Luís Outeiro, Xavier Úbeda and Marc Farrés

GRAM (Grup de Recerca Ambiental Mediterrània). Departament de Geografia Física i AGR. Universitat de Barcelona. C/ Montalegre, 6, 08001 Barcelona louteiro@gmail.com

PLOT SCALE DESIGN OF SAMPLING STRUCTURES WITH SOIL BULK DENSITY MEASUREMENTS: SEMIVARIOGRAM TO DETECT FOREST MANAGEMENT DISTURBANCES

Shape, size, spacing, distribution and density of the sampling network receive, in our opinion, small attention on the spatial statistics literature. Soil researchers should make a quick, accurate, and cost-efficient decision on which sampling structure best satisfies their objectives. Systematic versus random sampling becomes one of the first decisions that the researcher makes when designing sampling plots. Success in this task and the accurate description of the sampling spatial structure is a preliminary step towards spatial prediction or stochastic simulation. The design was based on three different sampling structures with different shape and distribution of sampling points: systematic-quadrangular (Campas plot), systematic-triangular (Masia plot) and nested (Cantallops plot). With the different soil variables (Litter, Fermentation and Humus layers and bulk density) measured, we fit variograms of the bulk density to study the efficiency of the sampling scheme. Then, we used another criterion for cross-validation which gives a rough idea of the accuracy of variogram fitting and thus the precision attained with kriging at unsampled points. The highest regression coefficient, and therefore the most satisfactory sampling scheme, is found for the Masia plot. The results of the spatial analyses also enable to detect which variables of the topsoil are significantly impacted due to the forest management operation.

Keywords: forest management, thinning, geostatistics, soil.

1. Introduction

1.1. Sampling locations

Statistical based monitoring plans require environmental scientist to collect samples from an environment at statistically determined locations. Ideally, each sampling location should be selected at random. Also, the number of samples must be defined with a maximum-accepted level of error in the results. In reality, sample location and number of samples must be considered in concert with several other important aspects unique to environmental science. For example, costs associated with sampling and analysis often limits the application of rigorous statistics in environmental monitoring.

1.2. Sampling interval

Usually scientist and researchers question themselves on how far apart or how close together we have to take samples of a variable in order to identify a correlation length. As stated by Nielsen and Wendroth (2003) we don’t sample to obtain a nice correlation length. Rather we wish to sample a conceptual spatial process in order to learn and describe relevant relations within an ecosystem. The spatial or temporal correlation length represents a measure of success or control how well we have designed our sampling scheme. The same authors state that when designing a sampling scheme, the population variance should be considered. But not all these variance components are know “a priori” and not are constant in time. However, the variance strongly depends on the domain size. Hence, the identification of a correlation range with a certain spatial sampling design depends on the domain size.

1.3. Size of plots

Depending on the scope of the task, some authors consider that the size of small, replicated plots in treatment experiments is selected on the basis of many practical considerations- the parcel of land available for the experiment; the number of treatments and replicates considered necessary; the width and length of each plot relative to the activity developed with the portion of terrain we want to study, etc. The reliability of the experimental semivariogram is affected by the size of the sampling scheme (or its inverse, the density of data), and the configuration or design of the sample. As the size of the sample increases such scatter decreases and the form of the semivariogram becomes clearer. Evidently the larger is the sample from which the semivariogram is computed, the more precisely is it estimated. Webster and Oliver, 2001 stated that sample projects based on fewer than 50 are often erratic sequences of experimental value with little or no evident structure. But the larger is the sample; the larger is the need of more logistics, more equipment, more personnel, and more resources that not always can be satisfied by the researchers due to the limited budget or the limited interval time to develop the study.

1.4. Forest and fire management operations

Land use changes are becoming a key factor in the study of wildfire. During last century, traditional land uses for forest and agricultural activities have declined and abandoned land has created a combustive situation that requires restoration. The current situation is definitely transitional where extensive farming interest contrast with abandoned farms and where forestry industry and state property interests contrast with the increase land use for tourism.

Abandoned agricultural fields and a mosaic landscape have increased vegetative fuel loads. On the other hand, urban and recreational uses near and in forest areas increase the potential impact on infrastructures, property, and people. Abandoned fields have allowed forest areas to grow in density of trees and increase their size by connecting once isolated areas. As the number of forest fires have increased as well as the magnitude, fire intensity, and the size. Thus, fires have now become the main manager of the forest, but the resilience of the ecosystem is uncertain (Terrades, 1998). Catalan forests have experienced an important change in the forest fires regime. The fire intensity is one important change. In order to avoid the increasing risk of these changes, stakeholders, land owners and authorities develop plans to tackle this problem. These plans are based on management operations which count with traditional tools as thinning, logging, harvesting or afforestation and others less traditional as prescribed fires or controlled burning. All of these tools provoke a distinct effect on soil upper layers which can be detrimental if not used with care. A spatial analysis with a deterministic approach (semivariogram) was chosen to study the connection between forest management operations (thinning with heavy machinery) and its impact on soil bulk density. Some studies shown that soil bulk density is one of the least variable physical properties (Page-Dumrose et al., 2006), thus the use of the semivariogram to study bulk density spatial variability can be a useful to detect disturbances which are manifest in the semivariogram structure.

1.5. Objectives

This study investigate three different types of sampling structures to find out the best suited to use in sampling projects with limitations in terms of budget and time. Based on the data collected for the latter objective, we aim to study the effects of forest management in the first layers of the soil. The three plots under study were managed at different years. Our hypothesis is that the most recent plot to be managed will show the highest values of bulk density due to the removing and mixing effect of forest management on the upper layers of soil.

2. Study area

The plots are located in the Gavarres massif (NE Iberia Peninsula). This massif declared Natural protected area in 1992 by the Catalan Government has an area of 350 km^2, and it is covered by a thick mantle of forest, mainly *Quercus suber* and *Pinus pinea* and *Pinus pinaster*. Granidiorites and thermic metamorphic rocks are the most representative bedrock and its soils are classified as Cambic Arenosol. Altitude of this massif ranges between 469 m.a.s.l and 148 m.a.s.l. For this reason, we located the plots at three different altitudes describing a transect. It is important to remark that we also look for areas were the management has being carried in distinct years.

Mean annual temperature is around 14ºC, with summer mean 24ºC and winter mean 6ºC. Mean annual rainfall is 675mm with high annual variability. During fall and the early winter months, easterly winds bring storm rainfall episodes which thin soils are highly exposed.

	Area (km2)	Aspect	Vegetation structures		Year of thinning
			Trees density (%)	Shrub density (%)	
Campas	0.5	SSE	60	45	2005
Masia	0.3	SE	80	65	2004
Cantallops	0.4	S	70	30	2006

Table 1. Main characteristics of the plots

3. Methodology

Soil samples were collected with a soil core of 10cm depth, and in areas where the soil was not developed and represent a thin layer was used instead a 5cm soil core.

The location of the plots was set aiming to characterize the whole heterogeneity of heights, management practices and biophysical conditions within the watershed (Table 1). As explained before in the study area section, plots were located following a transect to set a sampling scheme at three different heights. So we located one plot in the lowland next to the agriculture camp. One plot mid height taking into account that this plot remained without thinning since 2004. The last plot located at the highest altitude represent the most recent stage of management and because the steeper slopes compared with the other two plots. Bulk density (D): Divide the mass of the oven-dried soil sample by the volume of the sampling container. Actual density (d) = soil dried mass/volume moved; g/cm^3

3.1. Soil sampling structures

As already stated sampling scheme was designed regarding two types of shapes triangular and quadrangular, and attending two types of sampling method; systematic and nested (Figure 1). Matérn (1960) and Bellhouse (1977) suggested that the optimality of a spatial design for predicting the mean value is strongly related to the regularity of the design points.

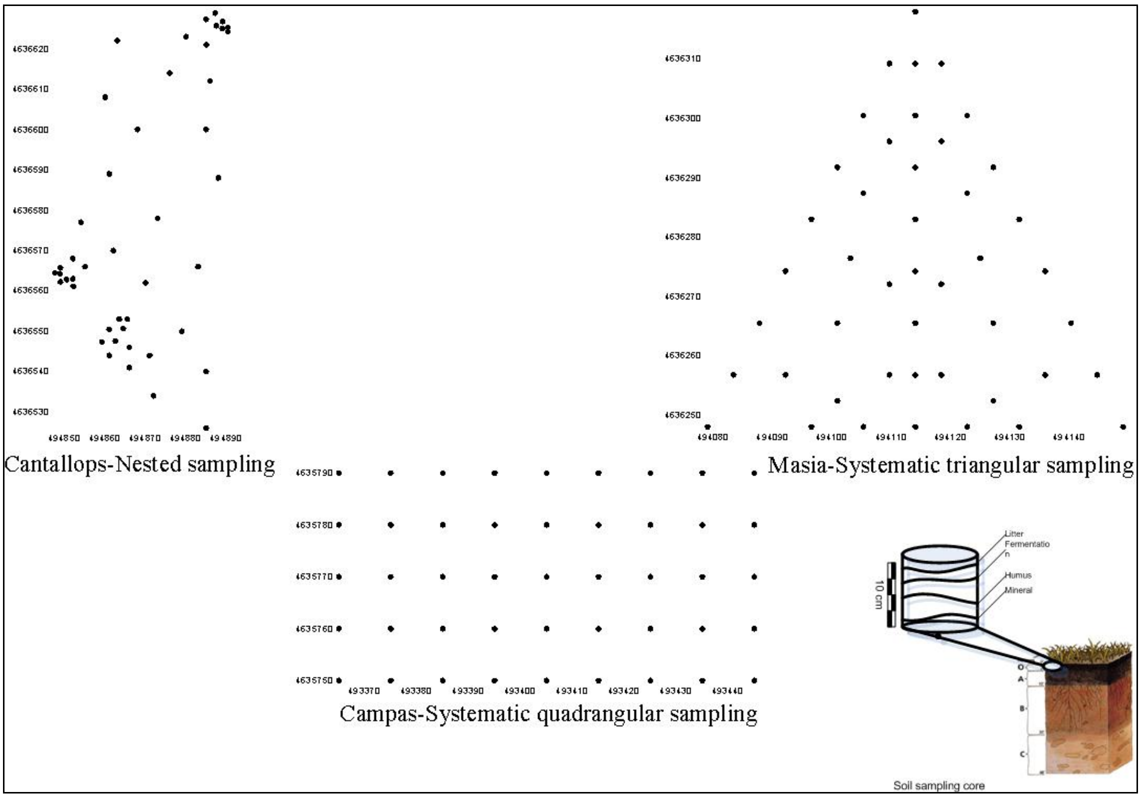

Figure 1. Sampling scheme of the three plots

3.1.1. Systematic

Systematic sampling is a statistical method involving the selection of every k^{th} element from a sampling frame, where k, the sampling interval. Using this procedure each element in the population has a known and equal probability of selection. This makes systematic sampling functionally similar to simple random sampling. It is however, much more efficient (if variance within systematic sample is more than variance of population) and much less expensive to carry out. The researcher must ensure that the chosen sampling interval does not hide a pattern. Any pattern would threaten randomness. A random starting point must also be selected. Systematic sampling is to be applied only if the given population is logically homogeneous, because systematic sample units are uniformly distributed over the population.

3.1.1.1. Systematic-triangular

This shape of the plot was used for the Masia plot. Sampling points were located following a triangular equilateral shape. It has been shown (Yfantis et al., 1987) that the optimal design locations are at the nodes of a triangular grid, however the presence of a boundary, especially an irregular boundary influence the optimal site locations. Stevens (2006) states that boundaries present estimation problems because points near boundaries do not have enough neighbors. This is the reason why compare this triangular shape with the quadrangular shape

3.1.1.2. Systematic-quadrangular

This shape was used in the Campas plot. This is one of the most common sampling schemes used in the literature. The grid of points is located uniformly within the sampling space following a regular distance in east and north directions.

3.1.2. Nested

This model is based on the notion that a population can be divided into classes at two or more categorical levels in a hierarchy. The population can be sampled using a multi-stage or nested scheme so as to estimate the variance at each level. Webster et al., (2006) show as this model offers a balanced between precision and cost because sample with separating distances increasing in geometric progression from stage to stage.

3.2 Semi-variogram

First step in geostatistical analysis is the deterministic inference based on semivariogram. This graph summarizes the spatial relations in the data. Also the semivariogram describes the variance of the region. Each calculated semivariance for a particular lag is only an estimate of a mean semivariance for that lag. The semivariogram is also subjected to errors, arising largely from sampling fluctutation. The true semivariogram representing the regional variation is continuous, and it is this semivariogram that we should really like to know (Webster and Oliver, 2001). Sources of fluctuation are due to two main reasons: error sampling(errors within the measurements of the variables. i.e. laboratory equipment, errors in observations) and structural error (the error within the lack of precision to obtain a spatial pattern from the sampling scheme, i.e sampling locations are too further apart to detect spatial patterns of the measured variable)

There are different parameters to define how well the observed data (experimental semivariogram often called) fits the theoretical semivariogram (the semivariogram that we model). In this study we use the r^2 provides an indication of how well the model fits the variogram data. And the proportion of C/C_0+C this statistic provides a measure of the proportion of sample variance (C_0+C) that is explained by spatially structured variance C. This value will be 1.0 for a variogram with no nugget variance (where the curve passes through the origin); conversely, it will be 0 where there is no spatially dependent variation at the range specified.

3.3 Minimum sampling interval distance

In Campas sampling scheme, all sampling points are a minimum of 9 meters further apart. In Masia sampling points are separated a minimum of 4 meters. In Cantallops all sampling points are a minimum of 1 meter further apart. This it means that the variation below this sampling interval distance is not detected by the sampling scheme.

4. Results and discussion

Despite bulk density of the three plots is very similar, some factors could explain the slight differences of mean values of bulk density on the three plots. The highest bulk density found in the Cantallops plot (Table 2) can be related with the forest management which highly altered the first layers of soil. Also the slope, this plot is located in the steepest area of the three plots which promote soil denudation and therefore the bulk density is affected by this factor. Lithology can play another important role because this plot is located in a schist area whereas the others are located in the granite area. This latter reason jointly with the forest management operation which removes the first layers may affect the availability of small rocks when sampling with the 10 cm soil core.

	N	Mean	CV	Min	Max	Skewness
Campas	45	1.11	0.15	0.75	1.44	0.03
Masia	45	1.02	0.24	0.50	1.50	-0.20
Cantallops	45	1.16	0.16	0.85	1.76	0.90

Table 2. Statistics of soil bulk density

4.1 Semi-variogram of the Campas plot

Best model suited in the semivariogram (Figure 2) shows an exponential structure with a nugget variance of 0.001m, therefore no spatial independence detected at the sampling interval of this plot.

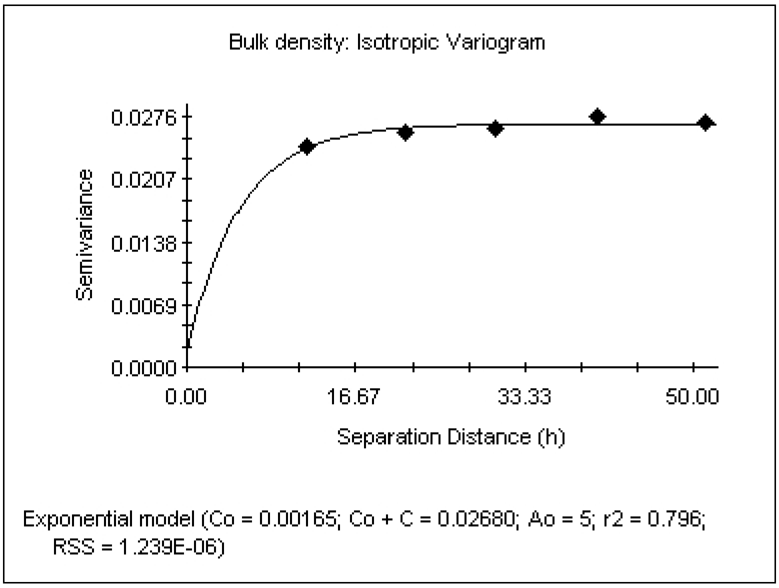

Figure 2. Semivariogram of Bulk density in the Campas plot

4.2. Semi-variogram of the Masia plot

In this sampling scheme we found the best model structure on the exponential model (Figure 3). We have no nugget variance (0.0006) which reveals again no spatial independence of the sampling interval at this plot.

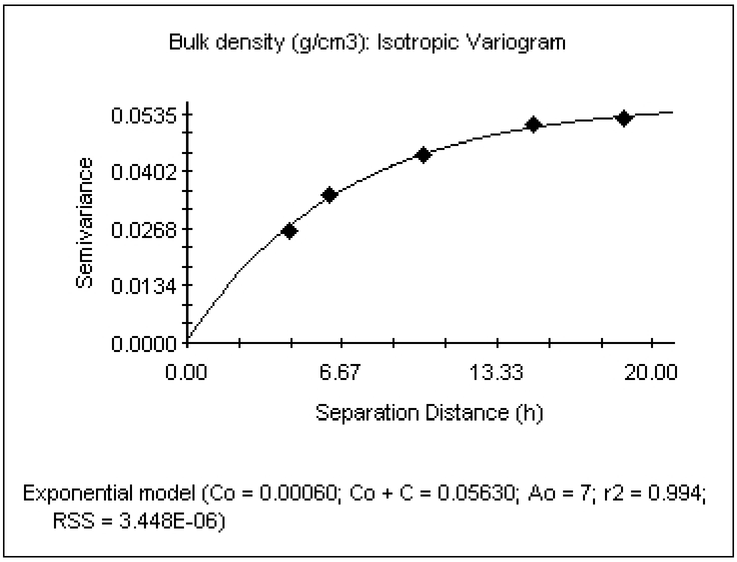

Figure 3. Semivariogram of Bulk density in the Masia plot

4.3. Semi-variogram of the Cantallops

In this plot the model and nugget variance fitted differ from the other two plots (Figure 4). Now it is the linear model and the nugget increase to 0.06. It is important to notice the lack of structure of this semivariogram due to the sill, the structural component, has the same value as the nugget variance. Webster and Oliver (2001) state that almost all properties of the soil and natural environment are continuous. If this is certain, then a semivariogram that appears as pure nugget has almost certainly failed to detect the spatially correlated variation because the sampling interval was greater than the range of spatial variation. The semivariogram of this plot has no detectable pattern in the variation. This statement is in some way contradictory with the purpose of the sampling scheme in this plot because has the minimum sampling interval distance among the three plots. One possible reason to explain this contradiction is the removal effect of the forest management machinery on the first layers of the soil. Cantallops plot experienced in March 2006 a forest thinning, which was the most recent of the three plots, to happen before the sampling in August 2007. Forest management with this kind of machinery provokes almost a tillage effect within forest soils. The mechanical machines penetrates 2-3cm into the soil upper layers breaking its structures and therefore provoking an impact on soil bulk density. In some area of the plot they remove soil upper layers whereas in other areas accumulates soil. In some instance this action provokes a patchy effect in soil distribution which can be detected with the lack of structure in the semivariogram model fitted of this plot.

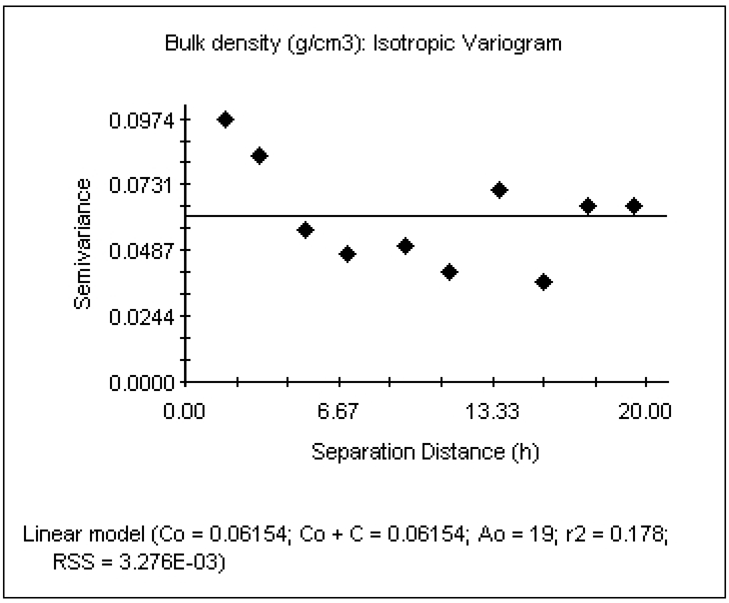

Figure 4. Semivariogram of Bulk density in the Cantallops plot

4.4. Cross validation

Cross-validation analysis is a means for evaluating effective parameters for kriging interpolations. In cross-validation analysis each measured point in a spatial domain is individually removed from the domain and its value estimated via kriging as though it were never there. In this way a graph can be constructed of the estimated vs. actual values for each sample location in the domain. The regression coefficient at the bottom of the graph represents a measure of the goodness of fit for the least-squares model describing the linear regression equation. A perfect 1:1 fit would have a regression coefficient (slope) of 1.00 and the best-fit line (the solid line in the graph below) would coincide with the dotted 45-degree line on the graph. Campas; the lowest correlation coefficient is found in the cross validation of the campas sampling scheme, 0.288. Masia; the values obtained of regression coefficient in the cross validation of this plot are the most satisfactory. r= 0.891. Cantallops; the value of the regression coefficient in the cross validation obtained for this plot -0.605. This fact denotes that where the predicted values increase, the actual data decrease its value and viceversa.

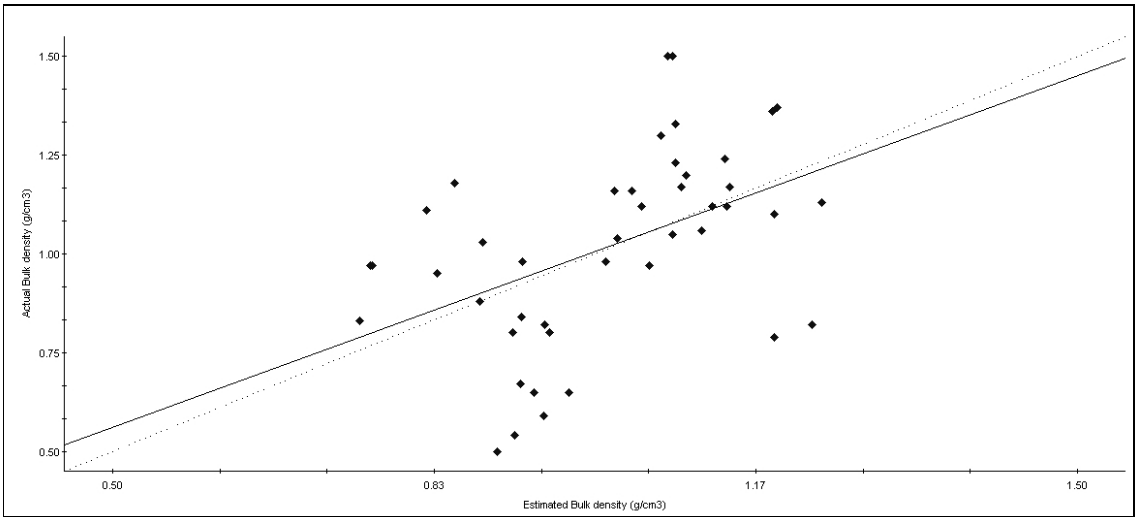

Figure 5. Cross validation: regression between observed vs actual for the Masia plot

5. Conclusions

Regarding the most satisfactory sampling scheme, the triangular systematic showed the best results not only for the robust of its semivariogram also because gave the best fit of correlation in the cross validation test. Nested and quadrangular sampling schemes did not achieve good results of cross validation when modeling with bulk density data. They might return better results when modeling other soil properties different from bulk density, which in our case was exposed to forest management during recent years. Semivariogram seems a good detector of soil disturbances when comparing two or more different plots exposed to forest management. Also it is important to highlight the capacity of the semivariogram to inform about the state of the recovery of that soil after any disturbance; flooding, forest fire, forest management, etc. The hypothesis launched previously was confirmed by the results of the bulk density measurements. In the Cantallops plot- the most recent to be thinned, 2006- the bulk density is higher than the in the other two plots. The removing and mixing effect of the forest management machinery when accomplished during periods of high probability risk of Mediterranean storm rainfall can trigger more sediment availability to be eroded downslope within the hydrological system.

6. Acknowledgments

This study was supported by Spanish Ministry of Science and Technology, project: CGL2006-11107-C02-02/BOS “Evaluation of the quality of Mediterranean soils affected by fire in a middle and large term” and European Regional Development Fund (FEDER) Funds. Also the first author of this study wants to acknowledge the support by the Ministry of Education and Science of Spain with a “FPU” grant since May 2007.

7. References

Bellhouse, D. R. (1977) Some spatial designs for sampling in to dimensions. *Biometrika*, 64, pp. 605-611

Matérn, B. (1960) *Spatial Veriation*. Stockholm .Sweden: Meddelanden fran Statens Skogsforskningsinstitut.

Nielsen, D. K., Wendroth, O. (2003) Spatial and temporal statistics. *Advances in Geoecology*. Catena Verlag. Amsterdam.

Page-Dumroese, D., Deborah, S., Jurgensen, M. F., Tiarks Allen, E., Ponder, F., Sanchez, Felipe G., Jr., Fleming, R. L., Kranabetter, J., Marty P., Robert F., Stone Douglas M., Elioff, John D and Scott, D. A. (2006). Soil physical property changes at the North American long-term soil productivity study sites: 1 and 5 years after compaction. *Canadian Journal of Forest research*. 36(3): 551-564.

Stevens, D. L. (2006) Spatial properties of design-based versus model-based approaches to environmental sampling. In: *International Symposium on Spatial accuracy assessment in natural resources and environmental sciences*. M Caetano and M. Painho (eds.), pp. 119-125.

Terrades, J. (1998) La dinámica de la vegetación y los puntos de vista actuales de la ecología en relación a la gestión del medio natural. In ‘*La gestión sostenible de los bosques’*. 3. (Centre Tecnològic Forestal: Solsona)

Webster, R., Welhama, S. J., Potts, J. M and Oliver, M. A. (2006) Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis of variance and residual maximum likelihood. *Computers & Geoscience*. 32, pp. 1320–1333

Webster, R., Oliver M. A., (2001) *Geostatistics for environmental scientist*. John Wiley and Sons. New York.

Yfantis, E.A., Flatman, G.T., Behar, J.V. (1987) Efficiency of kriging estimation for square, triangular, and hexagonal grids. *Math. Geol*. 19, pp. 183–205.