

## Characterisation of the spatial variability of waterlogging in the Blue River Basin (Argentina)

Y. Villacampa<sup>1</sup>, P. Sastre<sup>2</sup>, M. Gandini<sup>2</sup>, A. Cañibano<sup>2</sup>  
& M. Cortés<sup>1</sup>

<sup>1</sup>*Departamento de Matemática Aplicada,  
Universidad de Alicante, España*

<sup>2</sup>*Facultad de Agronomía, Universidad Nacional del Centro de la  
Provincia de Buenos Aires, Argentina*

### Abstract

Considering the soil as a complex, dynamic and heterogeneous system, chaos theory is a useful tool for gaining new knowledge into its variability. Fractal geometry has a close relationship with chaos theory, which describes dynamic and chaotic events with a development in time and space that can be predicted probabilistically. Furthermore, the concept of fractal dimension can also be applied to the study of spatial and temporal phenomena that possess and retain some kind of irregular geometric pattern within a range of scales. It is increasingly used to characterise the various properties of soil and to describe and model the characteristic functions of water retention and mobility in soil. This study is based on the need to estimate waterlogging of agro-livestock farms in the Blue River Basin (6,237 km<sup>2</sup>), with a view to recommending the type of land use for each plot. By analysing Landsat satellite images, those with no cloud cover were adjusted spatially in order to build mosaics for eight points in time covering the entire study area. The images were then rated and processed, resulting in eight binary masks of non-flooded / waterlogged areas.

This paper presents a method for calculating the fractal dimension based on the box counting process, in order to test its use in differentiating between varying degrees and distributions of soil moisture. The results indicate eight situations of soil moisture by means of the fractal dimensions calculated using this method. Using the parallelism test, the slopes of the simple adjusted linear



regression models were compared statistically, and significant differences were observed.

*Keywords: fractal, soil moisture, dynamic system.*

## 1 Introduction

According to Sommer et al. [11], soil is a complex, dynamic and heterogeneous system, defined as a natural, four-dimensional body (length, width, depth and time) variable in space and time. Being a complex system, it has two generic properties: invariance to scaling or self similarity and universality. In addition, in order to study complex systems it is necessary to analyse their main attributes, these being: entropy, lacunarity, degree of occupation of the space of interest, roughness and multifractality (Bullock and Brennen [1], Cheng [2], Logsdon et al. [6]).

The variability of the structural pattern of soil in space and time depends on the environmental conditions. In addition, said variability is a consequence of the natural heterogeneity and genesis of the materials of geological and biological origin, as well as the temporal variations of the processes that act on them. Of particular importance is erosion, climate changes, soil degradation, aquifer recharge, evapotranspiration, pollution, etc. (Zhang [14]).

Characterisation of spatial variability is useful with any attempt to model processes. Indeed, the study of heterogeneity on different scales was the prime incentive behind the introduction, in soil science, of fractal geometry, which at least in principle allows us to analyse its geometric complexity and characterisation with just a few numbers, the fractal indexes.

An object is considered to be fractal when its measurement dimension,  $D_b$ , is **greater than its topological dimension,  $D_t$  and less than its immersion dimension,  $D_i$** . Therefore,  $D_b$  contains the idea of the occupation of space. So, as  $D_b$  increases, the degree of space filling, disorder or irregularity also increases, as something continuous, able to assume non-whole values. All of this differs from the common classic idea of geometric or Euclidean dimension. To sum up, by determining the fractal dimension of an object, we are obtaining information about its capacity for occupation of space. (Pachsky et al. [7]).

Analysis of the spatial structure of soil humidity may a priori provide information needed for other studies, such as: what the sample size needs to be, how far apart samples need to be taken, if the humidity is invariable with regard to changes in scale, or how spatial interpolation should be carried out. In this study, the aim was to estimate the probability of flooding of agricultural-stock raising farms in the Arroyo del Azul waterway (6,237 km<sup>2</sup>), in order to recommend the appropriate land use for each plot.

## 2 Fractal dimension and box-counting dimension

The fractal dimension concept is used to study continuous temporal or spatial phenomena that show some kind of geometric pattern within a range of scales



and which, in this case, is appropriate for describing the spatial distribution of the humidity retention function of the soil.

One of the interpretations of dimension, possibly the most natural, is associated with the capacity of objects to occupy the Euclidean space in which they are immersed. Thus, quantifying fractals means defining the proportion of the physical space in which they are inscribed, that is filled by them. The fractal dimension is an indicator of how the fractal object occupies the space in which it is immersed, or what kind of variation predominates when dealing with long-term time series (Gaytán-Ortiz [4]). A fractal dimension of a time series with a value of less than 1.6 indicates the predominance of long-term variations, while values of 1.5 to 2 indicate the predominance of short-term variations and that the phenomenon is unpredictable.

One of the different versions of fractal dimension numerical calculation is called box-counting, which is related to the concept of auto-similar structure. The methodology consists of placing a grid with a mesh of sizes,  $\varepsilon$ , over the structure to be analysed, forming a covering of boxes. The number of boxes that contain values of the structure to be characterised is  $N(\varepsilon)$ , which varies with the value of  $\varepsilon$  chosen.

The box-counting dimension concept, which allows an algorithm to be deduced for numerical estimation, was developed by Kolmogorov [5], but has a predecessor in that described by Pontrjagin and Schnirelman [9]. Let  $N(\varepsilon)$  be the least number of radius sets less than or equal to  $\varepsilon$  needed to cover the compact set  $F \subset R^n$ . The lower and upper box-counting dimensions of set  $F$  are defined as eqn (1) and eqn (2), respectively:

$$\underline{\dim}_D(F) = -\lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log \varepsilon} \quad (1)$$

$$\overline{\dim}_D(F) = -\overline{\lim}_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log \varepsilon} \quad (2)$$

If the two amounts thus defined are the same, then we have the box-counting dimension of set  $F$ , eqn (3):

$$\dim_D(F) = -\lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log \varepsilon} \quad (3)$$

Let us look at a simpler geometrical interpretation of the box-counting dimension. Let  $N(\varepsilon)$  be the least number of balls of radius  $\varepsilon$  needed to cover compact set  $F \subset R^n$ . Then, the  $D$ -measure of  $F$  satisfies, approximately,  $B^D(F) \propto N(\varepsilon)\varepsilon^D$  then for certain  $c > 0$

$$N(\varepsilon) \approx \frac{c}{\varepsilon^D} \quad (4)$$

$$\begin{aligned} \log(N(\varepsilon)) &= \log c - D \log \varepsilon \\ \log(N(\varepsilon)) - D \log \frac{1}{\varepsilon} &= \log c \end{aligned} \quad (5)$$

$$\frac{\log(N(\varepsilon))}{\log \frac{1}{\varepsilon}} - D = \frac{\log c}{\log \frac{1}{\varepsilon}}$$



When  $\varepsilon \rightarrow 0^+$  the term on the right decreases indefinitely, which suggests the following definition for box-counting dimension, eqn (6):

$$D_B(F) = \lim_{\varepsilon \rightarrow 0} \frac{\log(N(\varepsilon))}{\log \frac{1}{\varepsilon}} \quad (6)$$

### 3 Algorithm for estimating the box-counting dimension

In the process of determining the box-counting dimension, we replace the evolution of the variable  $\varepsilon$  with a succession  $\{\varepsilon(k)\}$ . Let  $\{\varepsilon(k)\}$  be a succession of positive numbers that satisfy the conditions eqn (7):

$$\begin{aligned} \lim_{k \rightarrow \infty} \varepsilon(k) &= 0 \\ \lim_{k \rightarrow \infty} \frac{\log \varepsilon(k)}{\log \varepsilon(k-1)} &= 1 \end{aligned} \quad (7)$$

Thus,

$$D_B(F) = \lim_{k \rightarrow 0} \frac{\log N_{\varepsilon(k)}}{\log \left( \frac{1}{\varepsilon(k)} \right)} \quad (8)$$

With which we manage to make the infinite limit process discrete, eqn (8).

The balls can be replaced by another type of enclosure that better adapts to the construction of an algorithm. So, let  $n = 2$  and diadic boxes, a succession of boxes with sides  $2^{-k}$  configured as a succession of grids.

Let  $N'_k$  be the minimum number of boxes (square) with sides  $1/k$  needed to cover  $F$ . We define the box-counting dimension as the limit (if it exists):

$$D_B(F) = \lim_{k \rightarrow \infty} \frac{\log N'_k}{\log 2^k} \quad (9)$$

However, for physical world fractals, the infinite limit process, eqn (9), makes no sense.

To calculate an estimated  $D_B$ , we overlay the structure to be analysed with a grid of sizes,  $\varepsilon$ , creating a layer of boxes. The number of boxes that contain values of the structure to be characterised is  $N_{\varepsilon(k)}$ , which varies with the value of  $\varepsilon$  chosen. To make the corresponding calculation of the fractal dimension, we count the number of boxes  $N_{\varepsilon(k)}$  on different scales. The variation in scale from one step to the next depends on the size of the image and the multiples of that size. The value of the dimension corresponds to the slope of  $\log(N_{\varepsilon(k)})$  Vs.  $\log(1/\varepsilon)$ .

The important advantage of this methodology is that it can be applied systematically and simply when creating the layer of boxes. In addition, the corresponding recount can be carried out quickly, even with high-resolution images (Vicsek [12]).

## 4 Processing digital information and methodology

We started with Landsat 5 TM and 7 ETM+ image analysis (scenes 225-85 and 225-86), and selected those of a high quality (zero cloud cover). The images were geometrically corrected to the Gauss Krüger projection system (Datum Campo Inchauspe 1969, Argentina Zone 5) to start constructing mosaics in order to cover the whole study area. Mixed classifications were then made of the satellite images, processing them using masking and obtaining eight binary images of non-flooded/flooded areas.

The importance of the study lies in the need to estimate the probability of the flooding of farmland, in order to be able to recommend the land use of each plot. The area studied is the Arroyo del Azul basin, covering 6,237 km<sup>2</sup>. The Arroyo del Azul waterway has its source near the town of Chillar and stretches northwards, until it is lost before reaching the River Salado, known as the Arroyo Gualicho waterway in its lower reaches. The course is finally crossed and largely drained by Canal 11, which was adopted as the northern limit of the study area. The basin consists of a large plain (lower sub-basin) and a smaller area of low hills on the south (upper sub-basin), where the headwaters of the Arroyo del Azul are found. The upper area is connected to the plain by the foothills (middle sub-basin). The average slope of the land is 5% in the upper sub-basin and 0.2% in the lower sub-basin, while varying from 0.5 to 0.8% in the middle area. The overall behaviour of the region is that of a plains environment (Sala et al. [10]).

For each situation, we calculated the corresponding fractal dimension, using the box-counting method. We used the “parallelism” test (Zar 1984), to statistically compare the slopes of the eight adjusted simple linear regression models (one for each situation). The hypotheses of these tests are:  $H_0: \beta_1 = \beta_2 = \dots = \beta_k$  and  $H_1: \beta_i \neq \beta_j$ , using a conventional  $F$ -test with a specific  $F$  calculated ( $F_c$ ) where  $\beta_i$  represents the slope of the adjusted regression model (fractal dimension), with  $i = 1, 2, \dots, k$ . The statistic used is eqn (10):

$$F_c = \frac{(SC_c - SC_p)/(k-1)}{(SC_c/GL_p)}, \quad (10)$$

where:

- $SC_c$  = Common residual sum of squares;
- $SC_p$  = Grouped residual sum of squares (sum of the sums of squares for each regression);
- $GL_p$  = Grouped degrees of freedom;
- $k$  = Number of regressions.

To establish the differential effect of the degrees of humidity studied, represented by the fractal dimensions, we compared pairs of slope values by calculating a critical value,  $q = (b_i - b_j) / SE$ , where  $b_i$  and  $b_j$  are the slopes associated with the regression models of the degrees of soil humidity  $i$  and  $j$ , respectively, with eqn (11):

$$SE = \sqrt{\frac{\frac{SC_c + SC_j}{GL_i + GL_j}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}} \tag{11}$$

$q_{\alpha, \nu, p}$

with:  
 $\alpha$  = level of significance  
 $\nu$  = grouped degrees of residual freedom  
 $p$  = number of simple linear regression models (degrees of humidity).

5 Results and discussion

The table shows the estimates found for the adjusted regression lines  $\log(N(\varepsilon)) = \log c - D \log \varepsilon$ , the Adjusted  $R$  Square and their errors for the 8 cases or periods. In all cases, the significance tests (ANOVA) for the regression adjustments, showed highly significant values. The t-student tests, to check whether the coefficients were non-null, showed highly significant values, which means that the estimated coefficient values were different to zero in all the regressions.

Table 1: Regression adjustments.

Year	Adjusted Regression	Adjusted R Square	Std. Error of the Estimate	State H. A	Fractal Dimension
April 2001	$\log N(\varepsilon) = 10,124 - 2,046 \log \varepsilon$	,955	,354	Normal	2,046
April 2002	$\log N(\varepsilon) = 9,588 - 2,070 \log \varepsilon$	,997	,090	Normal	2,070
August 1993	$\log N(\varepsilon) = 9,241 - 2,026 \log \varepsilon$	,999	,056	Normal	2,026
March 1998	$\log N(\varepsilon) = 9,042 - 2,028 \log \varepsilon$	1,000	,032	Normal	2,028
May 1986	$\log N(\varepsilon) = 8,858 - 2,056 \log \varepsilon$	,995	,112	Normal	2,056
November 1997	$\log N(\varepsilon) = 8,542 - 2,182 \log \varepsilon$	,991	,171	El Niño	2,182
September 2001	$\log N(\varepsilon) = 7,349 - 1,890 \log \varepsilon$	,999	,044	Dry	1,890
November 2001	$\log N(\varepsilon) = 6,186 - 1,745 \log \varepsilon$	,977	,214	Dry	1,745

The “parallelism” test (Zar [13]) was applied to statistically compare the slopes of the eight adjusted simple linear regression models (one for each situation). The hypotheses of this test are:  $H_0: \beta_1 = \beta_2 = \dots = \beta_k$  and  $H_1: \beta_i \neq \beta_j$ , and the values calculated are:  $SC_c = C_1 - \frac{B_1^2}{A_1} = 167,48 - \frac{82,41^2}{41,10} = 2,21$ ,



$$SC_p = \sum_{i=1}^{K=8} SS_i = 1,59, , \quad \text{with} \quad GL_p = \sum_{i=1}^{k=8} (n_i - 2) = 7 * 8 = 56, ,$$

$$F_c = \frac{\frac{SC_c - SC_p}{\frac{SC_c}{GL_p}}}{\frac{k-1}{\frac{SC_c}{GL_p}}} = \frac{\frac{2,21 - 1,59}{2,21}}{\frac{8-1}{56}} = 2,244.$$

**Giving the result:** Probability  $F_c = 0.977763789$  and Probability  $F_{(2,244,7,56)} = 0.04383$ , so the slopes are not equal. This means that the fractal dimensions estimated for each period significantly differ from each other.

In order to establish the differential effect of the degrees of humidity studied, represented by the fractal dimensions, we compared pairs of slope values by calculating a critical value,  $= (b_i - b_j)/SE$ , where  $b_i$  and  $b_j$  are the slopes associated with the regression models of the degrees of soil humidity  $i$  and  $j$ , respectively, within eqn (11).

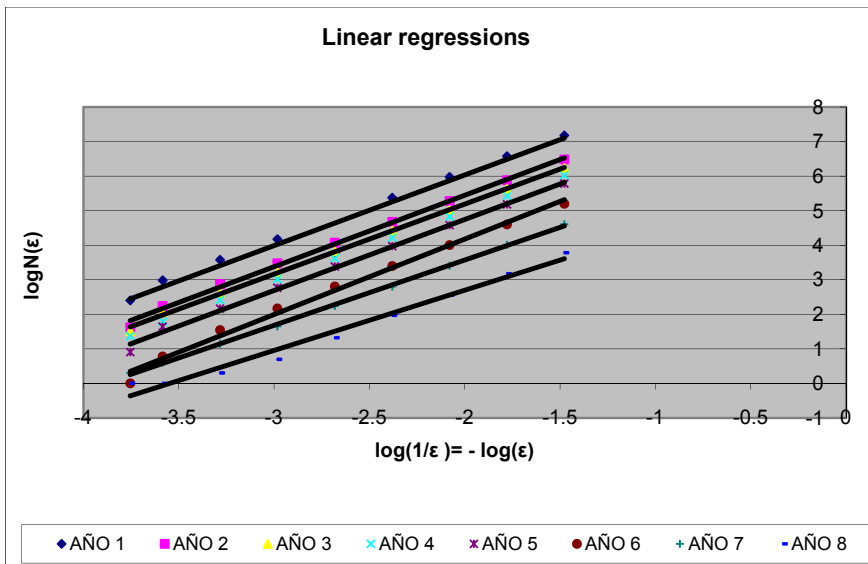


Figure 1: Slope values.

It can be seen that there are 3 groups of slopes that coincide with humidity statuses in the basin being studied and climatic phenomena such as “*El Niño*”. There is a group of intermediate values that corresponds to “normal” water status states of the basin and a high value (2.182) that is related to the “*El Niño*” event that affected the basin, flooding a large area. In addition, the slope values for November and September 2001 are related to a low humidity water status, in which shallow bodies of water dried out completely.

## 6 Conclusion

The fractal dimensions of a plain basin, calculated and analysed using satellite images, vary according to the different humidity statuses of the soil. This would indicate that the fractal dimension estimated in this way could be used to measure the probability of events leading to cases of flooding. This would be of great use for farmers, as they could then recommend the type of land use for each plot.

It is necessary to further this study, increasing the type and amount of situations considered.

## Acknowledgement

This paper has been supported by the project CTM2010-10878-E of Ministerio de Ciencia e Innovación, Spain. The authors wish to express their gratitude for such a grant.

## References

- [1] Bullock, S.S. & Brennen, G.K., Qudit surface codes and gauge theory with finite cyclic groups. *J. Phys. A: Math theor.* **40**, pp. 3481-3505, 2007.
- [2] Cheng, S.S. 2008a. Spatial scaling analyses of soil physical properties: a review of spectral and wavelet methods. *Vadose zone Journal*. Soil science Society of America. Madison. USA.
- [3] Cheng, Q. 2008b. Modeling local scaling properties for multiscale mapping, state key laboratory of Geological Processes and mineral resources. China Univ. of Geosciences.
- [4] Gaytán-Ortiz, L.A. 2007. Auto afinidad y ruido  $f-\beta$  de temperaturas extremas mensuales. Tesis de Maestría en Matemática Educativa. Universidad Autónoma de Zacatecas.
- [5] Kolmogorov, A. 1959. Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk. SSSR*, 124:754.
- [6] Logsdon S.D., E. Perfect and A. M. Tarquis. 2008. Multiscale soil investigations: physical concepts and mathematical techniques. Published in *Vadose zone Journals*. Madison, USA.
- [7] Pachesky Y., E. Perfect and M.A. Martín. 2006. Fractal Geometry Applied to Soil and Related Hierarchical Systems. *Geoderma*. 134 : 237-239.
- [8] Avances en Recursos Hidráulicos Número 15, Mayo De 2007. Medellín
- [9] Pontrjagin and Schnirelmal (1932).
- [10] Sala, J.M., Kruse, E, Aguglino L. 1987 Investigación hidrológica de La cuenca Del Arroyo AZUL. Informe 37. CIC. La Plata Argentina.
- [11] Sommer M., Gerke H. H., D. Deumlich. 2008. Modelling soil landscape genesis: a time split approach for hummocky agricultural landscapes.
- [12] Vicsek, T. 1992 Fractal growth phenomena Publisher: World Scientific (Singapore and New Jersey).





- [13] Zar, J.H., 1984. Bioestatistical analysis. Department of Biological Sciences, Northern Illinois University. EEUU. 842 P.
- [14] Zhang X. 2006. Consolidation theories for saturated-unsaturated soils and numerical simulation of residential buildings on expansive soils. Texas A&M University. <http://handle.tamu.edu/1969.1/2747>

