,GPS and geodetic applications., *Ziti, Greece* (in Greek)

Goovaerts, P. (1998). Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology and Fertility of soils, 27*(4), 315-334.

Huo, X.-N., Zhang, W.-W., Sun, D.-F., Li, H., Zhou, L.-D., & Li, B.-G., 2011, Spatial pattern analysis of heavy metals in Beijing agricultural soils based on spatial autocorrelation statistics. *International journal of environmental research and public health, 8*(6), 2074-2089.

Krivoruchko, K., 2011, *Spatial statistical data analysis for GIS users*: Esri Press Redlands, CA, USA.

Lakhankar, T., Jones, A. S., Combs, C. L., Sengupta, M., Vonder Haar, T. H., & Khanbilvardi, R., 2010, Analysis of large scale spatial variability of soil moisture using a geostatistical method. *Sensors, 10*(1), 913-932.

Lee, C. S.-l., Li, X., Shi, W., Cheung, S. C.-n., & Thornton, I., 2006, Metal contamination in urban, suburban, and country park soils of Hong Kong: a study based on GIS and multivariate statistics. *Science of the Total Environment, 356*(1), 45-61.

Lin, Y.-P., Chu, H.-J., Wu, C.-F., Chang, T.-K., & Chen, C.-Y., 2010, Hotspot analysis of spatial environmental pollutants using kernel density estimation and geostatistical techniques. *International journal of environmental research and public health, 8*(1), 75-88.

Longley, P., 2005, *Geographic information systems and science*: John Wiley & Sons.

Mitchell, A. (1999). *The ESRI Guide to GIS Analysis–Volume I: Geographic Patterns & Relationships*. ESRI. CA.

Mitchell, A., 2005, The ESRI Guide to GIS Analysis: Spatial Measurements and Statistics. Vol 2. Redlands: CA: ESRI Press

Nayanaka, V., Vitharana, W., & Mapa, R., 2010, Geostatistical analysis of soil properties to support spatial sampling in a paddy growing Alfisol. *Tropical Agricultural Research, 22*(1), 34-44.

Papadopoulos, A., Papadopoulos, F., Tziachris, P., Metaxa, I., & Iatrou, M., 2014a, Site specific agricultural soil management with the use of new technologies. *Global Nest Journal, 16*(1), 59-67.

Papadopoulos, A., Papadopoulos, F., Tziachris, P., Metaxa, I., & Iatrou, M., 2014b, Site Specific Management with the Use of a Digitized Soil Map for the Regional Unit of Kastoria. *Fork to Farm: International Journal of Innovative Research and Practice, 1*(1).

Paraschakis, I., Papadopoulou, M. & Patias, P., 1990, *Digital Cartography*. Thessaloniki, Greece, Ziti. (in Greek)

Pierce, F. J. & Clay, D., 2007, *GIS applications in agriculture*: CRC Press..

Sparks, D. L., Page, A., Helmke, P., Loeppert, R., Soltanpour, P., Tabatabai, M., Johnston, C., & Sumner, M., 1996, *Methods of soil analysis. Part 3-Chemical methods*: Soil Science Society of America Inc.