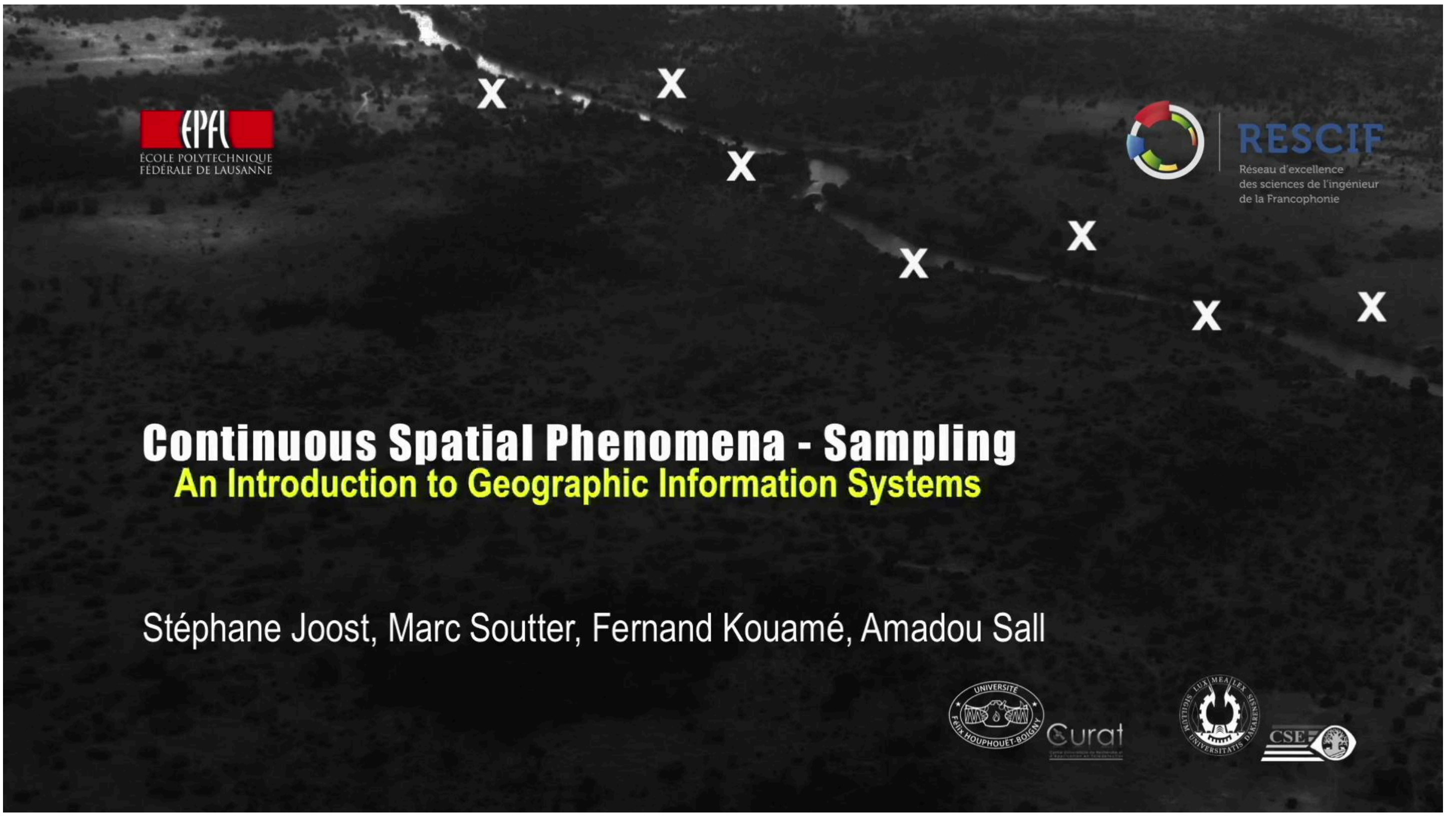
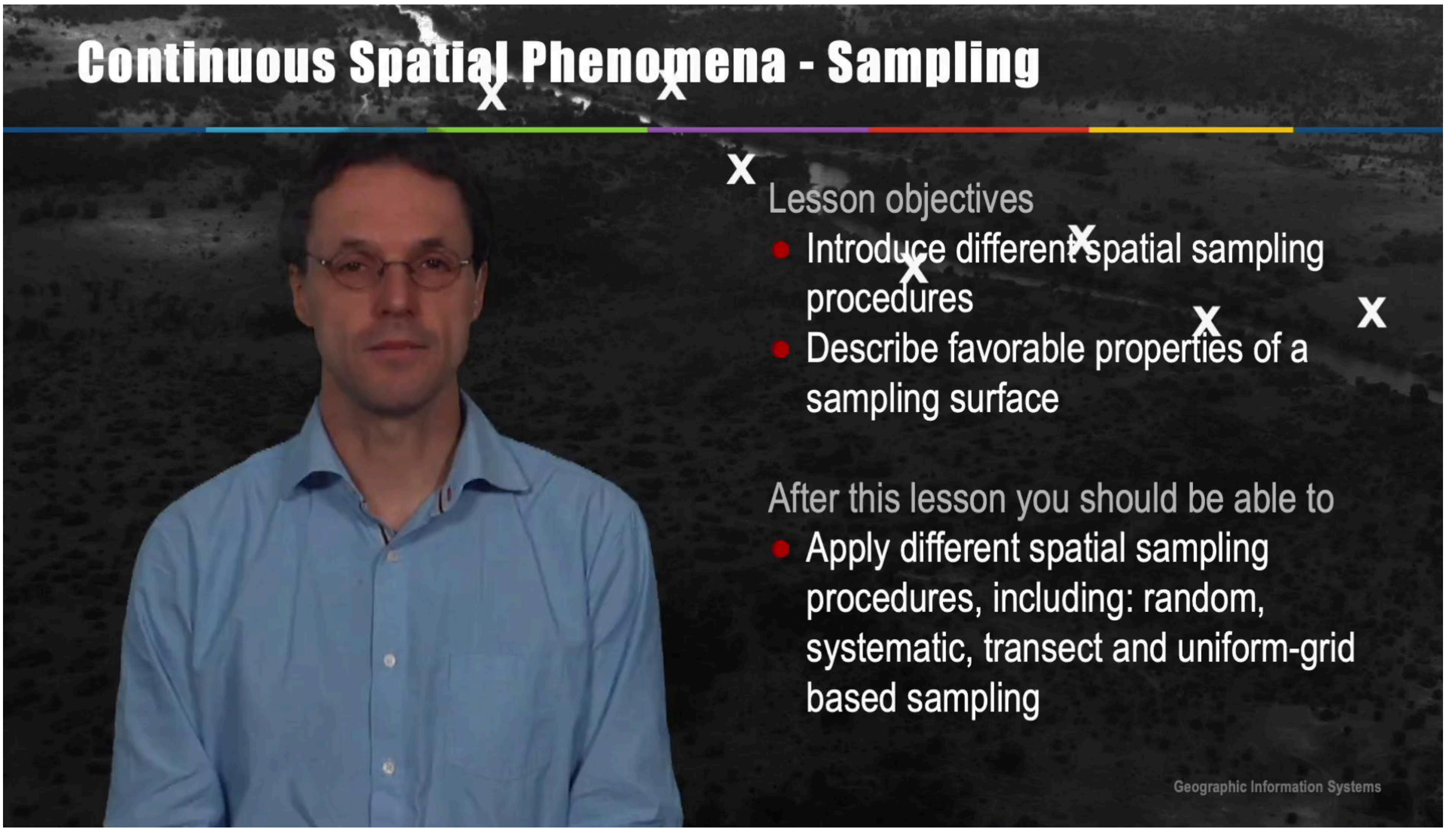


Search MOOC

Video

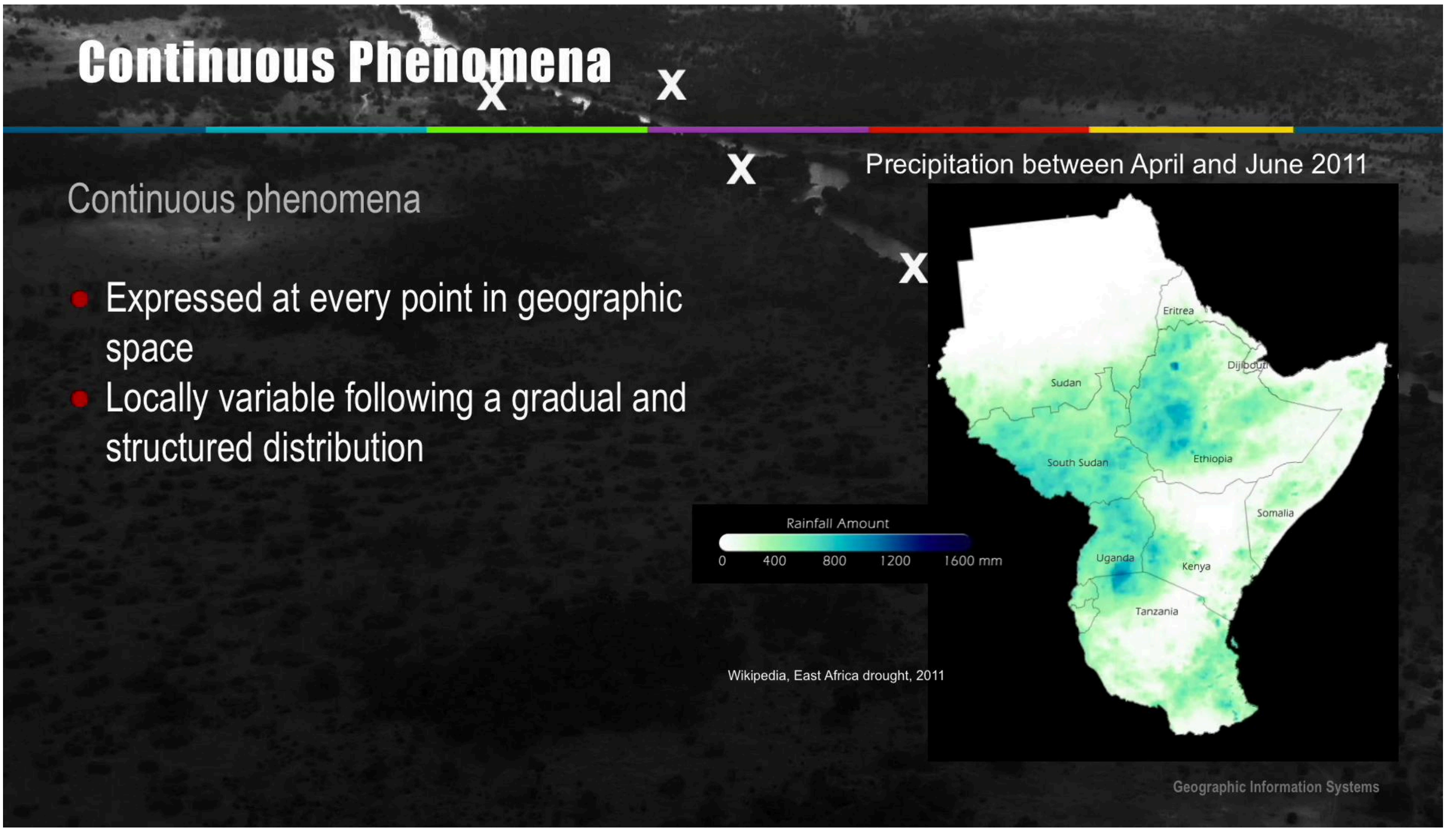


Welcome to this lesson on continuous variable sampling. This type of variable is defined in every point of space and it is very important to make a representative sampling. A phenomenon is said to be continuous if it is defined in any point in the geographical space and that its properties vary locally in a gradual and structured way. The altitude, the soil moisture and its heavy metal content are examples. And as Fernand Kouame suggested in his introduction this type of data must be acquired by sampling. Once the samples have been collected, we will proceed by spatial inference using interpolation to extend punctual data to the entire studied area. But the implementation of the sampling and interpolation operations is governed by hypotheses and specific laws which we will review. The goals of this lesson are to convey the basic notions which will allow to you to implement the most common sampling procedures. We will present these procedures as well as the properties that sampling is supposed to satisfy so that you can assimilate them and apply them judiciously to your own data.

Notes

Summary

0m 30s

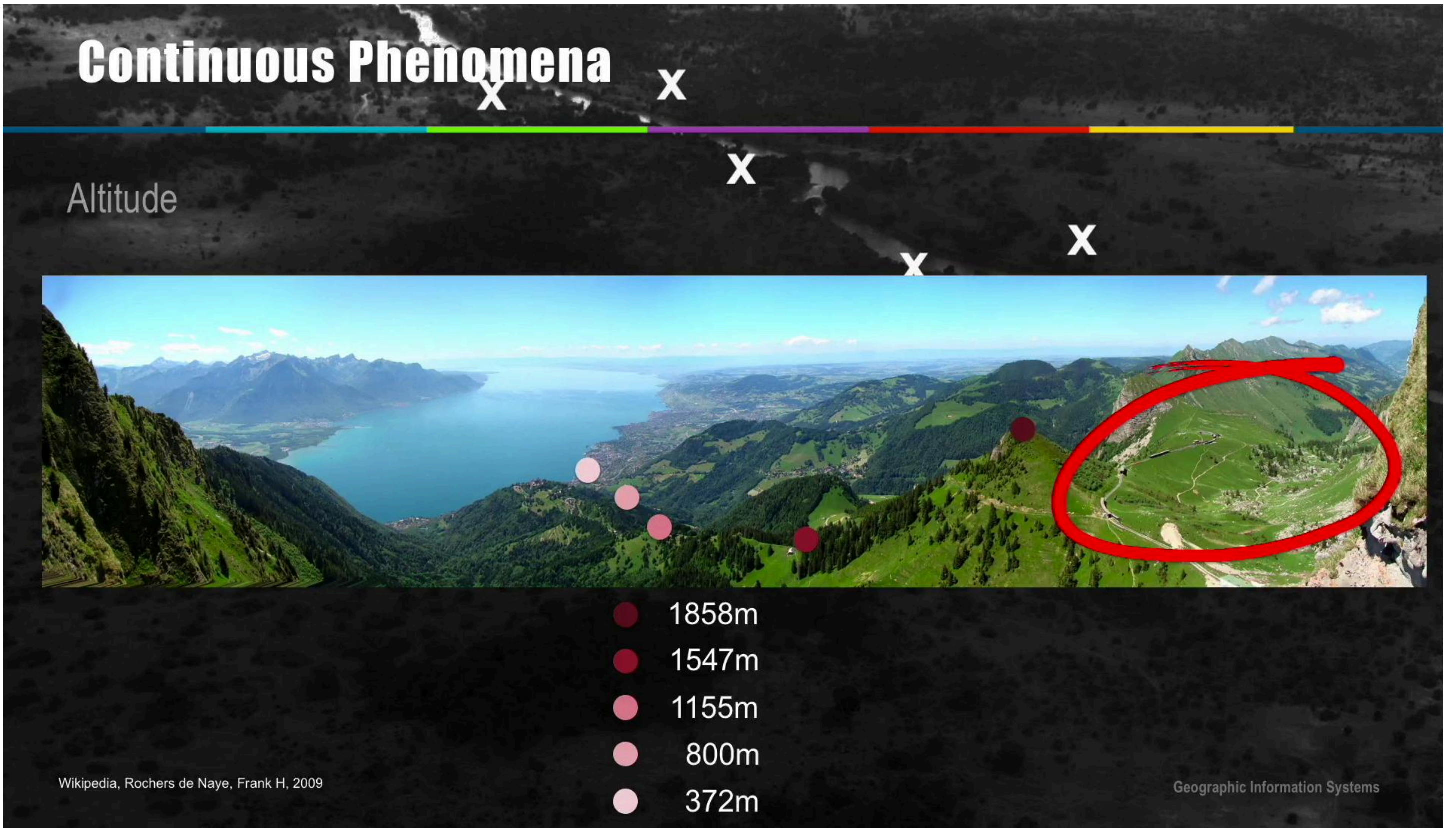


As mentioned in the introduction, continuous phenomena are defined in every point of space. They can vary locally, gradually and in a structured way, as is the case for precipitations shown on this map of East Africa.

Notes

Summary

1m 51s

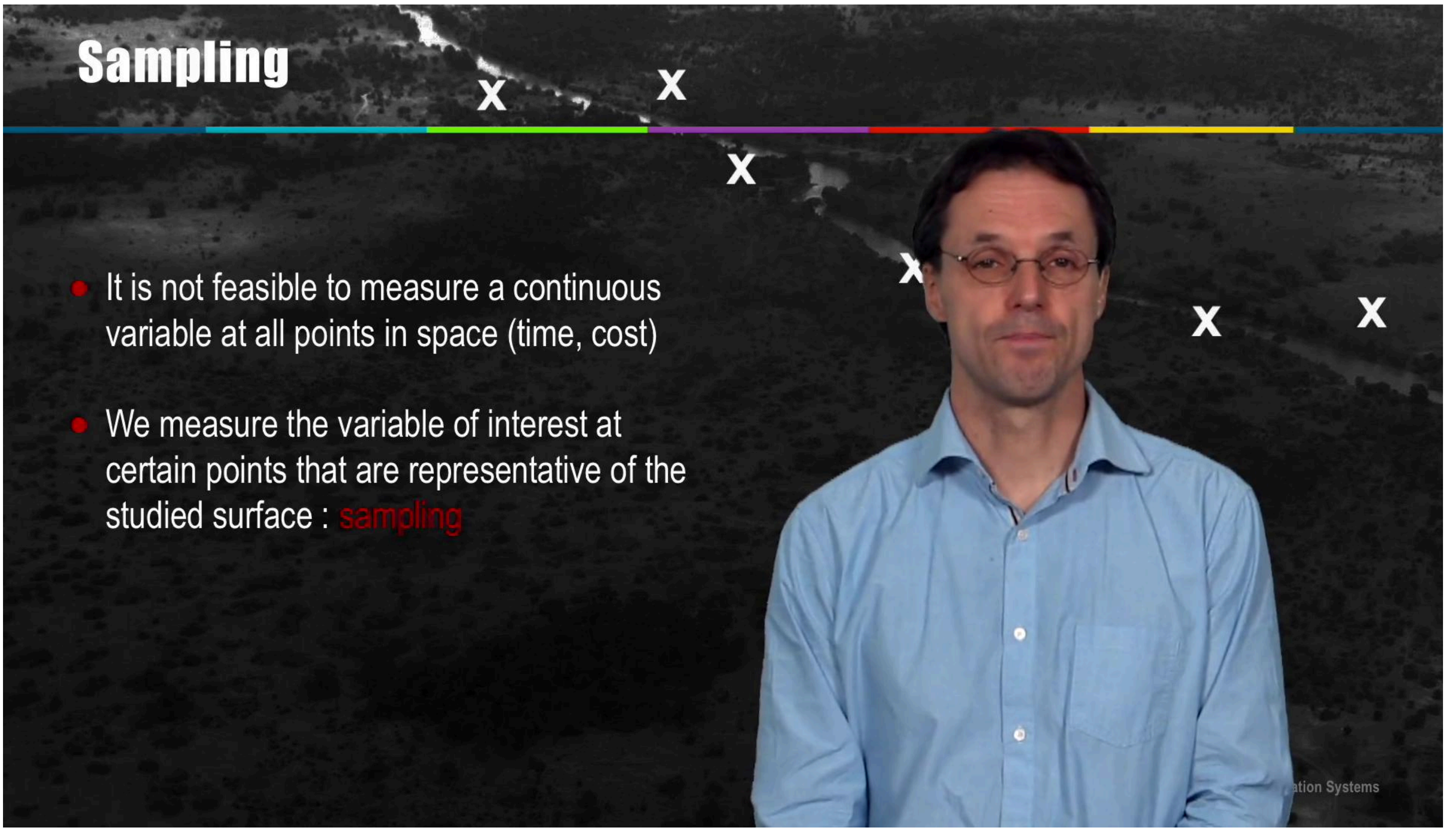


Or altitude, whose gradual elevation is represented schematically and discretely on this panorama East of Lake Geneva. The Naye valley, here on the right, will allow us to illustrate another example of continuous variable.

Notes

Summary

2m 05s

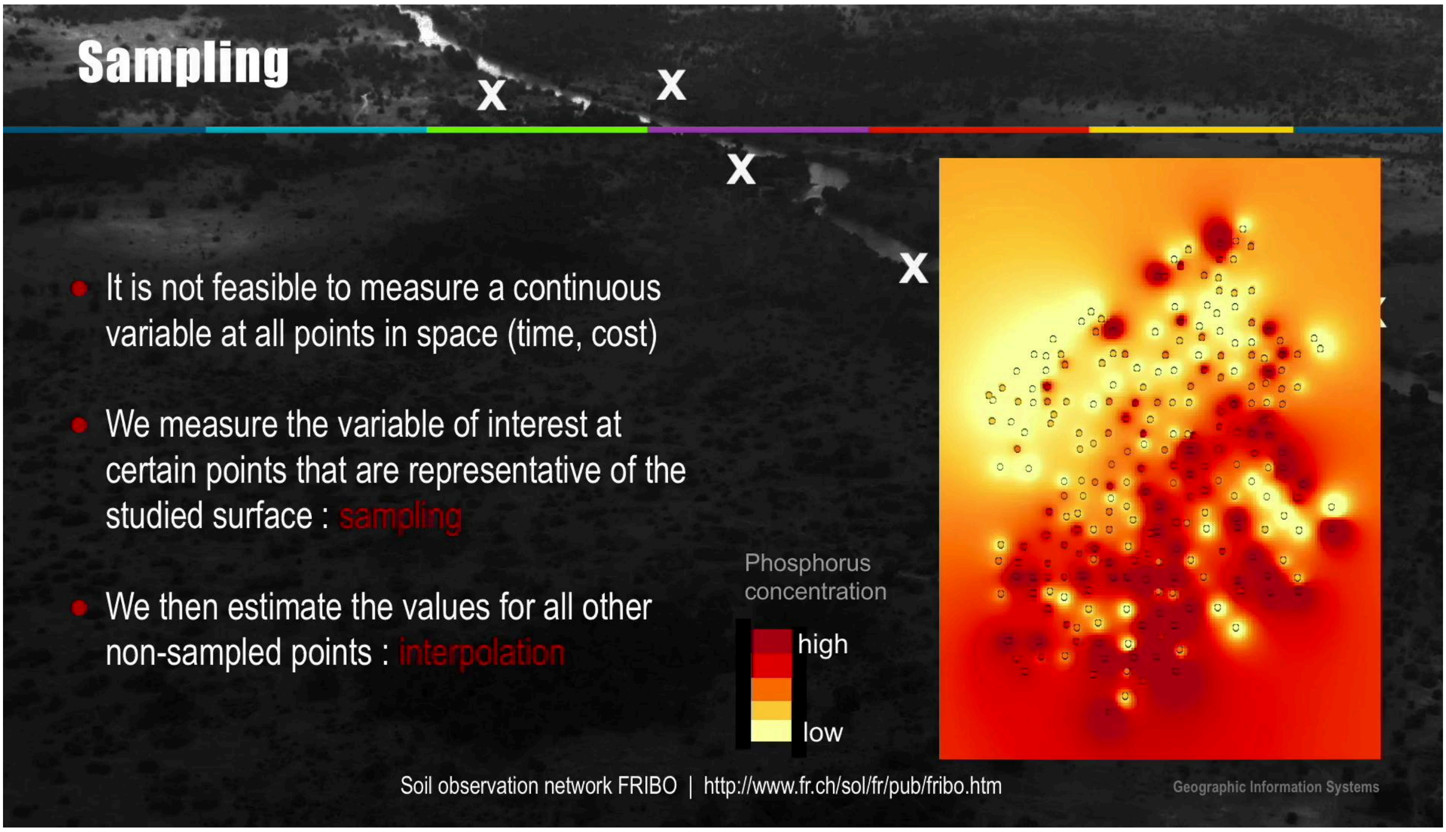


This is soil moisture, represented by a morphometric index which reproduces the concavity of the land, calculated here on the basis of a numerical model of altitude with a spatial resolution of 1 m. For time and cost reasons, it is obviously not possible to measure a variable of this type in any point of space. But for analytical purposes, it is nevertheless necessary to have these values. We will therefore perform measurements of the variable of interest for a reduced number of points that are considered representative. This is called sampling.

Notes

Summary

2m 21s

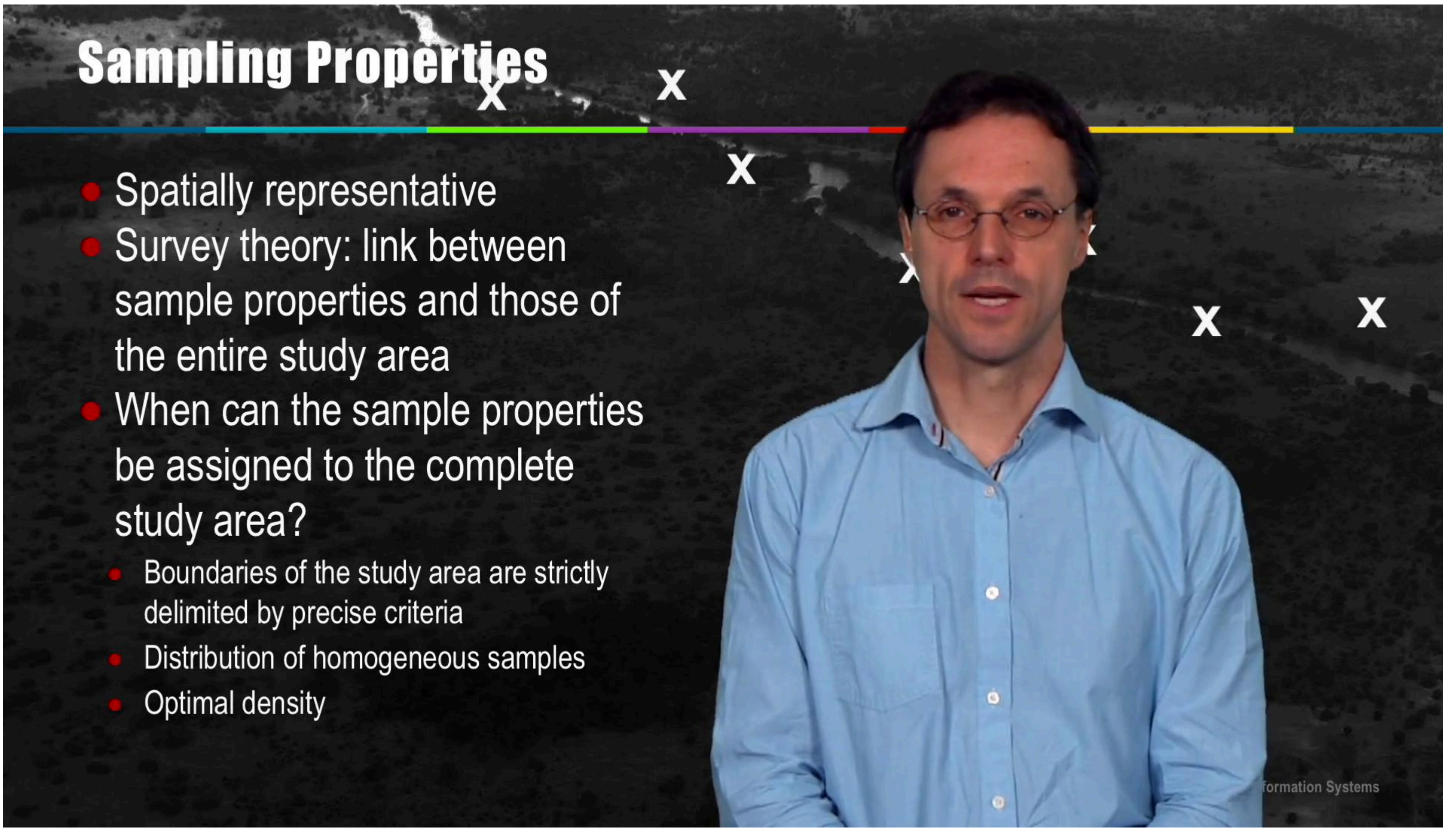


Here is for example the phosphorus concentration measured in the soil of 245 measurement stations distributed over the Canton of Friborg in Switzerland. The higher the content, the darker the red of the point which represents the actual measurement station. And from this distribution, we will deduce or infer the value for the rest of the geographical space. This is the approach of regionalization by interpolation. This approach can be applied only under the assumption that it is possible to estimate by inference with an acceptable precision the values taken by the variable in other points of the considered space.

Notes

Summary

3m 02s

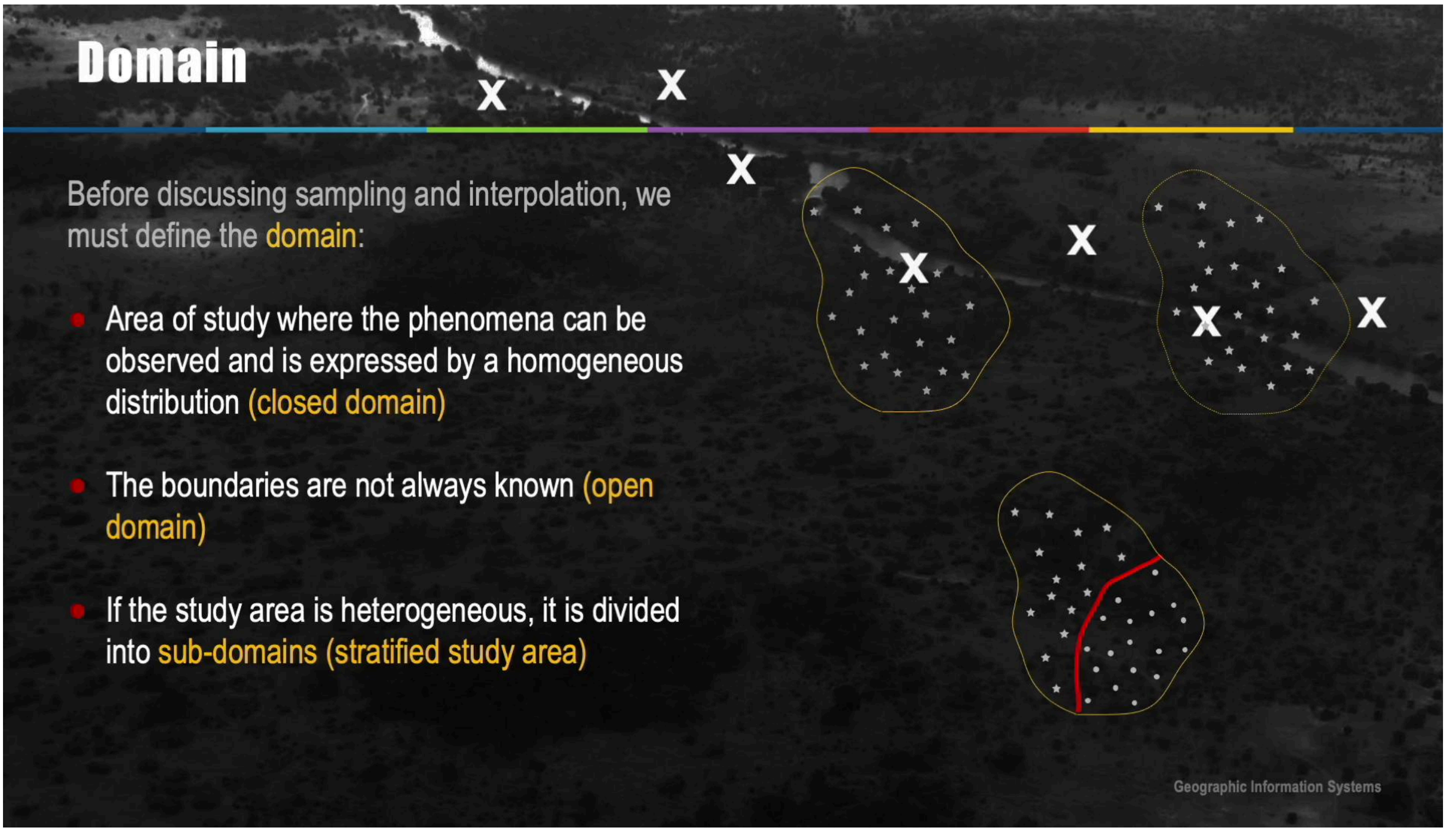


Sampling only makes sense if it produces a valid knowledge for the entire studied area. Its primary quality is therefore to be representative. Sampling is a matter of sampling theory, the problem is to study in all situations the relations between the properties of the sample and that of the whole population which corresponds in our case to the whole studied area. So, to what extent can we attribute to the whole area the properties of the sample? Here are the most important requirements. A sample is considered optimal if it meets the following two conditions: on one hand, the boundaries of the study area have been circumscribed according to very precise criteria and on the other hand, the distribution of the samples must be homogeneous over the entire study area and of optimum density, that is to say sufficient to represent the desired qualities.

Notes

Summary

3m 37s

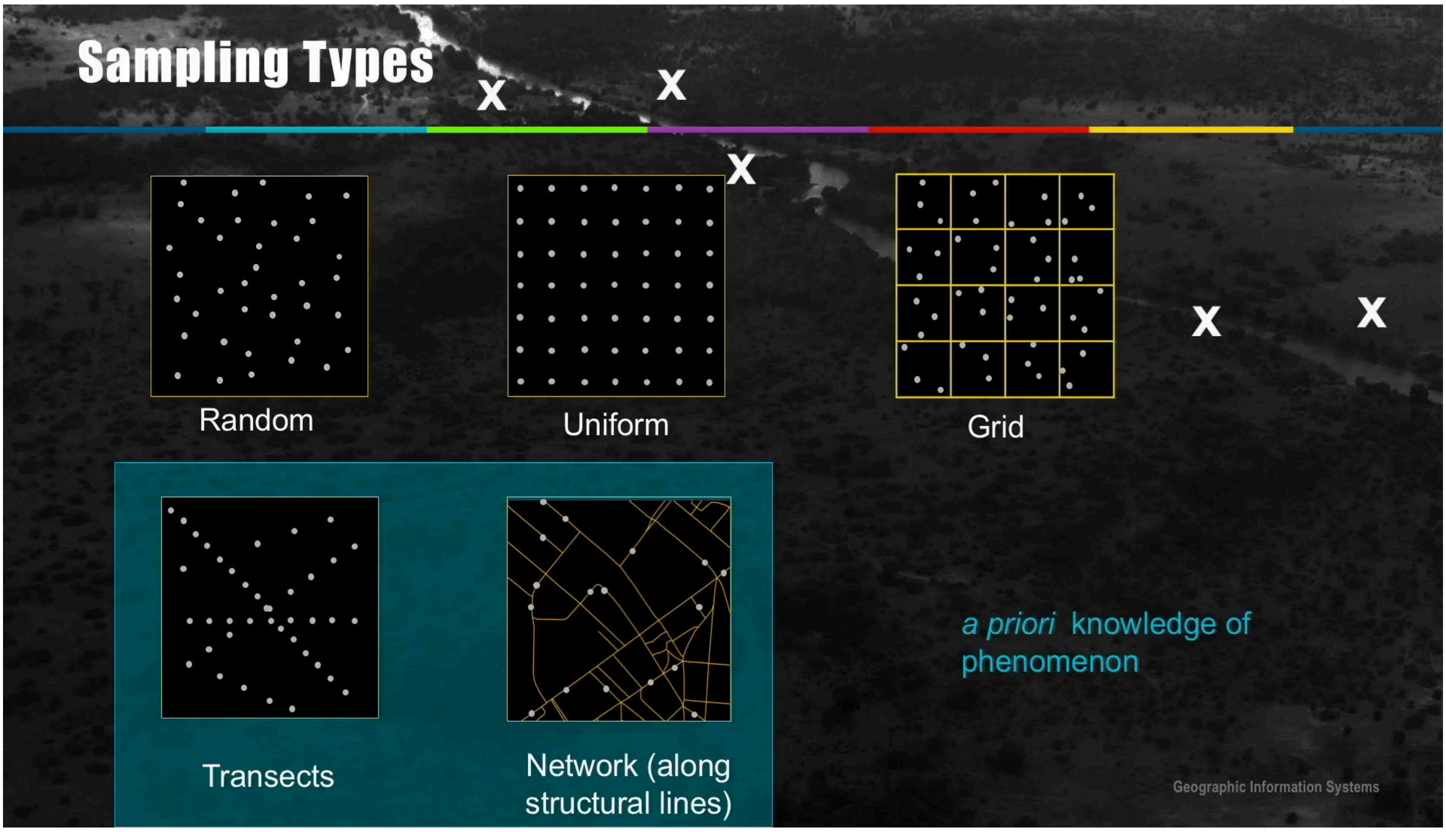


But in practice these conditions are often fulfilled according to an iterative procedure. The limits are gradually fixed According to the results of the surveys or the first measures. The distribution of samples can be arbitrary to start with and not uniformly distributed and the sampling will be able to be progressively densified by taking a sample in locations deemed relevant but this is not possible for all the phenomena analyzed. An important concept that should be used is that of domain. The domain covers the study area in which the phenomenon is supposed to exist and presents a similar and homogeneous behavior. Sometimes the limits of a domain are known and we will then speak of a delimited domain. But generally they are not or only approximate and we will speak of open domain. This distinction is important as an interpolation law will be able to be applied to the whole of a domain which is assumed to be homogeneous of a behavior of a phenomenon. If we observe heterogeneity, it will be necessary to divide the initial zone into several subdomains, each of which is supposed to present a homogeneous behavior so that the same interpolation law can be used. We thus carry out a stratification of the study region.

Notes

Summary

4m 32s

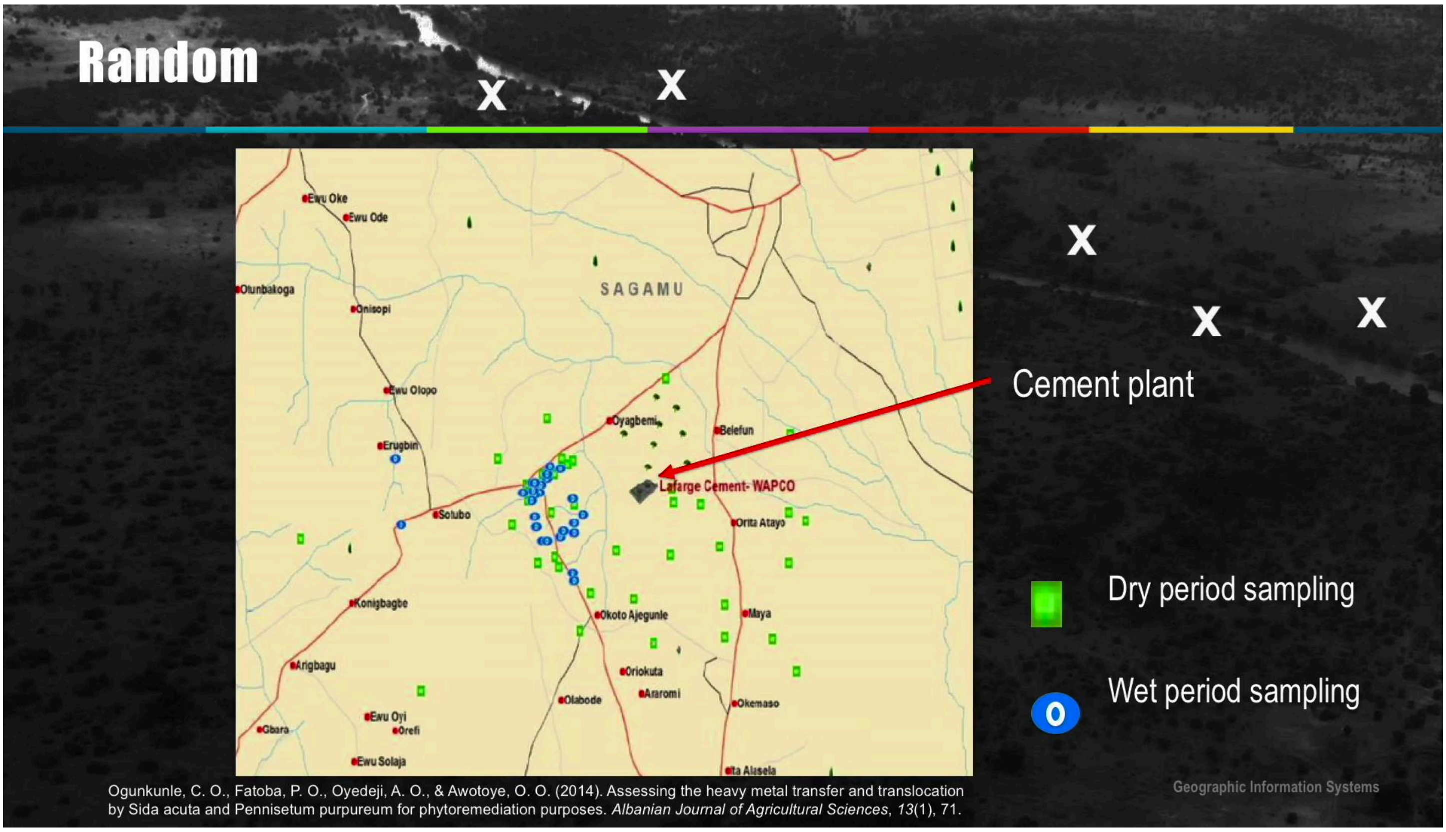


There are several ways to carry out a spatial sampling which depend mainly on the prior knowledge which we have on the behavior of the phenomenon studied or inversely the absence of the prior information. In the latter case, we will either apply a random sampling for which no specific scheme is applied or systematic, depending on variants of the regular point sowing or the regular grid or transects or structural lines. For invisible phenomena, as for example for a soil contamination we often follow an exploratory procedure in order to progressively apprehend the behavior of a phenomenon before proceeding to a complete analysis. For example, the sampling can be done in a regular pattern such as the transect or randomly but with a progressive densification of the measurement points. Sampling procedures conducted along one or more transects or along structural lines are procedures that require a prior knowledge of the phenomenon. For example, we will study some biological processes related to plant growth along an altitudinal transect. Indeed, we know that the altitude affects the duration of the growth period.

Notes

Summary

5m 57s

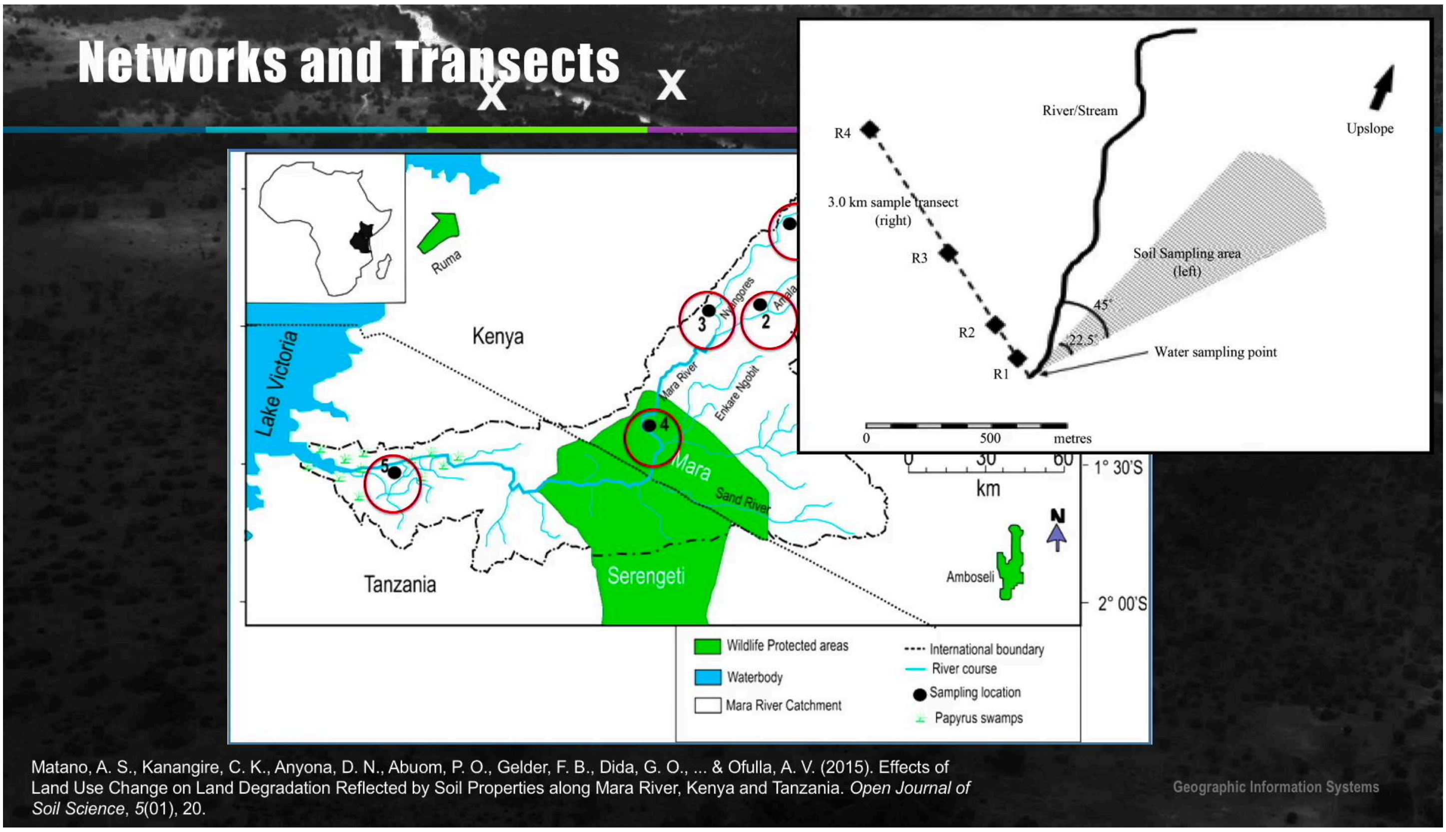


We present here an example which illustrates the sampling procedure used to study the concentration of heavy metals, such as lead, copper or zinc, in soils located around a cement production plant in South-West Nigeria. Indeed, depending on the degree of purity of the ingredients some heavy metals come into the composition of certain cements. To evaluate the importance of this pollution the authors have taken two plants which have the capacity to accumulate heavy metals with 15 cm of surface soil in 64 points and this within a radius of 3 km around the factory. The main constraint was temporal since it was necessary to obtain representative data for the dry season and for the rain season. On the other hand, the location of sampling points was determined in a completely random way.

Notes

Summary

7m 16s

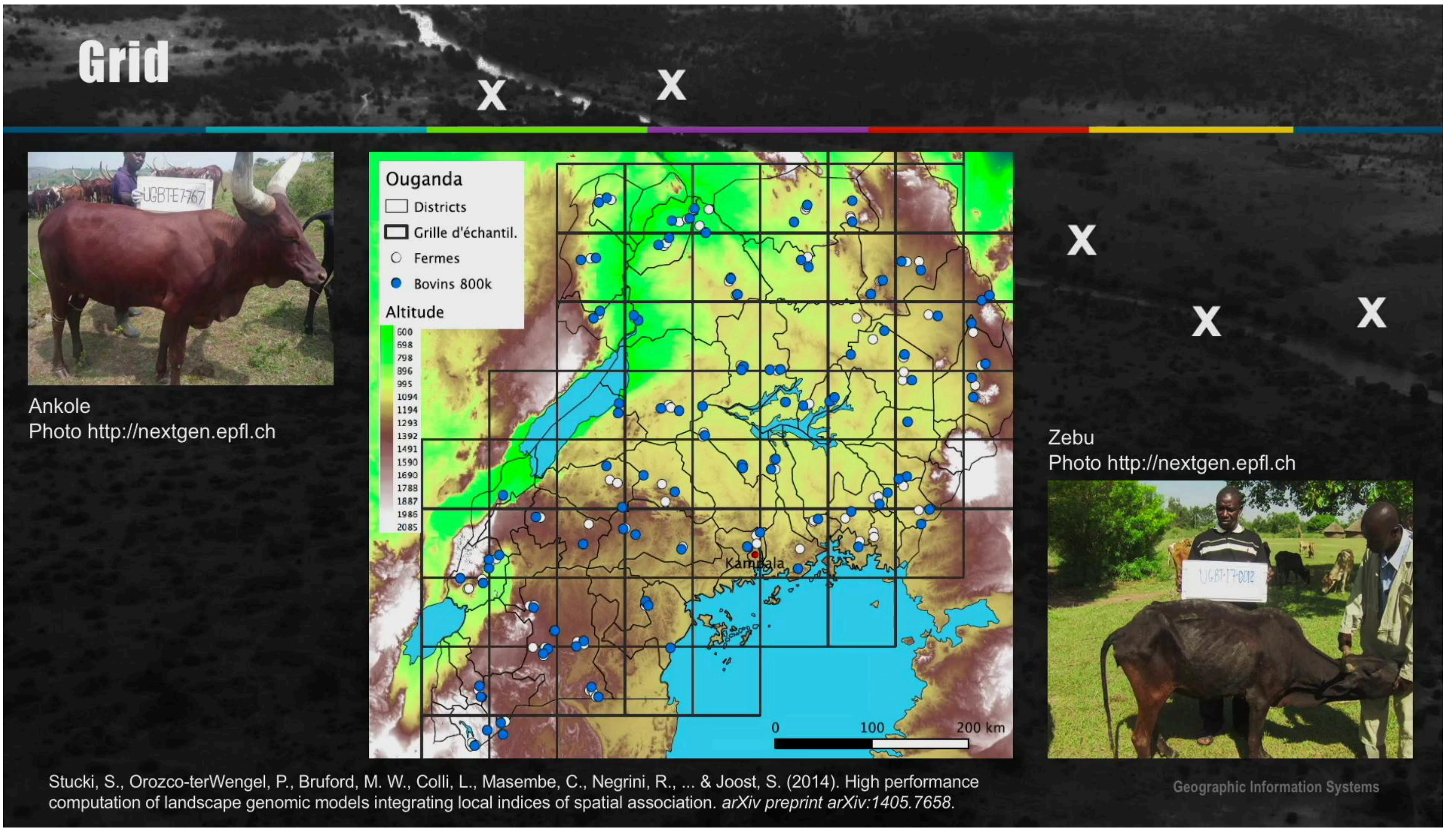


The results of the study show that the two plants in question Sida Acuta, which is considered a weed and Pennisetum Purpureum also called elephant grass, constitute good accumulators in particular of chromium and cadmium and that they could be used in remediation operations of polluted soils. Here is an example now which illustrates a sampling procedure carried out along a structure line, so a river here. The expansion of agricultural activities along the Mara River, between Kenya and Tanzania, has resulted in many changes in the soil cover. The savannah forests and grasslands have been converted to agricultural land. Several studies have shown that these changes could lead to a greater erosion of the soil, a reduction in nutrients and soil structure changes. The study presented here was therefore carried out in order to study the impact of land use around the Mara River on the physico-chemical qualities of soils, such as the carbon, nitrogen or phosphorus content or the soil acidity. Soil samples were collected in 5 different sites chosen along the river, in areas with different types of land use. In each site, 8 sampling points were then defined along a transect, 4 on each side of the river. Finally, we illustrate a sampling campaign carried out according to a regular grid.

Notes

Summary

8m 08s

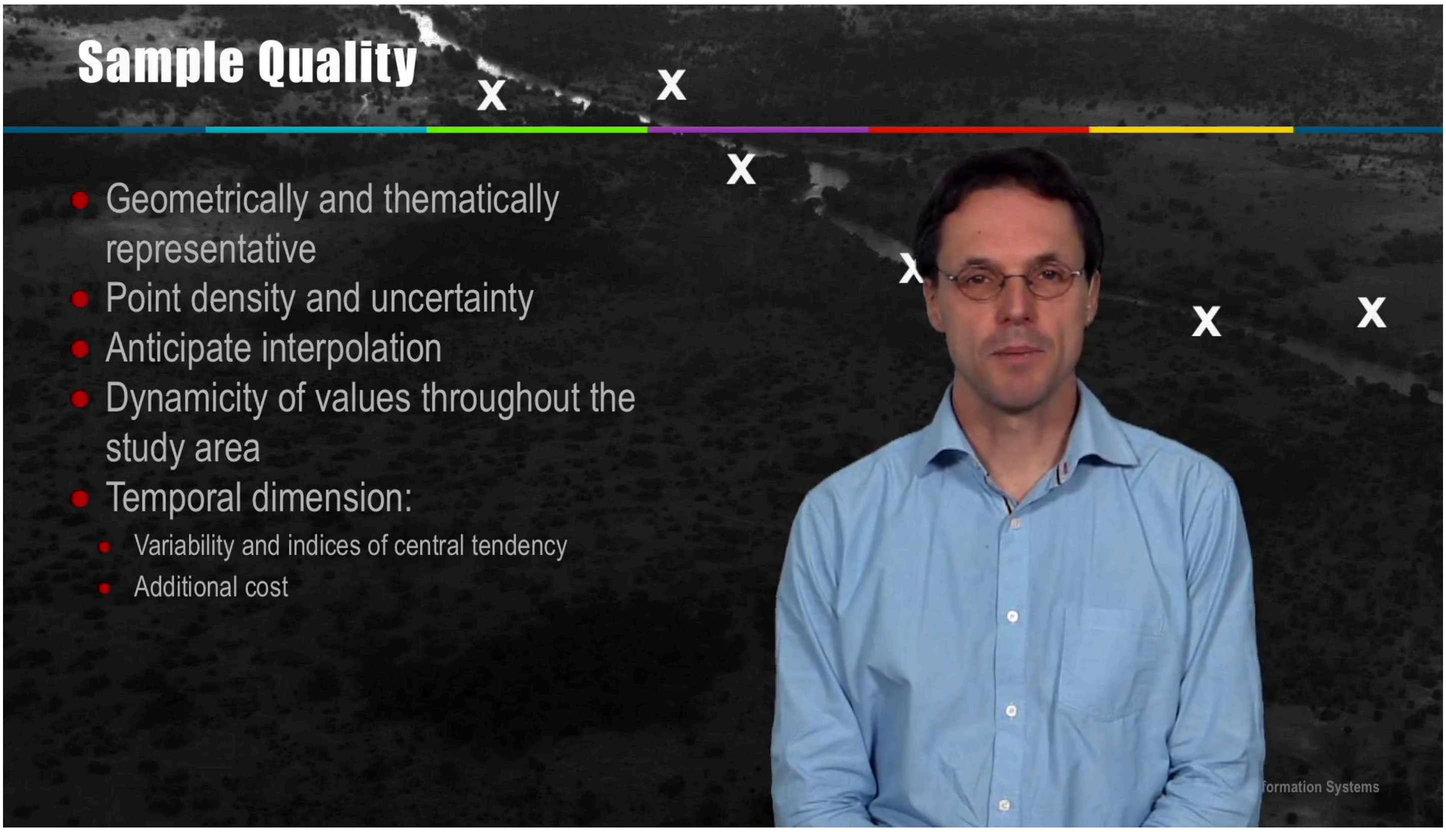


This is the analysis of genetic diversity and of the population structure of Ankole and Zebu cow breeds in Uganda. This study was carried out as part of a european project called "NextGen" and one of the aims was to identify the genes involved in the processes of resistance to certain tropical diseases such as the trypanosomiasis. To ensure the greatest possible spatial representativeness throughout the country 917 individuals, so about 4 per location were selected from 229 farms in Uganda. Each cell contains 4 farms, and the grid step is about 70 kilometers. Blue dots represent farms where an animal has been selected for a specific high density genotyping.

Notes

Summary

9m 32s



We have already mentioned that a spatial sample must necessarily represent the behavior of the phenomenon studied both in its geometric and thematic dimension. Regarding the geometry, it is useful to point out that the density of points sowing affects the degree of uncertainty associated with this distribution until an optimum size is obtained. Another aspect involving geometry concerns the constraints linked to the interpolation procedure often applied to measurements obtained by sampling and which will be discussed in the following lessons. It is necessary to ensure that the process of regionalization within the study area, proceeds well through interpolation and not through extrapolation. To ensure this, it is necessary to build a sample containing points outside the study area in a crown large enough to avoid edge effects. From a thematic point of view now the measured values should, as far as possible, account for the dynamics of the values present in the study region which is not easy for invisible variables. Finally, we must also consider the representativeness of the sampling in its temporal dimension in order to ensure that the thematic values express well the behavior of the phenomenon for a defined period of time.

Notes

Summary

10m 25s

In some cases, with a significant variability it will then be necessary to produce indices of central tendency accounting for the general behavior of the measured variable like monthly averages of air temperature for example. Depending on the phenomenon analyzed, taking into account the temporal dimension in the sampling can represent a significant additional cost.

Notes

Summary

11m 46s