Geographic Information Technology Training Alliance (GITTA) presents:

Discrete Spatial Distributions

Responsible persons: Claude Collet, Dominique Schneuwly, Regis Caloz

Table Of Content

1. Discrete Spatial Distributions

Introduction to the lesson

The distribution of properties of discontinuous phenomena into space produces spatial objects of type point, linear or zonal. The distribution of properties of land cover in a study area in is an illustrative example of the spatial division in **resulting objects**. In a complementary way, one can be interested in the spatial distribution of properties of a phenomenon in a collection of **preset objects**, such as for example the type of prevalent economic activity in a whole of communes. In these two situations illustrated in Figure 1.1, through the spatial analysis, one seeks to describe and to understand in which way the properties of the phenomenon are distributed in space: their **spatial organization**. These problems lead us to formulate more specific questions such as:

Learning Objectives

- How can we describe the spatial distribution of the properties of a topic in a study area ?
- Is this arrangement random or is there a spatial structure (organisation)?
- How do we describe the spatial arrangement of this set of thematic properties? What is the degree of spatial fragmentation?

1.1. Introduction

Discontinuous spatial distribution: production of resulting objects and properties of preset objects

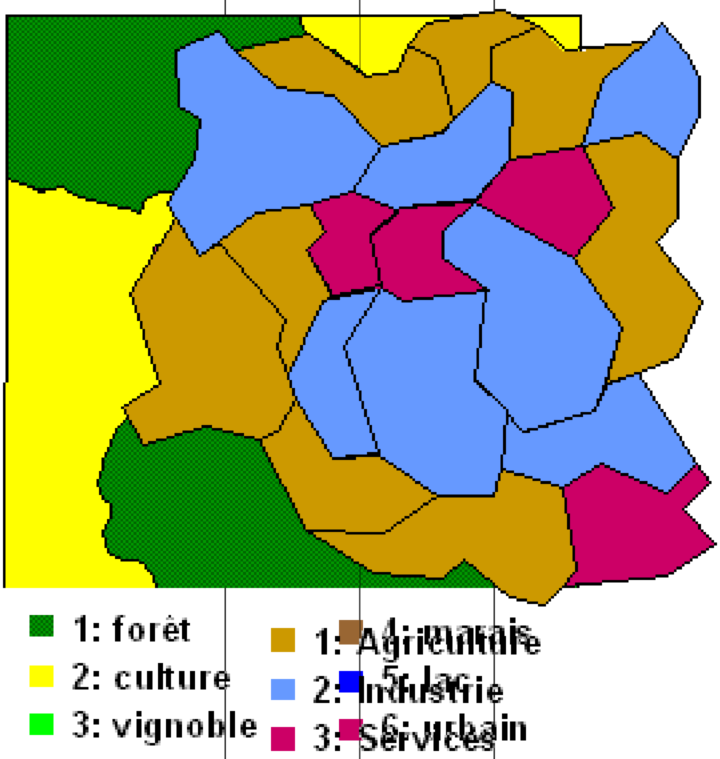
Resulting spatial objects	**Properties in preset objects (communes)**
	
Categories of land cover	Categories of major economic activity

Figure 1.1

The concepts of **structure** and **texture** summarize these interrogations on the spatial organization at two different scales: one **global** at the level of the whole study area and the other **more local** at the level of the neighborhood of the spatial entities.

1.1.1. Organization of the Lesson

Thus, this Lesson is organized into two Units:

- In the first Unit, one will be interested in methods allowing to describe the **spatial structure** present in the distribution of properties of a phenomenon. The description of this organization will be carried out by measuring the **spatial dependency**. This process is similar to that used for continuous spatial distributions (see Lesson Continous spatial variables of the module B-AN), but the developed indices of spatial dependency are adapted to the discontinuous aspect of the distributions considered.
- In the second Unit, we will approach methods of description of the spatial organization through concept of spatial arrangement. One will consider successively this arrangement at the global scale of the study area, then on a local scale of the neighborhood of the spatial entities. These descriptors supplement those presented in Units 2, 3 and 4 of Lesson 2 in the basic module in spatial analysis (B-AN). This concept of arrangement being particularly dense, it is necessary to control the significance of a great number of descriptors to try to account for the complexity of spatial arrangement, through its structure and of its texture.

Far from being exhaustive, this whole of methods constitutes a set of powerful tools to try to answer certain aspects of the fundamental interrogation on the spatial organization of the properties of discontinuous phenomena.

1.2. Spatial Dependency

Is the spatial distribution of the properties of a phenomenon in a study area random or is there a spatial structure (organisation)? This general question is the same one as that formulated for the continuous spatial distributions (see Unit 2 of Lesson 3 in the basic Module B-AN). Within the framework of a phenomenon continuously distributed in space, one can see that a dependence exists and that it is strong, thus ensuring the "spatial continuity" of properties. On the other hand a discontinuous spatial distribution contains precise discontinuities making it possible to delimit spatial features (point, linear or areal). It is observed that the property is the same one over the whole of the feature surface and that it is different beyond the reaches of this object. In spite of many observed discontinuities, it is however legitimate to raise the question of the presence of a spatial structure.

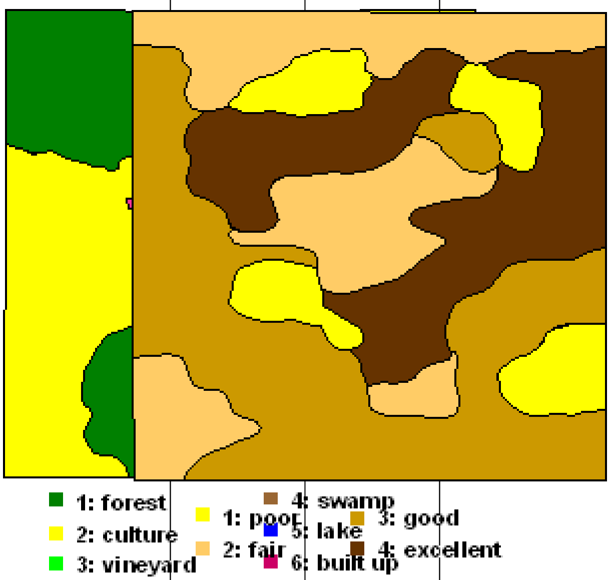
1.2.1. Introduction to unit Spatial Dependency

Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]

The question of the dependency is thus more complex to formulate in the case of predefined objects resulting from a spatial division. There are two distinct situations:

- The considered spatial features are **resulting** of the spatial distribution of properties of the phenomenon which one wishes to describe:
    - Example 1 (features produced by the distribution of landcover types; see Fig 2.2a): In this situation, the spatial dependency is absent by definition, the contiguous neighbours of each feature have different properties. These properties are **categories** expressed numerically at a **nominal** scale, for which only the identity or the difference has a meaning.
    - Example 2 (features produced by the distribution of the classes of soil quality for agriculture; see Fig 2.2b): Numerical properties of features express **hierarchical position** in the scale of aptitude, their contents are thus described as being of **“ordinal”** level. It is thus relevant to describe the way in which the level of soil quality varies according to the space.

Discontinuous spatial distributions for resulting features



Resulting spatial features	**Resulting spatial features**
Fig 2.2a: Landcover categories (nominal scale)	Fig 2.2b: Units of aptitude for agriculture (ordinal scale)

Figure 2.2

- The considered spatial features are defined a priori and one wishes to describe the spatial distribution of their properties for another particular phenomenon:
    - Example 3 (the spatial distribution of the major economical sector for administrative features such as the "districts" of a study area; see Fig 2.3a): In this situation, the spatial dependency can exist for a phenomenon measured on a nominal level because the properties are not necessarily related to the nature of the spatial features. Thus, the contiguous neighbours of each object "district" can have identical properties.
    - Example 4 (spatial features produced by the distribution of classes of aptitude for agriculture of soil units; see Fig 2.3b): The numerical properties of the features express **hierarchical position** on a scale of soil quality, but their contents are thus of **ordinal level** level. It is then relevant to describe the way in which the level of quality varies according to the spatial proximity.

Discontinuous spatial distributions for predefined features

Predefined spatial features (districts)	**Predefined spatial features (cells)**

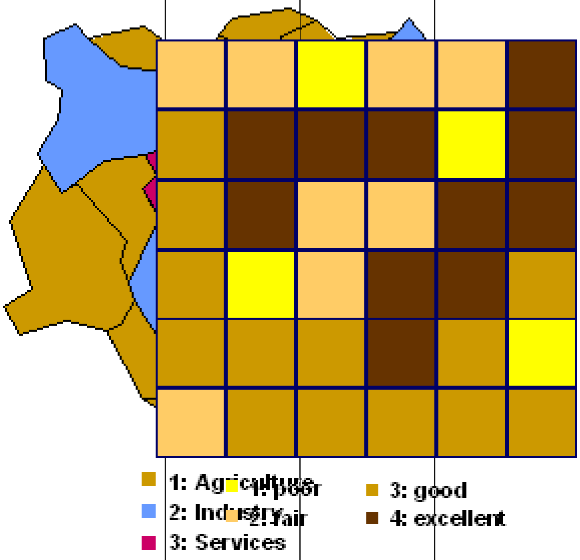



Fig 2.3a: Categories of major economic sector (nominal scale)	Fig 2.3b: Units of soil quality for agriculture (ordinal scale)

Figure 2.3

Thus, one can summarize the various situations in the following way:

- When the considered spatial features **result** from a spatial distribution of the properties of the phenomenon to be described, the spatial dependency (spatial autocorrelation) can exist only when this phenomenon is described on an ordinal level. The values of their properties are called classes.
- When the considered spatial features are **a priori** defined, they are independent of the spatial distribution of the phenomenon. A spatial dependency can thus exist, whatever the level of measurement of the phenomenon (nominal, ordinal or cardinal), it will be thus possible to describe it.

1.2.2. The concept of spatial dependency

In a very general way, the strength of the spatial dependency is a measure expressing the relationship between the variation of properties and the spatial proximity. In the case of a **continuous** spatial distribution, this relation can be expressed by a continuous function of the numerical difference of properties compared to the distance. Thus, one can see that the closer two places are in space, the more the difference in their property is weak (or the larger their similarity is). This continuous function also makes it possible to model the particular way in which distance acts on the importance of the difference (the two concepts of "range" and "function of distance decay" identified by the variogramme make it possible to account for spatial dependency). It will thus be shown that:

- The spatial proximity is measured by means of the Euclidean distance between pairs of places. This concept can be enriched by the directional contribution or the orientation in order to check the property of anisotropy.
- The variation of the properties is based on their numerical difference (interval of values) because the measurement level is cardinal and thus continuous.

1.2.2a Descriptors of spatial dependency for discontinuous distributions

Spatial proximity refers to point, linear or areal features. Particularly for the last two types, the proximity can be expressed simply by using only topological descriptors (contiguity, order of vicinity), because their size, form and orientation are variable. The most common descriptor is the **Join Count Statistic** Join Count Statistic (an index of contiguity or of adjacency).

Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]

The variation of the properties between contiguous objects is expressed in a different manner, according to the level of measurement of the considered phenomenon:

- at the nominal level, we will consider only **the similarity** or **the difference** of values for contiguous features.
- at the ordinal level, as the importance of the difference of values expresses a difference of ranks; this can be taken into account, in relation to the measuring unit of this hierarchy.
- at the cardinal level, it is possible to determine the difference of values of each contiguous features.

Thus we will retain two types of indices of spatial dependency, also called indices of spatial autocorrelation, based on the property of adjacency. The first is the *Join Count Statistic* (coefficient of adjacency), adapted to numerical properties measured at a nominal level. The seconds is *Moran's I Coefficient*, or alternatively, the *Gaery Ratio*.

1.2.3. The Join count statistic (at a nominal level)

At the nominal level, only **the presence** or **the absence** of a specific thematic property is considered. This characteristic can seem restricting and simplifying, but it makes it possible to process phenomena which are originally measured on an ordinal level or even cardinal level. The specific property which one wants to characterise the distribution and the spatial dependency can thus express:

- at the nominal level, a particular category or a set of categories, for example the presence of a socio-economic category or a type of plant association.
- at the ordinal level, a class (a rank) or a set of classes, for example the presence of the best agricultural soil classes.
- at the cardinal level, an interval of values, for example the presence of a significant rate of criminality.

The thematic property to be described is thus reduced to a variable of binomial level **(a binary variable)** containing only two values referring to the properties of presence/absence (e.g. yes/no, white/black, 0/1). The description of the spatial dependency is expressed through the similarity to one of the three possible types of distribution: **grouped, random** or **dispersed**. Figure 2.4 illustrates these three types of spatial distribution for the same study area whose 5 zones out of the 11 are concerned with the presence of a particular property.

Three types of spatial distribution of binary properties. Presence in 5 zones and absence in 6 zones

a) grouped	b) random	c) dispersed
Fig 2.4a: Strong positive spatial dependency	Fig 2.4b: No spatial dependency	Fig 2.4c: Strong negative spatial dependency

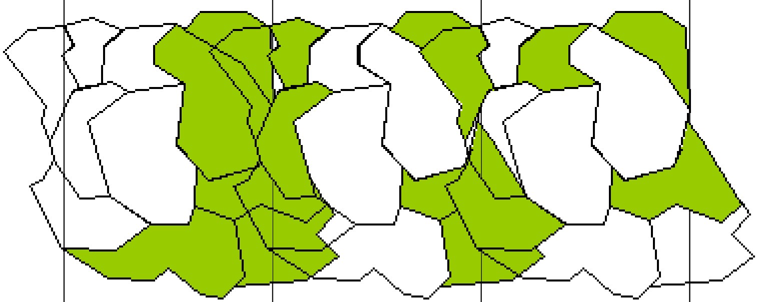


Figure 2.4

If the Binomial law or Poisson distribution enables us to determine the probability of an occurrence in the 11 zones of this area, it does not bring a response to the way in which these properties are spatially distributed. The join count statistic makes it possible to characterise the nature of this distribution according to three reference distribution: grouped, random or dispersed, and in consequence to deduce **the force** (degree) and **the direction** (positive or negative) of the spatial dependency. Thus, for the type:

- **"grouped"**; the spatial dependency is strong positively because the contiguity of zones with the property "occurring" is significant and consequently important too for zones of "absence".
- **"random"**; the spatial dependency is weak or even null because there is no similarity between the property of a zone and that of its neighbours.
- **"dispersed"**; the spatial dependency is strongly negative because the adjacency of the zones where the property occurs with that of the zones where if is absent is significant.

Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]

The join count statistic relates the number of observed connections between the zones of property "presence" and those of property "absence", with the theoretical number of connections of a random distribution. The definition of the theoretical number of connections of a random distribution is related to two factors:

- the **spatial arrangement** of features of the study area
- the **choice of the null hypothesis**

1.2.4. The spatial arrangement of features

According to the number and the shape of spatial features their adjacency results in many connections, independent of their thematic property. It is thus a question of describing this arrangement in the form of a matrix of adjacencies or graphically as on figure 2.5.

Description of the spatial arrangement of areas through their adjacencies

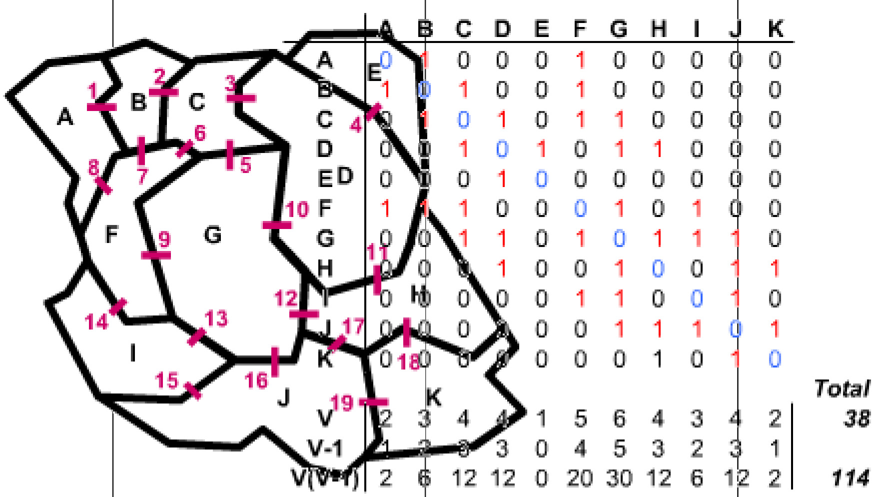
Identification of the connections	**Matrix of adjacencies**
	
15 contiguous zones with 19 connections (C)	V: number of neighbors of each area The total number of neighbors is equal to twice the number of connections (C)

Figure 2.5

1.2.4a Context of the null hypothesis

The choice of the null hypothesis expresses the way in which the properties "presence" and "absence" are assigned. From a statistical point of view, it is a question of determining if the study area is regarded as an independent sample (sampling with replacement, free sampling) or dependent (sampling without replacement, non-free sampling). The identification of one of these two situations is important because it will determine the nature of the theoretical distribution with which the observed distribution will be confronted.

A sample is considered independent when one knows a priori the probability p of the property "presence" - and thus of the number of "absence", independently of the situation observed in the area of study. For example, in a geomorphological region including the study area, one could determine that the probability of finding a soil of "good quality" for agriculture is 0.4 (p=0.4, therefore q=0.6), this number being independent of the number of zones having the property "good quality". Potentially, each zone has same probability of 0.4 of being regarded as "good quality", whatever the property already assigned to other zones in the study area. The estimated random theoretical distribution will express this particular situation by considering the parameters p and q.

A sample is considered dependent when the probability of occurrence of the property "presence" corresponds to the proportion observed in the study area. Returning again to the previously considered example, the situation of dependency would correspond to the selection of the n best zones of "good aptitude" for agriculture, among the t potential zones. The estimated random theoretical distribution will thus take into account these parameters n and t instead of p and q.

Generally, in practice, one gives the preference to a situation of non-free sampling if one cannot guarantee that the estimated probability of occurrence for the larger area is the same one as that in the study area. Moreover, the amount of "presence" and "absence" in a study area is generally observable and is thus given.

1.2.5. Estimate of the number of connections for a random distribution

According to the selected null assumption of independence or dependency, the theoretical number of connections between the areas "presence", P and "absence", A for a random spatial distribution, E_{PA}, is calculated as described in Table 2.1 below.

Number of connections E PA for a theoretical random spatial distribution

a) According to a null hypothesis of independence	b) According to a null hypothesis of dependency
$E_{PA} = 2Cpq$	$E_{PA} = 2CPA / n(n-1)$
C: total number of connections between zones p: probability of the property "presence" q: probability of the property "absence" p + q = 1.0	C: total number of connections between zones P: number of zones with the property "presence" A: number of zones with the property "absence" n: total number of zones in the study area

	n = P + A

Table 2.1

1.2.5a Variability of the number of connections for a random distribution

Given a estimated number of connections P/A for a random spatial distribution we can calculate the standard deviation value $_{\#PA}$. According to the choice of the null hypothesis, the calculation of $_{\#PA}$ can be carried out in the manner presented in Table 2.2 below.

Variability of the number of EPA for a theoretical random spatial distribution

a) According to a null hypothesis of independence
$\sigma_{PA} = \sqrt{\{[2C+\Sigma V(V-1)]pq - 4[C+\Sigma V(V-1)]\, p^2q^2\}}$
C: total number of connections between zones V: number of neighbors of each zone #V: sum of neighbors of all the zones, with #V = 2C p: probability of the property "presence" q: probability of the property "absence"
b) According to a null hypothesis of dependency
$\sigma_{PA} = \sqrt{\{\{E_{PA} + [\Sigma V(V-1)PA)]/n(n-1)\} + \{4[C(C-1) - \Sigma V(V-1)]P(P-1)A(A-1)\}/n(n-1)(n-2)(n-3)\}}$
C: total number of connections between zones V: number of neighbors of each zone #V: sum of neighbors of all the zones, with #V = 2C P: number of zones with the property "presence" A: number of zones with the property "absence" n: total number of zones in the study area

Table 2.2

1.2.5b Calculation of the observed join count

The observed join count statistic expresses the total number of connections C between the zones of property "presence" and those of "absence". It can be formulated as followed

$$O_{PA} = \Sigma C_{PA}$$

1.2.5c Test of a significant difference between the random and the observed distribution

It is now a question of defining the similarity of the spatial distribution of features with "presence" and "absence" between the real observed situation and the theoretically random situation. The use of statistical tests allows us to estimate, with a defined risk of error, if the difference between the number of connections observed O_{PA} and that of the theoretically and random E_{PA} is sufficiently large to be regarded as significant. The z statistic, which expresses the standardized difference, is defined by the following equation proposed in Table 2.3. It is the same for the two situations of dependent or independent null hypothesis.

Calculation of z statistic

$$z_{obs} = O_{PA} - E_{PA} / \sigma_{PA}$$

O_{PA}: number of connections P/A between the zones in the area of study
E_{PA}: number of connections P/A for a theoretical random distribution
$\#_{PA}$: standard deviation of the theoretical random distribution

Table 2.3

Two types of test can be applied, answering the question of similarity between the two distributions in a general or specific way, using a bilateral or unilateral test respectively:

- The bilateral test checks if the spatial distribution of zones of "presence" in the study area is significantly different from a "random" distribution. In the event of rejection of the null hypothesis, one determines that the observed distribution is simply random. The alternative assumption of a bilateral test is expressed thus, H_1: O_{PA} # E_{PA}.
- The unilateral test checks in a more specific way if the spatial distribution of the zones of „presence“ in the study area is significantly distributed as either "grouped", or "dispersed". One will be able to thus formulate one of the two following alternative hypothesis, H_1: $O_{PA} < E_{PA}$ or $H_{1'}$: $O_{PA} > E_{PA}$. In the event of rejection of the null hypothesis, one can determine that the observed distribution will get significantly closer to either a "grouped" distribution, or to a "dispersed" distribution.

1.2.6. Examples of calculation for three observed spatial distributions

Let us take again the 3 examples of spatial distribution presented at Figure 2.4. Intuitively we can express their spatial distributions as "grouped", "random" and "dispersed". Let us test the membership of these distributions using the index of adjacency according to two situations' of independent and dependent null assumption. Being given that, in these 3 examples, we consider the same study area and that the number of zones "presence" is identical, we can calculate the two common parameters E_{PA} and $\#_{PA}$ for the two hypothesis, on the basis of additional information provided by the table of Figure 2.5 (C=19, #V=38, #V(V-1)=114). Whereas C = total no. of connections and V = total no. of neighbours

- For **an independent null hypothesis**, with a probability of occurrence "presence" p equal to 0.4 (thus q=0.6), the equations of tables 2.1 and 2.2 enable us to write:

$$E_{PA} = 2 \times 19 \times 0.4 \times 0.6$$

$$\underline{\underline{E_{PA} = 9.12}}$$

$$\sigma_{PA} = \sqrt{\{[(38+14)x0.4x0.6]-[4(19+114)0.42x0.62]\}}$$

$$\sigma_{PA} = \sqrt{(36.48 - 30.643)}$$

$$\sigma_{PA} = \sqrt{5.8368}$$

$$\underline{\underline{\sigma_{PA} = 2.416}}$$

- For **a dependent null hypothesis**, knowing that the number of observed events "presence" in the study area is equal to 5 per 11 zones on the whole (thus P=5 and A=6), the equations of tables 2.1 and 2.2 enable us to write:

$$E_{PA} = (2 \times 19 \times 5 \times 6) / (11 \times 10)$$

$$\underline{\underline{E_{PA} = 10.36}}$$

$$\sigma_{PA} = \sqrt{\{10.36+[(114x5x6)/(11x10)]+4[(19x18)-114]x5x4x6x5/(11x10x9x8)-10.36^2\}}$$

$$\sigma_{PA} = \sqrt{(10.36 + 31.9 + 69.09 - 107.40)}$$

$$\sigma_{PA} = \sqrt{3.14}$$

$$\underline{\underline{\sigma_{PA} = 1.772}}$$

These common values being calculated, we now can operate the statistical tests relating to the two situations of independent and dependent null hypothesis considered. Let us proceed in turn for each of the three spatial distributions presented at Figure 2.4.

Tests of membership of the three types of spatial distribution of binary properties

Observed distribution	Theoretical independent distribution	Theoretical dependent distribution
a) grouped 5 connections PA  O_{PA} **= 5**	$Z_{obs} = O_{PA} - E_{PA} / \sigma_{PA}$ $= (5 - 9.12) / 2.416$ $= -1.705$ *Bilateral test with 5%* H_1: $O_{PA} \neq E_{PA}$ $Z_{crit} = -1.960$ thus $\lvert Z_{obs} \rvert < \lvert Z_{crit} \rvert$ thus H_0 cannot be rejected, thus the observed distribution **is not significantly different** from a random distribution.	$Z_{obs} = O_{PA} - E_{PA} / \sigma_{PA}$ $= (5 - 10.36) / 1.772$ $= -3.025$ *Bilateral test with 5%* H_1: $O_{PA} \neq E_{PA}$ $Z_{crit} = -1.960$ thus $\lvert Z_{obs} \rvert > \lvert Z_{crit} \rvert$ thus H_0 can be rejected, thus the observed distribution **is significantly different** from a random distribution. *Unilateral test with 2.5%* H_1: $O_{PA} < E_{PA}$ $Z_{crit} = -1.960$ thus $Z_{obs} < Z_{crit}$ thus H_0 can be rejected, thus the observed distribution is significantly different from a random distribution and is associated with **a grouped distribution**.

Table 2.4a

Observed distribution	Theoretical independent distribution	Theoretical dependent distribution
b) random 11 connections PA  O_{PA} **= 11**	$Z_{obs} = O_{PA} - E_{PA} / \sigma_{PA}$ $= (11 - 9.12)/2.416$ $= 0.778$ *Bilateral test with 5%* H_1: $O_{PA} \neq E_{PA}$ $Z_{crit} = -1.960$ thus $\lvert Z_{obs} \rvert < \lvert Z_{crit} \rvert$ thus H_0 cannot be rejected, thus the observed distribution **is not significantly different** from a random distribution.	$Z_{obs} = O_{PA} - E_{PA} / \sigma_{PA}$ $= (11 - 10.36)/1.772$ $= 0.361$ *Bilateral test with 5%* H_1: $O_{PA} \neq E_{PA}$ $Z_{crit} = -1.960$ thus $\lvert Z_{obs} \rvert < \lvert Z_{crit} \rvert$ thus H_0 cannot be rejected, thus the observed distribution **is not significantly different** from a random distribution.

Table 2.4b

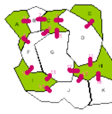
Observed distribution	Theoretical independent distribution	Theoretical dependent distribution
c) dispersed 14 connections PA  O_{PA} = 14	$Z_{obs} = O_{PA} - E_{PA} / \sigma_{PA}$ $= (14 - 9.12)/2.416$ $= 2.019$ *Bilateral test with 5%* H_1: $O_{PA} \neq E_{PA}$ $Z_{crit} = -1.960$ thus $\|Z_{obs}\| > \|Z_{crit}\|$ thus H_0 can be rejected, thus the observed distribution **is significantly different** from a random distribution. *Unilateral test with 2.5%* H_1: $O_{PA} > E_{PA}$ $Z_{crit} = 1.960$ thus $Z_{obs} > Z_{crit}$ thus H_0 can be rejected, thus the observed distribution is significantly different from a random distribution and is associated with **a dispersed distribution**.	$Z_{obs} = O_{PA} - E_{PA} / \sigma_{PA}$ $= (14 - 10.36)/1.772$ $= 2.054$ *Bilateral test with 5%* H_1: $O_{PA} \neq E_{PA}$ $Z_{crit} = -1.960$ thus $\|Z_{obs}\| > \|Z_{crit}\|$ thus H_0 can be rejected, thus the observed distribution **is significantly different** from a random distribution. *Unilateral test with 2.5%* H_1: $O_{PA} > E_{PA}$ $Z_{crit} = 1.960$ thus $Z_{obs} > Z_{crit}$ thus H_0 can be rejected, thus the observed distribution is significantly different from a random distribution and is associated with **a dispersed distribution**.

Table 2.4c

1.2.7. The Moran's coefficient of autocorrelation (at the ordinal and cardinal level)

Many phenomena can be measured on an ordinal or cardinal level. In order to preserve informational detail of spatial feature properties, it is often interesting to turn to a spatial autocorrelation index able to take into account these ranks or these intervals of values. As one will see the formulation of the Moran's coefficient and its application to an ordinal level of measurement is only meaningful if the rank difference has significance in its interpretation.
As for the join count statistics, the description of the spatial dependency can be expressed by the type of spatial distribution of properties within the study area: **grouped**, **random** or **dispersed**:

- The distribution is known as spatially **grouped** when properties of close value are contiguous. The spatial dependency is considered to be positively strong because the values vary in space in a "continuous" way appreciably. The spatial proximity involves a similarity of the properties (see Fig 2.6a).
- The distribution is known as spatially **random** when the distribution of properties in space is unspecified. The spatial dependency is considered to be null because there is no relation between the spatial proximity and the similarity of the properties (see Fig 2.6b).
- The distribution known as is spatially **random** when properties of very different value are contiguous. The spatial dependency is considered to be negatively strong because the values vary in space in a "discontinuous" way appreciably. The spatial proximity involves a great difference of the properties (see Fig 2.6c).

Figure 2.6 illustrates these 3 types of spatial distribution for the same study area made up of 7 zones (districts for example) on which are distributed 7 properties expressing the density of inhabitants per hectare.

Three types of spatial distribution of a set of 7 continuous properties

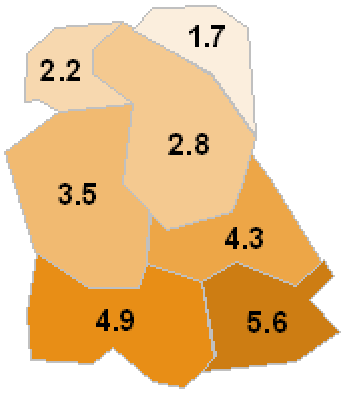
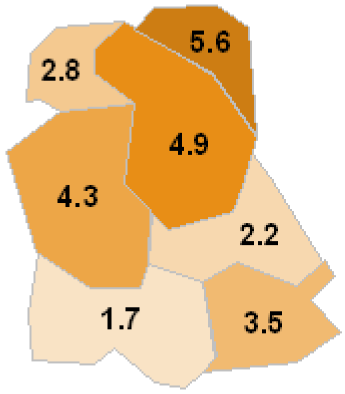
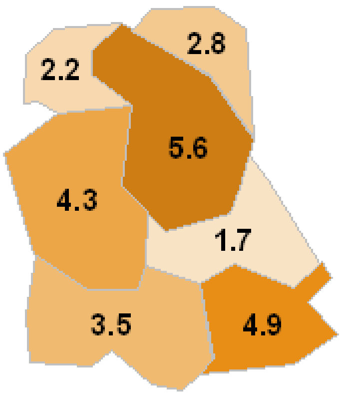
<table>
<tr><th>2.6a) grouped</th><th>2.6b) random</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Strong positive spatial dependency</td><td>No spatial dependency</td></tr>
<tr><th>2.6c) dispersed</th><th>2.6d) 7 properties to be distributed</th></tr>
<tr><td>
</td><td>x
1.7
2.2
2.8
3.5
4.3
4.9
5.6</td></tr>
<tr><td>Strong negative spatial dependency</td><td>7 values of population density per hectare</td></tr>
</table>

Figure 2.6

The Moran's autocorrelation coefficient, also called **Moran's I index**, makes it possible to characterize the nature of this distribution according to three types (grouped, random or dispersed) and in consequence to deduce the **force** (strength) and **the direction** (positive or negative) of the spatial dependency.

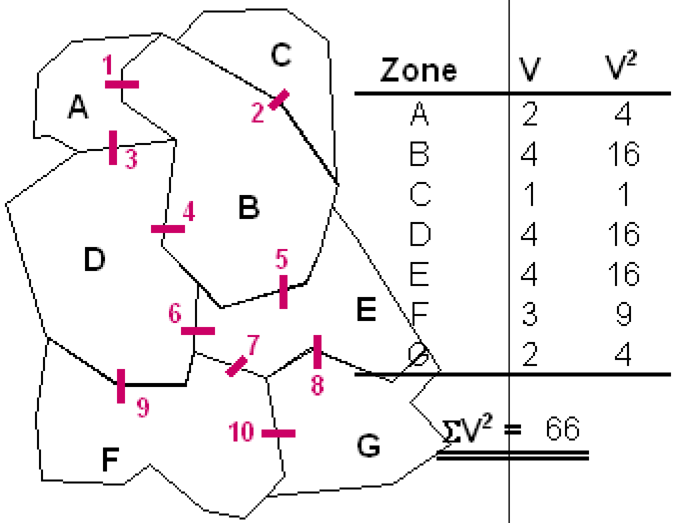
Moran's coefficient connects the differences in values between contiguous areas with reference to the total variability. Its value varies between –1 and +1. The force of the spatial autocorrelation is expressed by the value varying from 0 to 1, while the direction of the dependence is indicated by the sign, following the example of other coefficients of correlation.

Similar to the coefficient of adjacency, the definition of these differences of value between contiguous zones for a theoretical random distribution is related to two factors: **spatial arrangement** of zones in the study area on the one hand, and **the choice of null hypothesis** on the other.

1.2.8. The spatial arrangement of zones

As previously shown, according to the number and the shape of the spatial features, their adjacency results in a considerable number of connections, independent of their thematic properties. The identification of connections and the neighbours for each of the 7 districts presented on Figure 2.6 is illustrated below:

Description of the spatial arrangement of the 7 districts through their adjacencies

Identification of the adjacencies (connections)	**List of connections by zone**
	
7 contiguous zones with 10 connections (C)	V: number of neighbors of each zone The total number of neighbors is equal to twice the

	number of connections (C)

Figure 2.7

1.2.8a Context of the null assumption

In the case of distribution of properties measured on a cardinal level, the two alternatives of null hypothesis presented for the join count statistic remain similar. It is thus a question of determining if the study area is regarded as an independent sample *(free sampling)* or dependent *(non-free sampling)*. The identification of one of these two situations will make it possible to determine the nature of the theoretical distribution with which the observed distribution of the properties will be confronted.

On the assumption of an independent sample, one can see that the set of properties to be distributed in the study area (see figure 2.6d) results from a random sampling of a normal population; this is **the hypothesis of normality**.

On the assumption of a dependent sample, one considers the way of distributing the properties observed in the 7 zones, therefore one is interested in their spatial arrangement. In this case, the theoretical distribution will correspond to a random distribution; this is **the hypothesis of the random distribution**.

Generally, in practice, one gives the preference to a situation of dependent sampling because one usually determines the presence of the observed properties and one wishes to know the nature of their spatial distribution, i.e. to know if it is significantly different from a random distribution, therefore influenced by space.

1.2.9. Estimate of the number of connections for a random distribution

Whatever the selected null hypothesis (of independence or dependence), the theoretical value for a random distribution, E_I, is calculated the same manner, as presented in Table 2.5 below.

Calculation of the theoretical value EI

$E_I = -(1 / n-1)$
n: number of zones in the study area

Table 2.5

1.2.9a Variability of the theoretical value

With this estimated value EI, is associated a variability expressed by a value of standard deviation $\#_I$. According to the choice of the null hypothesis, the calculation of $\#_I$ is carried out in the way presented in Table 2.6 below. Once again it is more complex to calculate in the situation of dependency.

Variability of EI for a random theoretical distribution

a) According to a null assumption of independence	b) According to a null assumption of dependence
$\sigma_I = \sqrt{\{[n^2C + 3C^2 - n\Sigma V^2]}$	$\sigma_I = \sqrt{\{n[C(n^2+3-3n) + 3\,C^2 - n\Sigma V^2] - k[C(n^2-n) + 6C^2 - 2n\Sigma V^2\} / [C^2(n-1)(n-2)(n-3)]\}}$
C: total number of connections between zones V: number of neighbors of each zone n: total number of zones in the study area	C: total number of connections between zones V: number of neighbors of each zone n: total number of zones in the study area k: kurtosis of the observed distribution of the values

Table 2.6

1.2.9b Calculation of the Moran's index

The Moran's index expresses the importance of the difference of properties (values) between all the pairs (x_k, x_l) of contiguous zones

$$O_I = n\Sigma\ (x_k - x_{aver})(x_l - x_{aver})\ /\ C\Sigma\ (x - x_{aver})^2$$

1.2.9c Test of a significant difference between the random and the observed distribution

It is now a question of defining the similarity of the spatial distribution of the properties between the real observed situation and a theoretical random situation. The Z statistic expressing this standardized difference is defined by the equation presented in Table 2.7; it is the same one for the two situations of dependent or independent null hypothesis.

Calculation of z statistic

$Z_{obs} = O_I - E_I\ /\ \sigma_I$
O_I: value of Moran's I index calculated on the observed distribution E_I: value of Moran's I index estimated for a theoretical random spatial distribution #$_I$: standard deviation of the theoretical random distribution a

Table 2.7

Once again, two types of test can be applied to question the similarity between the two distributions in a general or specific way, using a bilateral or unilateral test respectively:

- The bilateral test checks to see if the observed spatial distribution of properties in the study area departs from a "random" distribution significantly. the event of rejection of the null hypothesis, one determines that the observed distribution is simply not random. The alternative hypothesis of a bilateral test is, H_1: O_I # E_I .
- The unilateral test checks in a more specific way that the observed spatial distribution of the properties in the study area is significantly closer to either a "grouped" or "dispersed" distribution. Thus one will be able to formulate one of the two following alternative hypotheses, H_1: $O_I > E_I$ or $H_{1'}$: $O_I < E_I$. In the event of rejection of the null hypotheses, one deduces that the observed distribution is significantly closer to either a "grouped" distribution, with a strong positive spatial autocorrelation, or to a "dispersed" distribution with a strong negative spatial autocorrelation.

1.2.10. Examples of calculation for three observed spatial distributions

Let us return to the 3 examples of spatial distribution presented in Figure 2.6. Intuitively we can express the spatial distributions as "grouped", "random" and "dispersed" respectively. Now let's test the membership of these distributions using the Moran's autocorrelation coefficient according to two situations of null hypothesis, independent and dependent. Given that in these 3 examples we consider the same study area and the same set of 7 property values, we can calculate the two values E_I and $\#_I$ common with the three distributions for each of the two hypothesis, on the basis of additional information provided on Figure 2.7 (n = 7, C = 10, $\#V^2$ = 66) and in Table 2.8 below.

Calculation of indices common to the three examples

x	$(x - x_{aver})$	$(x - x_{aver})^2$	$(x - x_{aver})^4$
1.7	-1.87	3.497	12.228
2.2	-1.37	1.877	3.523
2.8	-0.77	0.593	0.352
3.5	-0.07	0.005	0.000
4.3	0.73	0.533	0.284
4.9	1.33	1.769	3.129
5.6	2.03	4.121	16.982

Σx = 25 $\quad \Sigma(x - x_{aver})^2$ = 12.394

x_{aver} = 3.57 $\quad \Sigma(x - x_{aver})^4$ = 36.497

$$s_x = \sqrt{[\Sigma(x - xaver)^2 / n]} = \sqrt{(12.394 / 7)} = \sqrt{1.77} = 1.33$$
$$k_x = \Sigma(x - xaver)^4 / n(s_x^4) = 36.497 / (7*3.13) = 1.67$$

Table 2.8

For an **independent null hypothesis**, the equations of tables 2.5 and 2.6 enable us to write:

$$E_I = -(1/6)$$

$$\underline{\underline{E_I = -0.167}}$$

$$\sigma_I = \sqrt{\{[(49 \times 10) + (3 \times 100) - (7 \times 66] / 100(49-1)\}}$$

$$\sigma_I = \sqrt{\{(490 + 300 - 462) / 4800\}}$$

$$\sigma_I = \sqrt{\{1252 / 4800\}}$$

$$\sigma_I = \sqrt{0.261}$$

$$\underline{\underline{\sigma_I = 0.511}}$$

For an **dependent null hypothesis**, the equations of tables 2.5 and 2.6 enable us to write:

$$E_I = -(1/6)$$

$$\underline{\underline{E_I = -0.167}}$$

$$\sigma_I = \sqrt{\{\{7[10(49+3-21) + (3 \times 100) - (7 \times 66)] - 1.67[10(49-7) + (6 \times 100) - (14 \times 66)\} / [100(7-1)(7-2)(7-3)]\}}$$

$$\sigma_I = \sqrt{\{[7(310+300-462) - 1.67(420+600-924)] / 100 \times 6 \times 5 \times 4\}}$$

$$\sigma_I = \sqrt{\{(7504-160.32) / 12000\}}$$

$$\sigma_I = \sqrt{\{7343.68 / 12000\}}$$

$$\sigma_I = \sqrt{0.612}$$

$$\underline{\underline{\sigma_I = 0.782}}$$

The calculation of the Moran’s coefficient for each of the three observed distributions is carried out on the basis of values produced in tables 2.8 and 2.9.

Calculation of indices specific to each of the three spatial distributions

a) Grouped distribution

ID C	x_k	$(x_k - x_{aver})$	x_l	$(x_l - x_{aver})$	$(x_k - x_{aver})(x_l - x_{aver})$
1	2.2	-1.37	2.8	-0.77	1.055
2	2.8	-0.77	1.7	-1.87	1.440
3	2.2	-1.37	3.5	-0.07	0.096
4	2.8	-0.77	3.5	-0.07	0.054
5	2.8	-0.77	4.3	0.73	-0.562
6	3.5	-0.07	4.3	0.73	-0.051
7	4.3	0.73	4.9	1.33	0.971
8	4.3	0.73	5.6	2.03	1.482
9	3.5	-0.07	4.9	1.33	-0.093
10	4.9	1.33	5.6	2.03	2.700
				$\Sigma(x_k - x_{aver})(x_l - x_{aver}) =$	7.091

b) Random distribution

ID C	x_k	$(x_k - x_{aver})$	x_l	$(x_l - x_{aver})$	$(x_k - x_{aver})(x_l - x_{aver})$
1	2.8	-0.77	4.9	1.33	-1.024
2	4.9	1.33	5.6	2.03	2.700
3	2.8	-0.77	4.3	0.73	-0.562
4	4.9	1.33	4.3	0.73	0.971
5	4.9	1.33	2.2	-1.37	-1.822
6	4.3	0.73	2.2	-1.37	-1.000
7	2.2	-1.37	1.7	-1.87	2.562
8	2.2	-1.37	3.5	-0.07	0.096
9	4.3	0.73	1.7	-1.87	-1.365
10	1.7	-1.87	3.5	-0.07	0.131
				$\Sigma(x_k - x_{aver})(x_l - x_{aver}) =$	0.686

c) Dispersed Distribution

ID C	x_k	$(x_k - x_{aver})$	x_l	$(x_l - x_{aver})$	$(x_k - x_{aver})(x_l - x_{aver})$
1	2.2	-1.37	5.6	2.03	-2.781
2	5.6	2.03	2.8	-0.77	-1.563
3	2.2	-1.37	4.3	0.73	-1.000
4	5.6	2.03	4.3	0.73	1.482
5	5.6	2.03	1.7	-1.87	-3.796
6	4.3	0.73	1.7	-1.87	-1.365
7	1.7	-1.87	3.5	-0.07	0.131
8	1.7	-1.87	4.9	1.33	-2.487
9	4.3	0.73	3.5	-0.07	-0.051
10	3.5	-0.07	4.9	1.33	-0.093
				$\Sigma(x_k - x_{aver})(x_l - x_{aver}) =$	-11.524

Table 2.9

Thus, for the **grouped distribution**, the Moran's index value will be:

$$O_I = 7 \times 7.091/10 \times 12.394$$
$$O_I = 49.637/123.94$$
$$O_I = 0.4$$

For the **random distribution**, the Moran's index value will be:

$$O_I = 7 \times 0.686/10 \times 12.394$$
$$O_I = 4.802/123.94$$
$$O_I = 0.04$$

Lastly, for the **dispersed distribution**, the Moran's index value will be:

$$O_I = 7 X - 11.524/10x\ 12.394$$

$$O_I = -80.668/123.94$$

$$O_I = -\ 0.65$$

These coefficient values being calculated, we can now carry out the statistical tests relating to the two situations of independent and dependent null hypothesis considered. Let us proceed in turn for each of the three observed spatial distributions, presented in Figure 2.6.

Tests of membership of the three types of spatial distribution of cardinal properties

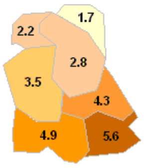
Observed distribution	Independent theoretical distribution	Dependent theoretical distribution
a) grouped  $O_I = 0.4$	$Z_{obs} = O_I - E_I / \sigma_I$ $= (0.4 + 0.167)/0.261$ $= 2.17$ *Bilateral test with 5%* H_1: $O_I \neq E_I$ $Z_{crit} = -1.960$ thus $\mid Z_{obs} \mid > \mid Z_{crit} \mid$, thus H_0 can be rejected, thus the observed distribution **is significantly different** from a random distribution. *Unilateral test with 2.5%* H_1: $O_I > E_I$ $Z_{crit} = -1.960$ thus $Z_{obs} > Z_{crit}$ thus H_0 can be rejected, thus the observed distribution is significantly different from a random distribution and is then connected with **a grouped distribution**.	$Z_{obs} = O_I - E_I / \sigma_I$ $= (0.4 + 0.167)/0.782$ $= 0.73$ *Bilateral test with 5%* H_1: $O_I \neq E_I$ $Z_{crit} = -1.960$ thus $\mid Z_{obs} \mid < \mid Z_{crit} \mid$, thus H_0 cannot be rejected, thus the observed distribution **is not significantly different** from a random distribution.

Table 2.10a

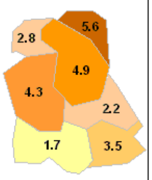
Observed distribution	Independent theoretical distribution	Dependent theoretical distribution
b) random  $O_I = 0.04$	$Z_{obs} = O_I - E_I / \sigma_I$ $= (0.04 + 0.167)/0.261$ $= 0.79$ *Bilateral test with 5%* H_1: $O_I \neq E_I$ $Z_{crit} = -1.960$ thus $\mid Z_{obs} \mid < \mid Z_{crit} \mid$, thus H_0 cannot be rejected, thus the observed distribution **is not significantly different from a random distribution**.	$Z_{obs} = O_I - E_I / \sigma_I$ $= (0.04 + 0.167)/0.782$ $= 0.26$ *Bilateral test with 5%* H_1: $O_I \neq E_I$ $Z_{crit} = -1.960$ thus $\mid Z_{obs} \mid < \mid Z_{crit} \mid$, thus H_0 cannot be rejected, thus the observed distribution **is not significantly different from a random distribution**.

Table 2.10b

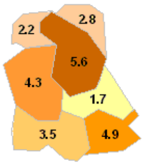
Observed distribution	Independent theoretical distribution	Dependent theoretical distribution
c) dispersed  $O_I = -0.65$	$Z_{obs} = O_I - E_I / \sigma_I$ $= (-0.65 + 0.167)/0.261$ $= -1.85$ *Bilateral test with 5%* H_1: $O_I \neq E_I$ $Z_{crit} = -1.960$ thus $\mid Z_{obs} \mid < \mid Z_{crit} \mid$, thus H_0 cannot be rejected, thus the observed distribution **is not significantly different** from a random distribution.	$Z_{obs} = O_I - E_I / \sigma_I$ $= (-0.65 + 0.167)/0.782$ $= -0.62$ *Bilateral test with 5%* H_1: $O_I \neq E_I$ $Z_{crit} = -1.960$ thus $\mid Z_{obs} \mid < \mid Z_{crit} \mid$, thus H_0 cannot be rejected, thus the observed distribution **is not significantly different** from a random distribution.

Table 2.10c

It is observed that with a risk of error of 5%, the 3 distributions are not significantly different from a random space distribution of the values, under the hypothesis of the dependency, while only the distribution considered as grouped is significantly similar to such a spatial distribution, under the hypothesis of an independent sampling.

1.3. Spatial Arrangement

Entry

In order to account for the complexity to describe the way in which the properties of a discontinuous phenomenon are distributed in space, many indices of **spatial arrangement** come to supplement the knowledge provided by the indices of spatial dependency. If the latter stressed the dependency of the thematic properties compared to the spatial proximity, the indices of arrangement try to account for the way in which the spatial objects are distributed in the study area.

1.3.1. Indices Of Arrangement

It would be illusory and quite tedious to propose an exhaustive list of indices of spatial arrangement suggested in the literature. The most productive fields of research in this respect are certainly those of the geography, ecology, regional sciences and the numerical image processing. From the methodological point of view, most significant is to seize **the aspects** of the spatial arrangement which one wishes to describe through indices, **the way of interpreting them**, as well as their **context of use.**
The context of application of the indices of arrangement is influenced by the following factors:

- The nature of spatial features:
    - For point and linear objects distributed in the study area, one will describe their spatial arrangement by a measurement of density per quadrat. We will not approach this situation within the framework of this Unit.
    - For **contiguous zonal objects** proposing a spatial division of the study area. It is on this type of context that we will pay our attention in this Unit.
- Mode of description of the spatial distribution:
    - Our major interest is certainly the description of the arrangement of spatial objects. In **object mode** the units of observation fully correspond to the spatial objects.
    - In **image mode**, the spatial objects correspond to the "zonal regions" produced by the aggregation of the units of observation, the contiguous cells sharing the same property. But it is also possible to be interested in the spatial arrangement at the level of the neighborhood of each cell.
- The scale of description of the spatial arrangement:
    - At the scale of the whole of the study area, one will account for the **"structure"** of the spatial arrangement, either globally for all the objects, or specifically by category or class of properties of these objects.

At the scale of the study area: “the structure”

On the level of the spatial objects	On the level of a group of objects with	On the level of the study area

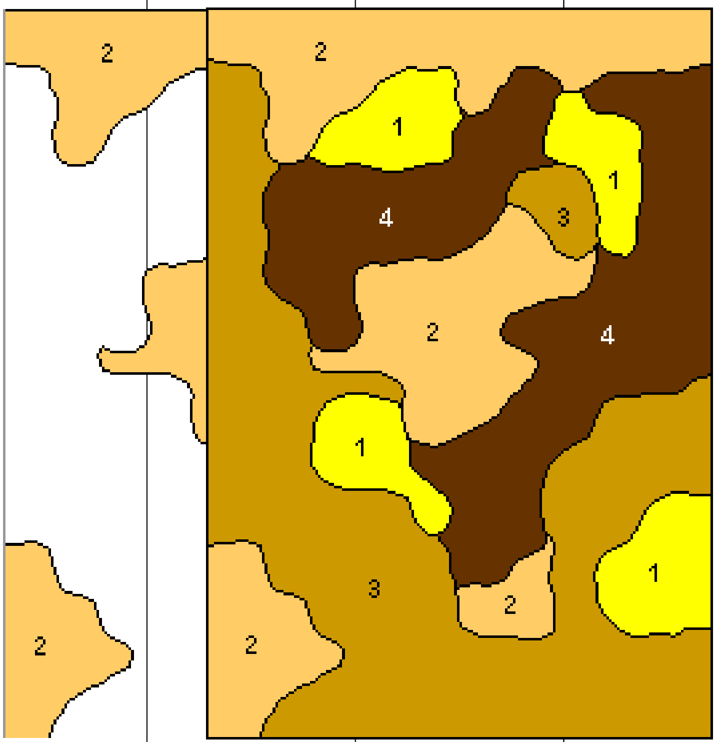
	a common thematic property	
		
12 spatial objects	4 objects with a class of average aptitude for agriculture	Distribution of the thematic properties of the 12 objects

Figure 3.1

- At the scale of the neighborhood of each spatial object, one will describe its spatial context.
- At the scale of the neighborhood of the unit of observation in image mode, one will be able to account for the **"texture"** of the spatial arrangement.

At the scale of neighborhood of the unit of observation: "the texture"

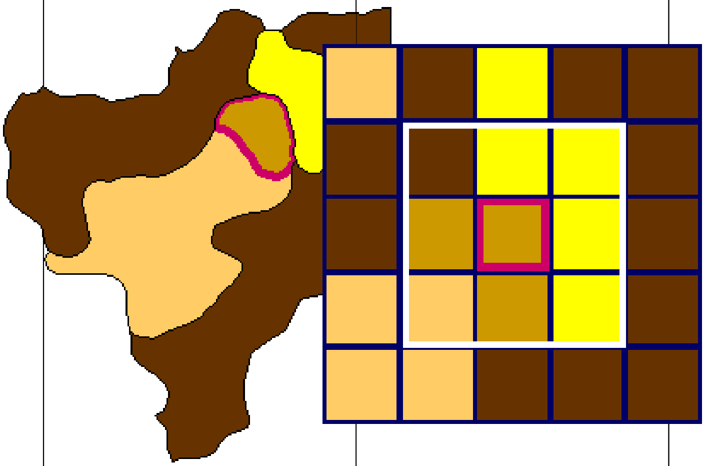
In object mode, for each unit of observation (object)	**In image mode, for each unit of observation (cell)**
	
Immediate neighborhood (of 1st order) of the object with enhanced limits in red	Immediate neighborhood (window 3x3) and of 2nd order (window 5x5) of the central cell

Figure 3.2

- Geometrical or thematic dimension of the distribution of the phenomenon:
  - **The geometrical indices** of arrangement describe the properties of size, of shape and the distribution in the space of a set of objects, corresponding to all those of the study area or to each categories or class of thematic properties present.
  - **The thematic indices** of arrangement account for the way in which the thematic properties are distributed on at scale of the study area or the neighborhood of the spatial entities. They are thus complementary to the geometrical indices because they do not consider the shape nor the size of spatial objects or regions. So they can apply as well to objects or to cells.

In order to present a synthetic and structured vision these descriptors of spatial arrangement for zonal units, we will organize this description according to the factors of "scale of description" and of "dimension" indices. In this Unit only major indices of spatial arrangement will be presented, those one can find in GIS or spatial analysis or image processing software (Idrisi, Fragstats, Envi…).

1.3.2. At the scale of the study area: "the structure"

Which are the number, the size, the shape, the localization of the spatial objects constituting the study area? Here is a type of questions relating to the spatial arrangement which concerns before all the geometrical aspect of the objects, considered either individually or in relation to the whole. One can supplement this structural description by integrating the thematic dimension of information. The thematic dimension arises either in term of selection criterion to return account of the global geometrical properties of each class or category of objects, or in a more significant way to deduce diversity of the thematic properties within the study area.

Table 3.1 summarizes the principal geometrical and thematic indices of structure, at the levels of the objects, of the classes or categories of objects and of the whole of the study area. This distinction in three levels corresponds to that presented by McGarical and Marks (1994) in their manual "FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure ". The authors name respectively them patches, classes and landscape. A certain number of these indices, in particular at the individual level of the objects and that of the categories, were already presented in the basic module of spatial analysis (B-AN), within Units 2 and 3 of Lesson 2 **"Spatial discrete variables"**.

Principal geometrical and thematic indices of structure
(terms in italic refer to the handbook of the FRAGSTATS software

		Description scale (level)		
Dimension	Indices	Object (*Patch*) (ref. to B-AN L2/U2)	Category/Class (*Class*) (ref. to B-AN L2/U3)	Study Area (*Landscape*)
Geometric	Area	Area (*AREA*)	Area (CA) Proportion (*%LAND*)	Area (TA)
	Perimeter	Perimeter (*PERIM*)	Perimeter (*TE*) Edge density (*ED*)	Perimeter (*TE*) Edge density (*ED*)
	Shape	Shape (*SHAPE*) Fractal (*FRACT*)	Average shape (*MSI*) Average fractal (*MPFD*)	Average shape (*MSI*) Average fractal (*MPFD*)
	Density		Number of objects (*NP*) Density of objects (*PD*) Average size (*MPS*)	Number of objects (*NP*) Density of objects (*PD*) Average size (*MPS*)
	Neighborhood	Distance (*NEAR*)	Mean distance (*MNN*) Standard distance (*NNSD*) Coef. of variation (*NNCV*)	Mean distance (*MNN*) Standard distance (*NNSD*) Coef. of variation (*NNCV*)
Thematic	Centrality			Mode (*MOD*) Median (*MED*) Mean (*AVG*)
	Diversité			Shannon's index (*SHDI*) Richness (*PR*) Richness density (*PRD*) Regularity (*SHEI*)

Table 3.1

We will illustrate all these indices with the situation presented at Figure 3.1, but only indices not yet presented at Lesson 2 of the basic module of **spatial analysis** (B-AN) will be described in a detailed way.

1.3.3. Indices of structure at the level of the objects

Among the five indices of Table 3.1, those of surface and perimeter are already known. Thus let us describe the three remainders.

Index of shape (SHAPE):
$$SHAPE = p_j / (2 \sqrt{(\Pi a_J)})$$
p_J: perimeter (m) of the object j a_J : surface (m^2) of the object j

Fractal index (FRACT):
$$FRACT = 2\ln p_J \ / \ln a_J$$
ln: natural log p_J: perimeter (m) of the object j a_J : surface (m^2) of the object j
Interpretation: Index value varies between 1 and 2. The more the shape of the object is complex, the more its value approaches 2.

Index of distance to nearest neighbor of the same category / class (NEAR):
$$NEAR = h_{ij}$$
h_{ij}: shortest distance (m) between the limit of the object i and that of the nearest object j of the same category

Value of the five indices of structure for the 12 objects of the study area illustrated in Figure 3.1

ID Object	ABILITY	AREA (ha)	PERIM (m)	SHAPE	FRACT	NEAR (m)
1	Poor	498	11400	1.28	1.03	1800
2	Poor	430	11600	1.4	1.04	1800
3	Poor	469	12200	1.41	1.04	3201.56
4	Poor	469	10200	1.18	1.02	3201.56
5	Average	1427	29600	1.96	1.08	1964.69
6	Average	1476	25600	1.67	1.06	1964.69
7	Average	288	9000	1.33	1.04	2549.51
8	Average	629	13000	1.3	1.03	2549.51
9	Good	5074	74200	2.6	1.11	3041.38
10	Good	263	8000	1.23	1.03	3041.38
11	Excellent	1612	27400	1.71	1.06	300
12	Excellent	2365	40600	2.09	1.09	300

Table 3.2

Exercises on Figure 3.1 and Table 3.2

1. On Figure 3.1 identify the objects:
   - whose surface is respectively smallest, largest
   - whose perimeter is respectively smallest, largest
   - whose shape is respectively simplest, most complex
   - the nearest of the same category
2. Compare your visual interpretation with the value of the indices in Table 3.2:
   - which are the simplest geometrical properties, and the most difficult to interpret?
   - which is the relation between the shape index value and that of the fractal index?
3. What can one observe in connection with the index of distance?

1.3.4. Indices of structure at the level of the categories / classes of objects

Among the indices of Table 3.1, those of surface and density are already known. Here the description of the other indices.

Index of density of contours (ED):

$$ED = \Sigma p_{ij} / A$$

#p_{ij} : sum of perimeters (m) of all the objects j pertaining to category i

A: surface (m^2) of the study area

Average index of shape (MSI):

$$MSI = \Sigma [p_{ij} / (2 \sqrt{(\Pi a_{ij})})] / n_i$$

n_i: number of objects of category / class i

Average fractal index (MPFD):

$$MPFD = \Sigma [2 \ln p_{ij} / \ln a_{ij}] / n_i$$

Interpretation: Index value MPFD varies between 1 and 2.
The more complex the average shape of the objects of the class is,
the more its value approaches 2.

A Index of average distance to neighbors of same category/class (MNN):

$$MNN = \Sigma h_{ij} / n_i$$

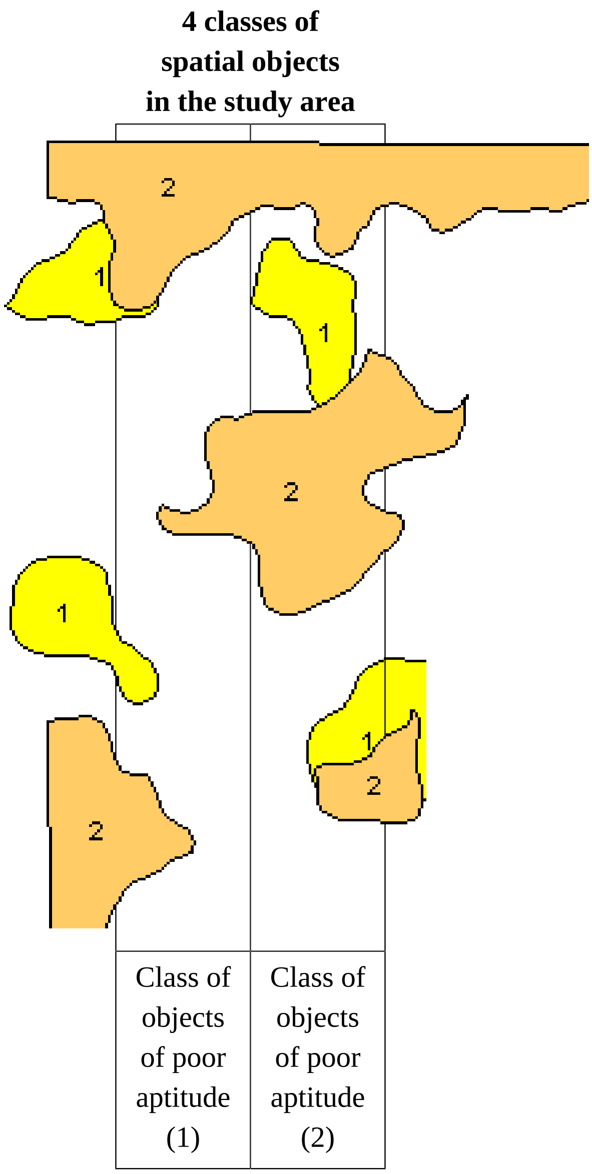
h_{ij}: shortest distance (m) between
the limit of the object i and that of the nearest object j of the same category.

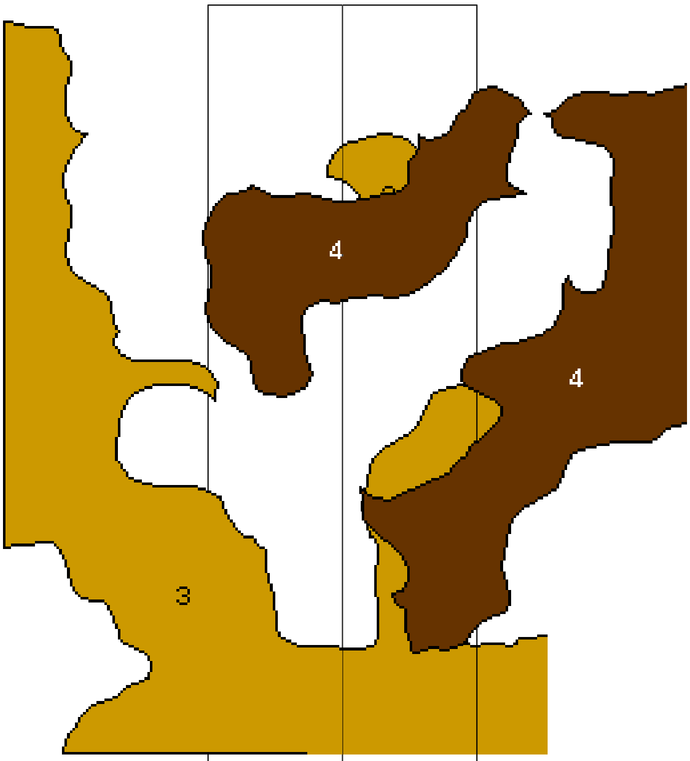
Index of dispersion distance to the neighbors of same category/class (NNSD):

$$NNSD = \sqrt{ \{ \Sigma [h_{ij} - (\Sigma h_{ij} / N_i)]^2 / n_i \} }$$

Coefficient of variation of the distance to the neighbors of same category/class (NNCV):

$$NNCV = (MNN / NNSD)\ 100$$


4 classes of
spatial objects
in the study area
2
1
1
2
1
2
1
2
Class of objects of poor aptitude (1)
Class of objects of poor aptitude (2)



Class of objects of poor aptitude (3)	Class of objects of poor aptitude (4)

Figure 3.3

Value of the indices of structure for the 4 classes of objects

in the study area illustrated in Figure 3.3

ID Class	1	2	3	4
ABILITY	Poor	Average	Good	Excellent
CA (ha)	1866	3820	5337	3977
%LAND	12.44	25.47	35.58	26.51
TE (m)	42500	60500	58500	61300
ED (m/ha)	2.83	4.03	3.9	4.09
MSI	1.32	1.56	1.92	1.9
MPFD	1.04	1.05	1.07	1.08
NP	4	4	2	2
PD (#/100ha)	0.03	0.03	0.01	0.01
MPS (ha)	466.5	955	2668.5	1988.5
MNN (m)	2500.8	2257.1	3041.38	300
NNSD (m)	700.78	292.41	0	0
NNCV (%)	28.02	12.96	0	0

Table 3.3

Exercises on Figure 3.3 and Table 3.3

1. On Figure 3.3 identify the classes:
   - whose surface is respectively smallest, largest
   - whose perimeter is respectively smallest, largest
   - whose shape is respectively simplest, most complex
   - whose average distance between the entities is largest
2. Compare your visual interpretation with the value of the indices in Table 3.3:
   - which are the geometrical properties simplest and most difficult to estimate?
   - which is the relation between the value of the average index of shape and that of the average fractal index?

1.3.5. Indices of structure at the level of the whole study area

By extension, all the definite geometrical indices for the level category/class in Table 3.1 can be applied to the whole of the objects of the study area. It is thus superfluous to again define them. It is on the other hand possible at this level to produce thematic indices of structure returning account either of the central tendency, or of the variability of the thematic properties in the whole of the study area. Among many proposed we will retain seven of them, including three of central tendency:

Index of majority, mode (MOD), nominal level:
$$MOD = v_{fmax}$$
v_{fmax}: value whose surface is in a majority in the whole of the area
Interpretation: It is the thematic value (category) most present in the area. In object mode, it is the category whose surface is in a majority, in mode image it is that of which the number of cells is largest.

Median index (MED), ordinal level:
$$MED = v_{med}$$
v_{med}:value of the central rank for the whole of the units of observation in the area
Interpretation: It is the thematic value of the unit of observation (UO) positioned in the middle of the ordered sequence of the values in the area. In object mode the units of observation are ordered according to of their property value, the median value is that of the object positioned in the middle of this sequence. It is noted that the respective surface of the objects does not contribute, but rather the number of feature relating to each property (importance of fragmentation). In image mode, as the Units of observation have the same surface, the median index considers the surface assigned to each property.

Index of average (AVG), cardinal level:
$$AVG = \sum_i (V_i * A_i) / \sum_i (A_i)$$
v_i: thematic value of the unit of observation i a_i: surface of the unit of observation i
Interpretation: This index is in fact an average weighted by the relative surface of each UO. In mode image, it is necessary to combine two layers (numerical grids), one identifying the areas (spatial objects) and the other describing the thematic property in each area.

Shannon's Index of diversity (SHDI):
$$SHDI = \sum (P_i * \ln P_i)$$
P_i: proportion of the surface of the study area occupied by category/class i ln: natural logarithm

Interpretation: Index SHDI is equal to 0 when the study area is made up of only one type of category or class (thematic homogeneity). Its value increases according to the number of values (richness, diversity) as well as to the uniformity tendency of the surfaces of each type.
An alternative of this index is proposed by the Simpson's index of diversity: SIDI = #(P_i ^{2}). The interest of this index is that its value varies between 0 and 1. As for the index of Shannon, it is equal to 0 when the thematic diversity is null. Its value tends towards 1 when the thematic diversity increases and that the surfaces of each type of category or class tend to the uniformity.

Index of richness (PR):
$$PR = c$$
c: number of categories or classes (types) in the study area
Interpretation: It is a simple indication of diversity expressing the thematic richness in the study area.

Index of density of richness (PRD):
$$PRD = (c\ /\ A)\ 10'000 * 100$$
c: number of categories or classes (types) in the study area A: surface (m^2) of the study area
Interpretation: This index expresses the density of richness per unit of 100 hectares. It is thus possible to compare the richness of study areas of different surfaces.

Shannon's index of regularity (SHEI):
$$SHEI = (\Sigma(P_i * \ln P_i))\ /\ \ln c$$
P_i:proportion of the surface of the study area occupied by category/class i c: number of categories or classes (types) in the study area ln: natural logarithm
Interpretation: SHEI Index value varies between 0 and 1. Its value is equal to 0 when the study area is made up of only one type of category or class (thematic homogeneity). Its value approaches 0 when the surfaces of each type of category or class are very different (irregularity, predominance of a type). The index is equal to 1 when the surfaces of each type are perfectly equal (regularity).

Distribution of the thematic classes in the study area	Value of indices of structure at the level of the whole study area

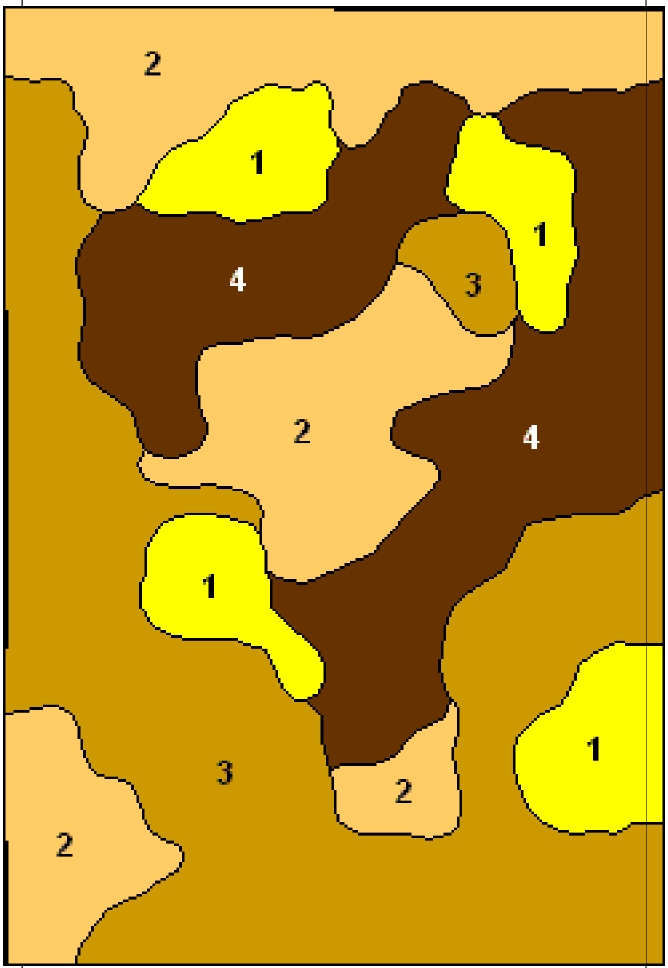


Figure 3.4

Geometrical Indices	
TA (ha)	15000
TE (m)	111400
ED (m/ha)	7.43
MSI	1.59
MPFD	1.05
NP	12
PD (#/100ha)	0.08
MPS (ha)	1250
MNN (m)	2142.9
NNSD (m)	969.45
NNCV (%)	45.24
Thematic Indices	
MOD	3
MED	2
SHDI	1.33
SIDI	0.72
PR	4
PRD (#/100ha)	0.03
SHEI	0.96

Table 3.4

Exercises on Figure 3.4 and Table 3.4:

1. Geometrical indices:
   - In the table of geometrical indices distinguish the most suited:
     - for the description of a study area?
     - for the comparison of the characteristics of several study areas?
2. Thematic indices:
   - Based on the observation of the map, which seems to be the modal class? Is this confirmed by the value of the thematic modal index?
   - Reconstitute the ordered sequence of values to determine the median value. Why this value is different from the modal value?

- How could one characterize in a few words the spatial arrangement of the properties of soil aptitude in the study area? Select the most relevant thematic indices among: SHDI, SIDI, PR, PRD, SHEI.

1.3.6. At the scale of the neighborhood: “the texture”

rks are arranged. This description of local arrangement will account for the differences of spatial organization between urban districts whose constituent elements are often identical: zones of individual or collective habitat, parks industrial and commercial areas, …
Just like the structure, the **texture** expresses the arrangement of properties (thematic) in space, but at **a larger scale**. This change of scale corresponds to a fitting of the spatial entities, namely the units of observation. In order to characterize for example the texture of build up areas, it will be necessary to handle the elements which constitute them, i.e. the units of observation defined at a larger scale. This constraint affects much more the object mode than the image mode. Indeed, we saw that in image mode, the functional spatial objects consist of a set of contiguous units of observation, the regions. It is consequently possible to describe at the same time the structure by spatial arrangement of the regions and their texture by arrangement of the cells.
This strict presentation of texture can fall under a broader concept of spatial arrangement of neighborhood at the same local scale. It is a question of operating**a contextual analysis** to relate the property of a unit of observation with those of its neighborhood, such as illustrated on figure 3.2. In this Unit, we will retain the vaster concept of spatial arrangement of neighborhood and will approach the methods of contextual analysis applied principally to information in image mode. In the literature, one finds these terms of analysis of texture, of context and of neighborhood applied to situations of various description of spatial arrangement. To simplify our matter and also to highlight the vast potential of these methods within the framework of the spatial analysis, we will regard these three terms as equivalents.

1.3.7. Principles of the contextual analysis

To connect the thematic properties of each unit of observation with those of their neighborhood, it is essential to define the neighborhood beforehand to be considered, as well as the numerical indicator expressing this relation. The neighborhood of analysis is expressed by the **moving window** which identifies successively the set of neighbor entities to be processed. It is based on the topological concept of adjacency of variable order. In image mode, the definition of the moving window was already presented and its application was illustrated in the Basic Module of Spatial Analysis (B-AN), in particular in Lesson 2. We point out simply here the parameters of definition of the neighborhood in image mode and the types of operators allowing to produce the indices of arrangement.

Parameters of definition of the neighborhood:

- *size of the window:* The moving window is admitted as square by definition. Its size is expressed in a number of columns and lines (CxL) which is generally odd so that the processed cell in each successive step is centered in the window.
- *the shape of the window:* Only certain cells in the moving window can be assigned as belonging to the neighborhood of the central cell. The selective activation of the cells makes it possible to define non square shape of neighborhood, dissymmetrical or directional.

- *proximity to the central cell:* The index to be produced can consider the whole of the neighborhood indifferently, or can take into account the order of the neighborhood or the distance to the central cell. In this second alternative, **weights** will be assigned to each cell of the neighborhood according to their distance to the center of the window. That will produce weighted indices of arrangement.

Parameters of definition of the neighborhood in a moving window

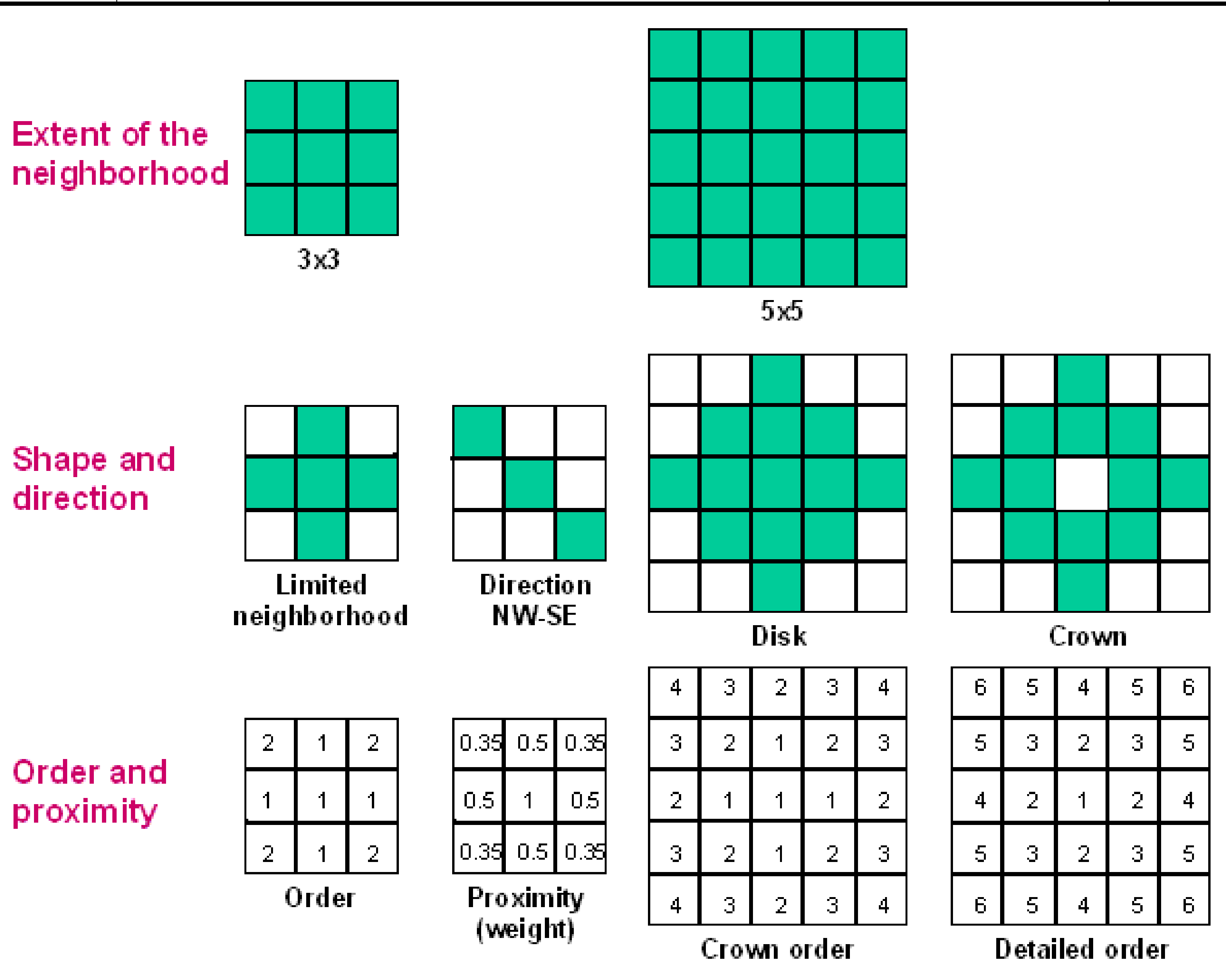


Figure 3.5

- *indices of central tendency:* They account for the general property (central) in the whole of the neighborhood considered. They result from statistical operators of central tendency (position) specific to each of the three levels of measurement of the thematic contents.

- *indices of variability:* They can be similar to the thematic indices of structure presented previously, but calculated on the defined neighborhood. They express the importance of the variation of properties in this neighborhood. One finds there also other indices specific to this scale of analysis.
- *indices of texture of first order:* They express in a global way the variability or the dispersion of the properties present within the neighborhood. They call upon statistical operators of dispersion adapted to the level of measurement of the thematic variable.
- *indices of texture of second order:* They return account in a finer way of the transitions from properties between all the pairs of cells ordered according to their proximity and optionally according to directions. Such indices are made up from matrices of co-occurrence. They describe the texture of properties principally measured on a cardinal level, therefore in image mode they are mainly associated with continuous spatial distributions. So they will not be described in this Unit, but the reader can familiarize himself with their principle of calculation and their potential of application using the following references: Caloz and Collet (2001, pp.262-269), Haralick (1979, pp.786-804).

Types of contextual indices and their relationship to the indices of structure

Type	**Contextual Indices**	*Level of measurement*	Equivalent Structural Indices
centrality	Mode (MOD) Médiane (MED) Moyenne (AVG)	*nominal* *ordinal* *cardinal*	Mode (MOD) Median (MED) Average (AVG)
variability	Heterogeneity (NDC) Center-neighborhood (CVN) Central class frequency (CCF) Relative richness (R) Fragmentation (F) Diversity (H) Predominance (D) Different pairs (BCM)	*nominal* *nominal* *nominal* *nominal* *nominal* *nominal* *nominal* *nominal*	Richness (PR) Shannon's index (SHDI) Regularity (SHEI)
1st order texture	Diversity (DIV) *nominal* Interquartile (IQ) *ordinal* Standard dev. (SD) *cardinal* Range (RNG) *cardinal*	*nominal* *ordinal* *cardinal* *cardinal*	Richness (PR)
2nd order texture	Contraste (CONT) Entropy (ENT) 2nd angular moment (2AM)	*cardinal* *cardinal* *cardinal*	

Table 3.5

1.3.8. Indices of central tendency

The prevalent property inside the neighborhood defined by the moving window is expressed in a different way according to the level of measurement. This measure of central tendency is described by the statistical indicators of mode, median and average. We will retain here only those relating to the nominal and ordinal levels of measurement:

Index of majority, mode (MOD), nominal level:
$$MOD = v_{fmax}$$
v_{fmax}: the most frequent value in the neighborhood
Interpretation: It is the value (category) most present in the neighborhood. If several modal values are present in the window, it will then be necessary to define a rule of priority among these values.

Median index (MED), ordinal level:
$$MED = v_{med}$$
v_{med}:value of the central rank for the whole of the units of observation in the area
Interpretation: It is the value of cell (UO) positioned in the middle of the ordered sequence of the values in the neighborhood.

1.3.9. Indices of variability

The variability of the values inside the neighborhood defined by the moving window can be expressed by a variety of indices. The following indices will be retained:

Index of heterogeneity (NDC), nominal level:
$$NDC = c$$
c: number of different values, categories or classes present in the neighborhood
Interpretation: It is a simple indication of diversityx expressing the thematic richness in the neighborhood. It is equivalent to the index of richness (PR) applied to the objects in the study area.

Index center-neighborhood (CVN), nominal level:
$$CVN = n_d$$
nd: number of values different from the central cell value
Interpretation: This index accounts for the insulation of the central property compared to its neighborhood. It expresses the degree of heterogeneity between the central cell and its neighborhood. Its value is relative to the number of potential properties in the image.

Index of frequency of the central class (CCF), nominal level:
$$CCF = f_c$$
fc: frequency of the central cell value present in the neighborhood
Interpretation: This index expresses the importance of the central value in the neighborhood. Its value varies between 1 (insulation) and n_p, the number of cells of the neighborhood. This maximum value characterizes a homogeneous neighborhood.

Index of relative richness (R), nominal level:
$$R = (c / c_{max}) * 100$$
c: number of different values present in the neighborhood c_{max}: number of different values present in the image
Interpretation: It is an indication of relative diversity compared to the whole of the values of the image. It is expressed in % and varies from 100 for a maximum diversity with a value close to 0 ($100/c_{max}$) for a homogeneous neighborhood.

Index of fragmentation (F), nominal level:
$$F = (c-1) / (n_p-1)$$
c: number of different values present in the neighborhood n_p: number of cells considered in the neighborhood (window)
Interpretation: It is an indication of diversity for the whole of the neighborhood. It varies from 0 for a homogeneous neighborhood with 1 for a completely heterogeneous neighborhood.

Index of diversity (H), nominal level:
$$H = \Sigma(P_i * \ln P_i)$$
#: sum of all the values present in the image P_i: proportion of each value i in the neighborhood ln: natural logarithm
Interpretation: The index H equals 0 when the neighborhood is composed of only one value (thematic homogeneity). Its value increases according to the number of properties (richness, diversity) as well as to the tendency to the uniformity of the surfaces of each one of them.

It corresponds to the index of diversity of Shannon (SHDI) applied to the neighborhood.

Index of predominance (D), nominal level:
$D = (\Sigma(P_i * \ln P_i)) / \ln c$ $D = H / H_{max}$
#: sum of all the values present in the image P_i: proportion of each value i in the neighborhood ln: natural logarithm c: number of different values present in the neighborhood H: Index of diversity H_{max}: maximum diversity
Interpretation: The index D varies between 0 and 1. Its value is equal to 0 when the neighborhood is composed of a single value (thematic homogeneity). Its value approaches 0 when the frequencies of each value are very different (irregularity, predominance of value). The index is equal to 1 when the frequencies are equal (regularity). It corresponds to the index of regularity of Shannon (SHEI) applied to the neighborhood. It is in fact a standardization of the index H.

Index of different pairs (BCM), nominal level:
$BCM = n_{Pd}$
n_{Pd}: a number of pairs of different values in the neighborhood considered
Interpretation: It is an indication of diversity for the whole of the neighborhood. It varies from 0 for a homogeneous neighborhood with a maximum value according to the importance of the neighborhood. The pairs of cells whose value is not identical are counted. All the pairs of cells of the neighborhood, even those noncontiguous, are considered.

Indices of variability applied to different local arrangements. Value of the central cell of the window for the various indices. The diversity of values present in the whole of the image (c_{max}) is set to 9.

Indices	1 1 1 1 **1** 1 1 1 1	1 2 3 4 **5** 6 7 8 9	1 1 3 1 **1** 3 2 2 3	1 1 1 2 **2** 2 3 3 3	2 1 2 1 **1** 1 2 1 2	2 3 4 1 **1** 5 6 7 8
	Homo-geneous	*Hetero-geneous*	*3 zones*	*3 equal zones*	*2 zones*	*8 zones*
NDC	1	9	3	3	2	8
CVN	0	8	5	6	4	7
CCF	9	1	4	3	5	2
R	11.11	100	33.33	33.33	22.22	88.88
F	0	1	0.25	0.25	0.125	0.875
H	0	2.197	1.061	1.098	0.687	2.043
D	0	1	0.97	1	0.99	0.98
BCM	0	36	26	27	20	35

Table 3.6

In conclusion, one notes that among these indices of variability each one characterizes a particular aspect of the arrangement of the properties in the neighborhood or then express this in a relative or absolute way. These indices are particularly adapted to a nominal level of measurement because they consider neither the hierarchy of the values nor their intervals.

Exercies on Table 3.6

1. In the Table 3.6 which index gives best account of the differences in arrangement for the 6 cases presented?
   - order this choice according to performances of the indices.
2. What is the influence of the values assigned to properties on the value of each produced index?
   - what can one conclude from it?

1.3.10. Indices of texture of first order

1.3.10a Indices of texture of first order

These indices of texture are based on the dispersion of the properties in the neighborhood, they are thus dependent on the level of measurement of information. The following indices will be retained:

Diversity (DIV), nominal level:
$$DIV = c$$
c: number of different values present in the neighborhood
Interpretation: It is a simple indication of diversity expressing the thematic richness in the neighborhood. It equivalent to the index of heterogeneity (NDC) previously described.

Interquartile (IQ), ordinal level:
$$IQ = Q_3 - Q_1$$
Q1: first quartile Q3: third quartile
Interpretation: It expresses the number of different values ranging between the first and the third quartile (the difference of rank), therefore the dispersion around the median value.
It is noted that only at the ordinal level the index of dispersion IQ comes to supplement the pallet of the contextual indices previously presented.

1.3.10b Exploitation of the contextual indices

In order to illustrate the potential and the complementarity, but also the redundancy of contextual indices presented in this Unit, let us apply to the description of the spatial arrangement of land cover types in an urban area. Figure 3.6 presents the spatial distribution of 5 categories of land cover (Forest/bush, Meadow, Building, Road, Pavement) in an urban area made up of 4 distinct zones: individual housing, collective habitat, industrial park and green area. The size of the cells is of approximately 15m on side. These thematic properties being categories, they correspond to a nominal level of measurement.

Urban area made up of 5 different zones

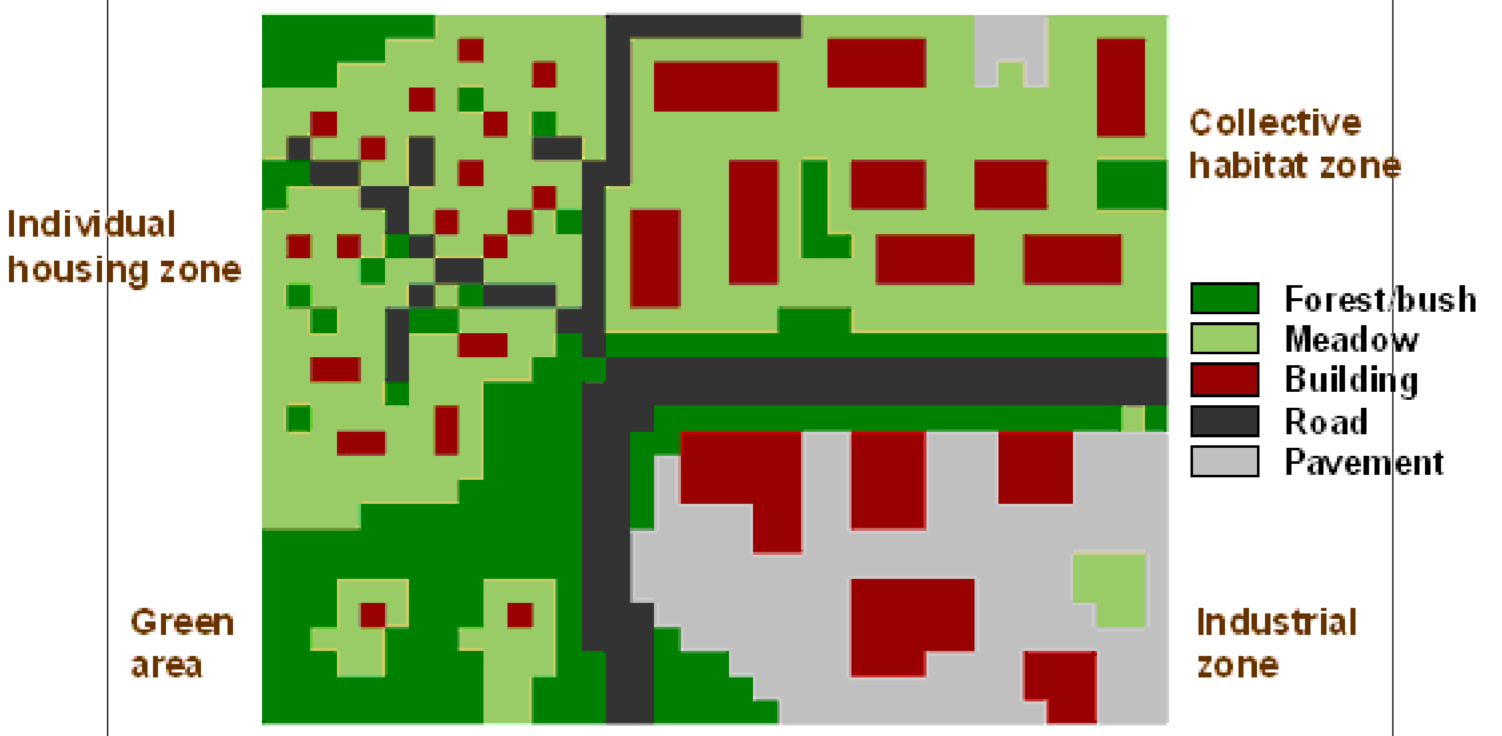
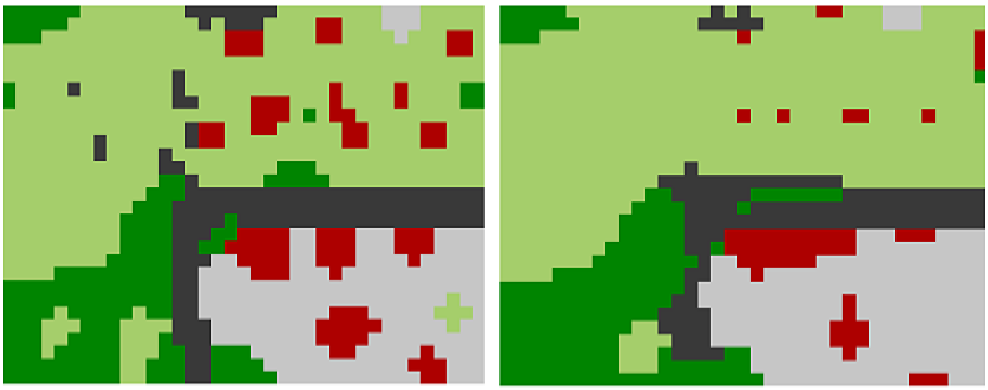


Figure 3.6

The description of the spatial arrangement of thematic properties of these units of observation (cells) can begin with the identification of the most represented land cover category in the neighborhood. We will thus choose to retain the modal value in the moving window. In order to experiment the effect of the neighborhood size on this *index of majority* (central tendency), we will choose a window size of 3x3 as well as 5x5.

Modal property of the neighborhood in windows of size 3x3 and 5x5



a) Neighborhood of 3x3 b) Neighborhood of 5x5

Figure 3.7

One can observe on figure 3.7 the prevalence of the meadow category for the 2 zones of habitat, of the forest/ bush type for the green area and of the pavement category for the industrial park. When a window of larger size (5x5) is applied, then the spatial smoothing is more marked, thus reducing the presence of vaster regions (objects) such as the buildings of big size in the industrial park and in the collective habitat.

Indices of texture for windows of size 3x3 and 5x5

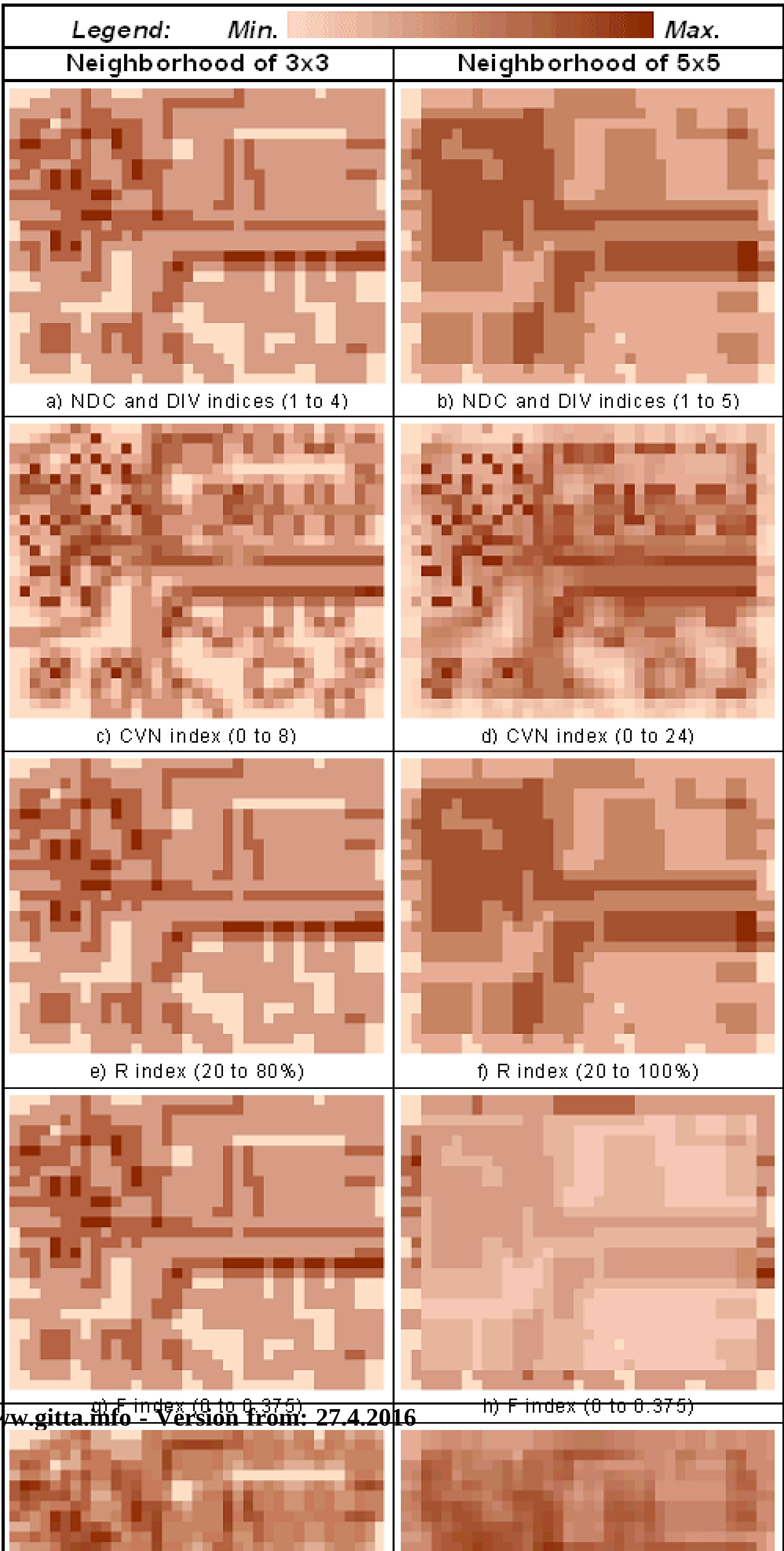


Figure 3.8

Exercises

-
    -
- Which is the influence of the values assigned with the properties on the value of each produced index?
    - what can one conclude from it?

1.4. Bibliography

- **Caloz R., Collet C.**, 2002. *In: Traitements numériques d'images de télédétection*. Précis de télédétection Volume 3, AUF: Presses de l'Université du Québec (PUQ), Sainte-Foy, Canada, 258-266.
- **Eastman J.R.**, 2001. *Idrisi 32, Release 2, Guide to GIS and Image Processing*. Volume 2. Clark Labs, Clark University, Worcester, USA.
- **Ebdon D.**, 1977. *In:* **B. Blackwell**, ed. *Statistics in Geography, A practical approach*. Oxford, the U.K., 128-141.
- **Griffith D.**, 1987. *Spatial autocorrelation, A primer*. AAG, Washington DC: Resources Publications in Geography.
- **Haralick R.M**, 1979. Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, 67/5, 786-266.
- **Jayet H.**, 1993. Analyse spatiale quantitative, une introduction. *Economica, Paris*, 53-108.
- **McGarigal K., Marks B.**, 1994. *FRAGSTATS : Spatial Pattern Analysis Program for Quantifying Landscape Structure, version 2.0*. Oregon State University, USA.
- **Monmonier M.S.**, 1974. Measures of Pattern Complexity for Choropleth Maps. *The American Cartographer*, 1/2, 159-169.
- **Murphy D.L.**, 1974. Estimating Neighborhood Variability with a Binary Comparison Matrix. *Photogrammetric Engineering and Remote Sensing*, 51/6, 667-674.
- **Turner M.G.**, 1989. Landscape Ecology: The Effect of Pattern on Process. *Annu. Rev. Ecol. Syst.*, 20, 171-197.