Τετράδια Ανάλυσης Δεδομένων

Τόμ. 20, Αρ. 1 (2024)

Τετράδια Ανάλυσης Δεδομένων - 20

The role of Spatial Autocorrelation on spatially correlated data for Hierarchical Cluster Analysis

Thomas Koutsos, George Menexes

Βιβλιογραφική αναφορά:

Koutsos, T., & Menexes, G. (2024). The role of Spatial Autocorrelation on spatially correlated data for Hierarchical Cluster Analysis. *Τετράδια Ανάλυσης Δεδομένων*, *20*(1), 109–119. ανακτήθηκε από https://ejournals.epublishing.ekt.gr/index.php/dab/article/view/33519

The role of Spatial Autocorrelation on spatially correlated data for Hierarchical Cluster Analysis

Thomas Koutsos[1], Georgios Menexes[1]

[1] School of Agriculture, Faculty of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki

KEYWORDS

Spatial analysis
Field Experiments
Agricultural Experimentation
Data clustering

CORRESPONDENCE

Thomas Koutsos
Aristotle University of Thessaloniki
Email: tkoutsos@agro.auth.gr

ABSTRACT

Ordinary multi-variate statistical analyses usually consider only the correlation between variables ignoring the location of the objects-data points. One of the core principles of Statistics is that observations and their corresponding values, either between or within various groups, are independent of each other. The fundamental concept of spatial dependence and spatial autocorrelation is often omitted in ordinary statistical analysis, although coordinates of measurements are usually available. Current study examines the results of two different methodological approaches to test the benefit of considering the “spatial information” of the measurements: (1) Hierarchical Cluster Analysis (HCA) with corrected data (replacement of missing values with sowing row mean value), and (2) Hierarchical Cluster Analysis with data after spatial interpolation applied. Spatial Autocorrelation Analysis (Univariate Local Moran’s *I*) was used to check the spatial autocorrelation of data in each case via LISA (Local Indicators of Spatial Association–LISA) cluster maps and thus make visual comparisons between the above two methodological schemes. Both HCA analysis and LISA cluster maps show that considering the “spatial” location of the measurements can lead to different results than those from an ordinary statistical analysis without spatial correlated data. Comparing LISA cluster maps to quantile maps (that show the real distribution of data) it can be deduced that considering the spatial information (via spatial interpolation) can lead to results closer to the real distribution of data.

Introduction

In ordinary multi-variate statistical analyses, such as Hierarchical Cluster Analysis (HCA) the location of the objects-data points is not considered, and the analysis is based on the correlation between the variables. However, in geosciences, measurements are treated as values of a variable with specific coordinates on the earth surface and not as a list of values with no geo-reference information. Therefore, it is interesting to examine if considering the “spatial” dimension of data in statistical analyses can provide results closer to reality and if this can be considered as a better approach compared to the original statistical approach.

Spatial analysis is the process of extracting or creating new information from spatial data. Using a variety of techniques, and geostatistical methods, spatial analysis shows the correlation within variables across georeferenced space. Therefore, it refers to the process of analyzing and identifying patterns in spatial data associated to their respective geospatial information. Spatial analysis reveals information hidden in data based on their spatial location combining different attributes, locations, and relationships of features (Cliff & Ord, 1973; Getis, 2008). Considering this geospatial information, spatial analysis can provide new insights to help to comprehend better the reasons for the given variation in data. There are many examples in different scientific fields that show how spatial analysis takes into account the geospatial information to offer new information such as studies for the spatial autocorrelation driven by abiotic and biotic factors found in an ecosystem (Diniz-Filho et al., 2003; Dormann, 2007), studies concerning socioeconomic patterns and processes (Grengs, 2001) or even recent studies dealing with COVID-19 cases (Zhou et al., 2022) that clearly illustrate the usefulness of the spatial approach by analyzing and tracing individual cases based on their location or analyze wider geographical areas having issues.

Spatially correlated data are geospatial data with "spatial autocorrelation". The term "Spatial autocorrelation" refers to the presence of a pattern in the spatial variation of the values in a variable (Anselin, 1994). Adjacent observations with similar high or low values have a positive spatial autocorrelation. If adjacent observations have contrasting values although they are very close to each other then there is a negative spatial autocorrelation. The investigation of whether spatial data have a positive autocorrelation is important because it provides useful insights to understand the reasons behind the observed spatial variation. Among many measures of spatial association, Moran's I statistic is the most widely used for spatial autocorrelation (Anselin, 1994; Anselin et al., 2006). Some other also well-known measures are Geary's c (global), Getis and Ord's G (global), Getis and Ord's G_i and G_i^* (local), Anselin's I_i and c_i (local indicators of spatial association (LISA)), and Ord and Getis' O (a local representation considering global autocorrelation) (Getis, 2008). The presence of spatial autocorrelation implies information redundancy and therefore, it is more a "drilling-down" process to discover new information and provide data for further analysis. Spatial autocorrelation is due to any of the following reasons: (a) spatial variation in measurements; (b) measurement errors; (c) spatial diffusion, spillover, interaction, and dispersal processes; (d) spatial variation in one variable due to another; (e) incorrect model specification (Haining, 2001).

In current study, Spatial Autocorrelation Analysis (Univariate Local Moran's I) was used to check the spatial autocorrelation of data in each case via cluster LISA maps and thus make visual comparisons between the above two methodological schemes. Both HCA analysis and LISA cluster maps show that considering the "spatial" location of the measurements can lead to different results than those from an ordinary statistical analysis without spatial correlated data. Comparing LISA cluster maps to quantile maps (that show the real distribution of data) it can be deduced that considering the spatial information can lead to results closer to the real distribution of data. Two different methodological approaches were examined to test the benefit of considering the "spatial information" of the measurements: (1) Hierarchical Cluster Analysis (HCA) with corrected data (replacement of missing values with sowing row mean value), and (2) Hierarchical Cluster Analysis with data after spatial interpolation applied. After short transformation in clustering values, Spatial Autocorrelation is used to check the spatial dependence in data and visually check the difference between the above two methodological approaches.

Materials and methods

In current study we compare two different methodological approaches to test the benefit of considering the "spatial information" of the measurements: (1) Hierarchical Cluster Analysis (HCA) with corrected data (filling the missing values with sowing row mean value), and (2) Hierarchical Cluster Analysis (squared Euclidian distance among plants' z-scores, Ward's joining method) with data after spatial interpolation applied. HCA was performed using IBM SPSS Statistics ver. 23.0. The data used to test the above hypothesis was taken from an experiment set in Aristotle University of Thessaloniki (AUTH) Farm (latitude: 40°32'1.75"N longitude: 22°59'26.98"E) using a maize crop (AGN720) in a three-ha field area during 2016. Fertilization was set to 200 kg N and 100 kg P/ha and irrigation plan was set according to the requirements of the crop plants. The field was sown with a 4-row pneumatic sowing machine Gaspardo in April and after crop emergence, a randomly taken plot (4 m × 4.25 m) with 6 sowing rows and 25 plants per row has been selected for the analysis. The distance between the crop rows was 80 cm, while the distance between the plants in the same row was 17 cm. Therefore, there were 25 plants per row and 150 plants in total.

The individual maize plants in the plot (Figure 1, a), for the purpose of the study, were considered as the units of the target statistical population from which samples had been extracted. In some spots, plants did not emerge leading to an original set of (spatially related) data with "missing values" or "bad quality". At the silage stage of maize (14 weeks after sowing), all plants were harvested from the plot and the silage yield (fresh weight - FW, dry weight - DW, and ear weight - EW) of each plant was recorded. These three variables (FW, DW and EW) are also tested to confirm if they are strongly correlated or not. The silage stage was determined by breaking the ears of maize and visually evaluating the kernels' stage of development. The distribution of values of these three crop parameters (FW, DW, EW) in the plot is given in Table 1 (b, c, d) using Natural Breaks maps (Jenks, 1967), a data classification method where classes are based on "natural" groupings inherent in the data.

Data was re-clustered based on *z*-score values to provide a different grouping needed in Spatial Autocorrelation Analysis. Records correspond to HCA clusters with high *z*-scores were assigned a value of "3" to indicate data in clusters with high performance (high values), while records that correspond to HCA clusters with *z*-scores near zero were assigned a value of "2" (values close to mean value), and records that correspond to HCA clusters with low *z*-scores were assigned a value of "1" to indicate clusters with low performance (low values). The above re-clustering process was needed for the Spatial Autocorrelation Analysis.

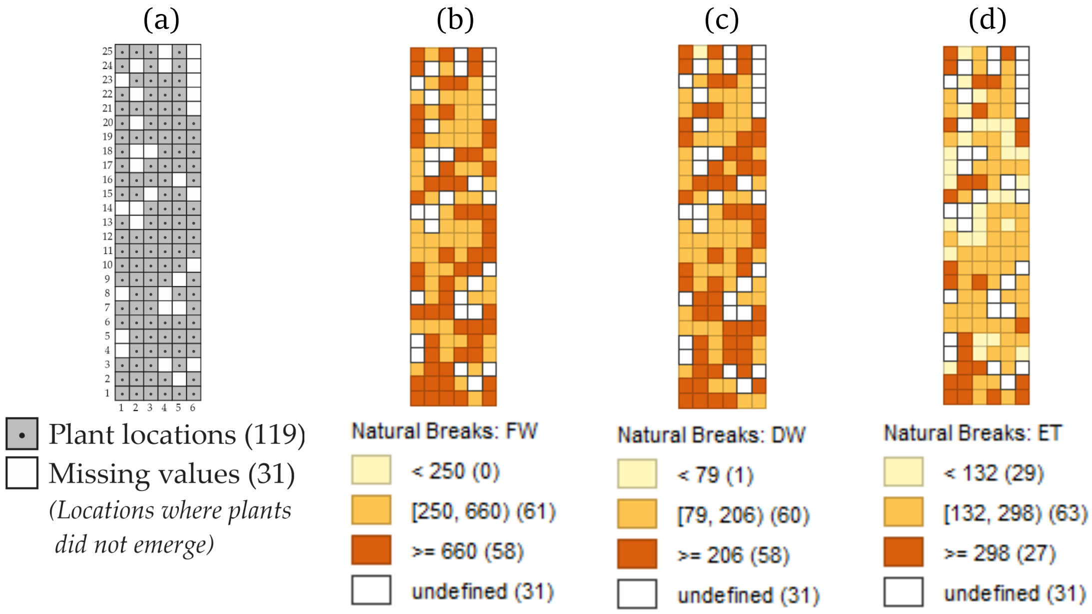

Figure 1. The experimental maize plot: (a) location of plants (black dots in "white" squares represent all harvested plants in plot (119 plants in total) and "grey" squares represent places, where plants did not emerge (39 in total); (b) Natural Breaks map for fresh weight (FW) data; (c) Natural Breaks map for dry weight (DW) data; (d) Natural Breaks map for ear weight (EW) data. The "Undefined" cluster in Natural Breaks maps corresponds to missing values.

Our analysis includes the following stages: ***Stage 1***: Fill missing values using some heuristic method (e.g., replaced with the sowing row mean); ***Stage 2***: Interpolation for missing values plus correction of original data considering the spatial correlation (Kriging based); ***Stage 3***: On the corrected and filled data (from Stage 2) perform various ordinary analyses, e.g., Hierarchical Cluster Analysis; ***Stage 4***: On the corrected and filled data (from Stage 2) perform spatial "cluster analysis"- spatial autocorrelation analysis; ***Stage 5***: Compare the results of various typologies (clusters). Are the Hierarchical Cluster Analysis results (using corrected data from Stage 2) in line with Ordinary Kriging results?

Results

The preliminary analysis on the data availability revealed that for the given grid of 6 rows with 25 plants per row instead of 150 plants (measurements) in total only 119 plants were emerged leaving 31 empty spots as missing values. The initial check on the crop yield data showed a strong correlation (Pearson's r>0.80) between the three selected crop parameters of silage yield (FW, DW, and EW), as expected for a maize hybrid.

The results per stage (as described in materials and methods section) are the following:

Stage 1: Fill missing values with sowing row mean value. The row mean value was calculated and used to fill up the missing values (locations where plants did not emerge).

Stage 2: Spatial interpolation was used to estimate missing values plus correction of original data considering their spatial correlation. A grid of 6x25 was used and the interpolation method applied was Ordinary

Kriging. Variogram analysis performed on the FW data and the main variogram parameters used for performing Kriging were the following (Shahinzadeh et al., 2022): (1) *General Fitting*: Least Squares (target precision: 0.0001%); 2) *Experimental*: Estimator type: Variogram; Lag size = 0.55; Number of lags=15; direction=0; Tolerance=90; (3) *Variagram component*: Component type: exponential; Partial sill=53470; Range=0.95; Aniso=1.999, 50.94). The Root Mean Square (RMS) calculated to check the performance of the interpolatio method was 233.4. Similar analysis was performed for DW and EW crop silage data to calculate (missing) values at locations where plants did not emerge.

(a) (b) (c)

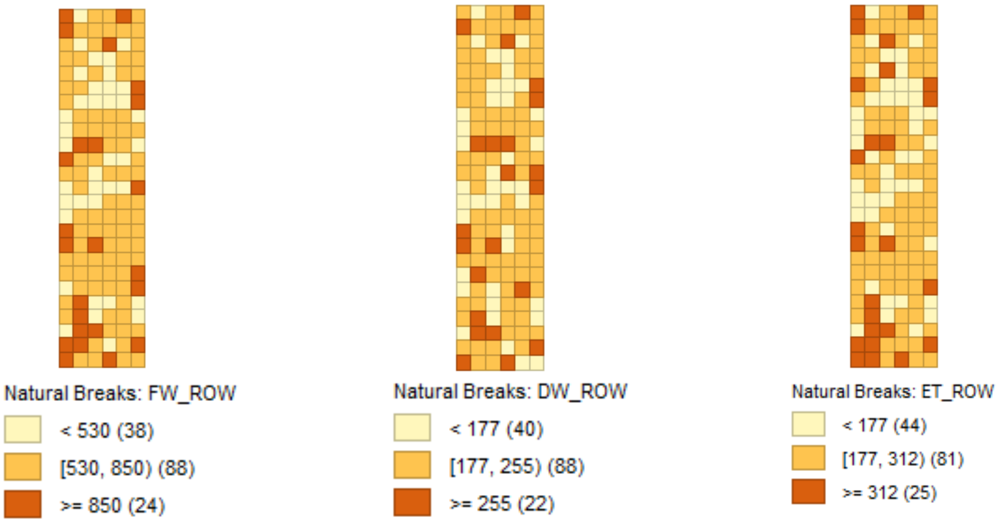

Figure 2. Natural Breaks maps for the distribution of data in the plot: (a) fresh weight (FW), (b) dry weight (DW), and (c) ear weight (EW), after filling the missing values with row mean value.

Comparing the results from stage 1 and 2, the natural break maps for the distribution of data for the same crop yield parameter look almost identical; however, there are still some minor differences that can be spotted easily if the maps are examined thoroughly. The clusters applied in each case are automatically proposed by the spatial software used (GeoDA).

Stage 3: On the corrected and filled data (from Stage 2) perform various ordinary statistical analysis. Hierarchical Cluster Analysis was performed on data in two cases (a) missing values were replaced with the sowing row mean (without considering the spatial correlation of measurements), and (b) missing values were calculated/estimated via spatial interpolation and all original data were "corrected" considering their spatial correlation. Row means were calculated in Excel and missing values were replaced by these values. In case of spatial interpolation, the missing values were estimated by applying Ordinary Kriging (using a grid of 25x6) and then find the estimated values at locations where the missing values were spotted.

(a) (b) (c)

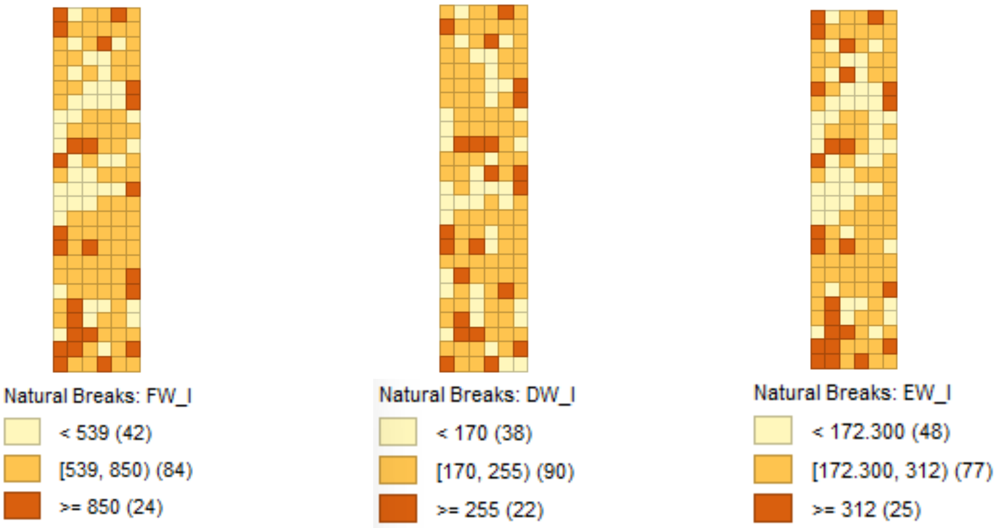

Figure 3. Natural Breaks maps for the distribution of data in the plot: (a) fresh weight (FW), (b) dry weight (DW), and (c) ear weight (EW), after applying Kriging interpolation.

Based on the silage data (FW, DW, EW) after filling up the missing values, Hierarchical Cluster Analysis (HCA) was performed in both cases. The results from Hierarchical Cluster Analysis (HCA) revealed 3 clusters (C1-3) in case the missing values are replaced by sowing row mean and 5 clusters (C1-5) in case the missing values were replaced by estimated values via (Ordinary Kriging) spatial interpolation (Figure 4).

Stage 4: On the corrected and filled data (from Stage 2) spatial "cluster analysis"- spatial autocorrelation analysis was performed. For this step, a re-clustering transformation of the HCA results was needed for the needs of spatial autocorrelation analysis. Based on *z*-scores it was assumed that clusters with high values can be marked as "high-performed" clusters, while clusters with low *z*-score values can be considered as "low-performed" clusters. Finally, for clusters with *z*-score values near zero or $|z| \leq 0.5$ standard deviation it was assumed that they can be treated as clusters with values near average (Figure 5).

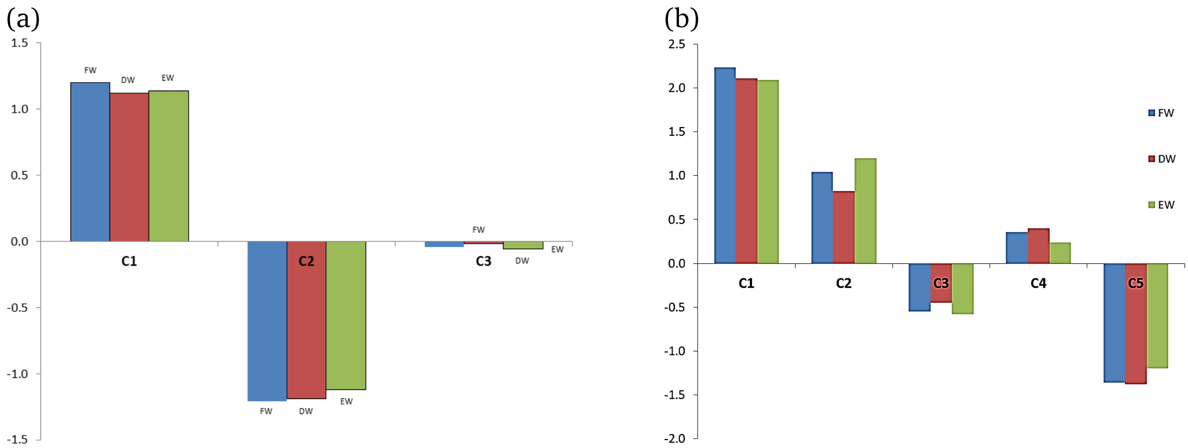

Figure 4. Hierarchical Cluster Analysis (HCA) results: (a) missing values replaced by sowing row mean; (b) missing values calculated via interpolation, C1-5 the HCA clusters, where FW: fresh weight, DW: dry weight, and EW: ear weight. The vertical axis corresponds to mean *z*-scores.

Therefore, all records were re-clustered based on the re-clustering approach (Figure 5) to transform HCA clusters into new ones for the needs of spatial autocorrelation analysis. In case of filling the missing values with sowing row mean, clusters C1 was considered as "high-performed" (marked with dark grey color), while cluster C2 was identified as "low-performed" (marked with light grey color), and C3 as near mean value (marked with white color). In case of filling the missing values via interpolation, clusters C1 and C2 were considered too as "high-performed" (marked with dark grey color), clusters C3 and C5 were identified as "low-performed" (marked with light grey color), and C4 as near mean value (marked with white color). In Figure 6, the location of each record (plant measurement) based on which cluster belongs, where: (a) the original HCA clusters in case missing values are filled up with row mean, and (b) the corresponding re-cluster map; (c) the original HCA cluster in case missing values were calculated using interpolation, and (d) the corresponding re-cluster map.

Using this new grouping of data based on the *z*-scores of HCA results, spatial autocorrelation analysis was performed to check the spatial dependence of data and visualize the areas where high values are surrounded by high values too (high-high cluster) and are where low values are surrounded by low values, respectively.

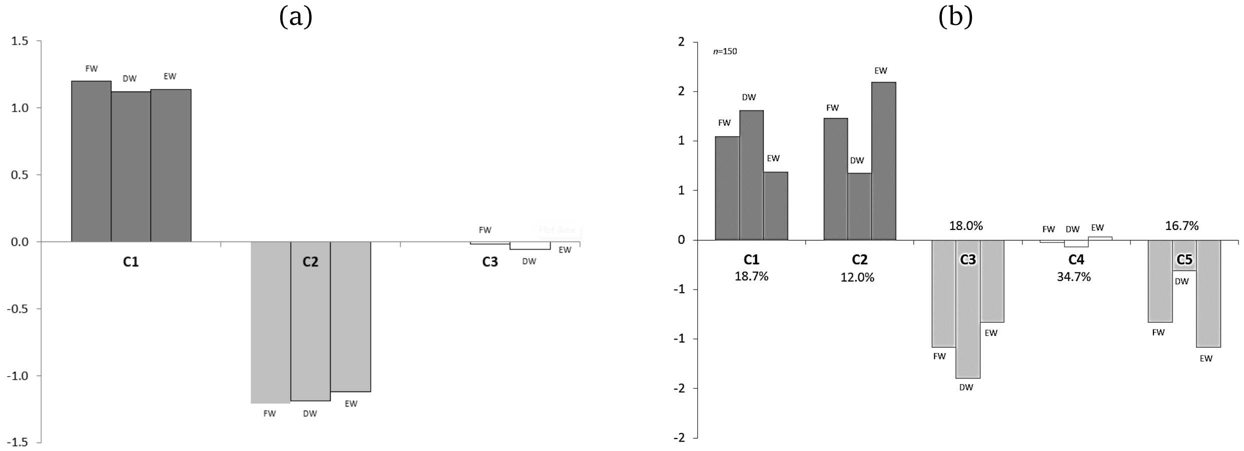

Figure 5. Re-clustering Hierarchical Cluster Analysis (HCA) results: (a) missing values replaced by sowing row mean; (b) missing values calculated via interpolation, C1-5 the HCA clusters. Dark grey color indicates high-performed clusters, light grey color corresponds to low performed cluster, and white color clusters represents clusters with mean values near total mean. The vertical axis corresponds to mean *z*-scores. Percentages indicate what percent of the total data each cluster represents.

Spatial autocorrelation analysis was used as an assessment tool to compare HCA results with the form of LISA cluster maps with the corresponding ones for each crop parameter (FW, DW, and EW). Apart the LISA cluster maps (Univariate Local Moran's *I*), the Global Moran's I autocorrelation index was calculated to show how dispersed or clustered are the spatial data used. It must be stressed that LISA cluster maps of HCA are a combined result considering the effect of the three crop parameters, compared to the LISA cluster maps of each parameter individually. However, the use of LISA cluster maps can be considered as very useful for the identification of spatial clusters in all cases.

The results of the exploratory spatial analysis on data where missing values were filled with sowing row mean value or calculated via interpolation are given in Figures 7 and 8, respectively. In case the missing values were replaced by the sowing row mean (Figure 7), the comparison of LISA cluster maps between HCA results and the data of each crop parameter indicated that: (1) the spatial clusters for measurements with high values surrounded by high values and the spatial clusters for measurements with low values surrounded by low values are spotted in almost the same location (a high-high spatial cluster at the lower part of the area, while two low-low spatial clusters at the center and upper part of the area); (2) Global Moran's *I* autocorrelation index of data in each crop parameter (FW: 0.084; DW:0.013; EW:0.0125) shows a slight positive autocorrelation (slight clustered data), while for HCA the value of index is 0.028 (line almost parallel), which means that considering the data of all the three parameters leads to the conclusion that there is no spatial autocorrelation in data.

Missing values replaced by row mean		Missing values estimated via interpolation	
(a)	(b)	(c)	(d)

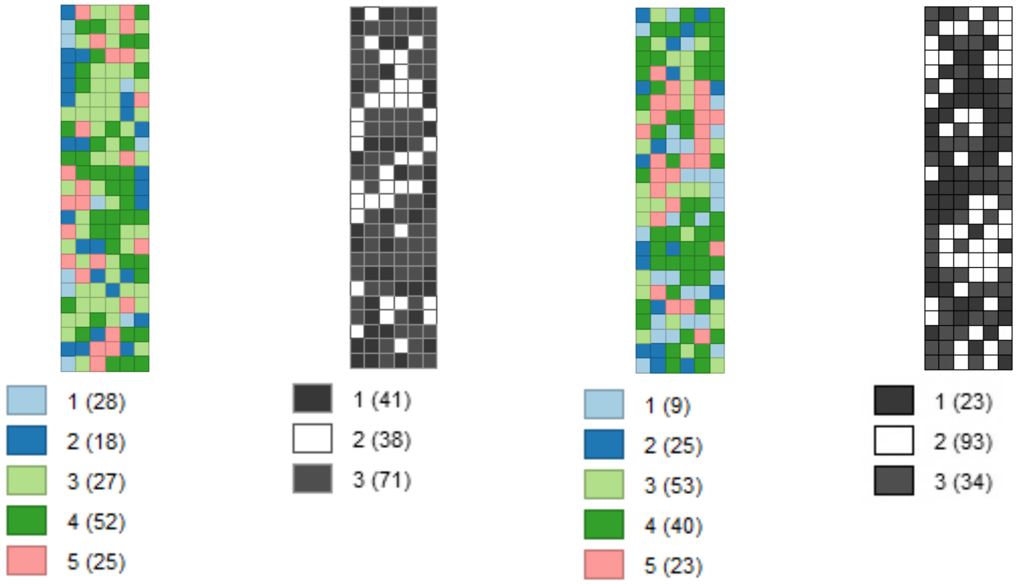

Figure 6. Re-clustering HCA clusters: (a) HCA clusters after filling up missing values with sowing row mean value; (b) Re-clustering HCA clusters in case missing values were filled with sowing row mean value; (c) HCA clusters after filling up missing values using interpolation (Kriging); (d) Re-clustering HCA clusters in case missing values were estimated via interpolation.

In case the missing values are estimated via spatial interpolation (Figure 8), the results with the form of LISA cluster maps are in line the previous results where the missing values were replaced by sowing row mean. Spatial clusters (high-high and low-low) are spotted at almost the same locations. Values for the Global Moran's *I* autocorrelation index in case of spatial interpolation are also very close to those where the missing values were replaced by the sowing row mean, showing a slightly more positive autocorrelation in data.

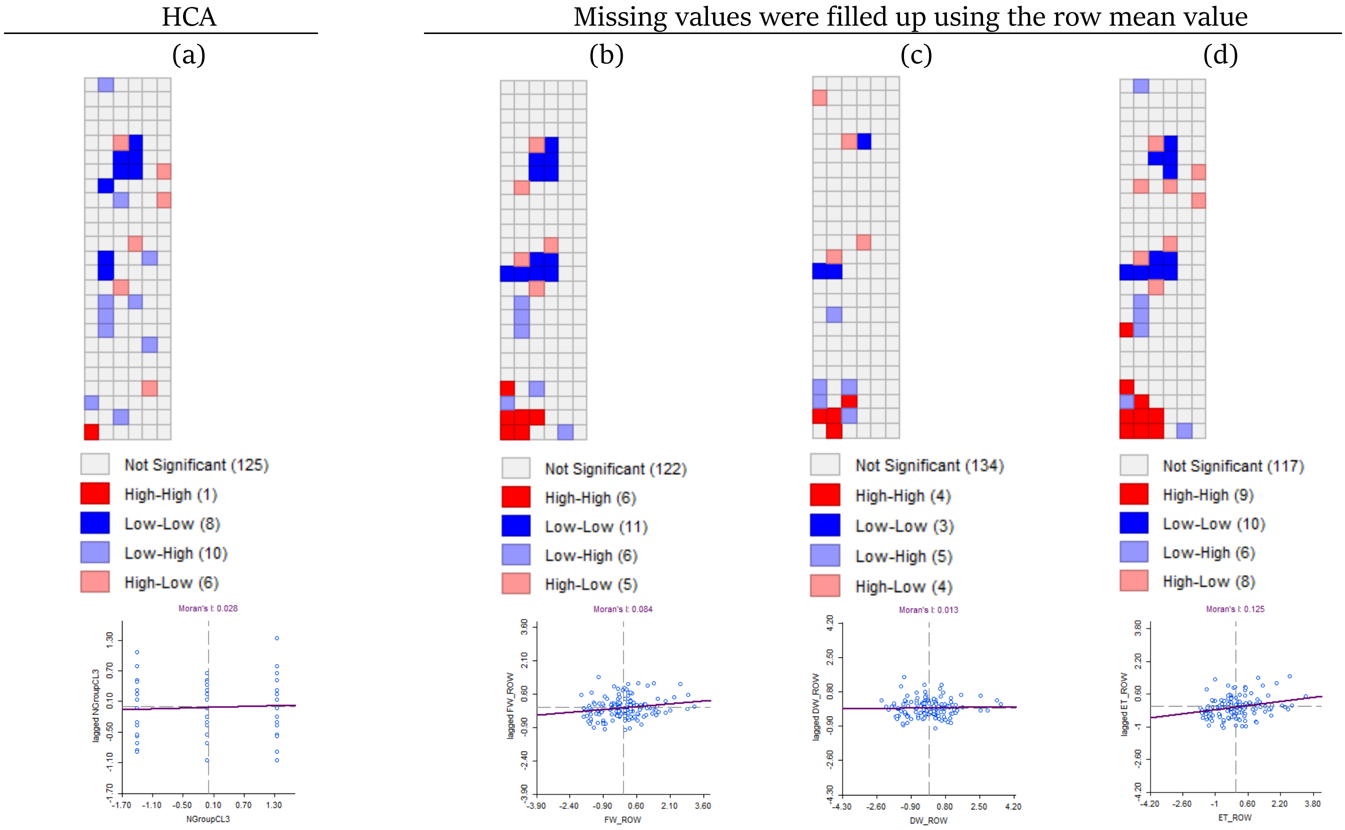

Figure 7. Cluster LISA maps (Univariate Local Moran's *I*) as results of the spatial autocorrelation analysis on re-clustered HCA results, where missing values were filled up using the sowing row mean value: (a) HCA re-

clustered data, and on data with missing values replaced by sowing row mean value (b) for FW; (c) for DW data; (d) for EW data.

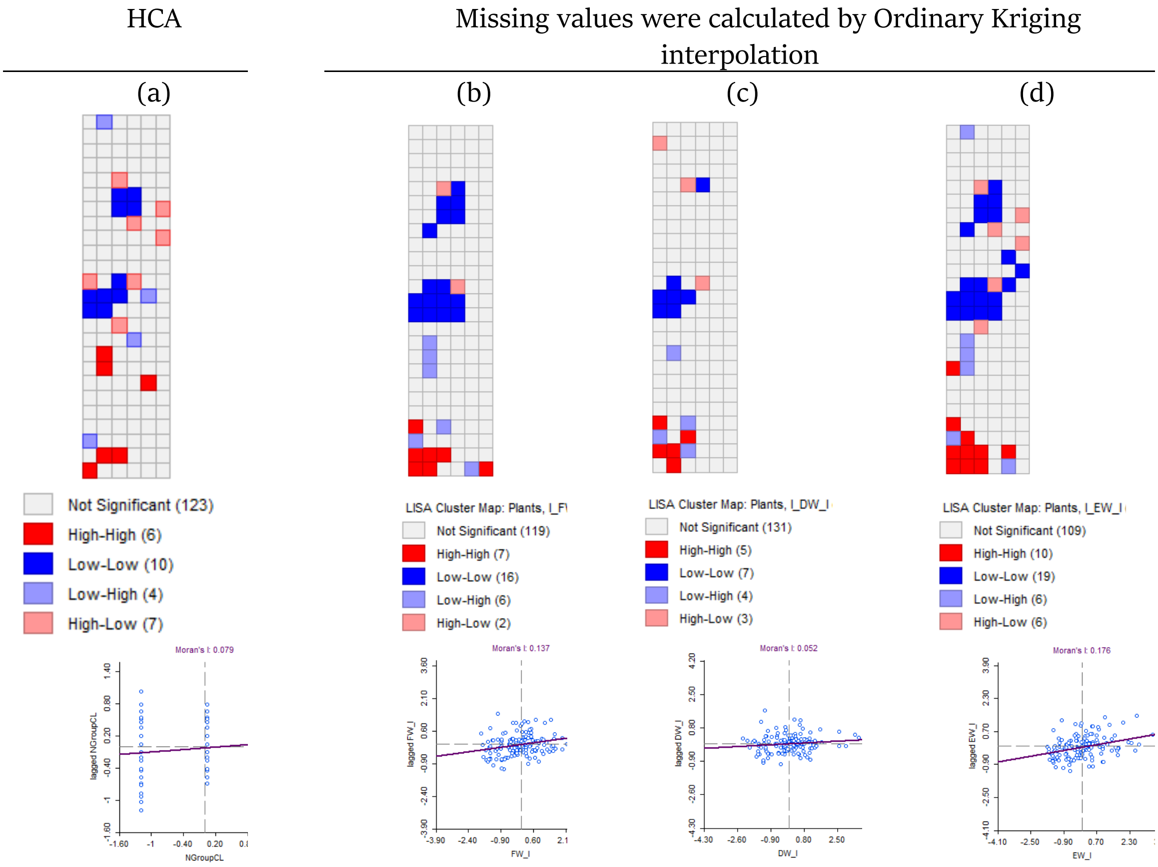

Figure 8. Cluster LISA maps (Univariate Local Moran's *I*) as results of the spatial autocorrelation analysis on re-clustered HCA results, where missing values were calculated via interpolation: (a) HCA re-clustered data, and on interpolated data (b) for FW; (c) for DW data; (d) for EW data.

Stage 5: Compare the results of clusters for the same parameter (HCA, FW, DW, EW). Two main differences can be spotted when comparing the results of clusters in case the missing values are replaced by sowing row mean and in case they are estimated by spatial interpolation (1) spatial interpolation shows data being more clustered in corresponding LISA cluster maps for data of each crop parameter; (2) spatial clusters in HCA results are also "enhanced" after applying spatial interpolation, showing that data are more clustered too. This can be explained by the fact that spatial interpolation estimates a value in a non-sampled location considering the neighboring measurements leading to a more "smoothed" area of data (having similar values) compared to the sowing row mean that may appear as a peak in the location of a missing value. This also explains why after applying spatial interpolation for the estimation of missing values spatial data seems to be more clustered. Thus, it can be deduced that spatial interpolation can "enhance" the existence of spatial clusters and improve the results of HCA.

Discusssion

Current study examines the results of two different methodological approaches to test the benefit of considering the "spatial information" of the measurements: (1) Hierarchical Cluster Analysis (HCA) with corrected data (replacement of missing values with sowing row mean value), and (2) Hierarchical Cluster Analysis with data after spatial interpolation applied. Based on these results it can be deduced that spatial interpolation can effectively replace commonly used practices of replacing missing values by sowing row mean or sowing column mean value. Interpolation estimates the values of the surrounding measurements and not only considering the measurements in the same row or column.

Spatial Autocorrelation Analysis (Univariate Local Moran's *I*) was used to present and reveal the differences between HCA results with corrected data (replacement of missing values with sowing row mean value) and HCA results after interpolation applied. Comparing LISA cluster maps proved to be a good tool to evaluate different conditions with spatial clustered data and it was confirmed that spatial interpolation "enhances" the spatial clusters existing in data. Therefore, by using spatial interpolation the spatial information of data can be used not only to estimate the missing values but also to achieve more accurate and representative data closer to their real distribution. As a result, compared to the practices used so far, the proposed method can help prepare spatially clustered data for use in HCA.

Conclusions

The implementation of spatial interpolation to data before HCA managed to improve the quality of data by effectively estimating the missing values in data (locations where plants did not emerge), considering not only the values in a row (sowing row mean value) but all the surrounding measurements to estimate a representative value. In addition, LISA cluster maps revealed that the implementation of Ordinary Kriging can "enhance" the spatial clusters providing better data for HCA.

Therefore, the main conclusions of the current study are the following:

- Spatial interpolation (Ordinary Kriging) can improve data quality by estimating missing values in experimental data considering all neighboring measurements rather than using means to fill up the empty spots.
- Spatial interpolation can "enhance" the existence of spatial clusters when used for estimating missing values.
- Spatial interpolation can provide more complete and representative data for better HCA results.

Based on the study findings the implementation of spatial interpolation can replace other practices in which missing values are replaced by the sowing row or column mean to effectively prepare data for HCA.

Acknowledgements

We would like to thank the University students A. Pesios and A. Chatzopoulos for providing the data and the staff of the A.U.Th. Farm for their assistance throughout the experiment.

Bibliography

Anselin, L. (1994). ''Local Indicators of Spatial Association–LISA.'' Geographical Analysis 27, 93–115.

Anselin, L., I. Syabri, and Y. Kho. (2006). ''GeoDa: An Introduction to Spatial Data Analysis.'' Geographical Analysis 38, 5–22.

Cliff, A.D., and Ord, J.K. (1973) Spatial Autocorrelation. London: Pion.

Diniz-Filho, J.A.F., Bini, L.M., Hawkins (2003). B.A. Spatial autocorrelation and red herrings in geographical ecology, *Global Ecology and Biogeography 12*, 53-64.

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography 16,129–138.

Haining, R.P. (2001) Spatial Autocorrelation. *International Encyclopedia of the Social & Behavioral Sciences*, 14763-14768. https://doi.org/10.1016/B0-08-043076-7/02511-0.

Jenks, G. F. (1967). "The Data Model Concept in Statistical Mapping", International Yearbook of Cartography 7: 186-190.

Getis, A. (2008). A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective. *Geographical Analysis 40,* 297–309.

Grengs, J. (2001). "Does Public Transit Counteract the Segregation of Carless Households? Measuring Spatial Patterns of Accessibility." Transportation Research Record: Journal of the Transportation Research Board 1753: 3–10.

Shahinzadeh, N., Babaeinejad, T., Mohsenifar, K., Ghanavati, N. (2022) Spatial variability of soil properties determined by the interpolation methods in the agricultural lands. *Modeling Earth Systems and Environment* 8:4897–4907.https://doi.org/10.1007/s40808-022-01402-w.

Zhou, L., Li, S., Li, C., Shen, G., Tang, H., Zhu, P., Han, H., Li, B. (2022). Spatial congruency or mismatch? Analyzing the COVID-19 potential infection risk and urban density as businesses reopen. *Cities (123)*, 103615.

ΤΕΤΡΑΔΙΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ | DATA ANALYSIS BULLETIN, 20(1), 109-119

ΕΜΠΕΙΡΙΚΗ ΕΡΓΑΣΙΑ | RESEARCH PAPER

Ο ρόλος της Χωρικής Αυτοσυσχέτισης σε χωρικά συσχετιζόμενα δεδομένα για την Ιεραρχική Ανάλυση σε Συστάδες

Θωμάς Κουτσός[1], Γεώργιος Μενεξές[2]

[1] Τμήμα Γεωπονίας, Σχολή Γεωπονίας, Δασολογίας και Φυσικού Περιβάλλοντος, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Χωρική ανάλυση
Πειράματα αγρού
Αγροτικός Πειραματισμός
Ομαδοποίηση δεδομένων

ΣΤΟΙΧΕΙΑ ΕΠΙΚΟΙΝΩΝΙΑΣ

Θωμάς Κουτσός
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Email: tkoutsos@agro.auth.gr

ΠΕΡΙΛΗΨΗ

Οι συνηθισμένες πολυπαραγοντικές στατιστικές αναλύσεις συνήθως λαμβάνουν υπόψη μόνο τη συσχέτιση μεταξύ των μεταβλητών, αγνοώντας τη θέση των αντικειμένων-σημείων δεδομένων. Μία από τις βασικές αρχές της Στατιστικής είναι ότι οι παρατηρήσεις και οι αντίστοιχες τιμές τους, είτε μεταξύ είτε εντός διαφόρων ομάδων, είναι ανεξάρτητες η μία από την άλλη. Η έννοια της χωρικής εξάρτησης και της χωρικής αυτοσυσχέτισης συχνά παραλείπεται στη συνήθη στατιστική ανάλυση, αν και οι γεωγραφικές ή άλλου είδους συντεταγμένες των μετρήσεων είναι συνήθως διαθέσιμες. Η τρέχουσα μελέτη εξετάζει τα αποτελέσματα δύο διαφορετικών μεθοδολογικών προσεγγίσεων για να ελέγξει το όφελος από την εξέταση των «χωρικών πληροφοριών» των μετρήσεων: (1) Ιεραρχική Ανάλυση σε Συστάδες (HCA) με διορθωμένα δεδομένα (αντικατάσταση τιμών που λείπουν με το μέσο όρο της αντίστοιχης γραμμής σποράς) και (2) Ιεραρχική Ανάλυση σε Συστάδες με δεδομένα μετά από χωρική παρεμβολή. Η Ανάλυση Χωρικής Αυτοσυσχέτισης (Univariate Local Moran's *I*) χρησιμοποιήθηκε για να ελεγχθεί η χωρική αυτοσυσχέτιση των δεδομένων σε κάθε περίπτωση μέσω χαρτών συμπλέγματος LISA (Local Indicators of Spatial Association) και έτσι να γίνουν οπτικές συγκρίσεις μεταξύ των δύο παραπάνω μεθοδολογικών σχημάτων. Τόσο η ανάλυση HCA όσο και οι χάρτες συστάδων LISA δείχνουν ότι η εξέταση της «χωρικής» θέσης των μετρήσεων μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα από αυτά μιας συνηθισμένης στατιστικής ανάλυσης χωρίς χωρικά συσχετισμένα δεδομένα. Συγκρίνοντας τους χάρτες συστάδων LISA με χάρτες ποσοστοιχιών (που δείχνουν την πραγματική κατανομή των δεδομένων) μπορεί να εξαχθεί το συμπέρασμα ότι ότι η εξέταση των χωρικών πληροφοριών (μέσω χωρικής παρεμβολής) μπορεί να οδηγήσει σε αποτελέσματα πιο κοντά στην πραγματική κατανομή των δεδομένων.