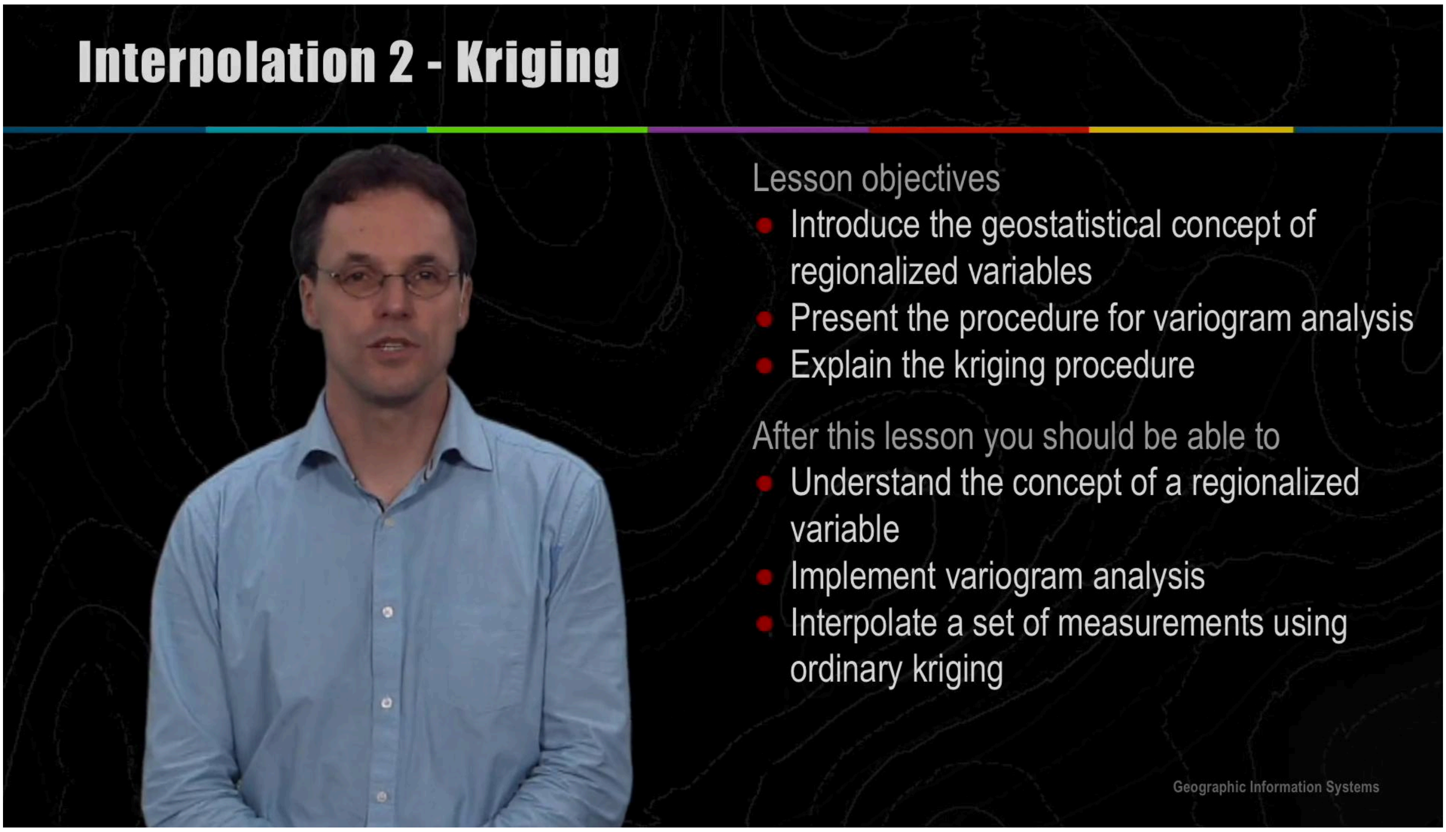
Search MOOC

Video

EPFL

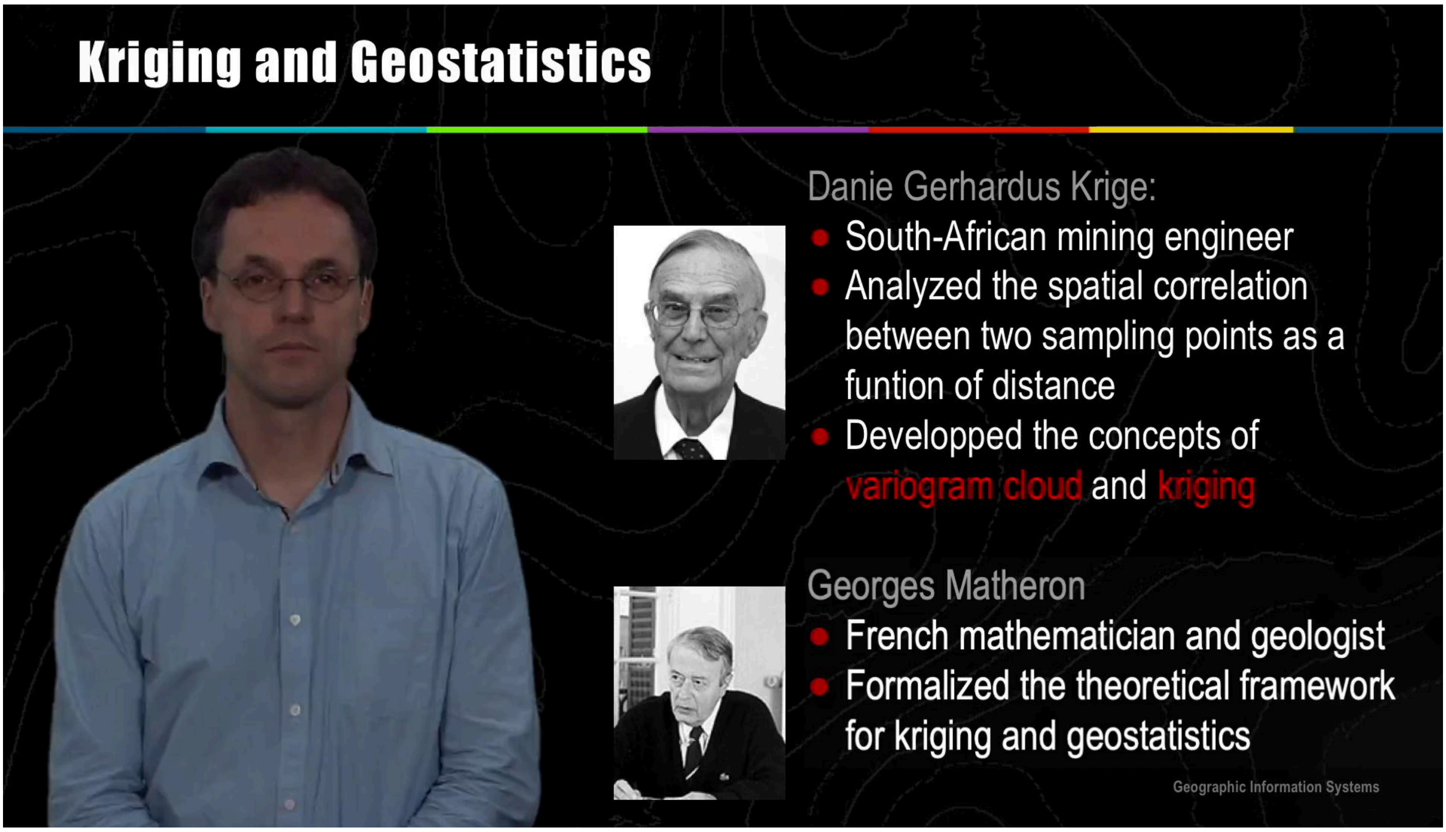


Hello and welcome to this lesson dedicated to geostatistics and kriging, a method of advanced interpolation. We have recently seen that the uncertainty of the results provided by the local deterministic interpolation methods was a problem. More in-depth knowledge on the nature of the spatial distribution were necessary and as early as the 1940s, it is on this subject that Danie Krige and Georges Matheron focused their research works. Their respective contributions have allowed to develop and consolidate the concept of regional variable. The first goal of this lesson is to present the concept of regionalized variable, introduced to describe the particular nature of a variable that characterizes a continuous spatial phenomenon. This type of variable plays a fundamental role in geostatistics. Next, we will introduce concepts and key tools like the variographic cloud or the variogram, before explaining the functionning of the kriging This information will allow you to acquire basic knowledge of geostatistics and to apply a standard variographic analysis procedure and kriging interpolation to any set of measures which characterizes a continuous spatial phenomenon.

Notes

Summary

0m 32s

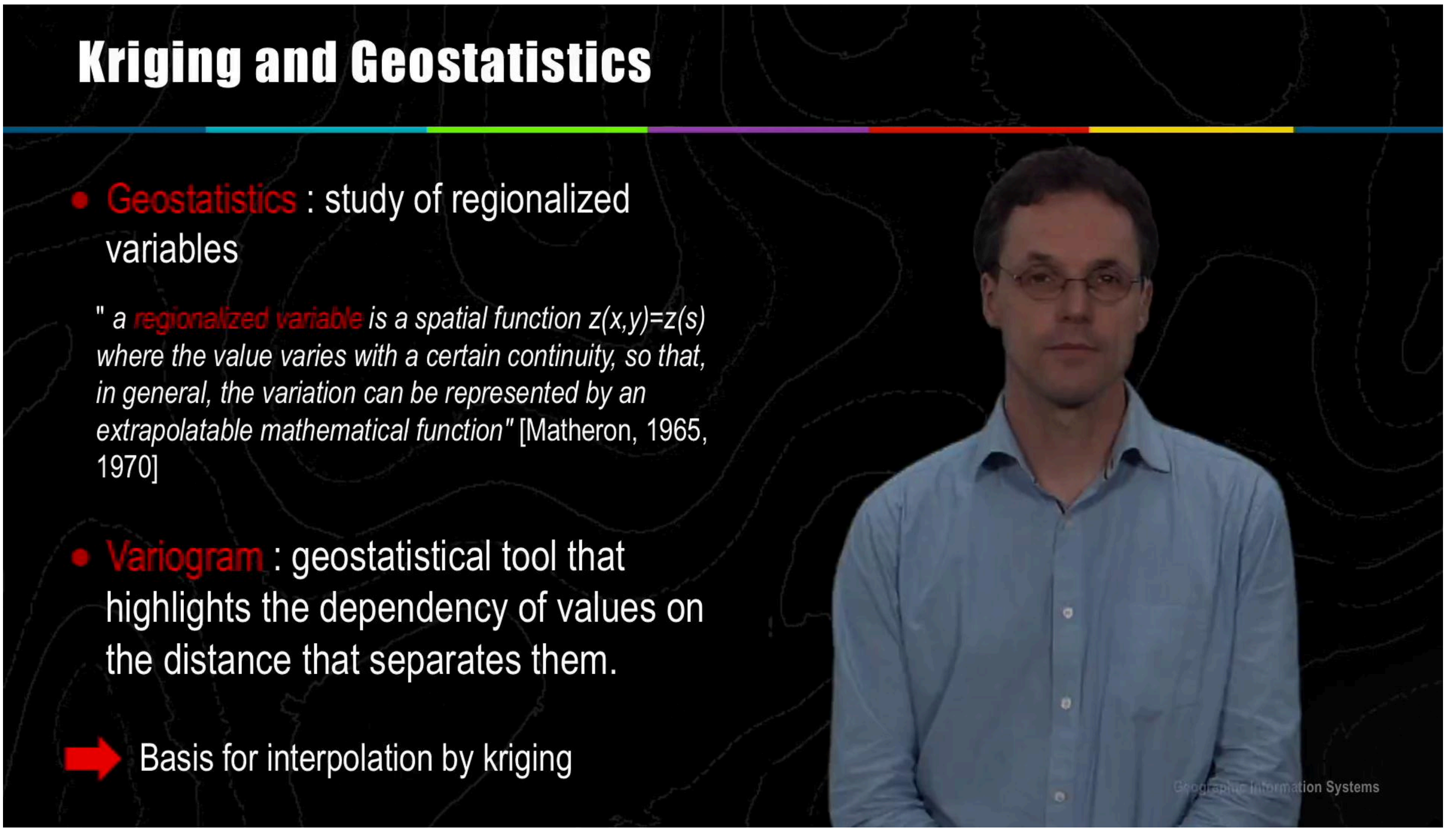


We will start with a brief introduction to geostatistics, then we will follow the course of a procedure of empirical variographic analysis, which will allow to introduce progressively the notions of variographic cloud, of experimental variogram, of surface and directional variogram and of the theoretical variogram, before proceeding to the kriging. To what extent can we estimate a value at any point of a domain by interpolation? And if the answer is yes, what interpolation law is it judicious to apply? Is it linear? Is it quadratic or even of another nature ? These questions are answered if we are able to propose a complete and rigorous theory which allows to model the relation between the behavior of the sample and that of the phenomenon on the whole domain. This problematic is a matter of geostatistics which has its roots in the study of random functions, developed in particular by Kolmogorov and Weiner during the 1930s and 1940s. But we attribute the first works to the South-African mining engineer Danie Krige. Krige examined the spatial correlations between two sampled points depending on the distance between them and he has proposed the variogram to represent them. His approach was then taken up by the French mathematician, Geogres Matheron, at the beginning of 1960s to establish the theoretical foundations.

Notes

Summary

1m 42s

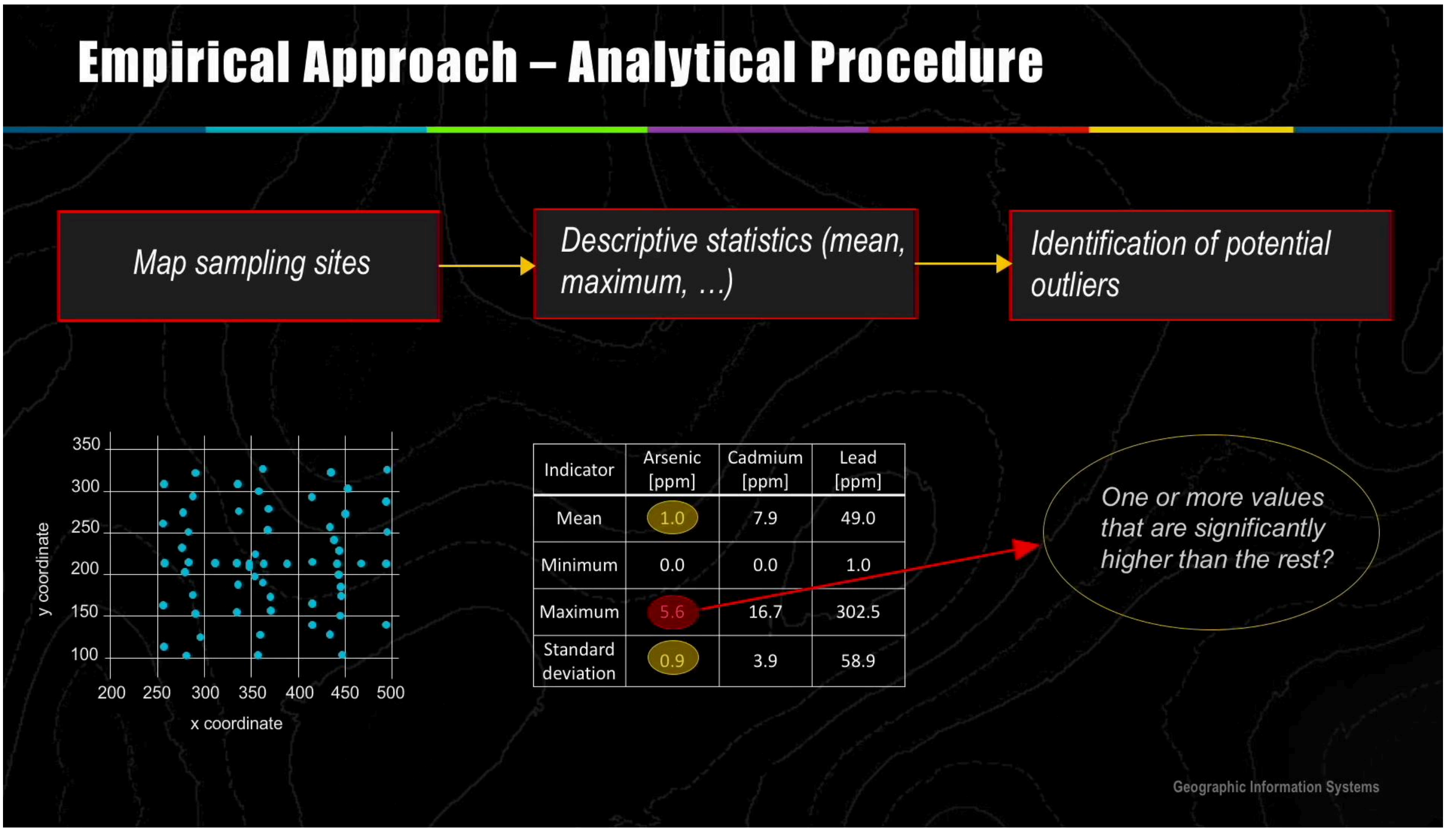


Matheron introduced the concept of regionalized variable to describe the particular nature of a variable characterizing a continuous spatial phenomenon and he appealed to the notion of random function to develop a rigorous and complete theory. The concept of regionalized variable and variogram are the two basic tools at the service of kriging interpolation techniques. We will discover their role and functioning by gradually applying a variographic analysis procedure to a set of empirical data.

Notes

Summary

3m 07s

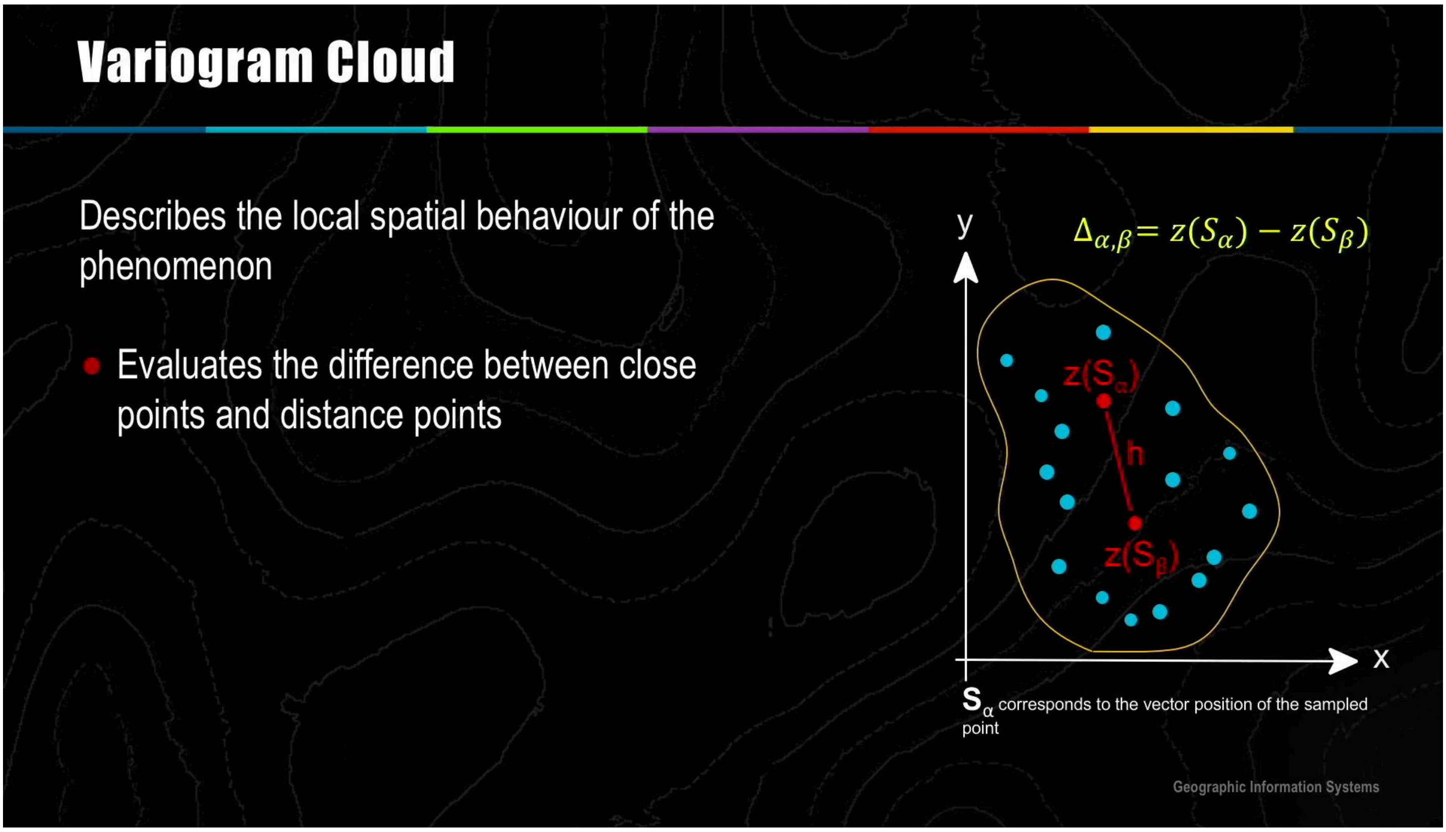


A sampling campaign enabled to collect soil samples and to measure the arsenic, cadmium and lead contents, expressed in parts per million or ppm. Any analysis of this type begins by a descriptive characterization of the dataset, based on statistical indicators. Firstly, we need to get an idea of the data available, with the aim of describing the statistical characteristics of the variable, to observe if the variable behaves in a homogeneous way on the whole domain and also to verify if the variable satisfies the requirements of the statistical tools used for regionalization. We will start by mapping the sampled domain, then calculate a number of statistical indicators which provide a first view of the properties of the measured values. This step allows to detect possible inconsistencies or outlier values that may result from measurement errors.

Notes

Summary

3m 39s

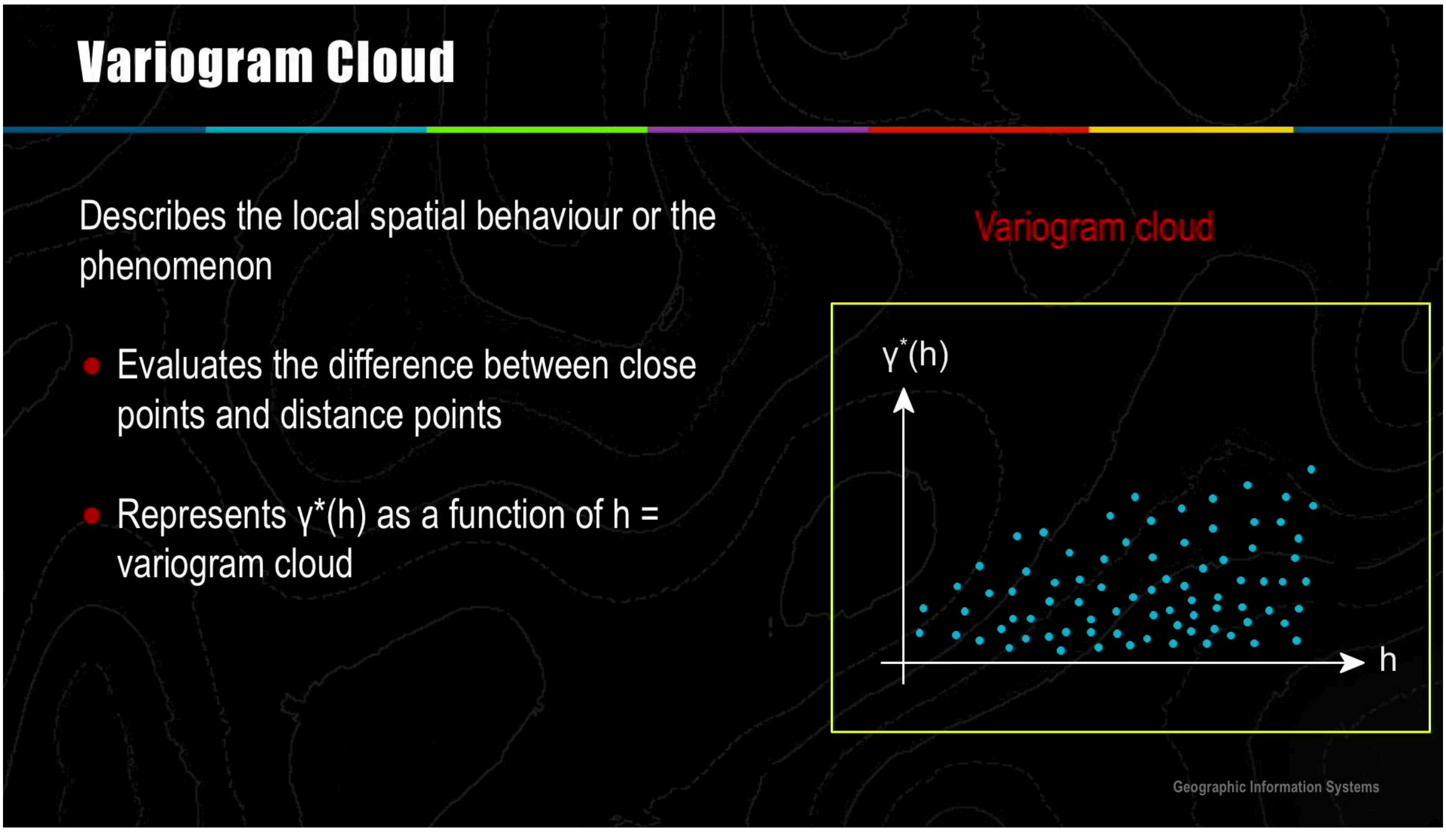


Here in the table, observing a maximum variable for the arsenic, very far from the average, suggests the presence of one or more outlier measures. The indicators of descriptive statistics do not give any indication on the local behavior of the interest variable on domain D. They do not take into account the localization of sampled sites and are only global indicators. It is therefore necessary to highlight the spatial and local behavior of the phenomenon, this is what is called the variographic cloud which allows us to perform this operation. To study the local variation, the principle of the variographic cloud is to analyze the differences of values between pairs of sites close, then distant, and then finally between all the pairs of points in the domain. One possibility is the calculation of the difference between pairs of points and to draw the average, it is the formula that appears in yellow, at the top, on the right. Here, Δ (z) can take positive or negative values so that the average can be null independently of the amplitude of the variation of the variable, it is therefore not a good indicator of the behavior of the variable. On the other hand, the square of the difference enables to avoid this inconvenience.

Notes

Summary

4m 37s

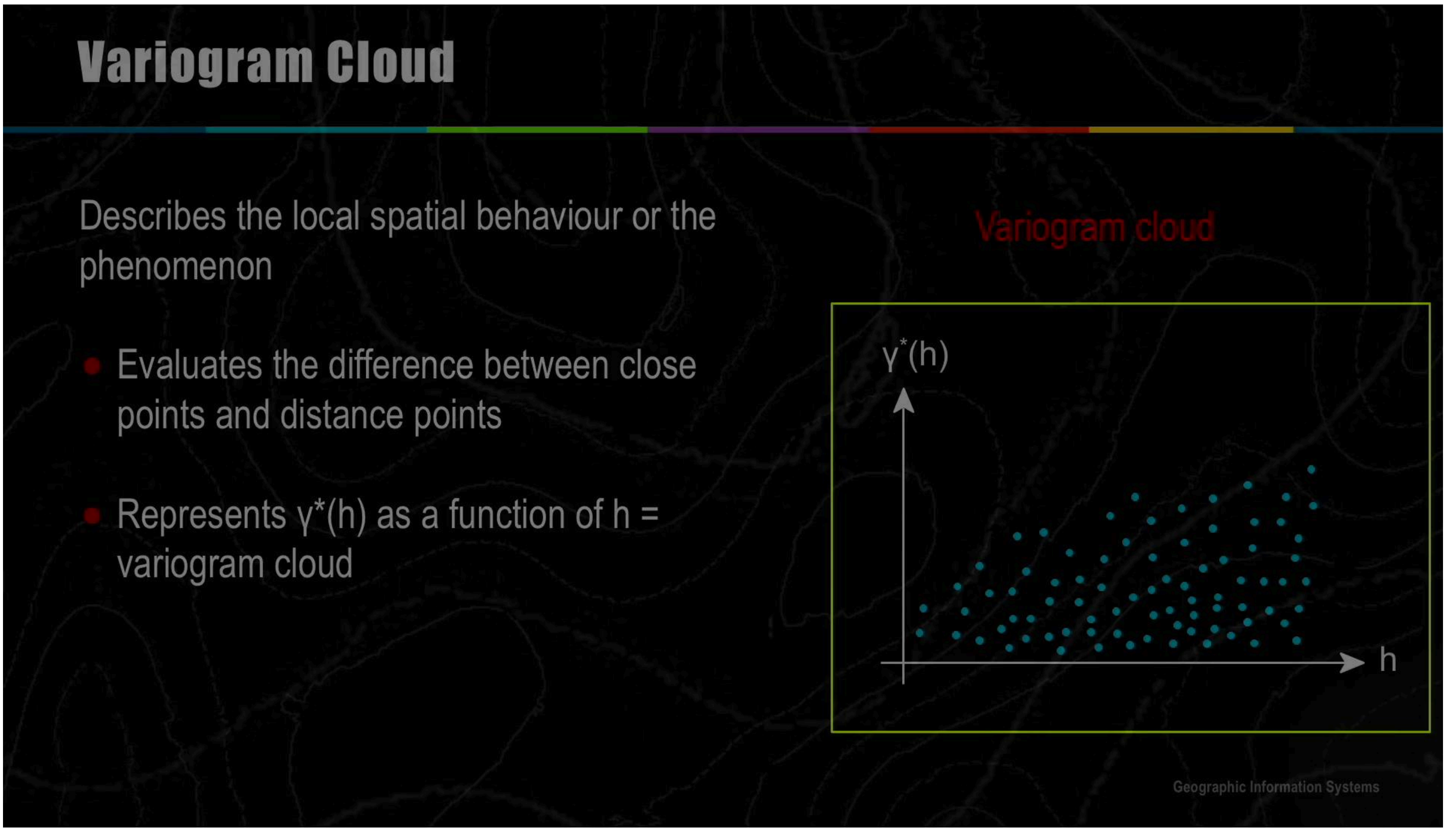


We then create the indicator $\gamma^*(h)$ where h is the distance between the measuring points α and β, and which is constituted of squares of differences which we divide by two to express the variation on one point only, this is the quadratic half-difference. To analyze the behavior of all pairs of points, we use a dispersion diagram of the quadratic deviations, actually called variographic cloud. This diagram shows the valuesof γ * (h) in relation to h.

Notes

Summary

6m 00s

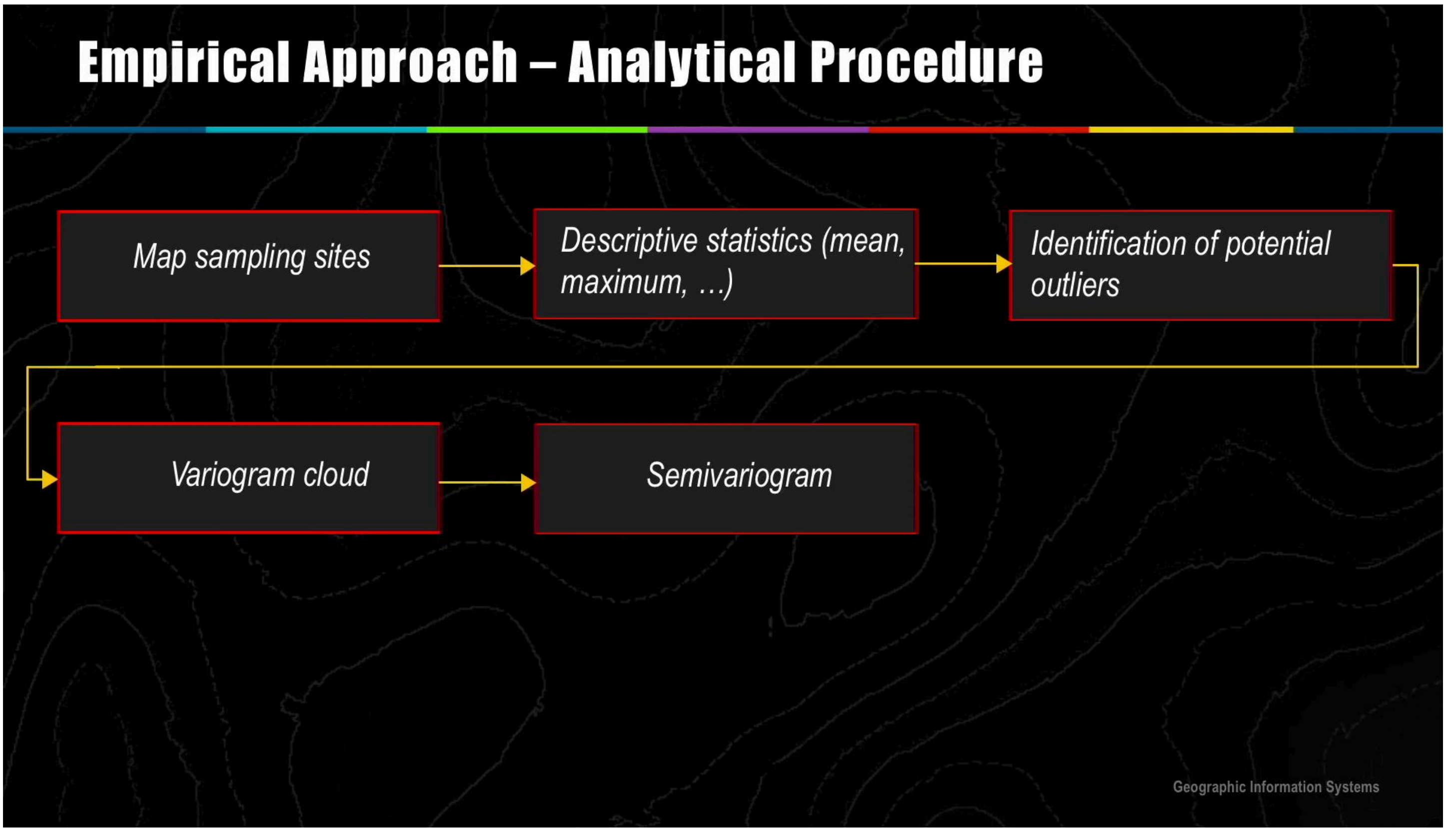


The pairs of points are represented only once. It is important to note that this approach does not take into account the orientation of the pairs of points, the variographic cloud is said to be omni-directional.

Notes

Summary

6m 29s

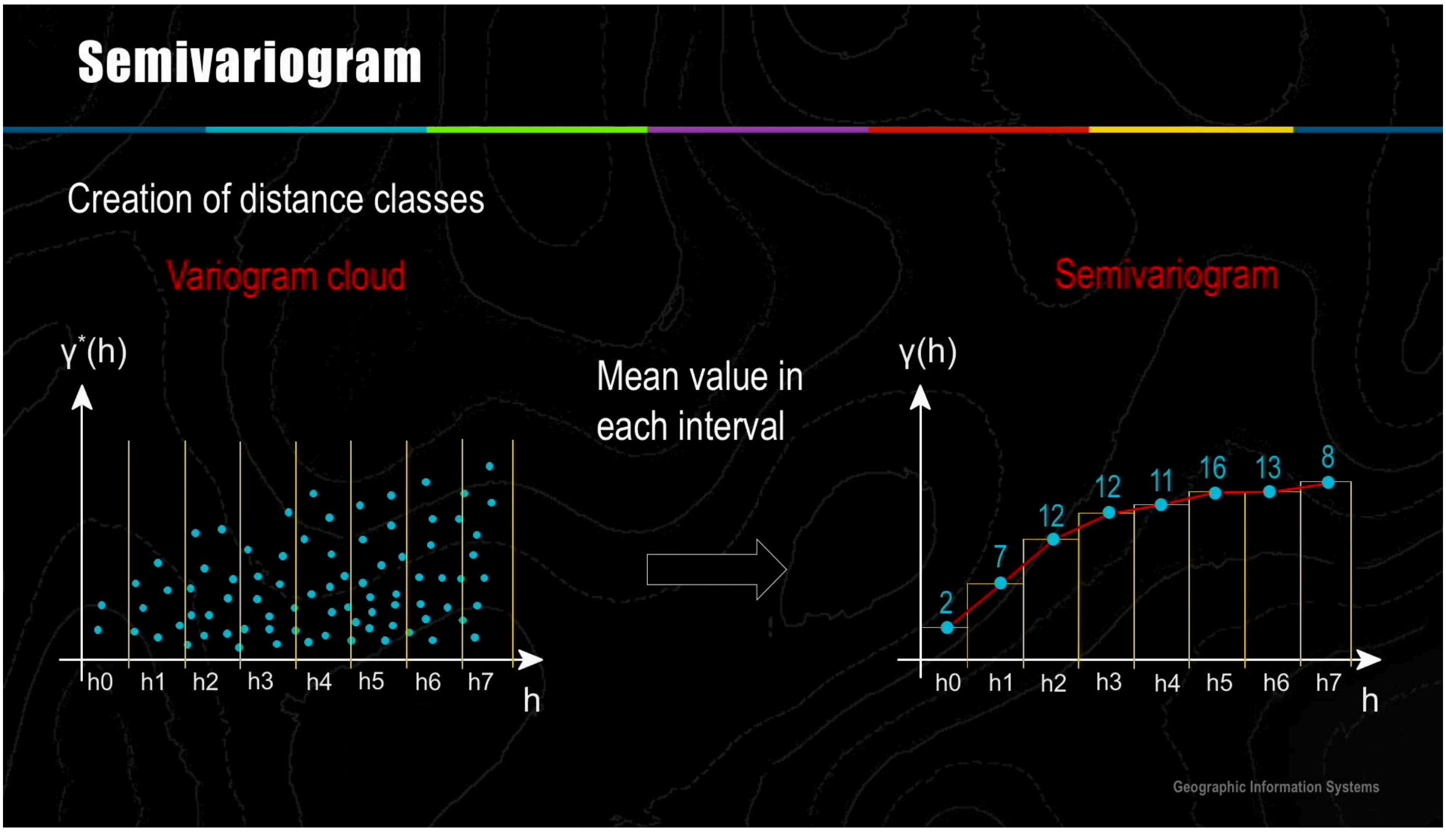


Let's move on to the next step in the analysis process. The variographic cloud produced a large number of pairs and it is not easy to draw a teaching on the dispersion of quadratic half-deviations.

Notes

Summary

6m 51s

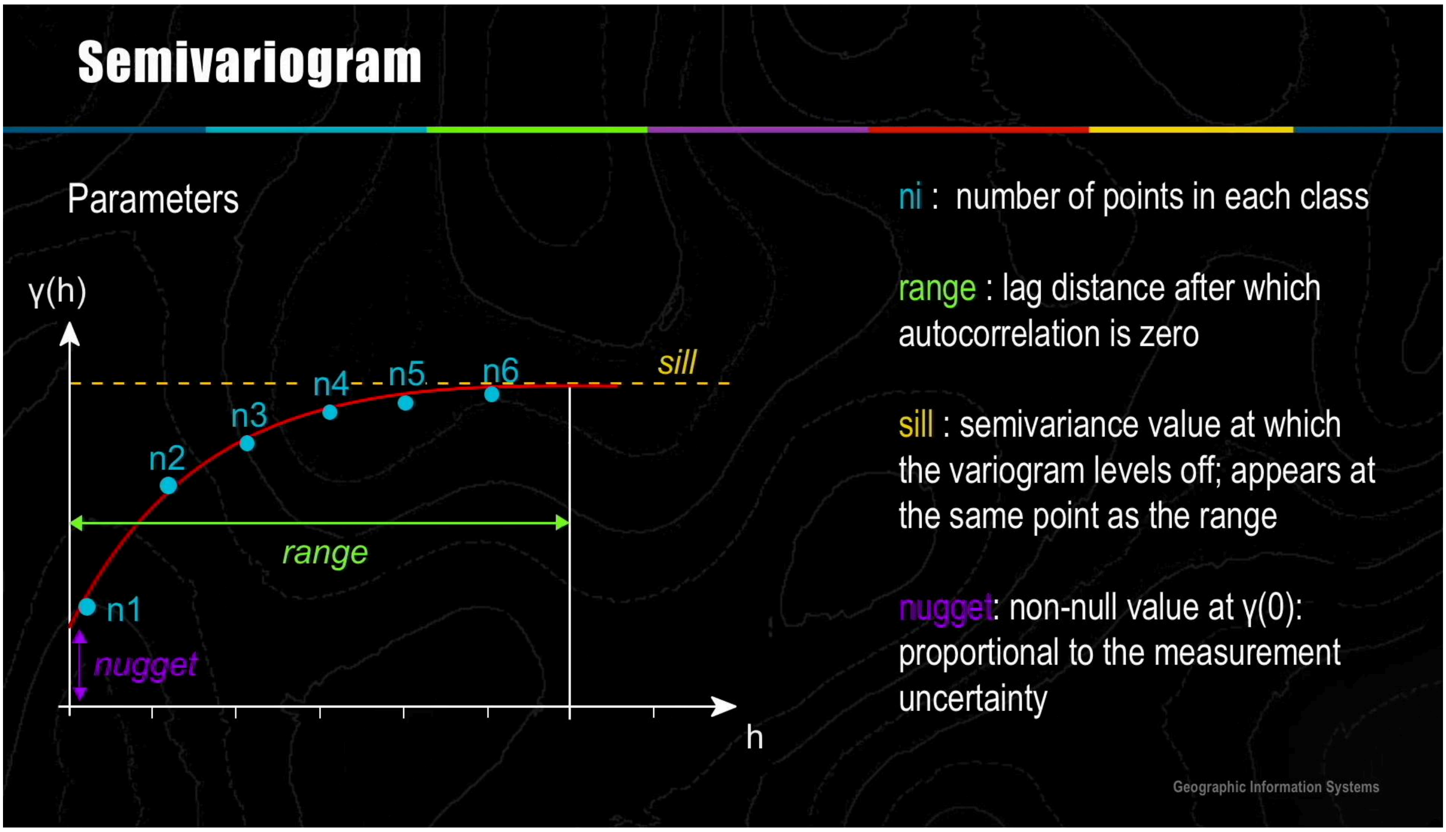


This is achieved by segmenting the domain into distance classes, symbolized here by the orange bars, similarly to a histogram. A lag Δ (h) is chosen, and for each interval from h (o) to h (n), we calculate the average of the values of γ * (h) We will see later on the basis of which criteria the value of the lag is established. From the moment we consider an average per distance classes, the quadratic dispersion is denoted by γ (h), without the asterisk, and takes the name of experimental variogram. The blue numbers, above the mean points of the histogram, denote the number of pairs used for calculation.

Notes

Summary

7m 03s

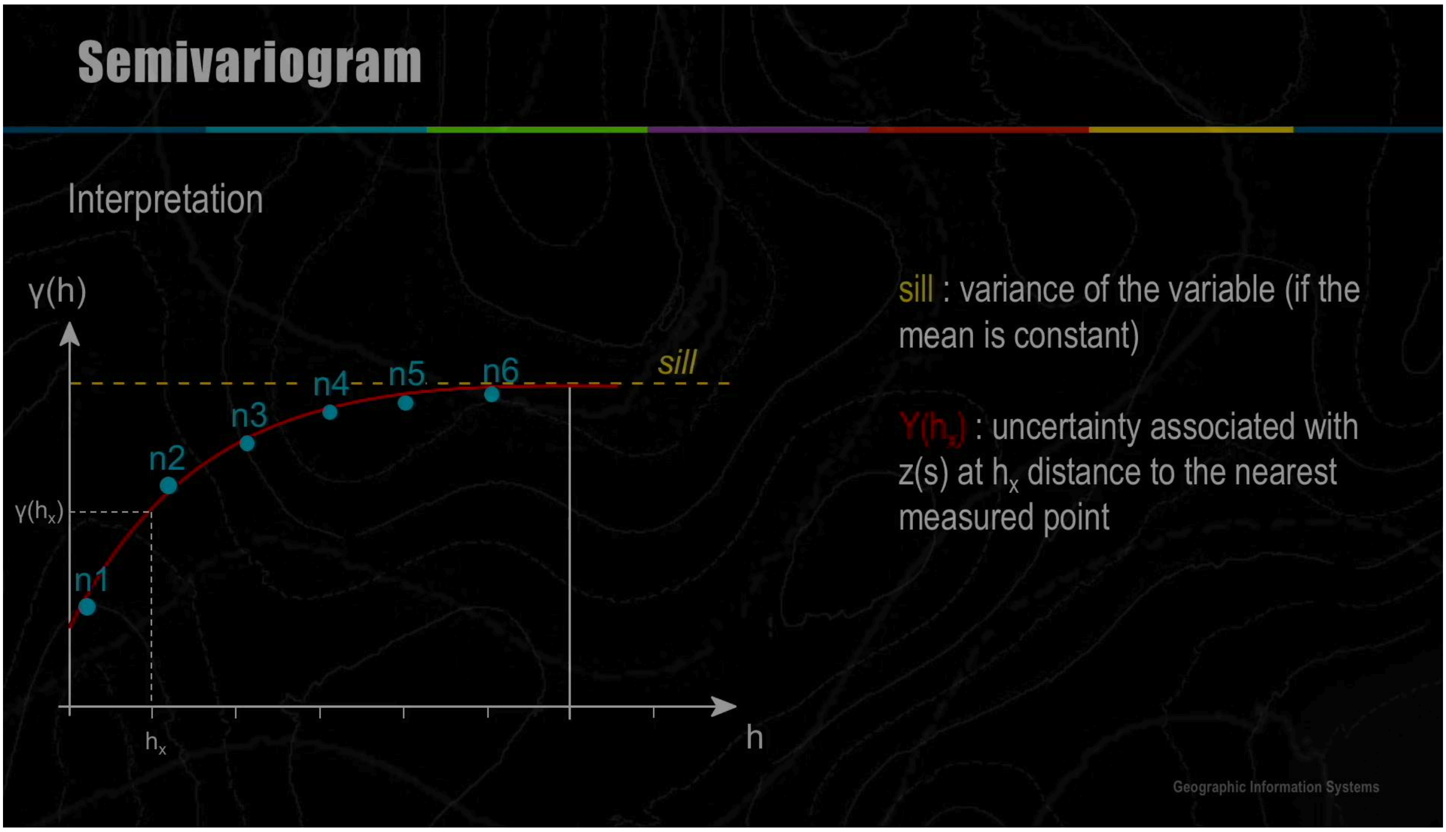


The higher the number of pairs, the more representative the lag average is. A value produced by too low a number of pairs is not acceptable in statistical terms. In this case, it is necessary to increase the size of the lag so that a larger number of pairs is included. The line connecting the points facilitates the interpretation of the variogram. The experimental variogram is characterized by various parameters: the range is the distance from which the auto-correlation disappears, the bearing is the plane part of the variogram, supposed to appear beyond the range, the nugget is determined by the intersection of the line of the experimental variogram with the ordinate and the nugget of the variogram represents the uncertainty about the measurement in each point.

Notes

Summary

7m 47s

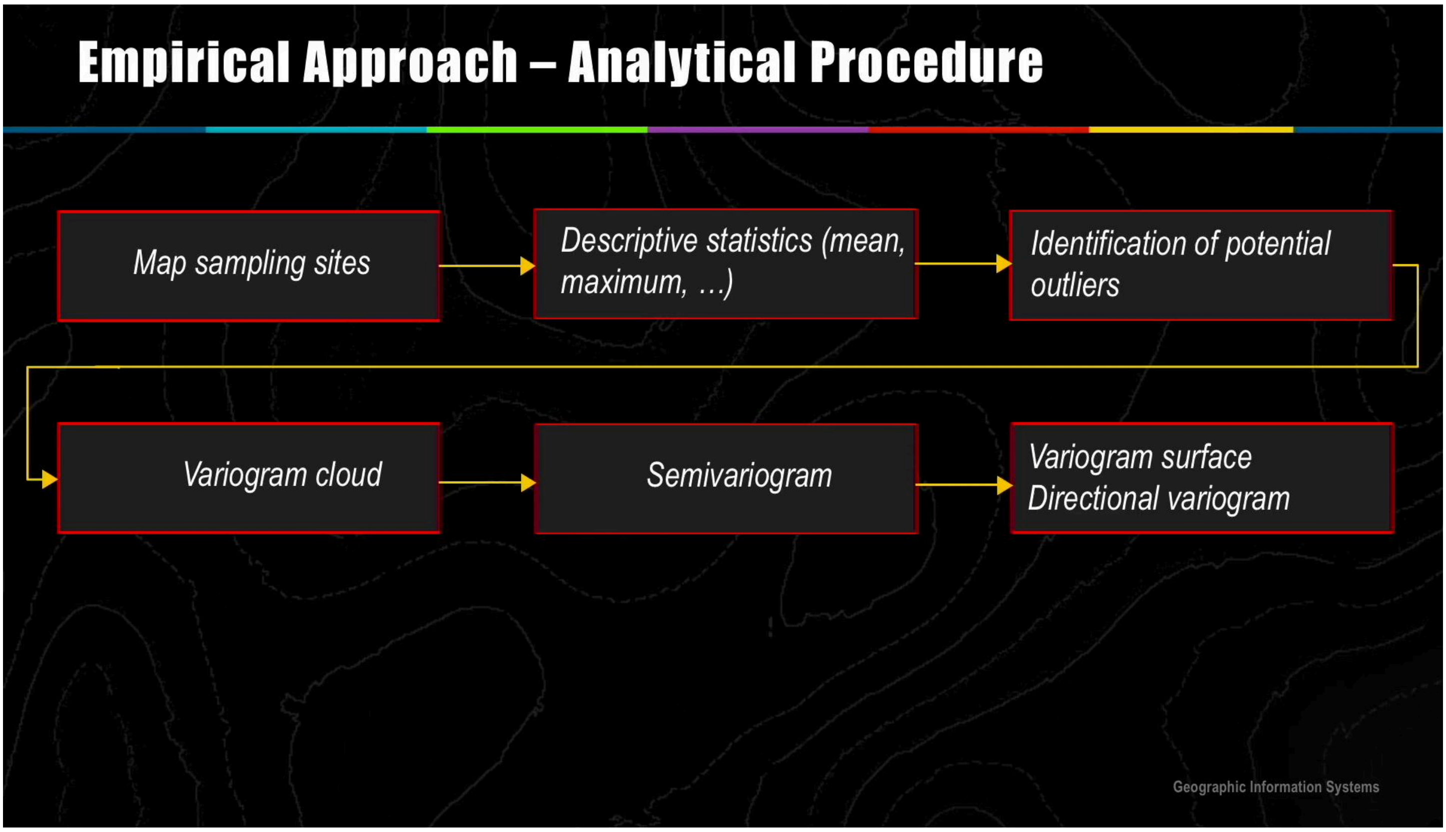


The name of nugget was proposed by Krige in the context of gold extraction in the mines of South Africa, the presence of a nugget in fact creating a discontinuity in the density of the ore. A pair of very close measurement sites produces therefore for one, a very low content, and for the other, a very high content, which produces uncertainty. Theoretically, the height of the bearing corresponds to the variance of the variable. If this assumption is respected, The value of γ (h(x)) can be interpreted as the uncertainty which affects a content z estimated of s and moved away from a distance h(x) from the nearest point of support. In practice, the coincidence bearing-variance is not often observed, but the proposed interpretation in terms of uncertainty remains valid.

Notes

Summary

8m 32s

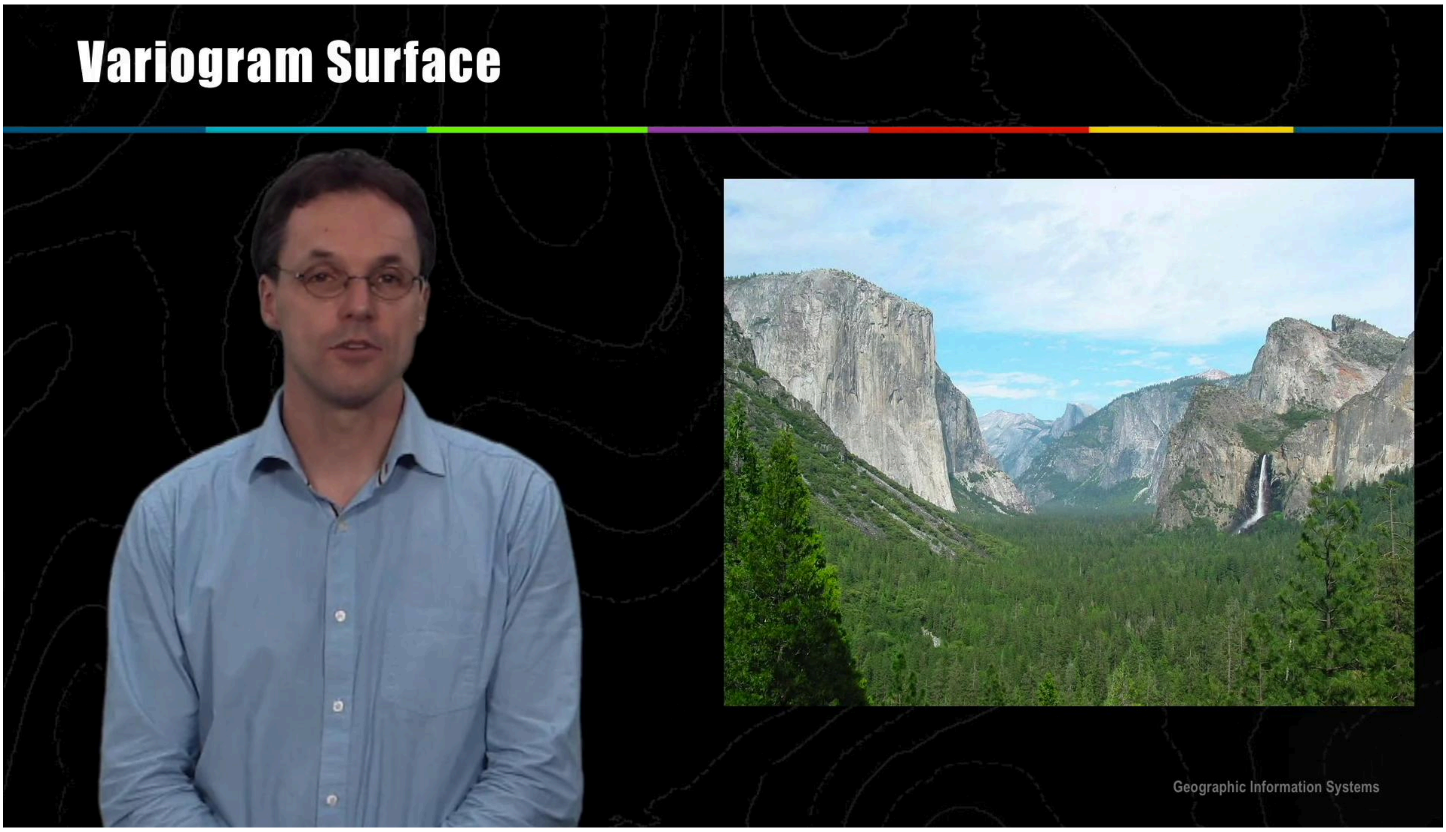


We will now explain How to take this orientation into account.

Notes

Summary

9m 38s

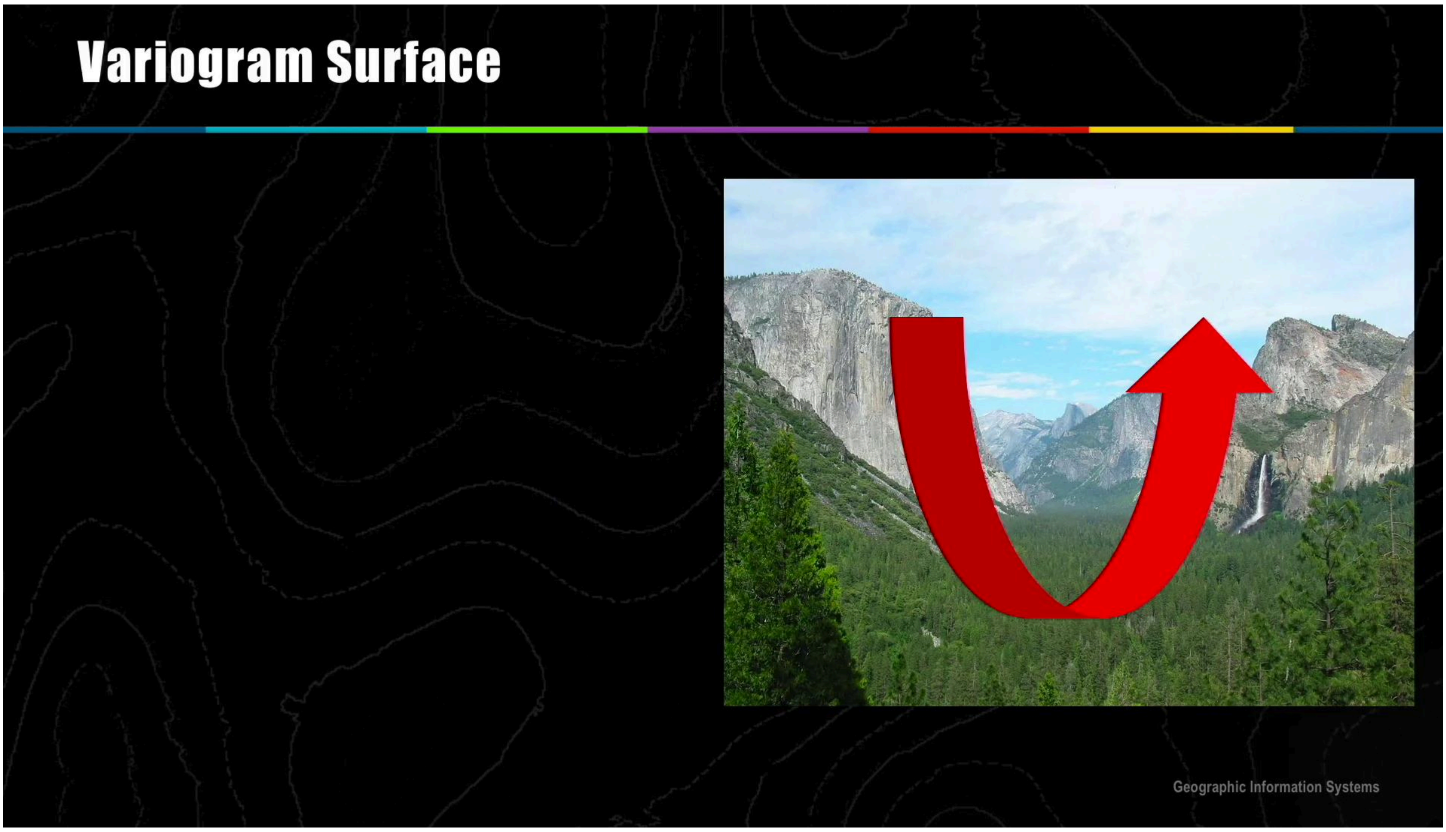


If the phenomenon manifested a different behavior according to the orientation, in other words if it was anisotropic, the omni-directional variogram would not reveal it. How then do we highlight such a potential behavior? Let's do a little thought experiment to start with in order to visualize an anisotropic regionalized phenomenon. The relief of a valley expressed by the altitude is a good example.

Notes

Summary

9m 42s

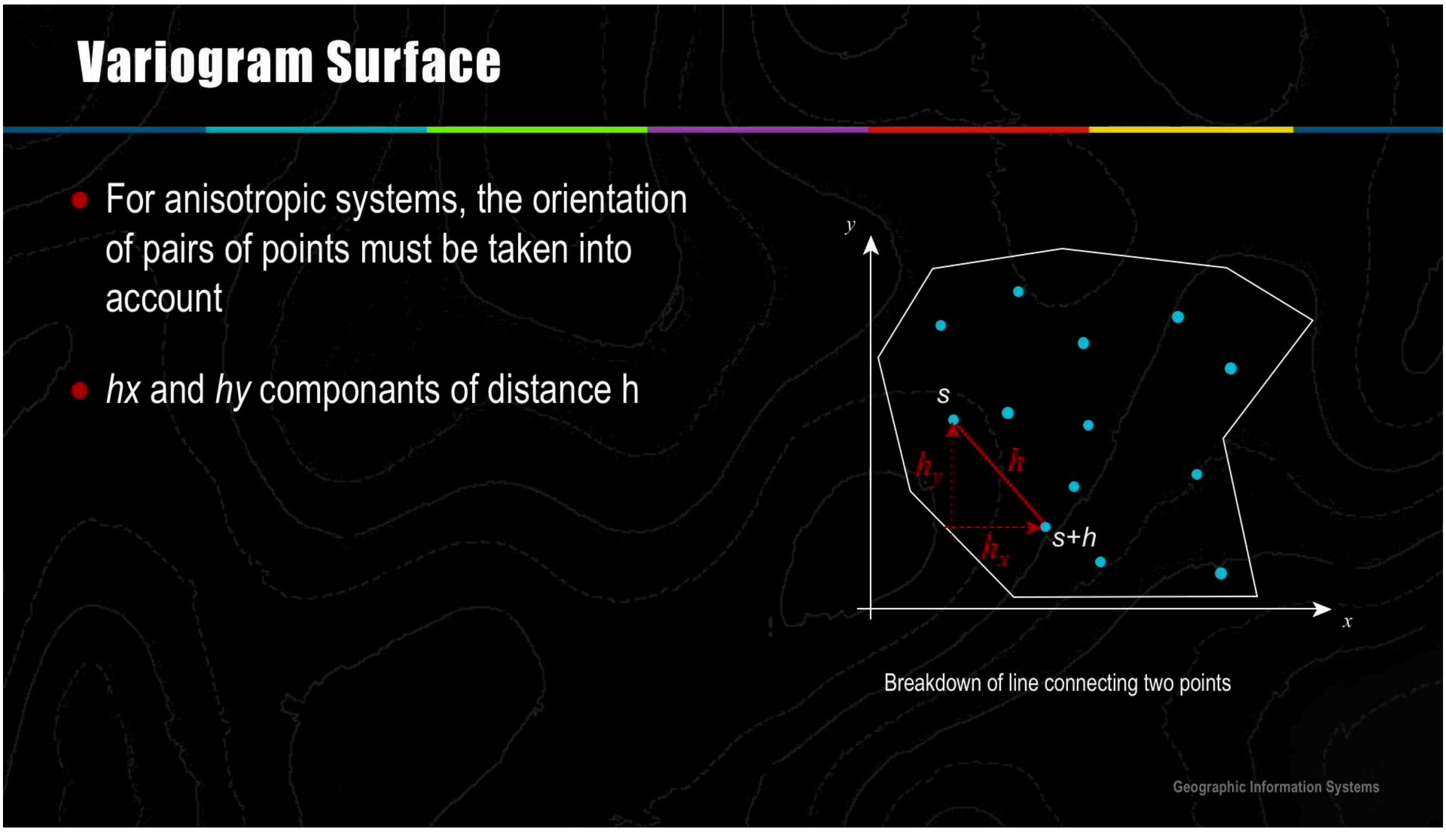


Its behavior is different if we observe it in the direction of the watercourse or on the contrary perpendicularly to the latter.

Notes

Summary

10m 09s

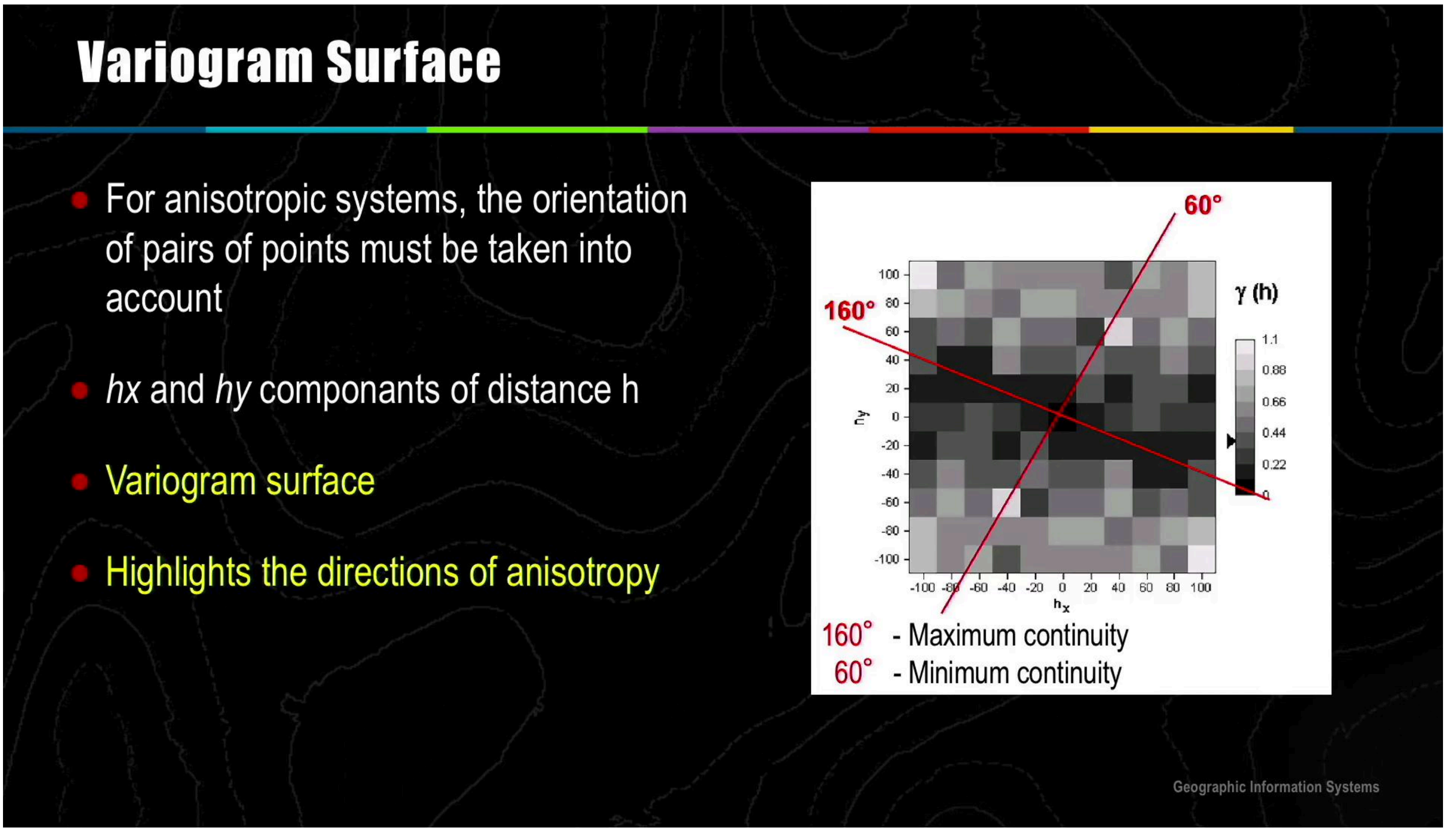


The solution is to represent the variographic cloud or the variogram on a diagram whose axes are the components h of x and h of y of the line connecting the pairs of points. The values of γ(h) are carried perpendicularly to the plan h of x, h of y.

Notes

Summary

10m 17s

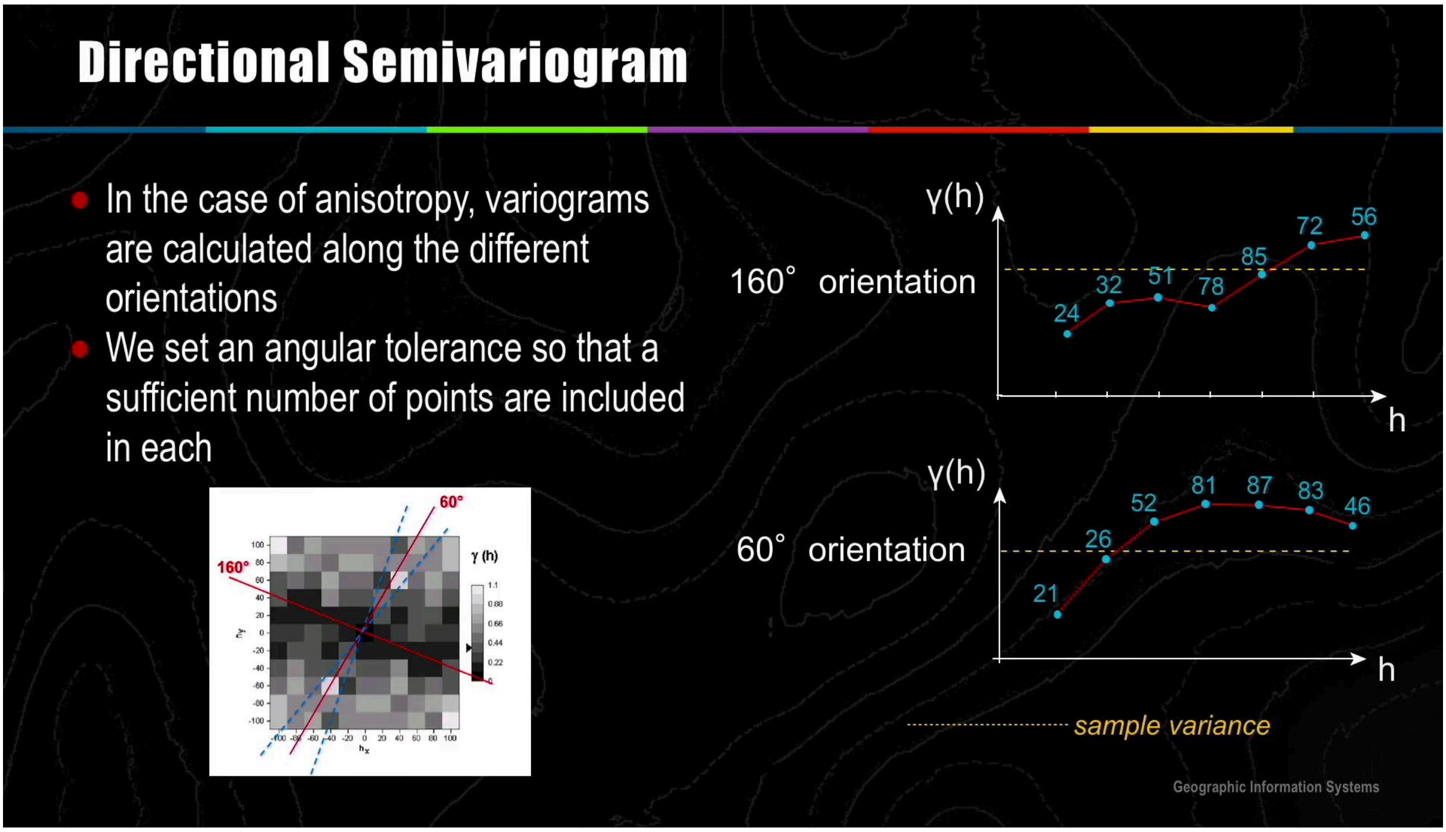


Such a variogram corresponds to a bivariate histogram. It is called a surface variogram. The figure here on the screen presents a surface variogram of the arsenic content which highlights the main directions of anisotropy of its spatial behavior. The γ(h) values are represented in shades of grey according to the scale on the right of the image. This variogram is constructed on the basis of five lags of 20 meters. The arsenic concentration presents an anisotropic behavior. It shows a maximum continuity in the 160 degree direction and minimum in the 60 degree direction. In general, if the surface variogram is uniformly distributed around its center, the behavior of the variable is isotropic. The isovalues form concentric circles. But if the isovalues form concentric ellipses, there is a geometric anisotropy and the bearing is the same in all directions. And if the bearing is not the same in all the directions, it is then called a zonal anisotropy.

Notes

Summary

10m 37s

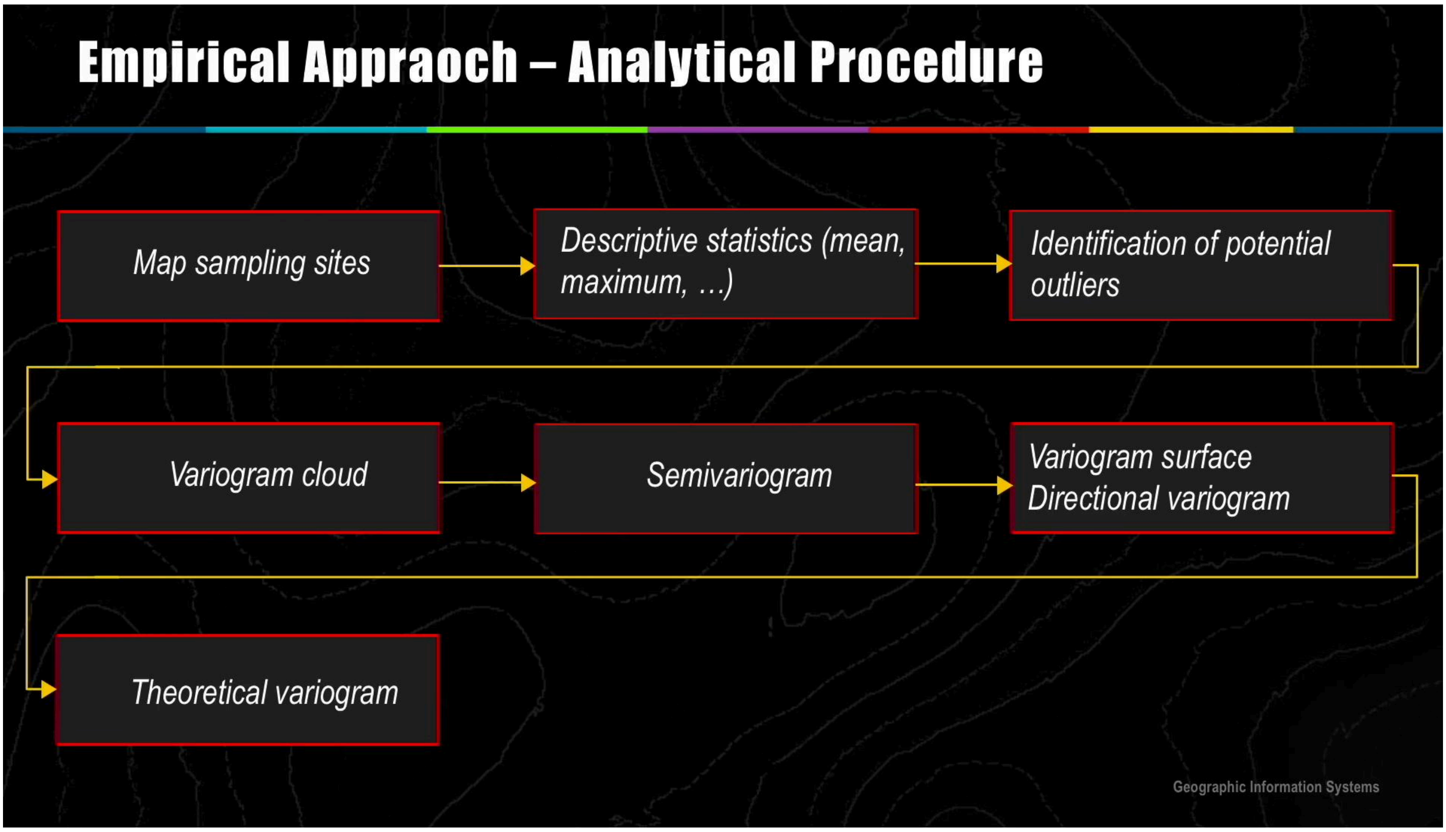


When an anisotropy is detected, we calculate variograms in the observed directions. In this case, in order to avoid a too restricted number of pairs to be taken into account, we define an angular tolerance. We will therefore include all the pairs whose orientation is within a main direction interval more or less a certain tolerance. For example, 60 degrees for the main direction and more or less 20 degrees. We have represented here the calculated variograms on one hand for the direction of the maximum continuity of the arsenic concentration, 160 degrees, and for its perpendicular direction, 60 degrees.

Notes

Summary

11m 42s

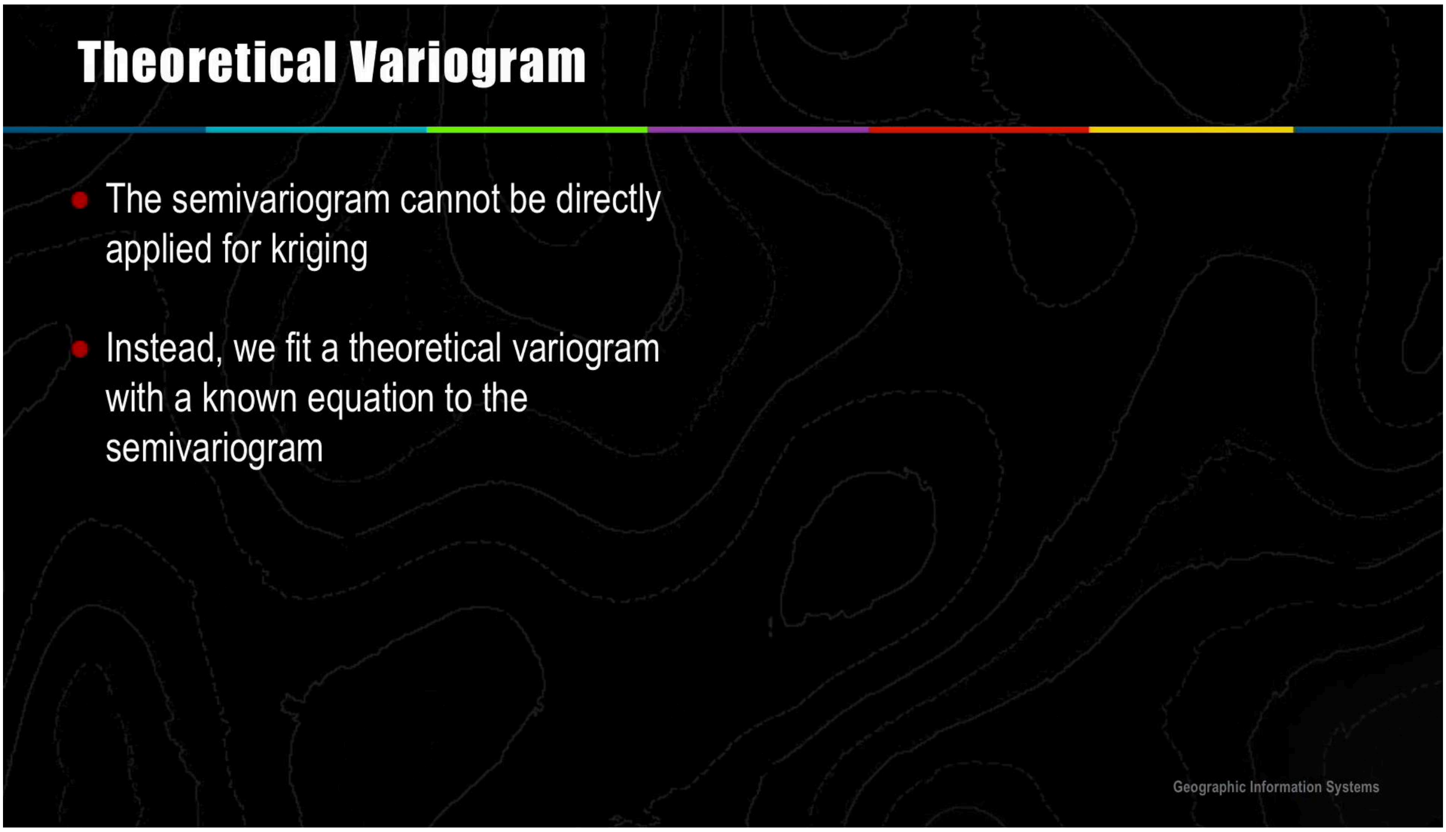


We have seen that the experimental variogram enables to express the dependence between the values measured between two points constituting a pair of sites depending on the distance between them. This dependence can be interpreted as a weight that we will attribute to the measurement point in an interpolation procedure. So, how do we exploit such a property?

Notes

Summary

12m 31s

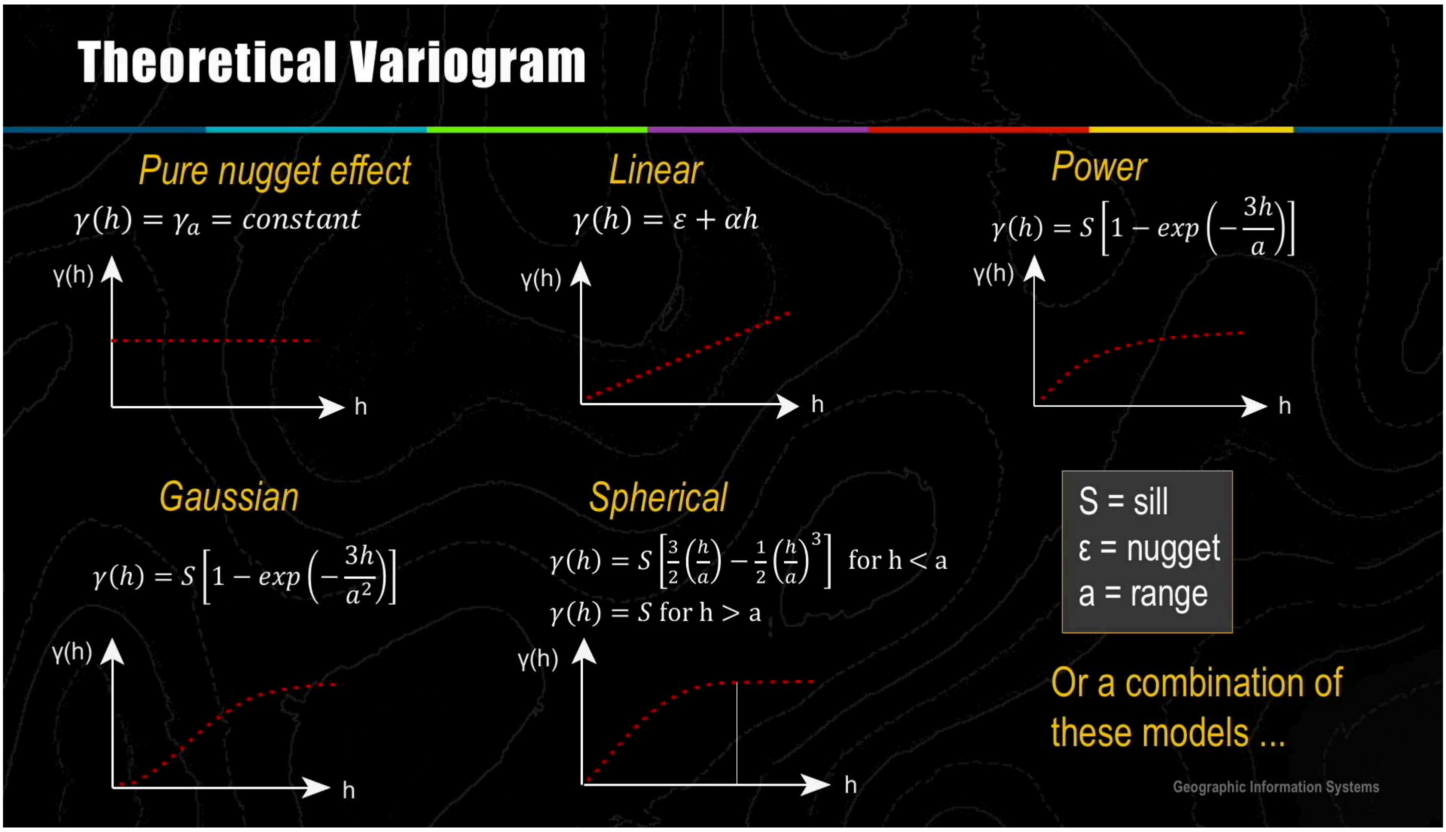


To implement the optimal interpolator which is the kriging and that we will present in a moment, the sampled spatial variables must be interpreted in terms of random variables. However, the constraints fixed on the random variables cannot directly exploit the information provided by the experimental variogram to match the reality and the solution of the kriging equations. The experimental variogram must be adjusted by a line which can only be the result of a limited number of authorized functions for the modelling of the theoretical variogram.

Notes

Summary

12m 53s

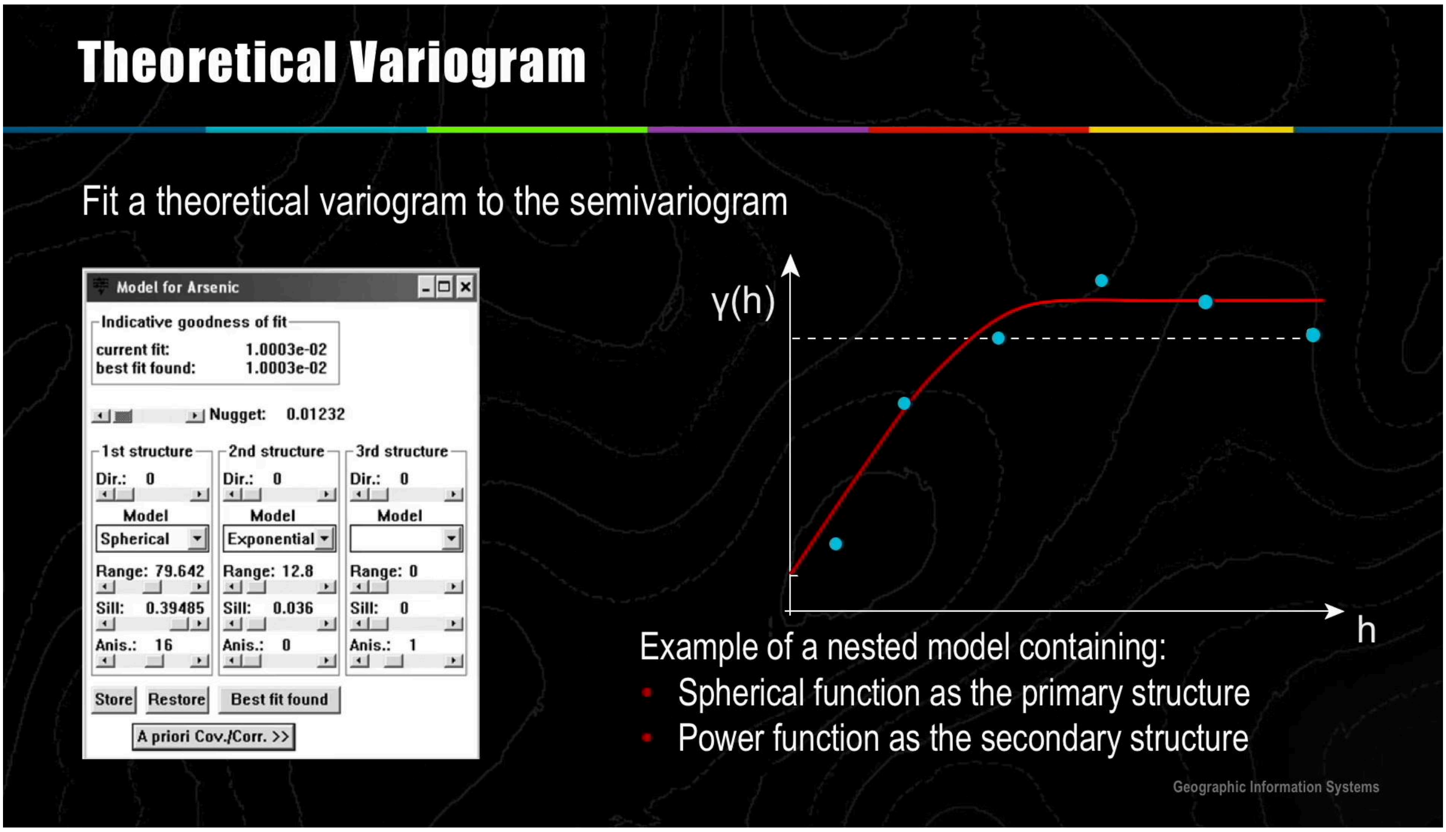


These authorized functions are the pure nugget model with a γ(h) which is constant, the linear model, the exponential model, the gaussian model and the spherical model. The combinations of these different functions are also permitted.

Notes

Summary

13m 30s

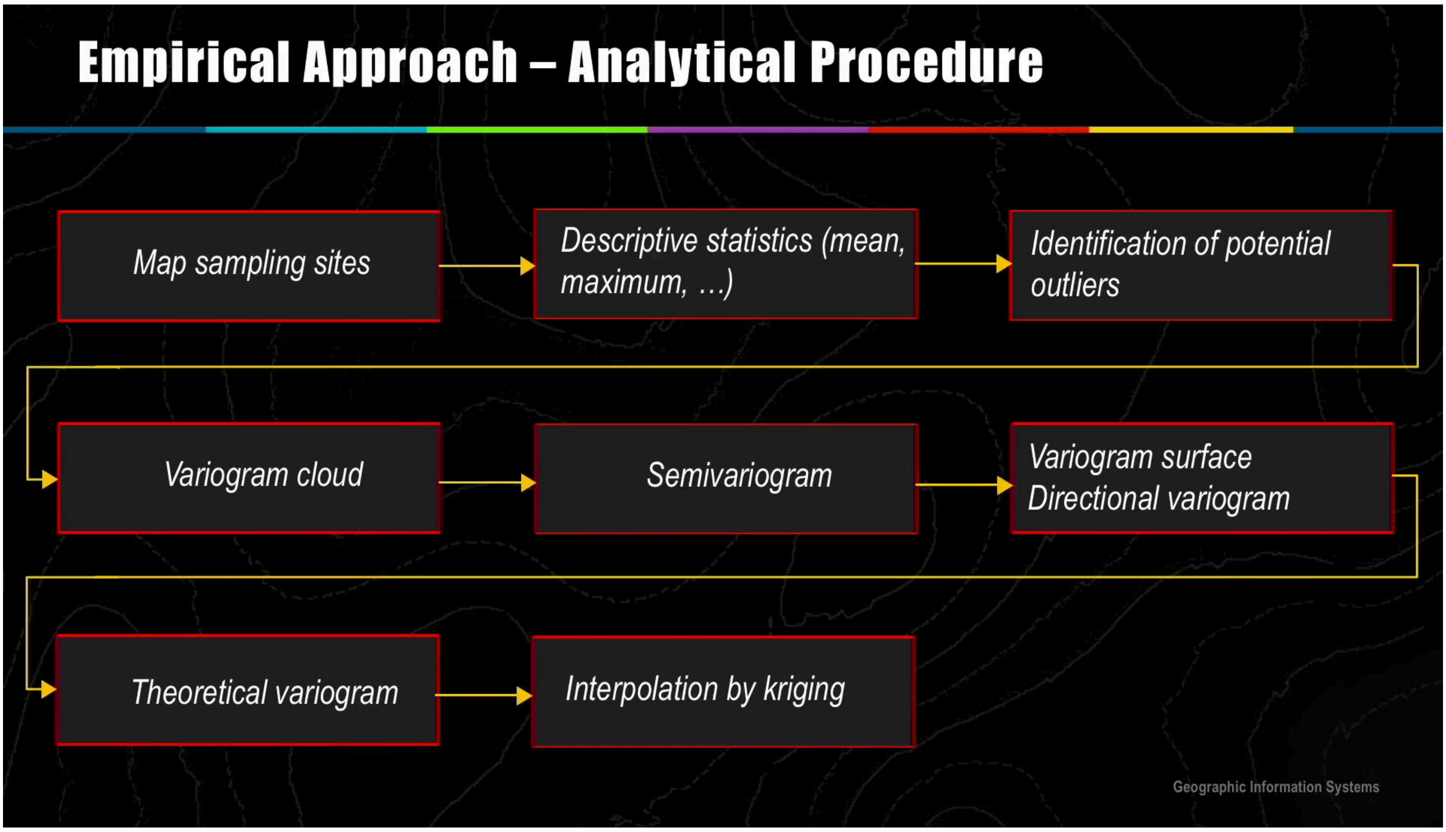


However, we must be aware that only one variogram is defined for the whole domain and for all the directions, even if the analysis has highlighted an anisotropy. In general, variography softwares provide an adjustment mode and calculate the parameters necessary to enable the interpolation by kriging. Here on the screen, we have an example of a parametrization interface allowing the adjustment by a theoretical variogram of the γ(h) values by a combination of functions, so a spherical function for the first structure and an exponential function for the second.

Notes

Summary

13m 49s

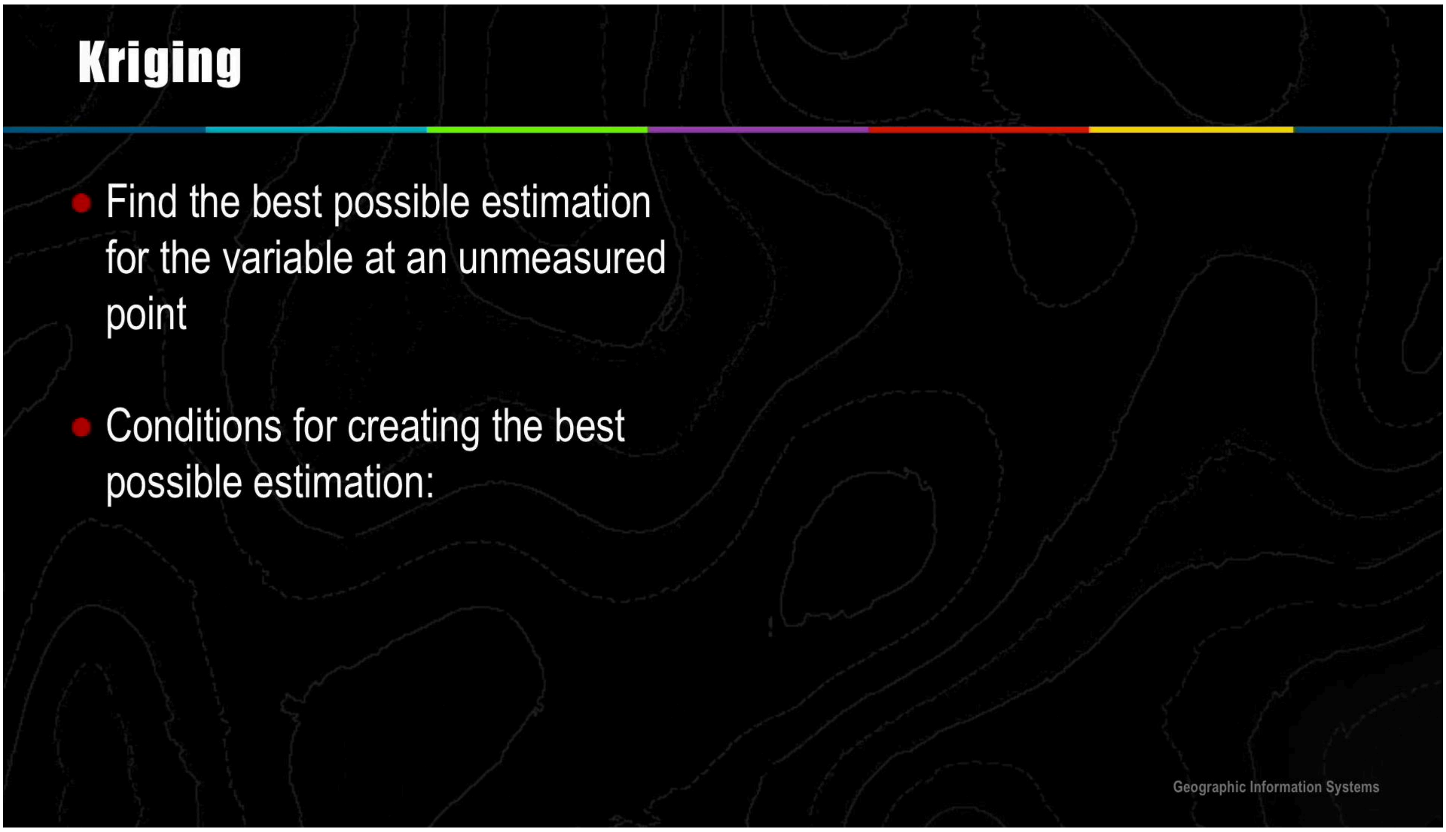


Once the various stages of the structural analysis have been completed, the results remain to be exploited by an interpolator capable of drawing the maximum profits and that is the kriging.

Notes

Summary

14m 33s

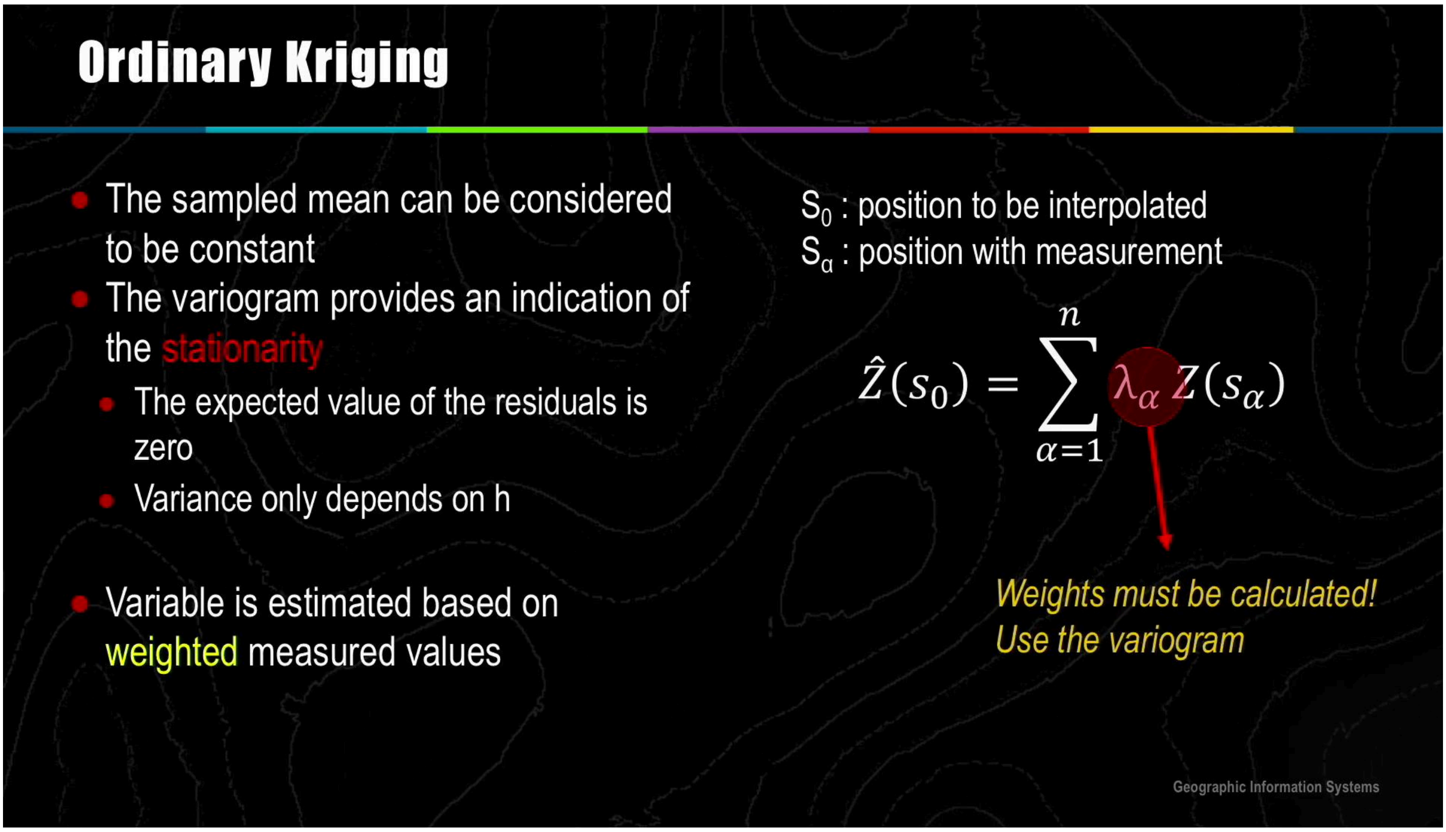


The problem of kriging consists in finding the best possible estimate of the regionalized variable z of s given the information available, that is to say given the determined values in a certain number of samples. From a theoretical point of view, the best estimator is one that satisfies two conditions.

Notes

Summary

14m 45s

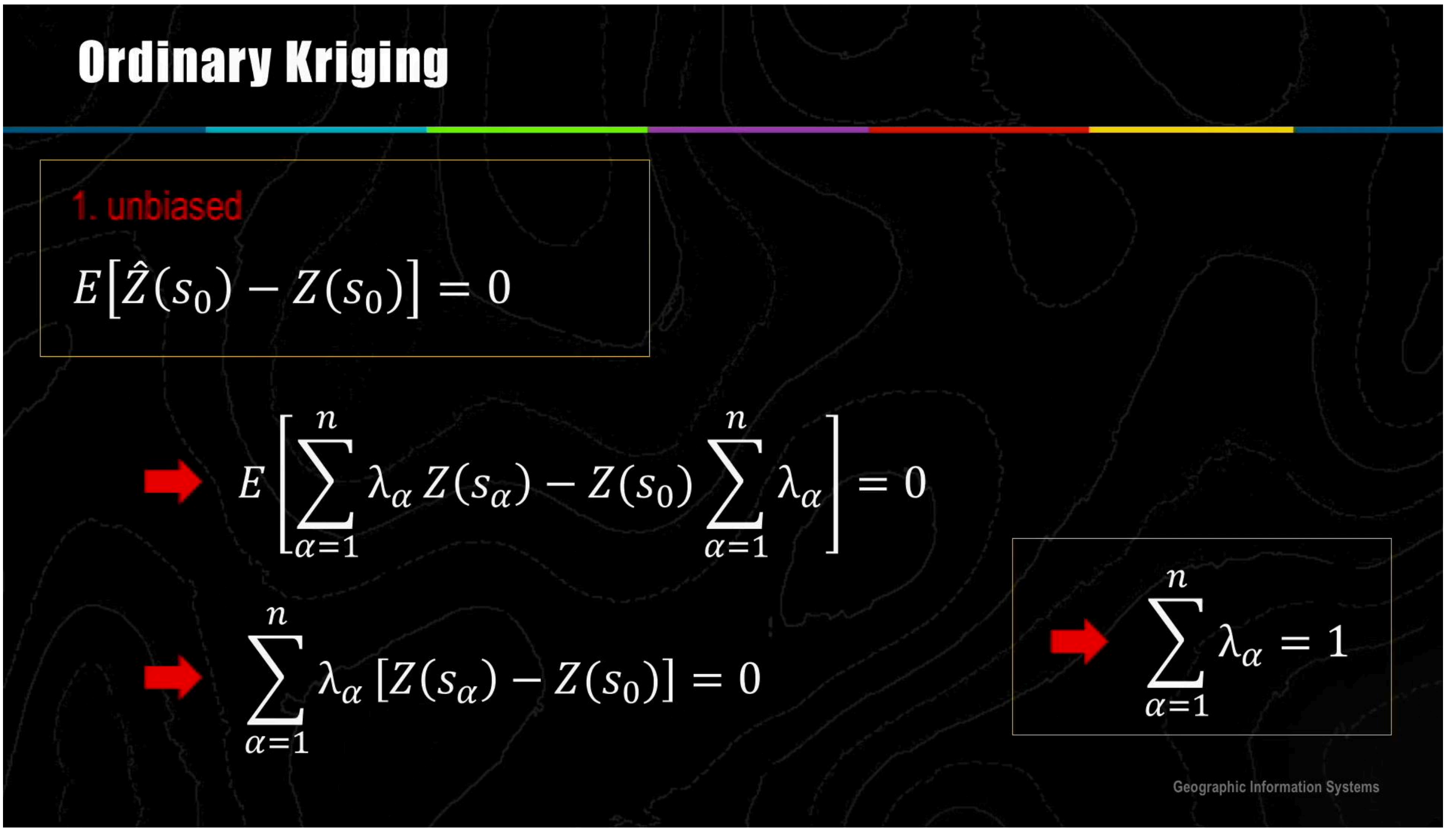


First, being unbiased, which corresponds to the residuals expectation which is zero Therefore, the mean of the deviations between the estimated values and the measured values is null. And on the other hand, to be optimal. This corresponds to minimizing the estimated variance. So the variance of the deviations between the estimated value and the measured value is minimum. This hypothesis enables then to calculate an estimation uncertainty on the interpolated sites. And it is this property that constitutes one of the great qualities of kriging. The ordinary kriging is used in the most common situation. We have a domain that has been sampled and whose average can be considered as constant. And the variographic study showed that the stationarity assumptions are respected. The mathematical expectation on the residuals is null and the corresponding variance depends only on h. The kriging is a linear interpolator. The estimated value is a weighted sum of the values taken by the sites in its neighborhood. It is therefore a particular case of the interpolator, inverse distance weighting, which we have already examined. An estimated value is therefore produced by the following equation where γ(α) are the unknowns.

Notes

Summary

15m 04s

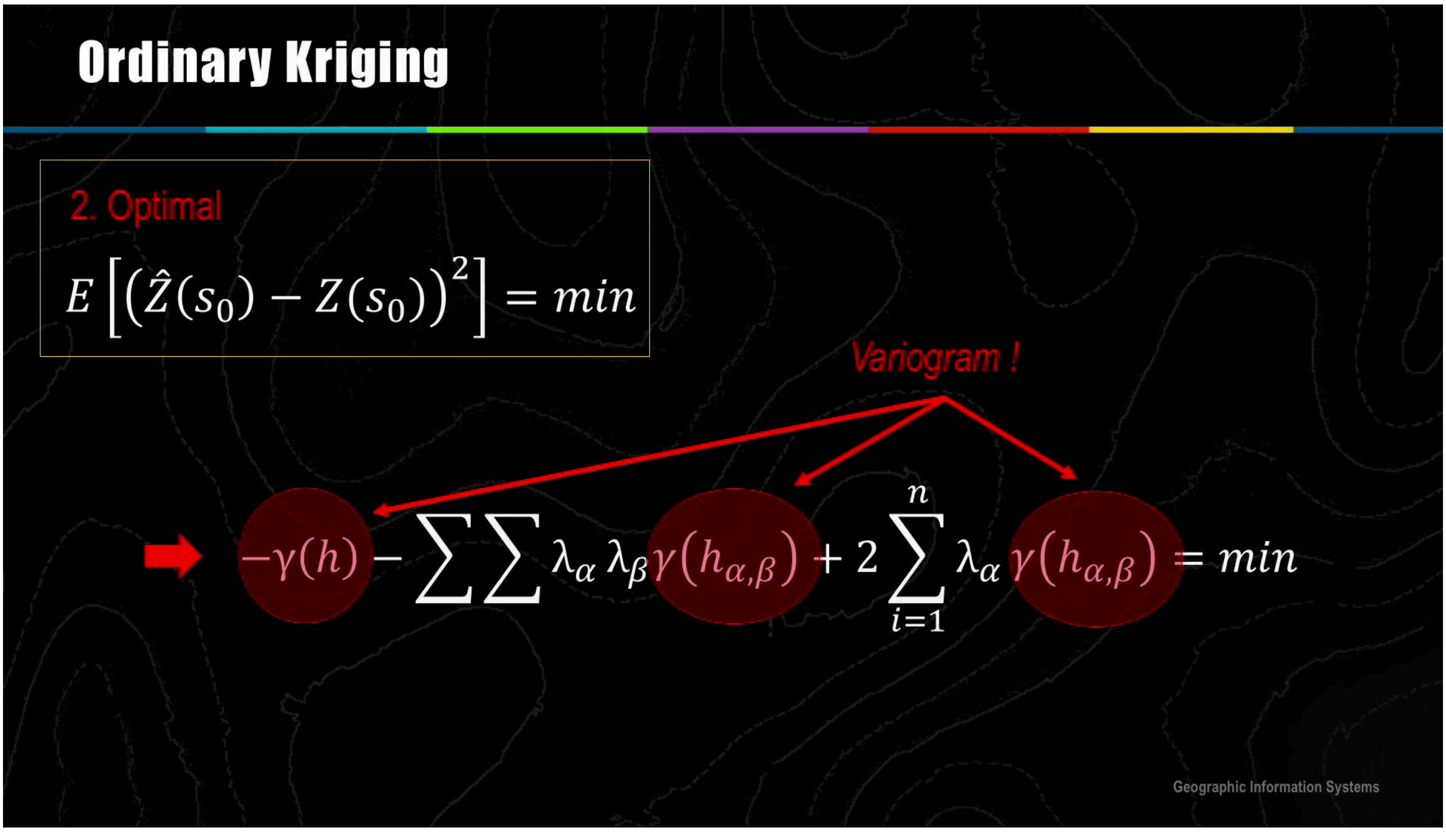


To determine these unknowns, we must apply the two equations that define the properties of kriging, so an unbiased estimator on one hand and optimal on the other hand. To begin with, we draw a first condition on the weights from the unbiased property. This sum of weights must be equal to 1. In fact, we show that if this sum is equal to 1, it checks the equation that defines the unbiased condition, so a residual expectation that is null.

Notes

Summary

16m 23s

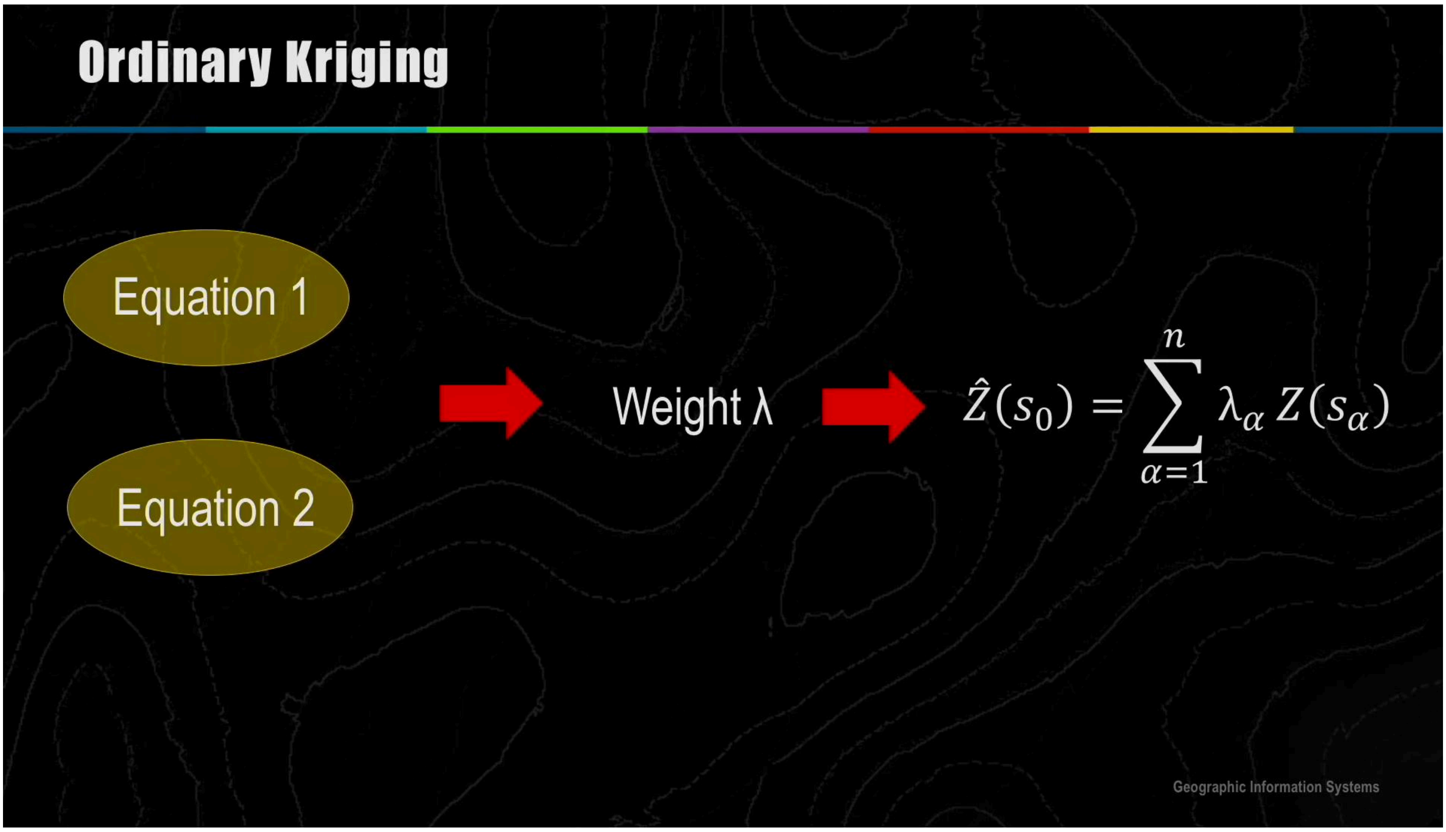


The estimator must also be optimal, so the estimated variance must be minimal. This variance is expressed by the γ(h) variogram in accordance with specific conditions defined by Matheron in the theoretical foundations of the structural analysis but in the details of which we will not go here. The important thing is to remember where here in red the information produced by the variogram intervenes in the equation which allows to minimize the estimated variance. The estimated minimum variance is reached when its derivative compared to the weight is null. The first step to solve the equation is therefore to calculate its derivative. And the final solution requires a particular method based on Lagrange parameters and its developments are not presented here.

Notes

Summary

16m 51s



In synthesis, the kriging owes its robustness to the fact that its functioning is conditioned by parameters produced by the variogram. The two conditions that we have just analyzed allow primarily to define the weighting related to the distance that characterizes the measurement points. And ultimately to infer intermediate values.

Notes

Summary

17m 39s