

Multiscaling of porous soils as heterogeneous complex networks

A. Santiago¹, J. P. Cárdenas¹, J. C. Losada², R. M. Benito¹, A. M. Tarquis³, and F. Borondo⁴

¹Grupo de Sistemas Complejos, Departamento de Física, E.T.S.I. Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid, Spain

²Grupo de Sistemas Complejos, Departamento de Tecnologías Aplicadas a la Edificación, E.U. Arquitectura Técnica, Universidad Politécnica de Madrid, 28040 Madrid, Spain

³Departamento de Matemática Aplicada, E.T.S.I. Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid, Spain

⁴Departamento de Química, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

Received: 7 August 2008 – Revised: 24 September 2008 – Accepted: 25 September 2008 – Published: 26 November 2008

Abstract. In this paper we present a complex network model based on a heterogeneous preferential attachment scheme to quantify the structure of porous soils. Under this perspective pores are represented by nodes and the space for the flow of fluids between them is represented by links. Pore properties such as position and size are described by fixed states in a metric space, while an affinity function is introduced to bias the attachment probabilities of links according to these properties. We perform an analytical study of the degree distributions in the soil model and show that under reasonable conditions all the model variants yield a multiscaling behavior in the connectivity degrees, leaving a empirically testable signature of heterogeneity in the topology of pore networks. We also show that the power-law scaling in the degree distribution is a robust trait of the soil model and analyze the influence of the parameters on the scaling exponents. We perform a numerical analysis of the soil model for a combination of parameters corresponding to empirical samples with different properties, and show that the simulation results exhibit a good agreement with the analytical predictions.

1 Introduction

The porous structure of soils has an important influence on the physical, chemical and biological processes that take place within them (Young and Crawford, 2004; Blair et al., 2007). The pores size distribution (PSD) is one important parameter to modeling these processes (Vogel and Roth, 2001;

Lin et al., 1999). However, the difficulty is in understanding the PSD at the same time that the pore geometry distribution and their behavior concerning the dynamics fluid transportation. Most of the theoretical approaches to the soil porosity can not explain them adequately because the available models are idealizations that fail to rescue the complex structure of the wiring and spatial location of pores (Bird et al., 2006) and moreover do not consider the mechanisms underlying the dynamics that leads to the emergence of such structure (Horgan and Ball, 1994).

During the last years the perspective of complex networks has successfully been applied to a wide array of scientific fields (Strogatz, 2001). Network theory describes complex systems from a purely topological point of view, abstracting away the dynamical processes that take such structure as substrate (Albert and Barabási, 2002; Newman, 2003). A *complex network* can thus be viewed as set of nodes and links with a non-trivial topology. The study of complex networks has shown the presence of common and hardly intuitive structural properties in both nature and man-made systems. These findings have created a large follow-up debate among physicists because of their ubiquity and peculiar statistical properties, markedly different from the random graphs (Erdős and Rényi, 1959). In particular, complex networks adjust some statistical properties to power laws, a characteristic feature of critical-point behavior and a fingerprint of self-organized systems (Bak et al., 1987).

The impact of this new approach has also affected the geosciences. Recent approximations describe the structure of soils as a network (Vogel and Roth, 2001). Using a network approach, Valentini et al. (2007) found that rock fractures exhibit the well known *small world* phenomenon (Watts and Strogatz, 1998) when those fractures are considered as nodes



Correspondence to: R. M. Benito
(rosamaria.benito@upm.es)

in a network (Valentini et al., 2007). A network approach has also been applied to the study of climate dynamics by Tsonis et al. (Tsonis et al., 2006, 2007, 2008; Tsonis and Swanson, 2008; Yamasaki et al., 2008). In this paper we propose a similar approach to quantify the pore architecture of a soil through a complex network model.

Network models are prescriptive systems that generate networks with certain topological traits, such as degree or clustering distributions. Dynamical network models (Dorogovtsev and Mendes, 2002) are stochastic discrete-time dynamical systems that evolve networks by the iterated addition and subtraction of nodes and links. In our study we interpret porous soils as heterogeneous networks where pores are represented by nodes with distinct properties (such as surface and spatial location) and the flow of fluids between them is represented by links connecting the nodes. The networks are generated by a dynamical model known as heterogeneous preferential attachment (PA) (Santiago and Benito, 2007a,b, 2008a,b), a generalization of the Barabási-Albert (BA) model (Barabási and Albert, 1999a) to heterogeneous networks.

The BA model is based on the mechanisms of growth and preferential attachment (Price, 1965, 1976) and provides a minimal account of the process leading to the emergence of *scale-free* networks (Barabási et al., 1999b). Such networks are characterized by a degree distribution with an asymptotic behavior according to a *power-law*, $P(k) \sim k^{-\gamma}$, where γ is known as the scaling exponent. This phenomenon leads to a non-negligible presence of highly connected nodes or *hubs*, a trait commonly referred as fat tails. In the BA model the process starts with a seed of arbitrary size and topology. A new node is added to the network at each step, bringing a fixed number m of links attached. These links are preferentially connected to the already existing nodes following the so-called *attachment rule*: the linking probability of a network node v_i is proportional to its degree k_i , $\Pi(v_i) = k_i / \sum_j k_j$. This step is iterated until a desired number N of nodes have been added to the network. Heterogeneous PA models incorporate the influence of node attributes in addition to the connectivity degree to the attachment rule.

The organization of the paper is as follows. In the next section we describe the formulation of the porous soil model. The analytical solution for the degree distribution of the pores generated by the model is presented in the third section. In the fourth section we present the results of the porous soil corresponding to two samples of soils of different porosity. In the last section we present our conclusions derived from this work.

2 Model formulation

The structure of a porous soil is modeled as a heterogeneous complex network where nodes v_i correspond to pores and links e_{ij} correspond to flow of fluids between them. The links

will be considered undirected $e_{ij} = e_{ji}$ and thus the connectivity degree k_i of a node v_i will be a measure of the number of pores that are directly connected with the associated pore. The properties of a pore are described by the node state (r_i, s_i) , which account for the position r_i of the pore center in the soil and the pore size s_i (surface or volume, depending on whether we are modeling the architecture of a 2-D or 3-D sample of the system).

The dynamics of the porous structure is modeled as a stochastic growth process known as dynamic network model (Dorogovtsev and Mendes, 2002). To motivate the rules that prescribe the evolution of the model, next we consider the influence of the pore properties on the probability for two pores being connected in a porous soil. Assume that at a given time a new pore is created in the medium, the likelihood that this pore will connect with any of the already existing pores will be proportional to the size of the older pores and inversely proportional to the distance between them. Likewise, the higher the number of connections accruing to an existing pore, the higher the likelihood that a new flow of fluids will intercept an existing connection and connect the new pore with the older one. Thus the attachment visibility of an existing network node $\pi(v_i)$ when a new node v_a is added will be proportional to k_i, s_i and $d^{-1}(r_i, r_a)$, where d is the Euclidean metric.

The previous considerations prompt us to model the dynamics of the porous structure of soils as a particular case of heterogeneous preferential attachment (PA) defined by three elements:

1. R is an arbitrary space. The elements $x \in R$ are referred as node states.
2. ρ is a nonnegative real function with unit measure over R referred as node state distribution.
3. σ is a nonnegative real function over R^2 referred as affinity of the interactions.

This formalism prescribes the evolution of a network according to the following rules:

- (i) The nodes v_i are characterized by their state $x_i \in R$. The node states describe intrinsic properties deemed constant in the timescale of evolution of the network.
- (ii) The growth process starts with a seed composed by N_0 nodes (with arbitrary states $x_i \in R$) and L_0 links.
- (iii) A new node v_a (with m links attached) is added to the network at each iteration. The number m is common for all the added nodes and remains constant during the evolution of the network. The newly added node is randomly assigned a state x_a following the distribution $\rho(x)$.

(iv) The m links attached to v_a are randomly connected to the network nodes following a distribution $\{\Pi(v_i)\}$ given by an extended *attachment rule*,

$$\Pi(v_i) = \frac{\pi(v_i)}{\sum_j \pi(v_j)}, \quad \pi(v_i) = k_i \cdot \sigma(x_i, x_a). \quad (1)$$

The *attachment kernel* or *visibility* π of a node v_i in the rule is given by the product of its degree k_i and its affinity σ with the newly added node v_a , which is itself a function of the states x_i and x_a . It thus can be seen that for each interaction σ biases the degree k_i of the candidate node. Steps (iii) and (iv) are iterated until a desired number of nodes has been added to the network. To sum up, the choice of the triple (R, ρ, σ) determines the form of heterogeneity in the attachment mechanism.

Consistently with the previous formalism, the porous soil model will be defined by a state space $R=M \times S$, where M is a Euclidean box with dimension 2 or 3 that represents the medium geometry and S is an interval of the real line that represents the spectrum of possible pore sizes; a state distribution $\rho(x)=\rho(r, s)$ that represents the probability for a new pore having a certain position r and size s ; and an affinity function

$$\sigma(x, y) = \frac{s_x^\alpha}{(\delta + |r_x - r_y|)^\beta}, \quad (2)$$

where δ is a small nonnegative offset that ensures the continuity of σ on R^2 , while α and β are free parameters that measure the relative importance of the size and distance of the pores in the affinity of the attachment mechanism. Defining the porosity of the medium $\phi=S_v/S_T=\sum_i s_i/\int_M dr$ as the fraction of void space in the material, then the desired number of nodes N that ends the growth process is the least number that yields $\phi>\phi_0$ for a certain threshold ϕ_0 .

3 Analysis

In this section we derive an analytical solution for the stationary degree distribution $P(k)$ of the proposed soil model. The solution is obtained by *rate equations* (Dorogovtsev et al., 2000; Krapivsky and Redner, 2001) which establish a balance in the flows of degree densities over a partition of the network. The exposition will be brief therefore we submit the reader to (Santiago and Benito, 2008a) for a more detailed discussion of the solution for the general class of PA models. Let us first define a sequence of functions $\{f(k, x, N)\}_{N>0}$ which measure the probability density of a randomly chosen node having degree k and state x in a network at the iteration $t=N$. The degree densities are local metrics, thus they uniformly converge when $N\rightarrow\infty$ to a stationary density function $f(k, x)$. Finally, the stationary degree distribution $P(k)$ measures the probability of a randomly chosen node having degree k in the thermodynamic limit.

We will denote by $V(x)=\{v_i, x_i=x\}$ the subset of nodes in the network with state $x=(r_x, s_x)$. Assuming that the assignment of states x_i is uncorrelated with the topology of the growing network, and that there are no linking events between existing nodes, the sequence $\{f\}$ can be modeled on each $V(x)$. For each x , the form of the equation will be $L_1-L_2=R_1-R_2$, where:

L_1 =density of nodes with degree k at $t=N+1$;

L_2 =density of nodes with degree k at $t=N$;

R_1 =increase in density due to nodes with degree $k-1$ that have gained a link at $t=N$;

R_2 =decrease in density due to nodes with degree k that have gained a link at $t=N$.

The resulting density rate equation for $k>m$ is

$$(N + 1)f(k, x, N + 1) - Nf(k, x, N) = m \left\langle \frac{\sigma(x, y)}{\psi(y, N)} \right\rangle_y [(k - 1)f(k - 1, x, N) - kf(k, x, N)]. \quad (3)$$

where $\psi(y, N)$ is defined as the partition factor

$$\psi(y, N) = \sum_{k'} k' \int_R \sigma(x, y) f(k', x, N) dx, \quad (4)$$

and the brackets mean averaging over the random variable y ,

$$\langle g \rangle_y = \int_R g \rho(y) dy. \quad (5)$$

For $k=m$ the resulting density rate equation is

$$(N + 1)f(m, x, N + 1) - Nf(m, x, N) = \rho(x) - m \left\langle \frac{\sigma(x, y)}{\psi(y, N)} \right\rangle_y m f(m, x, N). \quad (6)$$

There are no nodes with degree $k<m$, since when $N\rightarrow\infty$ all the links attached to newly added nodes find receptive nodes in the network, therefore the previous equations define all the possible cases in each iteration.

In the thermodynamic limit $N\rightarrow\infty$, $f(k, x, N + 1)=f(k, x, N)=f(k, x)$ and the rate equations become

$$f(k, x) = \begin{cases} m \left\langle \frac{\sigma(x, y)}{\psi(y)} \right\rangle_y [(k - 1)f(k - 1, x) - kf(k, x)] & \text{for } k > m \\ \rho(x) - m \left\langle \frac{\sigma(x, y)}{\psi(y)} \right\rangle_y m f(k, x) & \text{for } k = m \end{cases} \quad (7)$$

where $\psi(y)$ is the stationary partition factor, defined as

$$\psi(y) = \lim_{N\rightarrow\infty} \psi(y, N) = \sum_{k'} k' \int_R \sigma(x, y) f(k', x) dx. \quad (8)$$

Notice that the stationary rate equations are coupled via $\psi(y)$. In order to decouple Eq. 7, let us first assume that the variation of $\psi(y)$ over the subset $R_\rho = \{y \in R; \rho(y) > 0\}$ is small enough. This can be expressed as the following approximation criterium

$$\langle (\psi(y) - \bar{\psi})^2 \rangle_y < \varepsilon_1, \quad (9)$$

where $\bar{\psi} = \langle \psi(y) \rangle_y$ and ε_1 is an accuracy bound. In that case we may approximate the coupling term as a quotient of two *mean-field* measures,

$$\left\langle \frac{\sigma(x, y)}{\psi(y)} \right\rangle_y \simeq \frac{\langle \sigma(x, y) \rangle_y}{\langle \psi(y) \rangle_y} = \frac{w(x)}{\bar{\psi}}, \quad (10)$$

where $w(x)$ can be interpreted as the mean-field fitness of network nodes with state x ,

$$w(x) = \langle \sigma(x, y) \rangle_y = s_x^\alpha \int_R \frac{\rho(y)}{(\delta + |r_x - r_y|)^\beta} dy, \quad (11)$$

and ψ can be interpreted as the mean-field partition factor of the network,

$$\begin{aligned} \psi &= \langle \psi(y) \rangle_y = \\ &= \int_R \sum_{k'} k' \int_R \frac{s_x^\alpha}{(\delta + |r_x - r_y|)^\beta} f(k', x) dx \rho(y) dy, \end{aligned} \quad (12)$$

both defined for a given distribution $\rho(y)$ of the incoming node states. Furthermore, let us assume that the variation of $w(x)$ over R is small enough, which can be expressed as a second approximation criterium

$$\langle (w(x) - \bar{w})^2 \rangle_x < \varepsilon_2, \quad (13)$$

where $\bar{w} = \langle w(x) \rangle_x$ and ε_2 is a second accuracy bound. In that case the mean-field stationary partition factor ψ can be approximated by $\psi \simeq \sum_k k P(k) \bar{w} \simeq 2m \bar{w}$. Under the assumption that both approximation criteria (Eqs. 9 and 13) hold, the coupling factor may then be estimated as

$$\left\langle \frac{\sigma(x, y)}{\psi(y)} \right\rangle_y \simeq \frac{w(x)}{\bar{\psi}} \simeq \frac{w(x)}{2m \bar{w}} = \frac{\hat{w}(x)}{2m}, \quad (14)$$

where $\hat{w}(x) = w(x)/\bar{w}$ is a mean-field *normalized fitness*, which measures the fitness of network nodes with state x relative to the average fitness of all the network nodes, for a given distribution $\rho(x)$. The resulting decoupled system becomes then

$$\begin{aligned} f(k, x) &= \\ \hat{w}(x)[(k-1)f(k-1, x) - kf(k, x)]/2 &\text{ for } k > m, \\ \rho(x) - \hat{w}(x)m f(k, x)/2 &\text{ for } k = m. \end{aligned} \quad (15)$$

Solving Eq. 15 we obtain $f(m, x) = 2\rho/(\hat{w}m+2)$ and the solution for the stationary density for $k > m$ is

$$f(k, x) = \left(\prod_{j=m+1}^k \frac{\hat{w}(j-1)}{\hat{w}j+2} \right) \frac{2\rho}{\hat{w}m+2} \quad (16)$$

and integrating the density in Eq. 16 over the state space R the stationary degree distribution is

$$P(k) = \int_R \left(\prod_{j=m+1}^k \frac{\hat{w}(j-1)}{\hat{w}j+2} \right) \frac{2\rho}{\hat{w}m+2} dx. \quad (17)$$

The solution given by Eq. 17 is valid for any variant of the soil model, irrespective of the geometry of the medium M or the pore sizes S , the distribution of pore properties ρ or the dependence of the affinity σ on the pore properties via α and β . Nevertheless, the validity of the solution is restricted by the extent to which the approximation criteria in Eqs. 9 and 13 hold. The mean-field approximation can be considered accurate in the measure that the averages of $\sigma(x, y)$ along its two arguments do not exhibit large variations over R . This does not exclude the possibility of large fluctuations in the affinity $\sigma(x, y)$ over R^2 , provided that the variability of the averages is kept within the desired bounds.

Furthermore, we may assume that any pore in the soil has a non-zero size, thus from Eq. 11 it follows that the mean-field fitness $w(x)$ is strictly positive over R and the stationary degree density in Eq. 16 may be expressed for $k \geq m$ as

$$f(k, x) = \frac{2\rho/\hat{w}}{m+2/\hat{w}} \frac{B(k, 1+2/\hat{w})}{B(m, 1+2/\hat{w})}, \quad (18)$$

where Legendre's Beta function $B(y, z) = \int_0^1 t^{y-1}(1-t)^{z-1} dt$ for $y, z > 0$ satisfies the functional relation $\Gamma(a)/\Gamma(a+b) = B(a, b)/\Gamma(b)$ for Euler's Gamma function $\Gamma(x)$. Likewise, integrating the density in Eq. 18 over R we obtain an expression of the stationary degree distribution $P(k)$ for $k \geq m$ in terms of the Legendre's Beta function which is equivalent to Eq. 17.

Notice that the stationary densities for a given pore state in the soil model, as given by Eq. 18, follow the form of a Beta function with arguments $(k, 1+2/\hat{w})$. Given that the Beta function behaves as $B(y, z) \sim y^{-z}$ when $y \rightarrow \infty$, this implies that the degree densities exhibit a *multiscaling* according to power laws $k^{-\gamma(\hat{w})}$ along the continuous spectrum of normalized fitness \hat{w} , with scaling exponents $\gamma(\hat{w}) = 1+2/\hat{w}$ spanning themselves a continuum. The multiscaling phenomenon is a general property of heterogeneous PA networks (Santiago and Benito, 2007a, 2008a) that is exhibited by any variant of the proposed soil model, irrespective of the particular details. As \hat{w} increases (resp. decreases), the exponent γ of f decreases (resp. increases). Nodes with states more fit than the average ($\hat{w} > 1$) adopt densities with $\gamma < 3$, which exhibit a slower asymptotical decay, and tend to produce more hubs. Nodes with states less fit than the average ($\hat{w} < 1$) adopt densities with $\gamma > 3$, which exhibit a faster asymptotical decay.

Notice also that the mean-field fitness in Eq. 11 may be factorized into $w(x) = w_1(r_x) \cdot w_2(s_x)$, where the first term accounts for the fitness dependence on the pore position, and the second one accounts for the dependence on the pore size. We may use this fact to simplify the analysis of the influence

of the pore properties (r_x, s_x) on the scaling exponents in the density $f(k, x)$. The influence of the pore position r will depend largely on the probability distribution of the pore positions $\rho_M(r) = \int_S \rho(r, s) ds$. When the pore position is uniformly distributed over the medium M , for a given pore size s_0 the mean-field fitness w will be highest on the medium center and will decrease as the pore position approaches the boundary of M , since $w_1(r)$ is performing an averaging of the reciprocal of distance over M . However, as the distribution of the pore positions ρ_M becomes increasingly inhomogeneous, the situation of the pores with respect to the regions with higher densities will become progressively more important in the determination of the fitness factor w_1 . For highly variable distributions ρ_M this trait will prevail over the centrality of the pores in the medium geometry. Irrespective of the inhomogeneity of ρ_M , notice that as β is increased the variability of the factor w_1 over M will grow, as the contribution to w_1 by distant pores becomes progressively lower.

With regards to the influence of the pore size s , for a given pore position r_0 the fitness factor will increase with the size as $w_2 \sim s^\alpha$ irrespective of the distribution ρ of pore properties. Notice that again as α is increased the variability of the factor w_2 over S will grow. It should be also pointed out that with the general form of σ adopted for the soil model the scaling exponents of $f(k, x)$ will exhibit themselves a power-law scaling in their dependence on s according to $\gamma(r_0, s) \sim s^{-\alpha}$. To sum up, the increase of either α or β results in an increase in the variability of w over R , and in the case of β such increase is more acute as the inhomogeneity of ρ_M increases. Given the dependence of the scaling exponent γ on the normalized fitness \bar{w} , the higher variability of w will translate into a larger spread of the distribution of exponents γ of the density components.

The described behavior evidences a *signature* of heterogeneous PA in the structure of porous soils, by which the density components $f(k, x)$ of pores in the underlying network will exhibit different scaling exponents according to the pore properties such as size or position. This signature may be empirically detected by selecting subsets of pores within the soil sample so that one of their properties takes a similar value, common to all the subsets, while the other property takes a similar value within each subset, but differing across the other subsets: $\{(r_i, s_0)\}_i$, or conversely, $\{(r_0, s_i)\}_i$. If the partial degree distributions for these subsets $\{P(k|x_i)\}$ exhibit power laws with different exponents, it is then feasible to assume that this pore property (either position or size) is biasing the attachment mechanism. Under equal circumstances, a higher variation in the empirically obtained exponents would point to a higher exponent α or β in the affinity function σ . Furthermore, to check the existence of heterogeneity it would be sufficient to find one such subset of nodes whose partial distribution $\{P(k|x_i)\}$ exhibits a power-law behavior with an exponent different to the one in the global distribution $P(k)$.

With regards to the stationary degree distribution $P(k)$ of the soil model, Eq. 17 shows that it is obtained by the integration of the density components, which form a set of power-laws with varying exponents $\gamma(x)$. As we have seen, little fluctuations of w over R will yield density components with similar scaling exponents and thus degree distributions similar to the homogeneous PA case, the so-called Barabási-Albert model. On the other hand, large fluctuations of w over R (as in the case of large α or large β parameters, particularly with inhomogeneous ρ distributions) will yield a wider spectrum of exponents γ and thus a larger deviation of the degree distribution from the Barabási-Albert model. The asymptotic behavior of $P(k)$ will be dominated by the slowest decaying components, associated to pore properties with highest mean-field fitness w . Furthermore, given that \hat{w} will be distributed by definition around 1, the highest value of the spectrum will necessarily verify $\hat{w}_{\max} \geq 1$ and therefore the degree distribution $P(k)$ of all the variants of the soil model will exhibit scaling exponents satisfying $1 < \gamma \leq 3$.

Finally a caveat should be raised about the accuracy of the preceding discussion regarding small pore networks. The analytical results have been derived for the thermodynamic limit, where the degree densities (and thus the degree distribution) reach stationarity. Strictly this does not apply in the context of pore networks, where there exist intrinsic limitations to the growth process. Given that each pore has a non-negligible size and the medium size is constant, the porosity of generated soils grows monotonically with time, eventually reaching the desired porosity of the modeled system. This circumstance yields a requisite size of the generated network for a medium sample of a given size. Being the average pore size $\bar{s} = \int_R s \rho(r, s) dx$, then the requisite network size will be given by $N_0 \simeq \phi_0 S_T / \bar{s}$, where S_T is the total medium size and ϕ_0 is the desired porosity. As the requisite size N_0 becomes lower the finite-size effects will become more important, in particular when $N_0 \ll 10^3$. This may be the case when the medium porosity is very low or when the pore size is large relative to the medium size. The finite-size effects will yield a growing discrepancy between some network metrics and the analytical results, such as a decrease in the scaling exponent γ of the degree distribution and the presence of an exponential cutoff regime. This problem could nevertheless be circumvented by modeling a larger soil sample.

4 Model application

In this section we present results concerning the numerical simulation of the porous soil model using as input data from intact soil samples with different physical properties, obtained from the same profile but at different depths. We simulate pore networks with the same porosity, size distribution and spatial location of the empirical 2-D soil samples.

Table 1. Physical properties of two soil samples (D1 and D3) obtained in the same profile at different depths. Note: C. Sand and F. Sand correspond to coarse and fine particles of sand respectively. The names D1 and D3 indicate the sequence of cuts in depth, thus, in this soil, there is an intermediate cut D2 between D1 and D3.

Soil	Horizon	Depth (cm)	Particle (%)			
			C.sand	F.sand	Silt	Clay
D1	A2	10–35	62	24	3	11
D3	Bt2	98–152	21	40	4	35

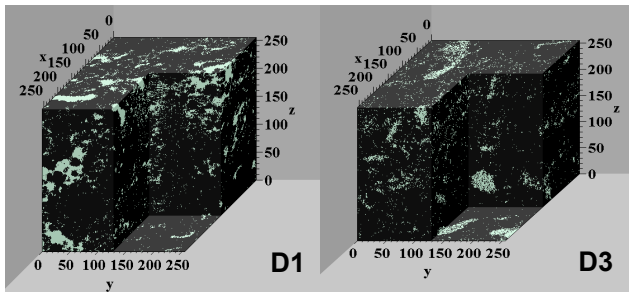


Fig. 1. 3-D representation of soils D1 (left) and D3 (right) obtained using Computed Tomography (CT).

4.1 Empirical analysis of soil samples

The soil samples used in this study were collected from two horizons of an Argissolo (EMBRAPA SOLOS, 2006). Physical characteristics of relevance to the current study, can be observed in Table 1. This soil is characteristic of the coastal tablelands of northeast Brazil, formed on the Tertiary Barreiras group of formations in Pernambuco state (Itapirema Experimental Station) presenting a hardsetting behavior.

The images of the soil samples D1 and D3 were obtained using an EVS (now GE Medical) MS-8 MicroCT scanner (Fig. 1). Though some samples required paring to fit the 64 mm diameter imaging tubes, field orientation was maintained. Imaging parameters were 155 keV and 25 A. Proprietary software (GE Medical) was used to reconstruct the 16-bit 3-D imagery from the sequence of axial views. The resulting voxel size was $45.1 \mu\text{m}$. File sizes ranged from 70 to 200 Mb, which made subsequent processing of the entire volume practically infeasible. Accordingly, three subvolumes were extracted from each of the two original volumes (using GE Medical Microview) and care was taken to ensure no overlay of the subvolumes. The subvolumes measured $256 \times 256 \times 256$ units, corresponding to about 16.8 million voxels. A 3-D Gaussian filter in MicroView (GE Healthcare, 2006) was also run on each subvolume to reduce noise and beam-hardening artifacts, typical of CT imagery.

CT imagery of soil, like other digital imagery, typically contains a large proportion of mixed-voxels (voxels whose digital number is the weighted average of more than one con-

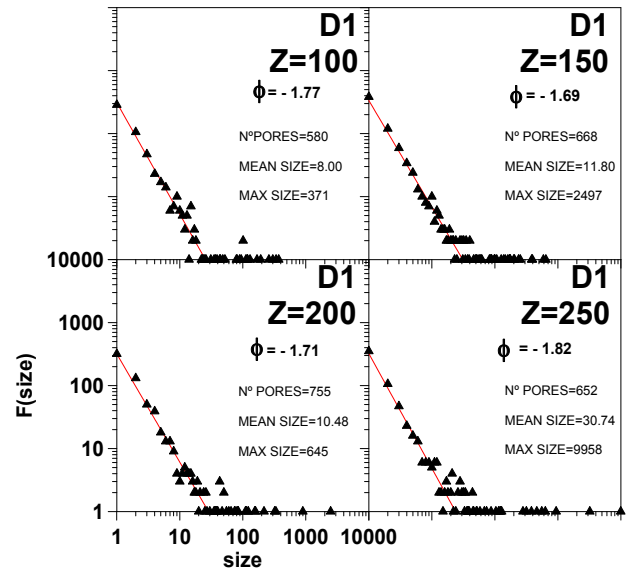


Fig. 2. Pore size distribution, $F(\text{size})$, for soil sample D1 obtained at different depths (i.e. longitudinal cuts Z). The graphics also display the number of pores, mean size of pores (pixels^2), maximum size of pores (pixels^2) and scaling exponent ϕ for each distribution. The straight lines correspond to the power law fittings.

stituent - such as a solid/air interface). To facilitate identification of constituent peaks in the gray-scale histogram, a 3-D filter executed in NIH ImageJ (Rasband, 2006) was run on each subvolume to mask voxels which differed by more than 0.1% from the surrounding neighborhood of 124 voxels ($5 \times 5 \times 5$ unit volume). Full details of this technique can be found in (Elliot and Heck, 2007). Histograms of the unmasked voxels were subsequently ported into OriginPro (Origin Lab Corporation, 2006); after smoothing the histograms (adjacent averaging of 25 levels), peaks were identified in the Peak Fitting Module. The major peak with the lowest mean digital number was taken to be that corresponding to the void space; the next major peak was considered to be solid material. Based on the central tendency and dispersions (assuming Gaussian distributions) of the two peaks, one for air ($\mu_{CT\text{air}}$) and one for solid ($\mu_{CT\text{solid}}$), a threshold value was identified as the equi-probability value for air and solid.

As our model considers the pore size (surface= S) we analyze the distribution of pore sizes in both soil samples making longitudinal cuts of 256×256 units (pixels) in the 3-D CT imagery following the algorithm proposed by Vogel and Roth (2001). This algorithm basically consists in marking the adjacent black pixels in two consecutive lines on the entire plane and then performing an overall study to recognize the pore entities and in that way to obtain sizes and locations. The PSD, given by $F(\text{size})$, for D1 can be observed in Fig. 2 for four profile cuts at different depths ($Z=100, 150, 200, 250$). As can be seen, all the cuts display a distribution that

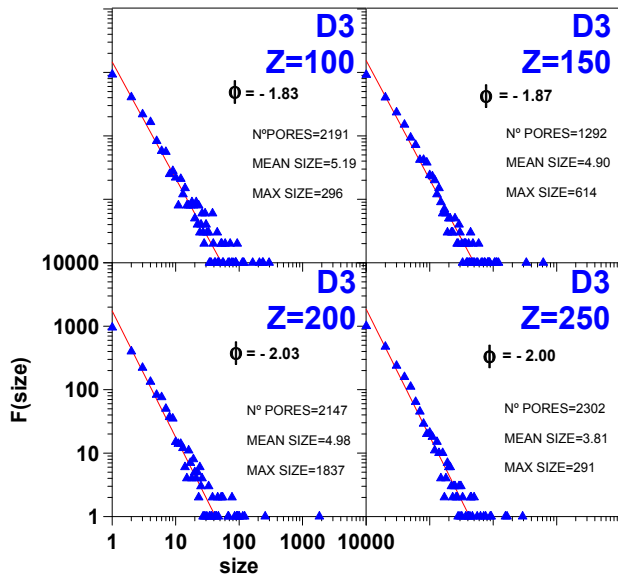


Fig. 3. Same as Fig. 2 for soil sample D3.

follows a power law denoting a scale free character in the distribution of pore sizes, $f(s) \sim s^{-\phi}$.

The same behaviour can be observed in the deepest soil sample D3 exhibited by Fig. 3, nevertheless, in this scenario the number of pores and its size are opposite to D1. The differences in the 3-D CT imagery of both soil samples can be observed in Fig. 1. Due to the fact that our model considers also the spatial location of pores, we analyzed the spatial distribution of the centers of mass of pores (CMP) (Vogel and Roth, 2001) in the same cut previously mentioned for both soil samples. Figure 4A shows the distribution of the pore centers in four superimposed cuts for the sample soil D1 (the results for D3, not shown, are similar). The spatial distributions of CMP were analyzed to study their uniformity at each depth considered. We counted the number of CMP found within each box, varying the box length size from 2, 4, 8, up to 32 pixels. In the longest size length used 64 data points were obtained hence no bigger sizes were used to assure enough data available for statistics. Then, the variance for each side length was calculated and is shown in Fig. 4B. The results evidence that the spatial distribution of pores is reasonably uniform over the surfaces for all the cuts considered.

4.2 Results of the model

Figure 5 shows two pore networks generated by the soil model assuming parameters corresponding to soil samples D1 and D3. Each new node added to the growing network is assigned a characteristic surface and spatial location. Following the empirical results obtained regarding the PSD in soil samples D1 and D3 (see Fig. 6), the pore surfaces in both networks are distributed according to power-laws with scal-

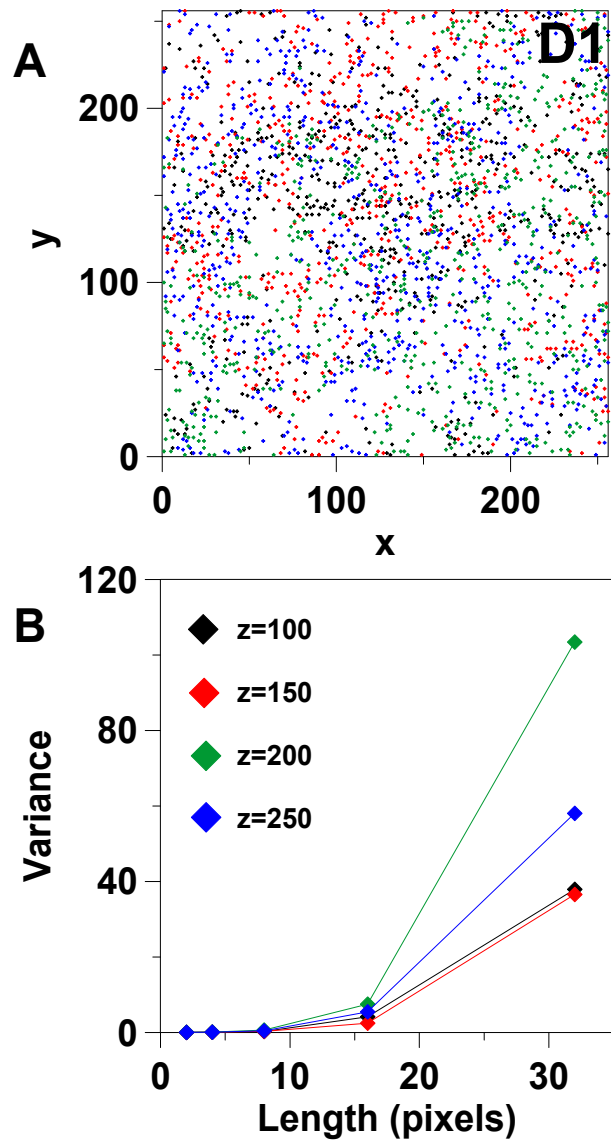


Fig. 4. (A): Spatial location of the mass center of pores in four longitudinal cuts Z (different colors) for the soil sample D1. (B): Variance of the number of CMP found in boxes of different length in different longitudinal cuts Z.

ing exponents $\phi_{D1} = -1.7$ (D1) and $\phi_{D3} = -2.0$ (D3). Likewise the maximum pore size in the distributions is bounded by 9958 pixels² (D1) and 1837 pixels² (D3), and in both cases we consider the minimum pore size equal to 1 pixel². It should be pointed out that the nodes in the networks correspond to the centers of mass of pores and the size of pores is not represented. The spatial location of the pore centers is uniform for both networks, again in accordance with the empirical results obtained in soil samples D1 and D3 (see Fig. 4). The surfaces of the soil samples simulated are chosen differently in order to yield the porosity coefficients observed considering Z=11 cuts (10.3% for D1 and 15.8% for

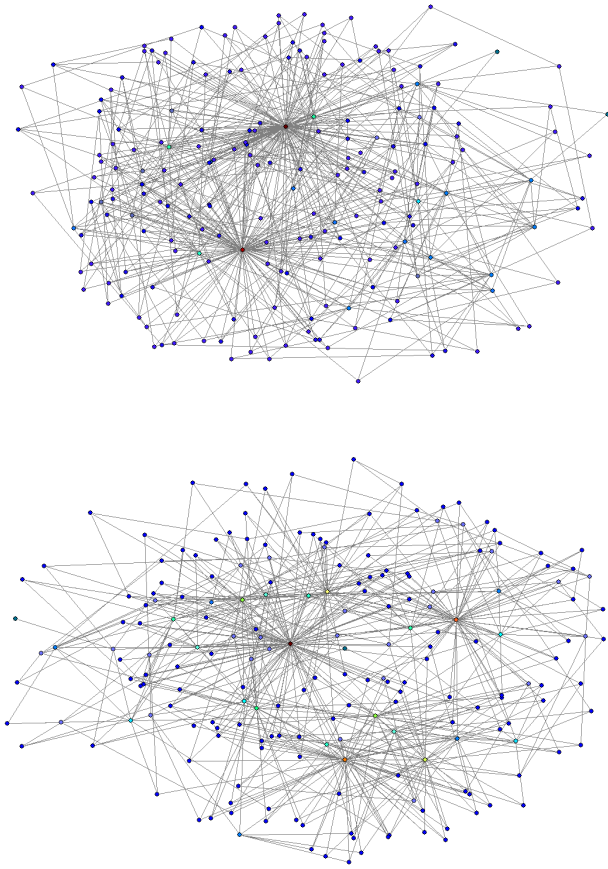


Fig. 5. Layouts of two pore networks generated by the soil model corresponding to soil sample D1 (top) and D3 (bottom). Node color corresponds to the connectivity degree k : warm colors mean high degrees, cold colors mean low degrees. The spatial position of pores in these layouts has been chosen according to the Kamada-Kawai algorithm.

D3) irrespective of the network size. For instance, given a network size of $N=10\,000$ pores we simulate soil surfaces of $332\,6976\text{ pixels}^2$ (D1) and $473\,344\text{ pixels}^2$ (D3).

The parameters α and β were also chosen differently in the simulated networks. Considering that the soil sample D3 has a higher percentage of clay (see Table 1) and a lower size of pores (see Fig. 1) we can assume that it is more compact. For this reason we assign to the distance between pores a higher importance in our model. Thus, for the soil D3 we used $\alpha=1.0$ whereas in D3 $\alpha=0.5$. On the other hand, due to the fact that the soil D1 has a higher percentage of sand (see Table 1) and a higher size of pores (see Fig. 1) we can assume that the bigger particles in the sandy soil prevent higher connectivities in pores formed by sand, contrary to the pores formed by clay. In a clay pore its surface (perimeter in our 2-D simulation) can be occupied by a higher number of particles, contrary to a sand pore of similar size, thus the number of potential connections (space that separates two particles) of a clay pore is higher in comparison with a sand pore. Be-

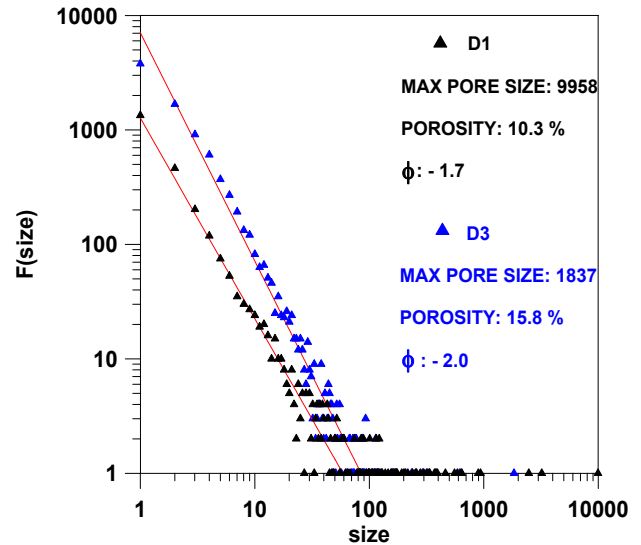


Fig. 6. Pore size distribution, $F(\text{size})$, for soil samples D1 (black triangles) and D3 (blue triangles) obtained at different depths (i.e. longitudinal cuts $Z=11$). Inside the figures can be observed the maximum size of pores (pixels²) and the scaling exponent ϕ for each distribution. The straight lines correspond to the power law fittings.

cause the surface of the pore is more important in a clay soil, in our model we use $\beta=1.0$ for D1 and $\beta=2.0$ for D3.

In the model both networks start with a seed of $N_0=10$ pores connected by $L_0=9$ links. In both cases the the number of potential links for each new node added is $m=3$ and it remains unaltered during the evolution of the network. The aggregation process is iterated until 200 nodes have been added to the network (the low number is chosen for the sake of visibility). The simulation results in Fig. 5 show that in the top network the model parameters tend to prevent the formation of hubs more drastically than the bottom one. Figure 7 depicts the cartesian pore networks generated by the model representing the soil D1 (top) and the soil D3 (bottom). The cartesian layouts also evidence that the size of the pores tends to be a more significant trait in the final connectivity of the pore than the pore position, as evidenced by a large share of highly connected nodes being located in peripheral positions of the soil surface. This insight may be explained by the fact that pore sizes follow highly inhomogeneous distributions, while pore positions follow completely homogeneous distributions. The choice of the affinity parameters α and β also play a role in this trait, as the introduction of a superlinear β tends to further decrease the correlation between position and connectivity in the cartesian layouts, as shown by the bottom network in Fig. 7.

Figure 8 depicts the degree distributions $P(k)$ obtained through numerical simulation of the D1 and D3 networks with a size of $N=10\,000$ pores. The results show that the distributions do not fit well a power-law along the studied

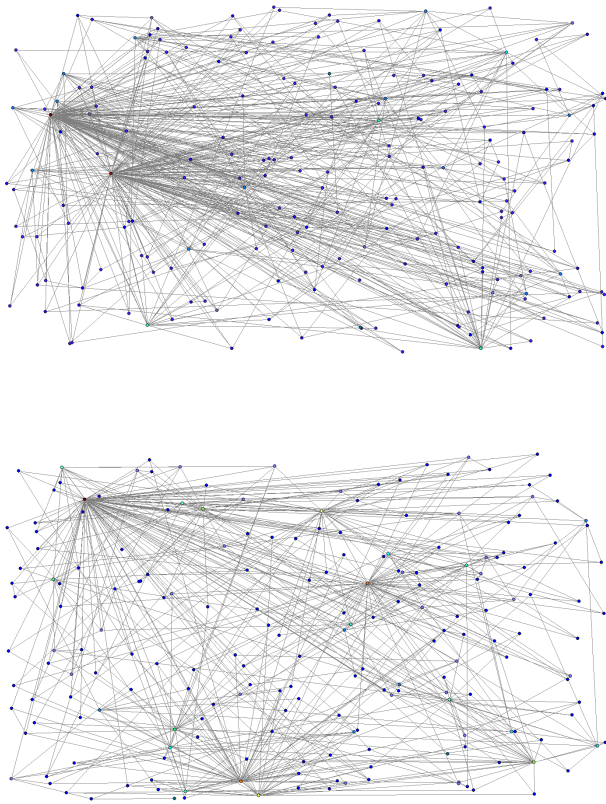


Fig. 7. Cartesian networks of pores generated by the soil model considering real data from soil samples. The top network corresponds to soil sample D1 and the bottom one to sample D3. Color code is as in Fig. 5.

interval, however they evidence a progressively better agreement for higher degrees. This can be explained by the fact that the power-law components in the density spectrum with lower scaling exponents dominate the distribution decay, so that the asymptotic behavior of $P(k)$ over arbitrarily high degrees would fit the power-law as analytically predicted. The behavior of the distribution for lower degrees can be explained by the inhomogeneity of the distribution of pore sizes, which yields a non-negligible presence of very large pores in the network. The distributions also evidence that the sample D3 tends to yield a higher probability of hubs than the sample D1, as was expected from the network layouts represented by Figs. 5 and 7. We have to remark that the connectivity pattern between pores was not observed empirically, the scale-free distributions obtained (both numerically and analytically) are the result of the implementation of a reasonable attachment rule $\prod(v_i)$ and the adoption of real parameters of soil samples. The distributions $P(k)$ generated by the soil model can be tested in future research.

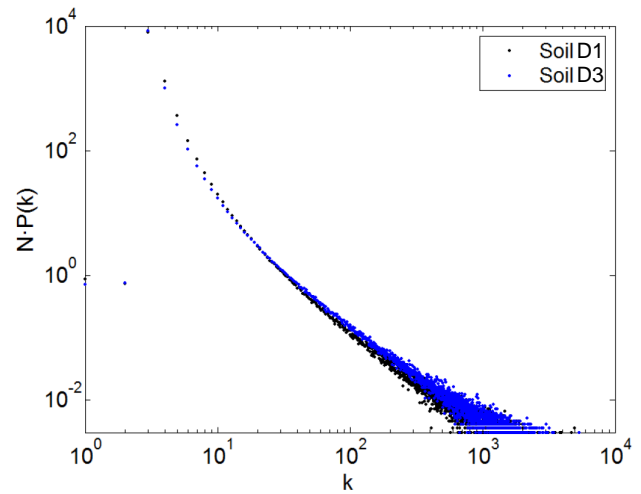


Fig. 8. Distributions of the number of nodes with degree k , $N \cdot P(k)$, obtained with the soil model for the soil samples D1 (black) and D3 (blue). The network size is $N=10\,000$ in both cases.

5 Conclusions

In summary, we have presented a complex network model based on a heterogeneous PA scheme in order to quantify the structure of porous soils. The proposed soil model allows the specification of different medium geometries and pore dynamics through the choice of different underlying spaces, distributions of pore properties and affinity functions. We have obtained analytical solutions for the degree densities and degree distribution of the pore networks generated by the model in the thermodynamic limit. We have shown that these networks exhibit a multiscaling of their degree densities according to power laws with exponents spanning a continuum, and that such phenomenon leaves a signature of heterogeneity in the topology of pore networks that can be empirically tested in real soil samples.

We have also shown the relationship between the variability in the scaling exponents and the parameters regulating the affinity function, as well as the inhomogeneity of the distributions of pore properties, and the consequences of this on the asymptotic behavior of the degree distribution. In particular, it is worth emphasizing that the degree distributions of all the variants of the soil model exhibit power-law behaviors with exponents within the limits empirically observed in real networks. We have performed a numerical analysis of the soil model for a combination of parameters corresponding to empirical samples with different properties. We have shown that the simulation results exhibit a good agreement with the analytical predictions, concerning the scaling behavior of the degree metrics as well as the ranges of values of the scaling exponents obtained.

Acknowledgements. This work has been supported by the Spanish MEC under Project “i-MATH” No. CSD2006-00032 and Project No. MTM2006-15533, GESAN and Comunidad de Madrid under Project TAGRALIA-CM-P-AGR-000187-0505.

Edited by: A. Tsonis

Reviewed by: J. Garcia Miranda and another anonymous referee

References

- Albert, R. and Barabási, A.-L.: Statistical mechanics of complex networks, *Rev. Mod. Phys.*, 74(1), 47–97, 2002.
- Bak, P., Tang, C., and Wiesenfeld, K.: Self-organized criticality: an explanation of $1/f$ noise, *Phys. Rev. Lett.*, 59, 381–384, 1987.
- Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science* 286, 509–512, 1999.
- Barabási, A.-L., Albert, R., and Jeong, H.: Mean-field theory for scale-free random networks, *Phys. A.*, 272, 173–197, 1999.
- Bird, N., Cruz Díaz, M., Saab, A., and Tarquis, A. M.: Fractal and multifractal analysis of pore-scale images of soil, *J. Hydrol.*, 322, 211–219, 2006.
- Blair, J. M., Falconer, R. E., Milne, A. C., Young, I. M., and Crawford, J. C.: Modeling three-dimensional microstructure in heterogeneous media, *Soil. Sci. Soc. Am. J.*, 71, 1807–1812, 2007.
- Cislerova, N.: Characterization of pore geometry, in: *Proceedings of the International Workshop on modeling of transport processes in soils at various scales in time and space*, Leuven, Belgium, 24–26 November 1999, 103–117, 1999.
- Dorogovtsev, S. N. and Mendes, J. F. F.: Evolution of networks, *Adv. Phys.*, 51, 1079–1187, 2002.
- Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N.: Structure of growing networks with preferential linking, *Phys. Rev. Lett.*, 85, 4633–4636, 2000.
- Elliot, T.R. and Heck, R. J.: A comparison of 2-D and 3-D thresholding of CT imagery, *Can. J. Soil Sci.*, 87(4), 405–412, 2007.
- EMBRAPA SOLOS: Sistema de Clasificacao de Solos, 2 Edicao, Empresa Brasileira de Pesquisa Agropecuaria. Solos, Rio de Janeiro, 2006.
- Erdős, P. and Rényi, A.: On random graphs, *Publ. Math. Debrecen*, 6, 290–297, 1959.
- GE Healthcare: MicroView 2.1.2 – MicroCT Visualization and Analysis, London, Canada, 2006.
- Horgan, G. W. and Ball, B. C.: Simulating diffusion in a Boolean model of soil pores, *Eur. J. Soil Sci.*, 45, 483–491, 1994.
- Krapivsky, P. L. and Redner, S.: Organization of growing random networks, *Phys. Rev. E.*, 63, 066123, doi:10.1103/PhysRevE.63.066123, 2001.
- Lin, H. S., McInnes, K. J., Wilding, L. P., and Hallmark, C. T.: Effects of soil morphology on hydraulic properties. I. Quantification of soil morphology, *Soil Sci. Soc. Am. J.*, 63, 948–954, 1999.
- Newman, M. E. J.: The structure and function of complex networks, *SIAM Review*, 45, 167–256, 2003.
- Origin Lab Corporation: Origin Pro 7.5. Northampton, MA, 2006.
- Price, D. J. S.: Networks of scientific papers, *Science*, 149, 510–515, 1965.
- Price, D. J. S.: A general theory of bibliometric and other cumulative advantage processes, *J. Amer. Soc. Inform. Sci.*, 27, 292–306, 1976.
- Rasband, W.: ImageJ 1.36. National Institutes of Health, USA, 2006.
- Santiago, A. and Benito, R. M.: Emergence of multiscaling in heterogeneous complex networks, *Int. J. Mod. Phys. C.*, 18, 1–17, 2007a.
- Santiago, A. and Benito, R. M.: Connectivity degrees in the threshold preferential attachment model, *Phys. A.* 387, 2365–2376, 2007b.
- Santiago, A. and Benito, R. M.: An extended formalism for preferential attachment in heterogeneous complex networks, *Europhys. Lett.*, 82, 58004, doi:10.1209/0295-5075-82-58004, 2008a.
- Santiago, A. and Benito, R. M.: Improved clustering through heterogeneity in preferential attachment networks, *Int. J. Bifurc. Chaos*, accepted, 2008b.
- Strogatz, S. H.: Exploring complex networks, *Nature*, 410, 268–276, 2001.
- Tsonis, A. A., Swanson, K. L., and Roebber, P. J.: What do networks have to do with climate? *Bull. Amer. Meteor. Soc.*, 87, 585–595, 2006.
- Tsonis, A. A., Swanson, K. L., and Kravtsov, S.: A new dynamical mechanism for major climate shifts, *Geophys. Res. Lett.*, 34, L13705, doi:10.1029/GL030288, 2007.
- Tsonis, A. A., Swanson, K. L., and Wang, G.: On the Role of Atmospheric Teleconnections in Climate, *J. Climate*, 21, 2990–3001, 2008.
- Tsonis, A. A., and Swanson, K. L.: Topology and Predictability of El Niño and La Niña Networks, *Phys. Rev. Lett.*, 100, 228502, doi:10.1103/PhysRevLett.100.228502, 2008.
- Valentini, L., Perugini, D., and Poli, G.: The small-world topology of rock fracture networks, *Physica A.* 377, 323–328, 2007.
- Vogel, H.-J.: Morphological determination of pore connectivity as a function of pore size using serial sections, *Eur. J. Soil Sci.*, 48, 365–377, 1997.
- Vogel, H.J., and Roth, K.: Quantitative morphology and network representation of soil pore structure, *Adv. Water Res.*, 24, 233–242, 2001.
- Watts, D. and Strogatz, S.: Collective dynamics of small-world networks, *Nature*, 393, 440–442, 1998.
- Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate Networks around the Globe are Significantly Affected by El Niño, *Phys. Rev. Lett.* 100, 228501, doi:10.1103/PhysRevLett.100.228501, 2008.
- Young, I. M. and Crawford, J. W.: Interactions and self-organization in the soil-microbe complex, *Science*, 304, 1634–1637, 2004.