



UNIVERSITÀ
DI TRENTO

Department of
Cellular, Computational and Integrative Biology - CIBIO

MASTER DEGREE IN QUANTITATIVE AND COMPUTATIONAL BIOLOGY

**About biotic and moisture influence on soil structure and its
information theory-based complexity characterization**

Supervisors:

Riccardo Rigon
Luis Cunha

Graduant:

Enrico Chiesa

Signature

Date of discussion: 20/03/2025

Academic Year: 2024/2025

1 Abstract

Soil is, arguably, the most important system for life on the surface of earth. Its intricate porous structure retains water, that can be used by the variegated microbial, fungal and animal communities that live inside this tridimensional maze. It allows for water and nutrient intake by plants' roots, that in exchange shape the soil structure, exchanging nutrients and minerals with microbes and fungi, while contributing to the dynamic maintaining of soil architecture. The complex network of multiscale porous structures dictate the soil's ability to regulate water flux, disperse nutrients, and, ultimately, sustain life at any level. Soil is a complex system where biological, chemical ad physical processes are tightly intertwined, in a way that does not really allows for a separation between different functions, as every actor is deeply influenced by all the others at any moment. Very small perturbations of the system are likely to have a dramatic cascade of effects with repercussions on a large spatial and temporal scale. At the same time, thanks to its extraordinary internal variability, soil is one of the most resilient environments that we know of; biological communities inside soil can resist almost every environmental condition, an incredible variety of life forms is found in any type of soil that covers our planet. Apart from supporting life inside itself, soil is a key component in the regulation of nearly all processes in earth's critical zone, offering support and habitat to almost all terrestrial plants and animals, and mediating water cycle and flow though the environment.

Due to the very nature of its structure and functionality, it appears natural to study soil with an holistic and complexity-oriented approach, so to try and individuate some crucial quantities and processes that can be used as central parameters to model soil processes. This thesis investigates the impact of moisture conditions and earthworm activity on soil structure complexity using Information Theory approaches. While most diffused methods to analyze soil structure focus either on the aggregates or on the pore space, we will use an innovative approach, with the objective to address directly and quantitatively soil structural complexity. We will analyze soil architecture as captured through 3D X-ray Computed Tomography (CT) greylevels scans; by making use of tools from the Fisher-Shannon analysis of complex signals, we will characterize the complexity of the soil structure under different environmental and biological conditions.

Specifically, the study was carried out on soil samples maintained in controlled environment, comparing soils in different biological conditions: with and without the presence of two earthworm species with different physiological and behavioral characteristics (*A. rosea*, endogeic and *A. trapezoides*, anecic). The samples were subjected to two different environmental conditions for what concerns humidity levels: moisture was maintained at 45% or 55% of the maximum water-holding capacity, representing dry and wet conditions respectively. Our information theory approach utilized three key metrics: Shannon Entropy Power (SEP), Fisher Information Measure (FIM), and Fisher-Shannon Complexity (FSC). These metrics were calculated, starting from the probability density function of the intensity values of the greyscale images, for four different partitions of the soil volumes (Columns, Planes, Walls, and whole Volumes). This was done in order to identify the best partition to discriminate between the various conditions in our settings. The Walls and Volumes partitions provided the clearest differentiation between the diverse ecological and biological conditions.

Our results indicate the chosen complexity metrics are very good at discriminating between different soil conditions, in a way that is environmentally and biologically relevant. In particular, we found that moisture level strongly influence soil structure's entropy, with wetter soils that appear to be more disordered than the dry ones. Our measure of structural organization (FIM) of the soil architecture was able to identify with good precision, not only the presence of earthworm of a particular species, but also the degree of activity of those animals inside the sample, embodied by their final mass. While aligning with our current understanding of soil physical and biological processes, our analysis provides new, innovative tools for the non-disruptive study of soil systems' structural proprieties. Our approach avoids completely the direct manipulation of the soil samples, and also eliminate the necessity for the controversial binarization process of greyscale images, usually performed to retrieve some basic properties of the pore space. Our findings have also the potential to help qualitatively understand and measure how earthworms' activity influences hydraulic connectivity and enhances the bimodality of the Water Retention Curve (WRC).

Our work is a small, but clear step in the new direction of information theory based characterization and analysis of soil systems. In our opinion, this research have the potential to be very impactful, both from a theoretical and a practical point of view, inside the soil science community. We acknowledge the fact that our results need to be confirmed independently, and our analysis validated for different soil types, environmental and biological conditions, and spatial and temporal scales. We look forward to the future steps in this direction,

hoping that our research will help to set a good start, and convinced that an holistic, complexity-oriented approach is necessary to effectively address soil as the fundamental, complex system that it indubitably is.

Contents

1 Abstract	2
2 Introduction	6
2.1 Soil as a Complex System and Soil Models	6
2.1.1 Soil Structure	8
2.2 Hydrology	9
2.2.1 Historical Context of Hydrological Models	9
2.2.2 Pedotransfer Functions	11
2.2.3 Pore-space architecture and Hydraulic Properties	13
2.3 Earthworms as Ecosystem Engineers	14
2.3.1 Earthworm classification and Ecosystem Impact	14
2.4 Information Theory	15
2.4.1 Complexity and information theory in soil systems	15
2.5 Fisher-Shannon Information plane	16
3 Objectives	19
4 Materials and methods	20
4.1 Soil Samples and Earthworm characterization	20
4.1.1 Sample preparation and experiment description	20
4.2 X-ray Computed Tomography Images	22
4.3 Analysis on the Fisher-Shannon plane	24
4.4 Analysis of the pore space	26
5 Results	28
5.1 Statistical quantities trends with depth	28
5.2 Fisher-Shannon Information analysis	31
5.2.1 Fisher-Shannon Information Plane	31
5.2.2 Fisher-Shannon Complexity	32
5.2.3 Distance from the isocomplexity line	34
5.2.4 Moisture level and earthworm presence discrimination	35
5.3 Correlation between statistical measures and other soil quantities	38
5.4 Statistical measures and pore space characterization	40
6 Discussion	43
6.1 Depth influence on information theory measures	43
6.2 Considerations about the different partitions of the volume	44
6.3 Fisher-Shannon Information Analysis	45
6.3.1 Fisher-Shannon Complexity and Distance from the isocomplexity line	45
6.3.2 Differences between wet and dry conditions	46
6.3.3 Earthworm presence and absence	46
6.3.4 Considerations about Information Theory metrics	47
6.4 Fisher Information Measure and Earthworm Population	47
6.5 Pore Space	48
6.5.1 Minkowsky Functionals	48
6.6 Variability between the samples of the same category	49
6.7 Insights into earthworms' influence on water behavior in soil	50
6.8 Limitations	50
7 Conclusion and future perspectives	52
8 Bibliography	54

2 Introduction

In this section we will present and discuss some topics of soil science and information theory that will result fundamental for the understanding of the work carried out in this thesis project. Here we will describe tools from physics, biology and information theory that will enable us to draw meaningful considerations from the results of our study.

First, we will introduce soil as a complex system, describing briefly the intricate network of physical, chemical and biological factors that shape the spatial and functional organization of soils. We will introduce and explain some of the most recent and complete attempts at modeling soil functionality with a multidisciplinary approach, highlighting the strengths and weaknesses of each method. By doing so, we try to transmit the sense of how complex and complicated is to quantitatively describe such a chaotic and variably composite system. We will focus on the description and influence of soil structural organization on the system as a whole, considering the hypothesis of adopting structural parameters as the central quantities that regulate soil functionality.

We will also introduce the main concepts of soil hydrology, and the principal mathematical tools used to model and describe water movements inside porous structures and soil matrix in particular. First, we will briefly go over the history of hydrological models, we will then focus on the development of pedotransfer functions and specifically on the influence of the pore space characteristics on water behavior in soil. Making use of the tools described in this section, we will try to gain insights about the influence of worm activity on hydraulic conductivity and Water Retention Curves.

We will then give an introduction about earthworms' biology and ecology, and their role as 'ecosystem engineers', with the objective to gain some biological insight about the structure-altering activity of those animals. Since one of the objective of this work is also to quantify in innovative ways the impact of earthworm activity on soil architecture, it is crucial to understand the differentiated impact that different ecotypes of worm have on the system. Our introduction on the topic, therefore, serves as a tool to be able to draw biologically informed considerations from the results of our analysis.

Lastly, we will introduce and explain in detail some concepts of Information Theory, in particular, we will present the Fisher-Shannon method for the analysis of complex signals, giving the historical and theoretical context for the application of this kind of analysis. We will try to clarify the relevance of this kind of analysis in our context and to reinforce the need and opportunity for complexity-oriented analysis of soil systems.

2.1 Soil as a Complex System and Soil Models

The complex nature of soil has been widely recognized from some years now [77], the complicated and interconnected dynamics of biological, chemical and physical processes going on inside the soil matrix result in an intricate system with all the defining characteristic of complexity. This calls for the necessity of a systemic approach, that was already clear, for example, in [83]:

"The impression may be that our scientific knowledge on soil processes and how they produce emergent soil functions is pretty much settled, and it is only insufficient how to translate this knowledge into sustainable management practices. We are convinced that this is a misimpression [...] we stress the fact that our knowledge on soil processes is fragmented throughout various disciplines and the system perspective required to truly capture the reaction of soils to external forcing through land use and climate change is still in its infancy. This systemic approach is furthermore necessary considering the need to distinguish the enormous variety of different soil types in various geographic and climatic regions, all of whose functioning reacts specifically in response to external forcing [...]. Such a systemic approach, providing a clear perspective on how soil functions emerge from small-scale process interactions, is a prerequisite to actually understanding the basic controls and to developing science-based strategies towards sustainable soil management. This will also have an enormous potential for facilitating communication towards stakeholders and policy makers by replacing the cacophony generated by a disciplinarily fragmented research community with harmonized information on the soil system's behavior."

The various actors at work inside the soil are all linked one to another in an intricate way, originating complex, non-linear, chaotic dynamics of the system behavior, such as feedback loops, and emergence [77]. To make one example, microbial activity influences structural and chemical properties, altering nutrient availability that in turn regulates plants growth and the subsequent input of organic matter back in the soil, which is the central factor in determining the rate of microbial activity in the first place. Or again, the burrowing activity of

earthworms and other soil invertebrates alters the configuration of the pore space, and with that the capacity of the soil matrix to retain and distribute air, water and nutrients, deeply influencing the metabolism of microbes and fungi, as well as the rate and spatial configuration of root growth and decay, in much larger areas than the one directly modified by their actions. We could go on and on with countless of this cause-consequence systems, that are of course all interconnected with one another. On top of that, soil structure and functionality are easily influenced by a great number of external factors, such, for example, rain patterns, dry-wet cycles, temperature fluctuations, and also land management practices or anthropological activities [54]. Its remarkable complexity makes the soil a highly adaptable system, capable of self organization and resilience, allowing for recoveries even when the biology and structure of the system are highly degraded from impactful human activities such as intensive farming and grazing or mining [31].

The deeply coupled nature of all this processes brings up a clear necessity for complex models to describe underlying soil dynamics in order to gain some predictive power regarding macro properties of interest in a variety of contexts. Soil models are vastly used in different branches of environmental sciences, such as the diffusion of contaminants, the regulation of ecosystemal services like carbon sequestration and biogeochemical cycles [81], or the valuation of the ecological impact of human activities. Of course they are also crucial in agriculture, to evaluate the effect of different practices on the health and productivity of the system and to assess the viability of a particular crop in a specific area. Another important field of application of soil models is the hydrological one [47], in which structural models are used to simulate water fluxes and infiltration proprieties, assess soil stability and even evaluate impacts of climate change on the water cycle on a global scale[78].

The firsts models of soil date back to the 1960s, they were mostly focused on physical ad chemical processes: water and heat fluxes ([36]), organic carbon [63] and nutrients dynamics [39], leaving aside the complicated biological factors at play. Those firsts models consisted mostly in the analytical or numerical resolutions of partial differential equations (PDE) for dynamics inside a well-defined porous media. In the seventies we saw the rising of some models that were starting to consider some aspects of soil ecosystem dynamics, specifically in the frameworks of food webs and nutrient cycling dynamics[55].

In the following decades, some models also started to incorporate soil structure in the picture, eventually leading to a mathematically consistent formulation of its role in physical processes, like water flow and solute transport, at the beginning of the nineties by Gerke and van Genuchten [28]. The advancement of data-gathering techniques as well as computational power that took place in the recent years allowed for models to try and be more holistic, exploring deeply the connections between soil phenomena and the life in it.

To get a more comprehensive description of soil processes it appears necessary to tackle the problem on all the different temporal and spatial scales that are pertinent to the system, as well as from all the fields of study involved. A recent example that aims in this direction, trying to extend the consideration of biological processes involved in soil models, is Romul_Hum [40, 13, 14], which focuses on the process of humification (creation of humus) mediated by soil fauna, in combination with mineralisation (mainly microbially driven) as mechanisms for the transformation of organic matter in soil, and carbon stabilization. An innovative introduction in this model is the explicit quantification of some sources of organic carbon, which derive from faunal presence and activity, like waste and necromass.

Another recent approach comes from the model KEYLINK [19, 26], that explores more closely the interactions between physical and biological aspects of the system, adopting soil structure as a central parameter for describing carbon cycling and soil behaviors. This model integrates concepts from soil organic matter pools, structural models, and food web models to simulate the living soil at an ecosystem scale. The authors claim that the central factor regulating SOM dynamic is, instead of its intrinsic characteristics as recalcitrance, its physical accessibility, which is determined by aeration, aggregation and, ultimately, the pore-size distribution. The model characterize the soil structure dividing the pore space in 5 different size classes, based on their measurability, accessibility by soil fauna and hydrological relevance. Soil biota is clustered in different functional pools, each able to access only the SOM present in a subset of pore classes, according to the the organisms' dimension and behavioral patterns; their metabolism is also regulated by external factors such as temperature, pH and water content. The flow of carbon is modeled between 13 different pools, comprehending the ones that represent soil microbes, fungi and animals, roots, one for litter and one for SOM. Every functional pool has an effect not only on SOM and the other pools, but also on the pore-size distribution, representing the feedback cycle that exists between soil's biota, structure and functionality. This choice, of adopting soil structure as the central parameter, allows to link different soil processes of physical and biological nature that influence, and are

influenced by, soil structure; and to stress the relevance of SOM accessibility as the most significance parameter for estimating its long-term stabilization. In a future version of the model it will be necessary to incorporate the action of soil macrobiota on the pore space, considering explicitly and quantitatively the impact of engineers, like earthworms, through bioturbation, biopores creation and aggregation processes.

The KEYLINK model offers a significant step forward in soil modeling by emphasizing the role of soil structure and its dynamic interaction with soil biota. It moves beyond the limitations of traditional models by incorporating a mechanistic understanding of soil processes, and by highlighting the importance of soil functional diversity and trophic organization. The model is relatively simple, fast, and easily modifiable, making it a valuable tool for studying soil systems. Although the model needs further validation with more complete datasets that include soil structure, hydrology and soil biota, it is a good step towards a new generation of ecosystem soil models.

2.1.1 Soil Structure

Soil is effectively a highly heterogeneous, three-dimensional, porous system, its structure can be defined as the spatial arrangement of mineral particles, organic material and pore spaces [47]. Recently the term "structure" is being replaced by the term "architecture", to emphasize the strong relationship between the arrangement of soil constituents and the functions that its specific arrangement enables [82]. Soil architecture can also be considered as a complex, heterogeneous, biogeochemical interface that forms the basis for all essential soil functions such as plant growth, water storage and dynamics, nutrient cycling, biome metabolism, decomposition of contaminants, long-term storage of organic matter (i.e. carbon sequestration) and even productivity[60].

There is still considerable disagreement in literature regarding the best strategy to study and model soil architecture. Two fundamentally different approaches have been developed for the study and characterization of soil spatial arrangement: one focuses on the pore space and its network; and the other adopts, as central building blocks of the architecture, the aggregates (i.e. isolated soil fragments). One practical difference between the two methods is that the pore approach recognises the importance of spatial position within a large-scale context of undisturbed soil, while the aggregate approach focuses on the small-scale context of single fragments, disregarding their position inside the soil matrix. The state of soil structure characterization changed drastically starting from the early 1990s, thank to the application of X-ray computed tomography (CT) scans in soil sciences. This technique allows to directly access the 3D structure of undisturbed soil, replacing completely the analysis of 2D thin sections, and allowing to link more directly soil structure and functions, resulting in a big boost of the pore-space approaches. More recently, soil structure has been described in a more holistic way, as an "*heterogeneous organo-mineral soil matrix [...]. The resulting architecture is reflected by the spatial configuration of primary pores between particles, variably cemented by organic molecules and pervaded by a secondary pore network at a higher hierarchical level produced by well-known processes such as root growth, faunal activity, swell–shrink dynamics and freezing–thawing cycles.*"[82].

Other, more topological, approaches made use of fractal geometry to describe the soil pore space [56], providing a way to characterize the scaling properties of pore size distributions and of the surface area of the pore system. An historical overview of the different approaches in modeling and studying soil architecture is presented graphically in figure 1. Here successive approaches to soil modeling are arranged in a tree-like shape, highlighting whether the modeling technique develop from aggregate or pore space approaches and it yields quantitative or qualitative results. Some important breakthrough in soil science are indicated in correspondence of the modeling technique that they enabled.

When trying to address the problem from a more comprehensive point of view, it is important to understand how the fabric of mineral and organic compounds and the pore space change in time. The structure dynamic is of course very complex and it is regulated by a number of physical, biological and chemical processes that spans a huge temporal (from seconds to decades) and spatial (from microns to kilometers) range [47]. Processes-based models focuses on describing the relationship between soil structure and various biological agents (roots, earthworms and other macrofauna, microbial and fungal activities), they often refer to small spatial and temporal scales. Empirical models, on the other hand, make use of simplified descriptions and statistical relations to quantify soil structure changes, thus being more suitable for application at larger scales. Hybrid models try to merge the two approaches to gain a more complete view of this processes.

There are also some Agent-Based models, developed to investigate closely the behavior of single organisms and their impact on soil architecture and function, an example that consider earthworms is the model SWORM

[6]. Here, fractal theory is exploited to create a multi-scale virtual environment made of mineral particles, organic elements and voids (pore space), in which earthworms move according to species-specific dynamical rules. In doing so, they are actively shaping the pore structure of the soil with creation of biopores, organic matter transformation through ingestion and casting of soil portions, compaction of soil elements on the margins of the macropores, and a number of other impacts.

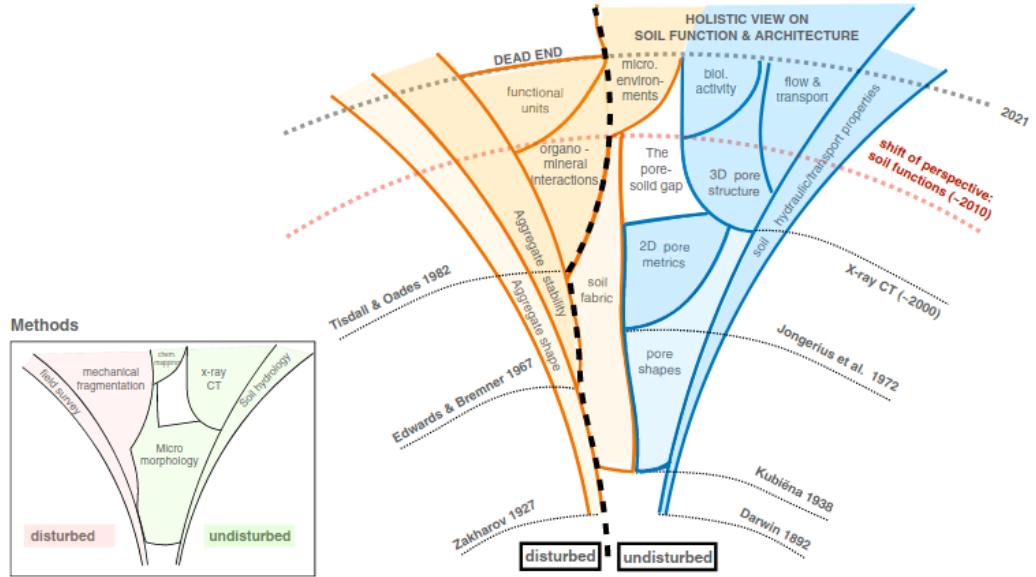


Figure 1: Historical development of exploration of soil structure. In blue the pore approach and in ochre the aggregate approach, dark and light colours indicate quantitative and qualitative analysis, respectively. Figure from [82]

2.2 Hydrology

2.2.1 Historical Context of Hydrological Models

Describing water behavior inside soil is an awfully complicated task. As mentioned before, the soil matrix is incredibly complex to characterize due to its intricate and multi-scale pore space, to the heterogeneity of the elements that comprise the medium and to their different interaction with water. As this was not complicated enough, it should also be taken into account the differentiated ability of the biome to maintain moisture levels in their vicinity, thanks to the production of biofilms, mucus and other strategies. Even though very challenging, predicting spatiotemporal variations in water fluxes and soil moisture content is crucial for a plethora of disciplines, such as hydrology, ecology, agriculture, geology, climate and soil science. Also, this factor affects almost every process in Earth's critical zone, from soil biogeochemistry to heat flow, from nutrient transport to plant and animal behavior.

For the importance and challenging character of the subject, the study of hydrology is one that has a long history [85]. The Richard Richardson Equation (RRE), fruit of the work of Richardson in 1922 and Richard in 1931 [59], describes the movement of water in a porous medium, it is a quasi-linear partial differential equation. It is based on the Darcy-Buckingham law [62], which is often used to describe water fluxes in soil and plants and have the form:

$$\vec{q} = -K(\Theta)\nabla(h) \quad (1)$$

that represents the volumetric flux \vec{q} in function of the unsaturated hydraulic conductivity of the medium K , function of the volumetric water content Θ , and the head potential h , which is defined as the energy per unit weight of pore water (i.e., divided by weight of water) and is the sum of the pressure head ψ and the head gradient ∇z , which is the geodetic head gradient, assumed $\nabla z = (001)^T$ for three dimensional systems. Considering an incompressible porous medium and an uniform fluid, the law of mass conservation yields:

$$\frac{\partial\Theta}{\partial t} + \nabla \cdot \vec{q} + S = 0 \quad (2)$$

where S is the sink term (typically root water intake). Combining the two equations we find the mixed-form RRE:

$$\frac{\partial \Theta}{\partial t} = \nabla \cdot (K(\Theta) \nabla(h)) - S \quad (3)$$

This formulation involve two unknown variables Θ and h , that are in reality linked by a relation $\Theta(h)$, known as the Water Retention Curve (WRC), this allows us to formulate the RRE in the head-based form, applying the chain rule on temporal derivative:

$$C(h) \frac{\partial h}{\partial t} = \nabla \cdot (K(\psi) \nabla(h)) - S \quad (4)$$

with $C(h) = \delta\Theta/\delta\psi$ known as the retention water capacity of the system. For the RRE to be applied, though, we need to establish a relationship between water content, water conductivity and head pressure, i.e. find the functional form of $\Theta(\psi)$ and $K(\psi)$. The first attempts in this sense were along two different lines, either based on a generalization of Kozeny's approach for saturated and unsaturated porous media, in which relative hydraulic conductivity K_r is a power function of the effective saturation; or making use of measured capillary head-water content curve to derive unsaturated conductivity.

In 1976, Mualem [49] proposed a new approach improving upon the latter category of models, taking into account not only the moisture content vs. capillary head pressure curve but also the measured values of hydraulic conductivity at saturation. This model is based on an approximate evaluation of the hydraulic conductivity of a pore domain with varying shape, and the expression for the relative hydraulic conductivity in function of the water content $K_r(\Theta)$ is derived in a simple integral form. This allowed Mualem to propose a model that was able to predict unsaturated hydraulic conductivities in soils better than the precedent ones, improvement on this approach were made in the eighties by van Genuchten, whom combined Mualem's pore-scale model with other models for water retention. The van Genuchten model (VGM) is still arguably the most popular method to retrieve information about Soil Hydraulic Properties (SHP) [85], and it is based on the concept of *capillary bundle*. In this view, the soil pore matrix is represented by a "bundle" of vertical parallel pores of different sizes (capillaries are interconnected to pairs in the model for Hydraulic Conductivity Curve (HCC)), and WRC is given by

$$\Theta(\psi) = \Theta_r + (\Theta_s - \Theta_r)S_e(\psi) \quad (5)$$

where $\Theta(\psi)$ represents the total water content in function of the head pressure, Θ_s the saturated water content and Θ_r represents the "residual" or "irreducible" water content (water present in the system at theoretically infinite pressure). Lastly, S_e is called "effective saturation", and it is a sigmoidal function of the pressure head ψ :

$$S_e(\psi) = [1 + (\alpha|\psi|)^n]^{-m} \quad (6)$$

with α inversely correlated with the air entry value of soil and n and m as dimensionless shape parameters related to the pore size distribution. This expression reflects a smooth unimodal pore size distribution, which is typical of well sorted materials but does not typically apply to real soils, which tend to have multi modal distributions. Usually this expression is considered constraining $m = 1 - 1/n$, yielding the expression for the unsaturated conductivity function $K(\psi)$ as:

$$K(\psi) = K_s K_r(\psi) = K_s S_e^\tau \left[1 - \left(1 - S_e^{1/m} \right)^m \right]^2 \quad (7)$$

The saturated regime is obtained with $\psi = 0$ and the conductivity in this case is represented by the term K_s , while $K_r(\psi)$ is the relative conductivity function that assume values between 0 and 1. In saturated conditions, we have $K_r(\psi = 0) = 1$ and, ideally, the water content is the same as the total porosity: $\Theta(\psi = 0) = \Theta_s \simeq \phi$. According to Mualem, the parameter τ may be positive or negative and accounts for the connection between pores and the flow path tortuosity, its value is determined by regression and Mualem find it to be $\tau = 0.5$, this value is used in the predominant cases but is very sensible to changes in soil structure and therefore should not be used lightly.

The VGM model has become widely used because of its flexibility in describing WRC data, its continuous differentiability over a full pressure-head range and because it does not require any measurement of the unsaturated Hydraulic Conductivity Curve (HCC); these characteristics made so that the model was included in a great

number of soil process modeling tools. Even though this widely employed model remains the norm, it comes with some important shortcomings: first of all the assumption of an unimodal pore size distribution, which is very limiting and not viable for most real soils. In fact, in the presence of different structural elements, such as aggregates, two different pore spaces can be identified: intra- and inter-aggregate pore space [52], leading to a bimodal distribution of the pore radii. In peat soils (usually uppermost soil sections, rich in organic matter, typically above 75%, characterized by and intense decomposition activity and sponge-like texture) the pore size distribution is often multi-modal due to plant structures, faunal activity and decomposition effects.

Neglecting multi-modality may have a significant impact, especially on the HCC, as the conductivity drops drastically with the emptying of the macropore's network; for this reason a number of methods have been developed to consider only conductivities measured at lower pressure heads, so as to refer only the soil matrix, disregarding the macropores impact. This kind of approaches are needed to adequately describe experimental data of water retention and hydraulic conductivity curves. It is still not clear whether an effective description of soil hydraulic properties can be achieved with a unimodal RRE or even by coupling a few of those, in order to represent multi-modal distributions. In fact, the presence of a network of large pores allows for turbulence to emerge, violating the hypothesis of laminar flow implied in the RRE.

Another criticality of the model lies in the capillary bundle approximation, which fail to describe water content and hydraulic conductivity at low pressures, in particular, the term Θ_r has little physical significance, since the water content approaches 0 in dry soil. Additionally it is well known that the WRC is not a single monotonic curve due to capillary hysteresis (capillary hysteresis refers to the non-uniqueness of the WRC and its dependence of the history of soil wetting and drying) resulting from pore-scale processes, mainly due to the irregular shape of pores, hysteresis of contact angles between water and solid soil particles and shrinking or swelling effects. Although hysteresis is widely recognized as a key process in soil water dynamic it is rarely accounted for in models due to the difficulty of both theoretically estimate and experimentally characterize this phenomena. The study of hysteresis is also influenced by dynamic non-equilibrium effects, that induce an apparent flow-rate dependence of the WRC under transient conditions. More precisely, when in transient conditions, the water phase equilibrate with the pressure head not immediately but as a long-term limit, after a considerable equilibration time. Similarly to hysteresis, this process is also induced by pore-scale properties, mainly related to pore space geometry and topology; rendering the effects of the two phenomena hard to tell apart. Accounting for hysteresis and dynamic non-equilibrium processes may improve significantly the translation from laboratory experiments to field-scale hydraulic properties, especially on short timescales (hours to days), while probably not being so impactful for much larger time scales (years to decade), where other factors, as land management and tillage practices, would play a more important role.

Another approach is the one adopted by Kosugi [41], that developed a combined soil water retention-hydraulic conductivity model starting from a unimodal log-normal pore size distribution and based on the generalized pore-scale model of Mualem. The basic idea behind this method is to relate pressure head ψ to pore radius r by the 'capillary pressure function': $\psi = A/r$, where A , in the case of air-water-soil systems, assumes the value $A = 0.149\text{cm}^2$. The log-normal pore radius distribution function reads:

$$\frac{\partial \Theta}{\partial r} = g(r) = \frac{\Theta_s - \Theta_r}{\sqrt{2\pi}\sigma r} \exp\left(-\frac{[\ln(r/r_m)]^2}{2\sigma^2}\right) \quad (8)$$

where r_m is the geometric mean pore radius and σ is the standard deviation of the log-normal distribution, related to the width of the distribution itself. In this view, all parameters of the hydraulic equation are determined by some quantities related to the pore size distribution.

2.2.2 Pedotransfer Functions

In spite of the tremendous advancement in the developing of measurement techniques, direct determination of SHP is still expensive, time consuming and overall impractical, especially for large-scale systems. To address this problem, in the last thirty years, a huge number of what are called Pedotransfer Functions (PTF) have been developed, with the objective to derive the parameters of hydraulic models, for example the van Genuchten-Mualem model, from easy to measure and readily available soil properties, called in this context "predictors". Most efforts have been focusing in the development of parametric PTF, in order to estimate the parameters over a range of pressure heads and obtain also a mathematical function for HCC to use in models. The classic PTF

predictors are sand, silt and clay content (or, more generally, some indication of the particle size distribution), bulk density (defined as the solid mass over the total volume of the sample) and organic carbon content (SOM); but extensive research has been carried out to identify additional soil specifics that could inform about PTFs. Additional information could include different indicators for soil structure (for example regarding the pore space), morphological and biological information about the system or also data about the water content at specific head pressures, so that to integrate the model with some experimental insight.

The two main categories of methods used to derive PTF are: statistical regression techniques (linear or not) in which a given parameter of the hydraulic model is given as a function of the predictors as in $\text{parameter} = F(\text{predictors})$, for example, for a linear model:

$$\text{Parameter} = a + b(\% \text{sand}) + c(\% \text{silt}) + d(\% \text{clay}) + e(\% \text{SOM}) + f(\text{bulkdensity}) \quad (9)$$

and, more recently, machine learning techniques such as neural networks, Vector Support Machines, random forests; with the latter approaches yielding better results due to their generality, as they don't need any a priori assumption about the functional form of the PTFs. Despite these efforts, PTFs are still a major cause of error in soil models, for a number of reasons, mainly related to the dataset used to construct the model. In fact, often, the predictors used to construct the PTF are not determined directly from the soil sample for which hydraulic properties are then analyzed, somehow neglecting the great variability present in soil structure even at small scales. Soils are usually divided in classes according to their aggregate composition, but this division is somewhat arbitrary and different soils belonging to the same class could have very different hydraulic properties due to other physical, geological, chemical and even biological factors. Furthermore, the databases, in most cases, contain only standardized information about the soil and lack other relevant aspects related to structural properties. Another critical point is the widespread use of empirical unimodal equations to relate the measured moisture retention curves and water conductivity curves (as in the case in VGM), which prevent the developing of more elaborated and accurate models in this sense. The last, and maybe most impactful [79], factor limiting the effectiveness of PTFs is the lack of accuracy and consistency in the measurement of hydraulic properties, since precise measurement of WRC and HCC remains very challenging at the extremes of the pressure head spectrum.

Another aspect that complicates the job to find the correct values for hydraulic parameters is that those often are non-unique [24], i.e. it exists more than one set of parameters that could lead to an acceptable fit of the WRC and HCC with experimental data. In order to obtain a unique set of hydraulic parameters, it would be necessary to have the complete data about both WRC and hydraulic conductivity $K(\Theta)$ across the whole range of water content Θ , from the oven-dry to the completely saturated state; but, as mentioned before, it is very rare to have such rich data sets. This brings up a series of challenges:

- Linked parameters: some of the hydraulic parameters are correlated, worsening the problem of non-uniqueness. This problem can be avoided by finding a mathematical formulation that features only independent parameters.
- Non-uniqueness solution: there are different sets of optimal hydraulic parameters, some of them can assume physically unacceptable values; this is a big limitation for physically based models.
- Optimal non-physical hydraulic parameters: The optimal combination of parameters to fit the experimental data could yield some absurd values, impossible to interpret in a real world context, rendering the model not feasible, especially if used in combination with other models.

In 2021, Fernandez-Galvez et al. [24] made an attempt to address this problems for a bimodal soil Kosugi hydraulic function by restraining the set of possible parameters within physically meaningful values. The authors choose to consider bimodal structured soil (dual porosity) because those represent the majority of real soil systems, considering here a bimodal WRC; the unimodal approach fails to describe the dynamics of two-stage drainage, or bimodal hydraulic behavior, induced by the presence of two different pore classes: fast flow is mediated by the macropore network at high pressures; matrix flow takes place inside the soil matrix, in the micropore network and persists at lower pressures. This relaxes the Maulem's bundle hypothesis, allowing for a partially independent evolution of the flows in the different domains. In order to use this description, we need to establish a pressure head threshold h_{MacMat} to consider only fast flow for $h \leq h_{MacMat}$ and only matrix flow in the $h > h_{MacMat}$ regime (here, pressure head values are taken positive for unsaturated soils). It follows that

also the WRC and hydraulic conductivity expressions should be split in two to get $\Theta(h) = \Theta_{Mat}(h) + \Theta_{Mac}(h)$ and $K(\Theta) = K_{Mat}(\Theta) + K_{Mac}(\Theta)$.

Of course dealing with a bimodal system implies that we will need two set of parameters, one for the macropore and one for the matrix regime, effectively doubling their number, exacerbating the problems mentioned above. The authors' approach consists in reducing the non-uniqueness of the optimized Kosugi hydraulic parameters by deriving the relationships between them from principles of soil physics, and then apply those to the bimodal WRC and conductivity functions.

In particular, the authors derived a relationship linking the residual water content Θ_r (that represent the water left in the system at "infinite" pressure) to the standard deviation of the log normal pore size distribution inside soil matrix (σ). This link between the two quantities is supported by empirical evidence, and the parameters involved in the functional relationship are also derived from experiments; this allows not only to reduce the number of parameters, but also to enforce a physical meaning upon σ , reducing its non-uniqueness.

Additionally, the group was able to relate the values of the mean h_{mMac} and standard deviation σ_{Mac} of the log-normal macropore size distribution to the threshold value h_{MacMat} that is assumed fixed. This allows to have only one parameter to be found concerning the fast flow: $\Theta_{sMacMat}$, that is the water content for which the soil matrix is completely saturated and the macropore space can start to be filled. Lastly, in this paper, the positive linear correlation between $\ln(h_m)$ and σ is exploited too, in order to reduce the non-uniqueness of hydraulic parameters. Specifically, a feasible range for h_m is derived from σ and the value of h_{MacMat} . At the end of this process the authors were able to construct bimodal Kosugi functions for water retention and hydraulic conductivity with only four parameters related to the pore size distribution instead of the starting eight: h_m ; σ ; K_s and $\Theta_{sMacMat}$.

2.2.3 Pore-space architecture and Hydraulic Properties

In this section we will consider the work that has been done by various scientists to unveil relationships between the soil pore space architecture and soil physical and hydraulic properties. As we mentioned before, the soil pore space regulates a huge number of processes, so much so that some models, like KEYLINK [19, 26], adopt pore-size distribution as the central parameter to describe all kinds of soil processes. We will focus here in particular on the relationships that have been found between pore geometry and hydraulic properties. When talking about soil architecture it is the norm to refer to data derived from X-ray computed tomography (CT) images, that allows to characterize the structure of an undisturbed soil up to really small scales. Of course the resolution at which the images are taken varies greatly, with a pixel dimension ranging from few micrometers to few millimeters, influencing dramatically the kind of considerations that can be made.

Some studies [42] focuses on the macropores, characterizing the preferential flow of nutrients on the basis of the geometry of their network, finding that an increased macroporosity leads to a less preferential transport, due to a higher connectivity of the bigger macropore network, and an increased diffusive flux through the soil matrix pore system. Other works, as the one of Smet et al. [71], aim at a more comprehensive description of hydraulic properties starting from a finer determination of porosity properties. In particular, this study employed Bayesian statistical methods to find a positive correlation of the saturated hydraulic conductivity K_s with the global connectivity of the micropore space, its fractal dimension (somehow a measure of the network complexity) and its anisotropy level (parameter that ranges from 0 to 1, where 0 is the value of a perfectly isotropic system). This suggests that soil ability to conduct water in saturated conditions is improved by inter-connectivity and complexity of the network. Additionally, the air permeability is positively correlated with both total porosity and number of connections, but even more strongly correlated with the average volume of the smallest pores, suggesting that this class of pores carry out a crucial role in air transport in soil. Furthermore, the authors found that the use of pore size distributions obtained from μCT images in addition to macroscopic data as input of the hydraulic models, increases the accuracy of the fit of the WRCs, especially near saturation, stressing the relevance of an accurate pore space description to obtain attendable predictions for water behavior. Another interesting aspect emerged in the study is a link between smallest and biggest pore volumes, suggesting an interdependency between different pore classes and pointing out the necessity to consider the soil system in its full multi-scale nature to properly represent water fluxes inside soils.

2.3 Earthworms as Ecosystem Engineers

2.3.1 Earthworm classification and Ecosystem Impact

Earthworms are ancient organisms that belong to the phylum Annelida, and the class Oligochaeta, there are approximately 8000 different species all around the globe and they play a crucial role in soil systems, reaching values between 40 and 90% of soil macrofauna biomass in many ecosystems [68, 7]. Their relevance in soil systems was already clear at the time of Darwin[1], the British naturalist studied some behaviors of earthworms, especially concerning earth movements between different soil layers and buring activity. Another important role of earthworms already studied by Darwin is the production of vegetable mould or humus, fundamental step in the transformation of organic matter inside the soil. Earthworms are usually divided according what is called Bouché's triangle, that classify them in thee different categories based on their typical behavior:

- **Epigenic:** they live on the soil surfaces, forming no permanent burrows and rarely ingest and transform soil, since they are more a surface species than a soil organism, their impact on soil function and structure is mostly limited to the transformation of organic matter above soil.
- **Endogeic:** geofagic species that live in the upper organo-mineral layers of soil and build mainly horizontal mazes of burrows, ingesting a lot of soil and producing casts with which the burrows are partially refilled. Their main impact concerns the creation of a macropore matrix that regulate the horizontal flow of water and air in the soil and the transformation of SOM into casts, that have different chemical and physical properties than the surrounding medium.
- **Anecic:** they feed from both soil and surface litter, carrying the latter down their long, vertical, thick burrows, and leaving their casts on the surface, they actively act as a layer mixer, integrating fresh organic matter in the deeper soil. Their semi-permanent burrows are a fast way of transporting water and air from the surface into the macropore matrix that they build for protection in the deeper layers, thus influencing greatly both physical and biological properties of the whole soil system.

Earthworm activity is crucially important, their ability to process a huge quantity of soil impacts physical, chemical and biological processes crucial for the equilibrium of the soil system as a whole; for this reason earthworms are often referred to as *Ecosystem engineers* [66]. The concept of ecosystem engineering has emerged a few decades ago to refer to the modulation of the availability of resources by the action of an organism. Engineers often create biogenic structures that may serve as habitats for different species and have a broader impact on the near environment, examples are beavers' dams or earthworms' burrows and aggregates.

Earthworms also increase the sequestration rate of organic carbon, but not only by changing its chemical nature during ingestion [4], it appears that the biggest impact is due to SOM physical sequestration, in the form of casts or stable microaggregates, protected by a coating of mucilaginous secretion that also increase aggregate formation and stability. The influence on soil structure is thus not limited to the creation of burrows: being major bioturbators in terrestrial ecosystems, earthworms largely contribute to the formation of stable microaggregates within macroaggregates, leading to a crumb soil structure that helps with the protection of organic carbon and the physical stabilization of organic matter [4].

Because of their transformative impact on soil structure, earthworms were thought to be the perfect tool in the regeneration of compacted soil, but recent studies [66] indicates that earthworm activity alone does not actually decrease bulk density of the soil matrix, it might even increase it around the burrows [11]. In fact, different species of earthworms in different condition will have a widely differentiated impact on compacted soil structure recovery. In some conditions, the introduction of earthworm could even trigger a process of compaction, since the earth moved out of the way to create the burrow system is pressed on the sides, destroying its porous structure, although it appears than micropores smaller than $150 \mu\text{m}$ are not affected by this activity [66]. This effect can be accentuated by the refilling of the burrows with casts, typical behavior for endogeic species, that usually are much more compact than surrounding soil.

We can than say that earthworm are able to regenerate the macropore network of compacted soil but they have no beneficial influence on the fine structural porosity; in more comprehensive experiments, though, was observed that, even in conditions in which earthworm action alone induced compaction, when in presence of leek roots and mycorrhizae, the burrowing activity did not cause structural porosity reductions [48].

It is clear at this point that earthworms are key species to unveil soil processes and dynamics. For this reason, a number of models have been developed to investigate various aspects of this animal's behavior and

impact in soil systems. Some efforts have been made to characterize population dynamics and spatial distributions of these organisms at field scale [5], taking into consideration earthworms' behavior, feeding ecology [17] and some environmental factors that modulate those, such as temperature, pH, humidity, interaction with microorganisms and also presence of contaminants, as earthworms are very sensitive and are regarded as a good indicator of soil health, as they tend to avoid polluted areas.

We will focus here on some models regarding earthworms-induced modifications of soil structure, since we are assuming this as the central action that determines this animal's impact on the ecosystem. This kind of models aim to describe alterations in some soil structure characteristics such as porosity, aggregation, macropore formation and overall soil density, as induced by earthworms' burrowing, casting and bioturbation activity.

The already cited SWORMS [6] (Soil Worm Model) is a Multi Agent based model that simulates the behavior of endogeic earthworms, considering the creation of biogenic structures as biopores and casts, taking into account effects of compaction and decompaction triggered by worm's activity. The model is very effective in simulating soil structure alterations due to the fractal approach in describing soil architecture, and can correctly reproduce compaction and decompaction effects of geofacile endogeic species. Although the model is still too simplistic, to the point that the mesocosm is considered as a close system, with no exchange of water, air or organic matter with the environment; it is a step forward in the direction of connecting physical and biological processes in soil with a more holistic approach.

Other models [22] focuses on simulating burrow systems of specific earthworms' species starting mainly from experimental data about the burrows' variegated topology, and insight about earthworm's behavior in a animal-oriented model. This effort was carried out as an attempt to link properties of the macro pore space, like inter connectivity, length, surface area or connectivity to the permeability of the soil and individuated connectivity (number of independent paths connecting surface and bottom of the test tube) as the most relevant parameter of the macropore network to explain permeability properties of the system.

X-ray microtomography has become a fundamental tool in the creation and validation of this structure-focused soil model, as it provides a way to see inside the soil without disrupting its delicate architecture, opening to the possibility of investigating strictly 3D properties as connectivity and continuity. New algorithms have been developed to identify also other (not burrows) bioturbated areas of soil [10] derived from compaction activity, cast production and dispersion, allowing to determine also the fraction of burrows that are refilled with casts, as an important factor to consider in modeling the burrow network.

2.4 Information Theory

2.4.1 Complexity and information theory in soil systems

The advancements in imaging techniques, X-ray Computed Tomography above all, helped tackling the problem of the opacity of soil mediums, allowing for the investigation of the system structural characteristics without disrupting the sample's integrity. The output of this kind of measures is a 3D grid of voxels (tridimensional pixels), each one with an associated "intensity" value. This figure is a proxy for the sample's density in that voxel's location, and, theoretically, it could be enough to discriminate between various types of solid soil particles and air or water-filled pore space. The access to information about the tridimensional spatial heterogeneity of soil allowed for the development of a number of approaches to qualify and process the information contained inside the soil scans. To this scope, various methods have been employed from the fields of statistical physics and information theory, in order to quantify the complexity and organization of soil architecture. Some theories focused on the analysis of the pore space, revealing its extraordinary complexity and modeling it with the use of multifractals and complex networks.

The multifractal approach characterize the pore space as, like the name suggests, geometric multifractal, which is a fractal that exhibits local fluctuations in density, and cannot, therefore, be characterized by a single dimension, but requires a spectrum of generalized fractal dimensions. Multifractal analysis of 3D images involves partitioning the space into non-overlapping cubes of multiple scales to construct samples with. Multifractals have been used both for the study of the properties of the macropore network[65], for example their dependency on depth or compression state; and for the construction of simulated soils and pore spaces[74].

Complex networks approaches also tackled this problems, but some efforts here have also been carried on in quantifying soil complexity, for example in the work of Samec et al. [64]. Here the network complexity is calculated as an entropy measure of the quantity $b_{\Delta k}$, which is the probability that a randomly chosen edge

links two nodes with degree difference Δk :

$$b_{\Delta k} = \frac{\text{Nuber of edges connecting nodes with degree difference } \Delta k}{\text{Total number of edges}} \quad \text{and entropy: } h = \sum_{\Delta k=0}^{k_{max}-1} b_{\Delta k} \log b_{\Delta k} \quad (10)$$

It was found that this measure of entropy is sensitive to differences in soil pore structure, in particular the value of h increase with the scaling exponent α of the pore size distribution expressed as $W(s) \propto s^{-\alpha}$.

The analysis of the pore space requires binary images, so to distinguish clearly the solid and fluid (water or air filled) phases, and focus on studying the properties of the latter. But until now no consensus has been reached regarding the appropriate threshold to choose in order to optimize the discrimination between the different spaces. In fact, several studies demonstrated that pore network properties are highly dependent on the threshold choice. In a work of Tarquis et al. [73] the authors confronted results obtained with four different thresholds assumed in various points of the probability distributions of solid and void intensities values. In particular, the group had a multifractal approach and calculated the generalized fractal dimensions D_q with $q = 0, 1, 2$ of the soil system for different thresholds. These generalized dimensions are called respectively *capacity*, *entropy* and *correlation* dimensions and represent the mass fractal dimension, the system's entropy and the correlations of the measures contained in boxes of various sizes. The difference of two generalized dimension is called $w = D_1 - D_2$ and is an indicator of the degree of multifractality of the system. In particular, low values of w indicate a low level of heterogeneity and complexity in the data, and vice versa. The authors found that increasing the value of the threshold, and consequently the porosity, reduce the range of generalized dimensions, effectively resulting in a reduced complexity of the pore space.

The strong dependence of the Multifractal analysis results upon the threshold choice suggests that this value should be carefully chosen, not only based on the formal analysis of the greyscale images but also on the patterns found in the images. It should also be noted that, no matter how the threshold is chosen, the process of binarization inevitably comes with the total loss of information about the densities of the voxels composing the solid portion of soil.

For these reasons, it has been suggested that the greyscale images should be used directly to quantify the information content of soil systems rather than binarised images obtained with arbitrary thresholding. Some research has been done in this sense in the context of multifractal analysis by Torre et al. in the last years. In a first work [75] the authors confronted multifractal analysis of 2D thin slices and the 3D whole volume of soil, finding that even when the descriptive statistics of gray values did not differ significantly between 2D and 3D images, the discrepancy in the MFA parameters was relevant. Furthermore, the values from the multifractal analysis of the slices were different depending on the slice and its direction, suggesting that a 3D sample is a better choice for this kind of study. Every sample studied exhibited a multifractal nature, but the 3D image showed a higher complexity in structure, supporting the hypothesis that this format is better to investigate soil complex nature. It has also been observed that lower gray values distribution has a strong influence in the scaling behavior. In another article of the same group applied similar analysis in order to discriminate between soils under a different tillage regimes; considering also the influence of image resolution and subdivision method in order to find the best procedure for the scope. The comparison of multifractal spectra allowed to differentiate clearly between the tillage method that creates a higher complexity in soil aggregates, by physically removing soil and the one that tends to destroy aggregates, homogenizing soil structure and weakening the fractal nature of soils. These results suggest indeed that the Multifractal approach can be applied directly to greyscale 3D images of soil in order to quantify its complexity and fractal characteristics.

2.5 Fisher-Shannon Information plane

In this work of thesis, we will follow the approach of Aguiar et al. [3] from a recent article. Here the authors decided to employ, for the first time, the Fisher-Shannon method to jointly quantify local and global properties of the probability density function of the intensity values of the 3D X-ray CT of soil samples. The input images, in these case, are cubes with sides 790 pixels long, each pixel has a resolution of $45\mu\text{m}$ and each voxel is characterized by an intensity value that reflects the density of the soil in that $45\mu\text{m}^3$ volume. In particular, the group calculated Shannon Entropy Power (SEP) defined as:

$$SEP = \frac{e^{2H_x}}{2\pi e} \quad \text{where} \quad H_x = - \int_{-\infty}^{+\infty} f(x) \log f(x) dx \quad (11)$$

and Fisher Information Measure (FIM), defined as the functional:

$$FIM = - \int_{-\infty}^{+\infty} \left(\frac{df(x)}{dx} \right)^2 \frac{1}{f(x)} dx \quad (12)$$

to quantify, respectively, the disorder and structural organization of the signal's variation, the signal $f(x)$ being here the probability distribution function (pdf) of the greylevels of the voxels.

Shannon (or Boltzmann) entropy has been for a long time considered the major tool to characterize and describe the informational behavior and complexity of physical systems; as it gives a global indication of the disorder present in a signal or system.

Fisher Information Measure is a statistical estimator introduced by Fisher in a work of 1925 [25], the British mathematician was here studying ways to make estimates more efficient, i.e. finding statistics which embody a large portion of the relevant information contained in the data. During the end of the last century various studies arose in theoretical physics regarding this measure, that have proven to be a very versatile tool to describe evolution laws of a number of physical systems. In particular, in a series of articles in the 1990s, Frieden and colleagues [27] demonstrated that minimizing the FIM functional of the $pdf(x)$, subjected to a physical constrain which is linear in the mean kinetic energy of the system, led to a temporal equilibrium solution which obeys the differential equation that regulate the dynamic of the system. In this context, Frieden was able to derive the Schrödinger (energy) wave equation, Klein-Gordon equation, Helmholtz wave equation, diffusion equation, Boltzmann law, and Maxwell-Boltzmann law from this one *classical* principle of disorder. These studies culminated in the formulation of the variational principle of extreme physical information (EPI), which determines simultaneously the Lagrangian and the physical ingredients of the concomitant scenario. As Plastino elegantly put it in a successive work [57], Frieden and colleagues were able to show "*that the Lagrangians of physics arise out of a mathematical game between an intelligent observer and Nature*". The convergence rate for the FIM is comparable to the one of maximum entropy calculations, suggesting that the FIM, like entropy, also defines a direction of time. The arrow in this case points towards the direction of decreasing accuracy for the determination of the mean value of a parameter. Soon more results in theoretical physics applications of FIM were found, for example, Plastino et al. [57] addressed the case of a General Liouville Equation, which does not have "intrinsic" irreversibility of time, but, at a coarse-grained description, exhibits a tendency to the loss of information in time. The group were able to demonstrate an H-theorem for the General Liouville Equation and provided a more detailed characterization of the FIM arrow of time in several processes and of this functional's relations with coarse graining processes. Those results led to the application of FIM to study a variety of complex systems, from ECG signals in order to predict epileptic episodes and irregular brain activity [45], to the numerical study of the electron distribution of a two-dimensional hydrogen atom [21]; from analysis of time series in geophysical systems, to application in ecology and astrophysics. FIM was also used to characterize winds and rainfall patterns [33], ocean temperature fluctuations as well as people behavior in crowds [44], proving to be a very powerful and multidisciplinary tool in the study of complex systems.

In this optic it makes sense therefore to start to explore the applicability of this concept also in soil science, especially in the characterization and description of soil structure, which is a perfect example of complex physical system. In particular we will be interested here in the signal analysis in the Fisher-Shannon information plane. Vignat et al. [80] found that both FIM and SEP may be required to characterize non-stationary behavior of a complex signal, in fact, both scaling and uncertainty properties suggest that SEP and FIM are intrinsically connected, so that the characterization of a signal should be improved when considering its position in the Fisher-Shannon information plane.

The product of SEP and FIM, called Fisher-Shannon complexity (FSC), was also calculated as a statistical indications of the complexity of the system. This quantity satisfies the *isoperimetric inequality* ($FSC \geq 1$) [20], where the identity is found for the Normal distribution. The analysis of these quantities were then carried on in the already cited Fisher-Shannon information plane (FS), which has FIM values on the horizontal axes and SEPs on the vertical one. Here, every sample is assigned to the corresponding point with coordinates (SEP, FIM) and the distance from the isocomplexity line (curve where $SEP \cdot FIM = 1$) is considered as a measure

of complexity. To calculate this distance effectively, a normalization is in order, in fact, the values of SEP are in the order of $10^5 - 10^6$ and FIM values are of the order $10^{-5} - 10^{-6}$, resulting in FSC values in the order of the unity. The disproportion in the measures' values is such that the distance from the isocomplexity line would always be basically equal to the difference in the dimension of FIM. To avoid this bias, the authors normalized the values of SEP and FIM dividing the former and multiplying the latter for the highest value of SEP for all samples, so to get values of FIM and SEP of the same order and not influence FSC values. The group calculated those quantities for every images constructing the pdfs in three different ways: considering all the voxels, considering each plane separately and averaging or considering all the vertical columns one by one and then averaging. This last approach was the one that gave the best results.

The researches were here able with this method to discriminate between soils from a sugarcane plantation and from a nearby Atlantic forest, finding that the latter are associated with a higher level of complexity. In particular, the values for the SEP are higher for the forest soil, indicating a higher degree of chaos in natural soils in respect to heavily managed ones; while the opposite is true for the FIM measure. The distance from the isocomplexity line resulted consistently higher for the soil of the Atlantic forest, confirming that a soil systems in natural regimes exhibit a higher structural complexity than the ones subjected to intensive agricultural practices. The authors also suggest that this completely model-independent measure could be used to get an overall indication of the health of soil, but to confirm this hypothesis more experiments needs to be done. Fitting in this line of research, the present work will try to establish if the same method can be used to discriminate between soils with and without earthworms, as the presence and abundance of those animals is often used as a qualitative measure of the soil's health state.

3 Objectives

The main goal of this work of thesis is to investigate the opportunity of using statistical tools from the field of information theory in the study of soil as a complex system. Specifically, we want to quantify the impact of moisture level and earthworm species-specific activity on soil architecture using complexity metrics derived from Fisher-Shannon information analysis of 3D X-ray CT scans. We will quantitatively assess the degree of disorder, structural organization and complexity of soil architecture in different environmental and biological conditions, in order to assess if these quantities are useful to achieve a classification of the samples in the various conditions.

Since this kind of analysis is almost a novelty in this field of research, and was attempted only once before for soil systems [3], we will have an exploratory approach in trying to understand how to best use these tools in our specific context and to answer our research questions. To this scope, the volumes object of our study will be partitioned in different ways, so to get a series of pdfs for various 'geometrical' elements of the parallelepiped: columns, planes, walls and the whole volume. Understanding how different partitions yield different results could help to have insight about the statistical metrics correlation with different factors that shape soil architecture as, for example, depth.

Furthermore, we will explore the correlations of the Fisher-Shannon metrics with standard information about soil physical, chemical and biological conditions that are typically considered in soil science, as, for example, SOM content, pH, Aggregate stability, Carbon and Nitrogen content, total final worm mass and others. The rational behind this research question is that soil is an interconnected system, where every element influence and is influenced by all the others, we hope to add foundation to the hypothesis that soil architecture is a sensible choice as the central parameter to model all soil processes.

We will also investigate how the complexity metrics correlate with the topological properties of the pore space, to assess the impact of porous structures on this analysis. Our objective is here also to consider whether the statistical analysis of greyscale 3D images can provide a qualitative alternative to the controversial though widespread process of binarization, base for the study of the pore space.

4 Materials and methods

In this chapter the methodology and materials used in the present study will be described and contextualized. We will start by describing the experiment from which the data object of our analysis were obtained, in order to allow reproducibility of our result and gain more context for the discussions about the findings. We will then discuss the manipulations of the X-ray CT data that was performed in our analysis. The focus here is on the methodology chosen to process the images using both specialized software (ImageJ, FIJI) and Python scripts written for this specific scope. We will also discuss in detail the choices made along the process, justifying them and adding context about the reasons that led us to proceed in that direction.

4.1 Soil Samples and Earthworm characterization

The analysis in this work was carried out on 24 soil samples, in different conditions regarding the population of earthworm present and the humidity level. The samples of soil were collected in February 2020 at *Herdade do Freixo do Meio*, an organic farm (Lat. $38^{\circ}42'N$, Long. $8^{\circ}19'W$) in Southern Portugal (Foros de Vale de Figueira, Montemor-o-Novo, Évora District). The soil was collected by the use of shovels at a depth between 0 and 20cm and was sieved at 5mm, conserving only the aggregates smaller than this size and discarding bigger aggregates, stones and organic material that might be present. The soil was then stored in plastic bags at room temperature. This process of course is very much impacting on the original structure of the soil, preserving the fine structure inside the aggregates but removing all the information about the coarser organization of the original system. This is not a big problem for our analysis since we want to assess the impact of earthworm in the soil structure and complexity. For our scope it is more than enough to just confront the samples without and with earthworms, as they all start from the same "structurally deconstructed" point. Of course, the experimental setup will not allow us to make consideration about microporosity properties of the soil samples, since the resolution of the X-ray CT images is quite coarse.

The soil has a sandy-loamy texture, in particular, from analysis made by J. Coutinho and T.N. da Luz from the "*Laboratório análises de solos e plantas*" of the Trás-os-montes e alto Douro University, on different samples of soil collected on the same site, we have some data regarding the particle composition, in particular, in table 1 are reported the average values for the particle size distributions.

Table 1: Particle size distribution

Fraction	Content (% W/W)
Coarse fragments (5mm – 2mm)	14.5
Coarse sand (2 – 0.2mm)	44.8
Fine sand (0.2 – 0.02mm)	22.0
Silt (0.02 – 0.002mm)	7.9
Clay (< 0.002mm)	10.8



Figure 2: Soil extraction site

From this analysis we got also preliminary information about the level of soil organic matter, which results quite high, about 5% of the total weight of the sample.

4.1.1 Sample preparation and experiment description

The experiment that produced the data that will be discussed in this thesis started in February 2021. Before starting the experiment, soil water content and maximum water content (MWC) were measured for nine sam-

ples. To restore microbial communities after one year of inactivity, an elutriate solution¹ was prepared using freshly collected soil in a 1:10 soil-to-water ratio (w:w). The soil moisture was then adjusted using this elutriate to reach either 45% (dry conditions) or 55% (wet conditions) of its maximum water-holding capacity. Even if it does not seem so, this 10% difference is very impactful on all soil dynamics, and while a 55% level of humidity represent normal conditions, 45% is a value that indicated a quite severe drought going on in the soil sample. The prepared soil was then placed in transparent PETG tubes (90.7 mm in diameter and 250 mm in height) with removable bottoms, allowing for easy handling and maintenance. Throughout the experiment, the vessels were regularly weighed and replenished with artificial rainwater, following Velthorst's (1993) method, on Mondays, Wednesdays, and Fridays if needed to maintain the humidity conditions approximately constant. Before introducing earthworms, the microcosms were acclimated for one week in a controlled chamber set at 20°C. Once this acclimation period ended, earthworms and litter (*A. poureti*) collected on the same site as soil were added, and the experiment proceeded for a duration of two months.

A total of 48 samples were prepared and studied, half of them were maintained in dry conditions (45% of MWC) and half in wet conditions (55% of MWC), which is closer to the level usually found in normal conditions for this type of soil. Twelve of those samples (6 dry and 6 wet) were left without earthworms as a control, in each one of the others, 6 earthworm individuals were added. Specifically, three groups of twelve vessels were inoculated, respectively, with one of two different species or both of them (3 individuals for each species in this last case). The two species of earthworms belong to different ecological groups and they are:

- **A. Rosea:** Endogeic species, that usually creates short and narrow burrows, that do not exhibit a preferential direction, the small diameter of the animals (about 1mm) and the tendency to refill the galleries with cast material makes it difficult to identify the burrow systems of this species[10].
- **A. Trapezoides:** geofacile, mostly Anecic species, that also have endogeic-like behavior in some conditions (not in our case); they tend to create larger burrows due to their wider bodies (about 2 – 3mm in diameter) and the reuse of the same tunnels makes them more resistant to degradation. The galleries tend to be vertical and therefore are very impactful for percolation and water distribution between layers.

We obtain in this way 6 replicates of the system in eight different conditions, that will be referred in the following as:

- SM-dry: Without earthworms, dry conditions
- SM-wet: Without earthworms, wet conditions
- R-dry: With *A. Rosea*, dry conditions
- R-wet: With *A. Rosea*, wet conditions
- T-dry: With *A. Trapezoides*, dry conditions
- T-wet: With *A. Trapezoides*, wet conditions
- RT-dry: With *A. Rosea* and *A. Trapezoides*, dry conditions
- RT-wet: With *A. Rosea* and *A. Trapezoides*, wet conditions

At the end of the two months period some soil quantities, as SOM, pH, Aggregate stability, Total Carbon and Total Nitrogen, reported in table 2, were measured. Specifically, the Soil Organic Matter content was estimated by loss of ignition, which means that the dry soil was weighted and then collocated in an oven at 500°C. At this temperature the organic matter burns completely, leaving only the mineral fraction of the soil to be weighted after; the measure is expressed by the fraction of weight loss over the total dry weight. Aggregate stability data were obtained with the wet sieving method, which enabled us to determine the fraction of stable aggregates over the total mass of aggregates. Other quantities that were estimated were pH (with a solution of KCl at 1M concentration), total Nitrogen and total Carbon content (expressed in percentages over the total weight) thanks to a Carbon Nitrogen elemental analyzer.

Three samples for each condition were also analyzed with X-Ray Computed Tomography, giving a total of 24 images of the soil in the eight different conditions.

Also, the number and mass of earthworms were measured at the end, this was done to assess the actual level of earthworm activity during the period of experimentation. In the table 3 this information is reported for the 24 samples for which the X-ray CT scans were performed. It is important to notice that in the samples with

¹Water is mixed with fresh soil to extract some of the alive microbial communities present, this solution is then used to water old soil samples in order to restore the microbial network of the old sample.

Sample	SOM (%)	pH (1M, KCl)	Aggregate Stability (%)	TC (%)	TN (%)
SM-dry-1	5.60	4.97	69.68	4.42	0.23
SM-dry-2	5.72	5.04	73.09	4.39	0.20
SM-dry-3	5.82	5.05	64.69	4.06	0.20
SM-wet-1	6.57	4.85	35.79	4.29	0.18
SM-wet-2	6.67	4.85	31.92	4.17	0.20
SM-wet-3	6.60	4.78	45.41	3.98	0.17
T-dry-1	6.24	5.03	58.23	4.22	0.19
T-dry-2	6.17	4.94	51.04	3.97	0.22
T-dry-3	6.01	4.95	56.32	4.35	0.22
T-wet-1	6.40	4.85	59.25	4.08	0.21
T-wet-2	6.60	4.88	47.90	4.11	0.20
T-wet-3	6.86	4.77	52.85	4.23	0.21
R-dry-1	6.13	4.94	54.63	4.24	0.22
R-dry-2	6.21	4.87	52.24	4.19	0.22
R-dry-3	6.32	4.82	63.11	4.24	0.21
R-wet-1	5.50	4.62	42.48	3.66	0.17
R-wet-2	5.33	4.64	40.75	3.55	0.17
R-wet-3	6.14	4.56	60.49	4.46	0.19
RT-dry-1	5.58	4.70	56.70	3.88	0.17
RT-dry-2	5.71	4.71	52.63	3.73	0.17
RT-dry-3	5.92	4.65	60.30	4.20	0.19
RT-wet-1	6.16	4.60	64.95	3.99	0.19
RT-wet-2	6.42	4.58	61.53	4.13	0.18
RT-wet-3	6.35	4.68	68.50	4.18	0.19

Table 2: Soil organic matter (SOM), pH, aggregate stability, total carbon (TC), and total nitrogen (TN) as measured at the end of the experiment for the samples analyzed with X-ray CT. The percentages refer to fractions of weight, specifically the weight of the compounds over the total weight of the dry sample.

A. Rosea in dry conditions almost all animals died out or entered a state of estivation ² quite early on in the experiment. The mortality of A. Rosea is not so high in the samples that has also A. trapezoides, indicating that this latter species might have a positive impact on the life cycle of the former one. Also those data confirm the strong impact of humidity on earthworm activity and life expectancy. From the data in table 3 is also easy to appreciate the differences in the sizes (weights) of the two species of earthworms, with A. trapezoides being roughly one order of magnitude greater than the one of A. rosea. Generally the anecic species is more resilient to humidity changes and the individuals manage to maintain a good level of activity, while the smaller endogeic species show less ability to auto regulate their internal humidity levels, resulting in a much higher rate of mortality and estivation.

4.2 X-ray Computed Tomography Images

The X-ray scans were realized with a machine created for medical purposes, courtesy of DIATON clinic, from Coimbra, in Portugal. The machine is a GE Medical Systems Revolution EVO (REV03), that was operated at a peak voltage of $120kV$, with the X-ray tube current at a value of $350mA$ and $1s$ of exposure time per rotation. The resolution of the resulting images is at $0.67mm$ per pixel, resulting in voxels with a volume of $\approx 0.3mm^3$. The images were organized as a series of vertical scans; the software Fiji (which is basically a version of ImageJ with some already installed plugins) was used to create stacks from the sequence of vertical

²sleep-like state, similar to hibernation for mammals; when the conditions are unfavorable the earthworms coil in a ball and cover themselves in mucus to keep the humidity level and to not dry out.

Moisture	Species	Worm abundance start (n)	Worm mass end (mg)	Mortality (n)	Estivation (n)
Dry	A. trapezoides	6	2895.1	0	2
Dry	A. trapezoides	6	3219.1	0	2
Dry	A. trapezoides	6	3255.9	0	2
Wet	A. trapezoides	6	3187.7	0	0
Wet	A. trapezoides	6	2925.5	2	1
Wet	A. trapezoides	6	3793.6	0	1
Dry	A. rosea	6	157.9	4	1
Dry	A. rosea	6	168.5	4	2
Dry	A. rosea	6	254.7	2	4
Wet	A. rosea	6	416.2	0	2
Wet	A. rosea	6	416.9	0	1
Wet	A. rosea	6	381.7	1	0
Dry	A. trap.+A. ros.	3+3	1217.3+61.4	0+2	3+1
Dry	A. trap.+A. ros.	3+3	1009+105.3	0+0	1+2
Dry	A. trap.+A. ros.	3+3	1481.8+135.5	0+0	3+1
Wet	A. trap.+A. ros.	3+3	1255.3+177.6	0+0	1+0
Wet	A. trap.+A. ros.	3+3	1627.4+160.9	0+0	1+1
Wet	A. trap.+A. ros.	3+3	946.7+171.7	0+1	1+0

Table 3: Worm abundance, mass, mortality, and estivation for the samples with earthworms analyzed with X-ray Computed Tomography.

scans, so to obtain a 3D volume of voxels, carefully constructed as to avoid any overlapping between pixels, so to avoid confusion about the voxels' values.

The stacks were reoriented to get a view from the top and, from the sample of soil, a region of interest was extracted from all the stacks. This region of interest was chose to be a 92X92X260 pixel parallelepiped volume at the center of the cylinder of soil. This procedure was necessary so to have samples all of the same sizes, in order to eliminate the differences in the sample sizes (The soil quantity is slightly different for each sample), and to focus on the topology of soil architecture far from the sample's borders. In fact, it is well known that earthworms alter their behavior when in contact with the PVC walls, profiting from the presence of the artificial element to create burrows more easily along the walls. Also, even without earthworm's influence, the arrangement of soil elements and of the pore space could be affected by the proximity of a rigid, impermeable solid surface. For these reasons then, the region of interest was choose to be in the center of the sample, reasonably far away from the walls, the surface and the bottom, so to carry on an analysis on a theoretically undisturbed region of the sample, in order to get results closer to the ones that would be obtained in a field experiment.

For the analysis of information theory a central slice 20 pixels thick was removed from the region of interest, creating samples with dimensions 92X92X240. This was done in order to not consider a layer of ceramic marbles, which was placed there to asses the vertical dislocation of soil matter induced by earthworms activity. The removal was carried on to, once again, try to avoid all the anthropogenic elements of the experiment; in particular, those ceramic marbles have a high reflective power and therefore results in the images as pixels with a much higher value of intensity than the surrounding soil. Keeping these layers of soil would inevitably affect greatly the distribution of the intensities, introducing a bias with both their presence and specific distribution, which is absent in natural soils. For the analysis of the burrows systems, since it is important to have the full volume in order to identify burrows connectivity and dimension, the region of interest was analyzed as a whole. The stacks were then saved as *.tiff* files for further analysis with Fiji or Python scripts.

4.3 Analysis on the Fisher-Shannon plane

The analysis of the images for the part of information theory part was carried out completely in Python (version 3.12), with a pipeline specifically written for this scope by the author. The full code, together with the dataframes obtained at the end of the analysis can be found on GitHub at the address in [15]. The *.tiff* files were uploaded as pandas dataframes (referred to as "dfs" from now on) with four columns, the firsts three being the spatial coordinates of the voxel ("x", "y", "z") and the fourth the intensity value of the voxel (here expresses in standard, default Housefield units (HU)). The absolute values of intensities are not relevant here, since our analysis will focus only on the distributions of those values. The resulting dataframes have then four columns and 2.031.360 rows, as this is the number of voxels in the 92X92X240 pixels volume.

The dfs were then reorganized in four different ways, to get four different partitions of the samples, rewriting all the values of intensity for every partition as a list in one column of the df:

- Columns: every column of soil (referring to one ("x", "y") couple) is analyzed separately; this partition resulted in 8.464 subsamples, each with 240 values of intensity, one for all the voxels in that column.
- Walls: every vertical layer parallel to the y axes (referring to one value of "x") is analyzed separately; this partition resulted in 92 subsamples, each with 22.080 values of intensity, one for all the voxels in that vertical plane.
- Planes: every horizontal layer of soil (referring to one value of "z") is analyzed separately; this partition resulted in 240 subsamples, each with 8.464 values of intensity, one for all the voxels in that horizontal plane.
- Volumes: the whole volume is analyzed; this resulted in samples with 2.031.360 values of intensity, one for every voxel in the volume.

Those different partitions were all evaluated in order to investigate if the partition method influences the results of this statistical analysis; therefore, the same process was applied to all the subsamples and all the samples. We will refer in the following as each column, wall, plane or volume as "subsample" indifferently, to describe the statistical analysis carried out on all of them.

The first passage was to derive a probability density function for each subsample; since we want this analysis to be completely model-independent, we impose no condition on the functional form of the pdf and we proceed with a non-parametric density estimation. We followed the same approach used by Aguiar and colleagues [3], adopting a kernel density estimator of the pdf given by

$$f_M(x) = \frac{1}{Nb} \sum_{i=1}^N K\left(\frac{x - x_i}{b}\right) \quad (13)$$

where N is the dimension of the subsample, $b > 0$ is the bandwidth parameter and $K(u)$ is the kernel function which must satisfy $K(u) > 0$ and $\int_{-\infty}^{+\infty} K(u)du = 1$ (in the continuous approximation, the integral is a sum if we stick to the discrete description). Following again the methodology of Aguiar and colleagues, we choose a Gaussian kernel, which is the most widely used one in this kind of studies, having therefore:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad \text{yielding} \quad f_M(x) = \frac{1}{Nb\sqrt{2\pi}} \sum_{i=1}^N e^{-\frac{(x-x_i)^2}{2b^2}} \quad (14)$$

This operations were carried out using the *gaussian_kde* function from the *scipy.stats* library, specifying a costume value for the bandwidth. The bandwidth controls the smoothness of the pdf and Aguiar and colleagues calculated it for every subsamples using Silverman's rule:

$$b = 0.9 \min\left(\sigma, \frac{IQR}{1.34}\right) n^{-1/5} \quad (15)$$

where σ is the standard deviation of the intensity distribution and IQR its interquartile range. The group used the average of all the bandwidths of a partition to calculate the pdfs for all the subsamples of that partition.

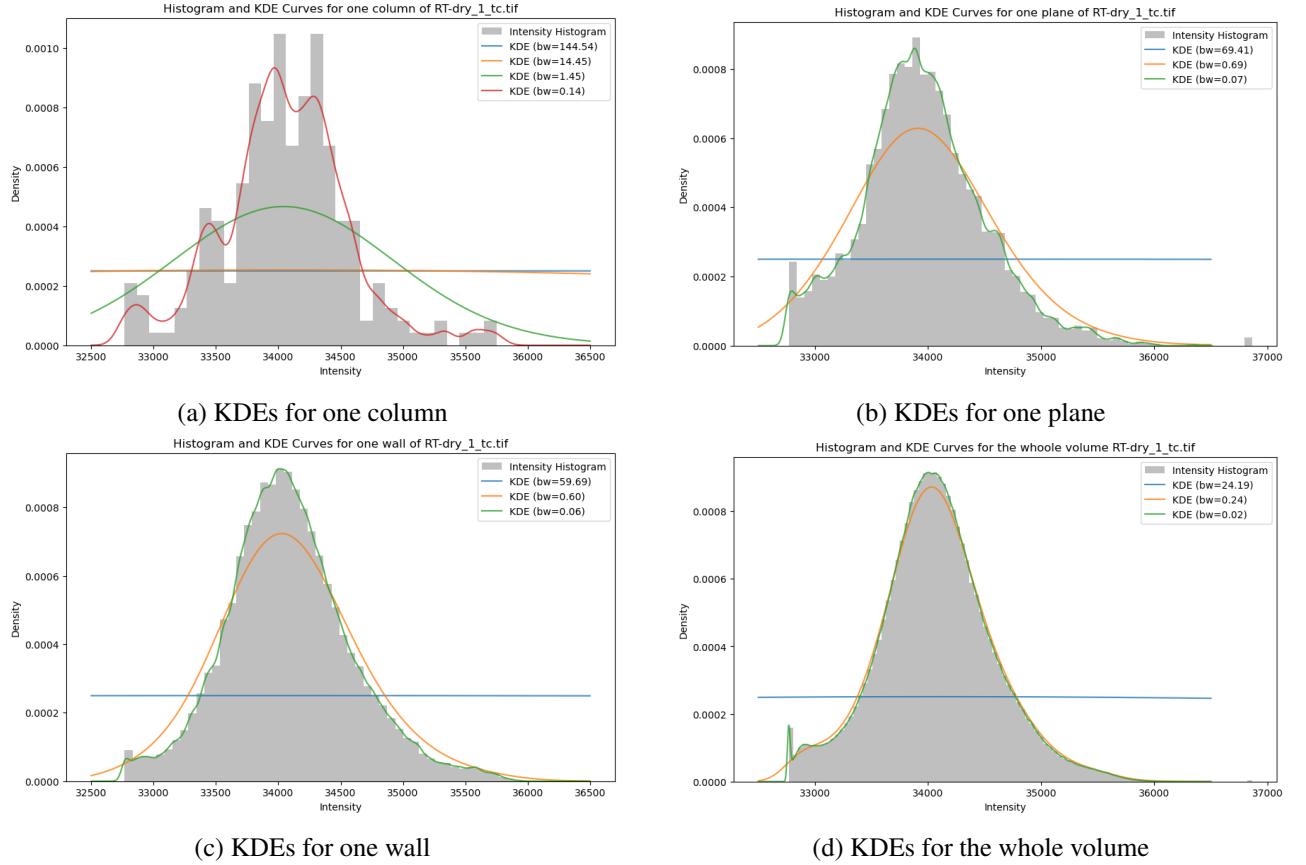


Figure 3: Probability density functions estimated with various bandwidths (starting from the fist Silverman's bandwidth) for the four different partitions of the sample.

In our case, though, this approach did not produce useful results, as the pdf curves resulted effectively as straight, horizontal lines, not being able to capture the real distributions of the values. We then modified the value of bandwidths obtained with Silverman's rule in order to obtain pdfs that follow closely the distributions without overfitting the dataset. As Heidenreich reluctantly puts it in a review on the topic *"Part of the community working on nonparametric statistics has accepted that there may not be a perfect procedure to select the optimal bandwidth"* [37], and this is even more true for large, highly irregular samples like ours [8]. In the absence of a robust protocol to find the best bandwidth, we proceeded by multiplying the value obtained with Silverman's rule for a reduction factor. As can be seen in figure 3 a very strong rescaling factor was necessary before the pdf assumed a form that was representative of the intensity distribution. This is probably due to the limited size of the dataset and of the subsamples. Another reason could be the high non-regularity of the distributions that, in most cases, present a quite chaotic profile, far from a regular distribution.

The final bandwidth choice was made, starting from the Silverman's rule value, conducting an exploration of the parameter over a spectrum, in order to identify a value for which the calculated Fisher-Shannon complexity respected the isoperimetric inequality ($FSC \leq 1$) and for which the pdf represented the distinctive features of the distributions. In particular, in the samples with earthworm presence, as can be seen in figure 3, only the distribution with the smallest bandwidth is able to represent the leftmost peak for the subsamples planes, walls and the whole volumes; this peak is crucial as it represents the pore-space volume. Furthermore, the smallest bandwidth is able to give a representative multimodal distribution for the columns datasets, allowing to consider the properties of the actual distribution of intensities. Last consideration is about the fact that our choice of the value of bandwidths, even if it is very small, doesn't seem to overfit the datasets, as it does not assume the characteristic form of a sum of separated peaks.

For those condition to be respected by every subsample, we scaled the Silverman's bandwidth, calculated as above, of a factor of 10^{-3} for all the subsamples.

Once the pdfs were obtained with the appropriate bandwidth, the values for Shannon Entropy Power, Fisher Information Measure and Fisher-Shannon Complexity were calculated as reported in the previous sections, ob-

taining one value for this quantities for each subsample. Those values were then averaged in order to obtain only one value of each quantity for each sample of soil. The calculations of the functionals of the pdf were carried out numerically using the Simpson's method for the computation of the integrals, specifically the *simps* function from the *scipy.integrate* library. At this point it was possible to represent each sample in the Fisher-Shannon Information plane, investigating their relative positions in this context. The distance from the isocomplexity line ($FIM = 1/SEP$) was suggested as a measure of complexity by Aguiar and colleagues and also, in a completely different context, by [article about people in the crowd]. To be able to use this indicator, we rescaled all the values of SEP and FIM, respectively dividing and multiplying for the highest value of the SEP for all the subsamples. In this way we obtained values for SEP and FIM that are of the same order of magnitude, and a distance in the transformed plane does not have anymore a preferential direction. Indicating as SEP'_i the SEP value of the projection point for a sample with coordinates (SEP', FIM') (the “/” symbol indicates that the quantities are rescaled as specified above) on the isocomplexity line, the corresponding FIM'_i value will be $FIM_i = 1/SEP_i$, and the distances between those two points will be:

$$d = \sqrt{(SEP' - SEP'_i)^2 + \left(FIM' - \frac{1}{SEP'_i}\right)^2} \quad (16)$$

Setting the derivative in respect to SEP'_i equal to 0 to find the closest point (SEP'_0, FIM'_0) on the isocomplexity line, yields the polynomial expression for $x = SEP'_0 = 1/FIM'_0$:

$$x^4 - x^3 SEP' + x FIM' - 1 = 0 \quad (17)$$

which was solved numerically for all samples using the *root_scalar* method from the *scipy.optimize* library.

4.4 Analysis of the pore space

The greyscale images give us some information about the porous structure of soil, but to quantify pores properties it is necessary to binarize the images to get a clear picture of solid and voids. Even though this process is a routine one in soil sciences, there is not yet a standardized procedure to achieve the best possible result from binarization. Cause for this is the great heterogeneity of quality in the images...

To get a good binarization of the images a range of thresholds was explored and a series of preprocessing approaches tested, we then get to what for us was the optimal procedure that we applied to all the images. A Macro was implemented in ImageJ in order to automatize the segmentation of all the scans, the *.tiff* files were processed as follows:

Firstly the images were transformed in *8bit* format, now the black pixels have value 0 and are associated to the pores and the maximum value is 255, this was done to assure that all the images had values in the same greyscale range and to not deal with excessively heavy images. After that, the contrast was augmented by 85%, and the command '*Sharpen*' was used to improve the separation sharpness. This function works by creating a blurred version of the image and subtracting it from the original, so to remove low frequency features and highlighting the high-frequency ones, like edges. Then, a 3D Median filter was also applied, this function is a non linear filtering technique used to reduce impulsive noise (like white uncorrelated noise noise) while preserving edges, since it does not average the values. It works by replacing each voxel's intensity with the median value of its neighborhood, that in our case was a cube with a side of three pixels.

At this point the scans were binarised with a fixed threshold, maintained at the same value for all the samples, in order to be able to make confrontation between them [67]. The voxels with intensity values equal or less than 10 were considered as belonging to the pore space and all the others as solid phase. A final modification to the already segmented images was done with the '*Despeckle*' function, that makes use of local statistical operations to eliminate speckle noise (random grainy pattern and salt-and-pepper). In this way we cleaned up the scans for small, isolated noise artifacts while preserving the overall structure of the edges.

In figure 4 it is possible to see the original image and the binary version side by side to get an understanding of the results obtained with this process; we can appreciate the main porous structures highlighted in the black and white picture as long as some of the finer void spaces. We show here also one 3D rendering of the porous structure of a sample with A. Trapezoides in wet conditions, to give a sense of how pervasive the pore space is even at this quite large resolution. The result of the process is a binary image in which the zeros correspond to

the solid phase and are represented in black, and the ones mark the void spaces and are pictured as white.

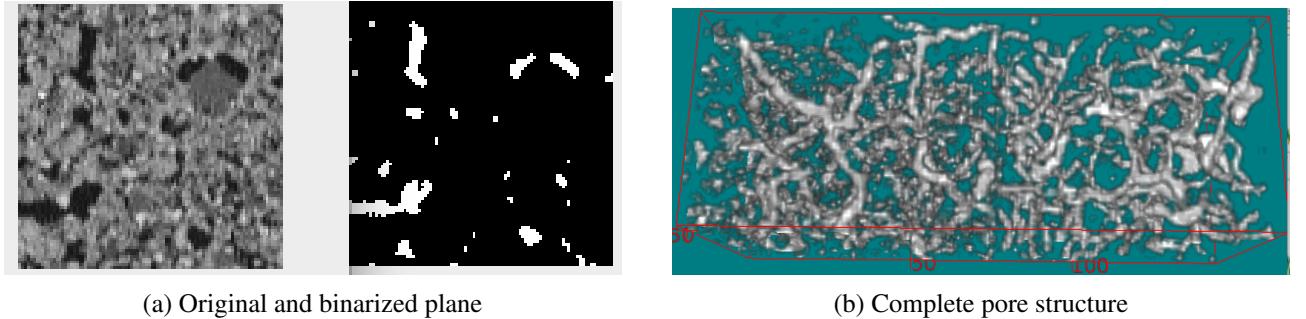


Figure 4: Original and binarised images of a plane of soil; 3D rendering of the pore space of a sample with the presence of *A. trapezoides*

The binarised *.tif* files were then analyzed in python and ImageJ to obtain a set of structural properties of the pore space following the widespread method of Weller et al. [86]. Four Minkowsky functionals (M_{0-3}) were calculated that comprise fundamental topological properties of the complex object that is the pore space. These quantities were calculated manually using tools from python's libraries *skimage.measure* and *scipy.ndimage*, the complete code can be found on GitHub [15]. The total pore volume M_0 is simply obtained by summing the values of all the voxels, obtaining the number of them that have value 1, and therefore belong to the pore space. The total surface area M_1 is a measure of the extension of the contact zone between solid and liquid media, and it is computed with the use of the *binary_erosion* function of the *scipy.ndimage* library. This method eliminates the voxels from the boundary of objects (pores, in our case), the resulting image is subjected to a XOR operation with the original binary image to obtain just the surfaces.

The integral of mean curvature M_2 is a quantity that characterize the shape of voids, it assumes negative values for concave surfaces, which is typical for packing voids between soil aggregates, and positive values for convex surfaces as bubbles and earthworm burrows. This functional was calculated with the *marching_cubes* method, which counts how many 0s and 1s are in the neighborhood of each voxel and from that draws an estimation of the curvature of the surfaces.

The Euler characteristic M_3 is just the number of isolated objects that comprise the pore space and is easily obtained with the *label* function of the *scipy.ndimage* library.

The Minkowsky functionals, except for M_3 which is obviously unitless, are in units of "pixels", and are also extensive properties, i.e. they depend on the actual size of the volume that is considered. To solve the first problem we just multiplied the three functionals respectively for the resolution cubed, squared and elevated to the minus one. Obtaining the volume in mm^3 , the surface in mm^2 and the curvature in mm^{-1} . Since our samples are all of the same size the fact that the properties are extensive is not impactful, we still decided to normalize the values for easier possible future confrontation with samples of different sizes.

5 Results

In this chapter we will present the results obtained with our analysis of the 3D X-ray scans of the soil samples, in particular we will show what was found regarding the statistical measures we used from information theory across the various samples. We will focus on exploring the connections of those values with the level of moisture, earthworms presence and activity and inherit properties of the soil samples as depth. We will also assess the correlations between the statistical quantities and some physical, chemical and biological properties of the samples as presented in tables 2 and 3. Finally we will present the analysis of the pore space with Minkowsky functionals and we will explore how those features correlate with the quantities obtained from the information theory analysis of the greyscale images.

5.1 Statistical quantities trends with depth

In this section we are going to analyze what was found about the variation of the values of the calculated statistical quantities across the vertical axis. This analysis is important as structural properties of soil vary considerably with depth, and this is a way to assess whether our statistical indicators are able to detect those changes.

The two figures 5 and 6 refer respectively to dry and wet conditions, both with and without the two species of animals. They contain plots of the values of the four statistical indicators (Shannon Entropy Power, Fisher Information Measure, Fisher-Shannon complexity and distance from the isocomplexity line in the Fisher-Shannon Information plane) for all the horizontal planes (the value for $z=0$ corresponds to the uppermost plane and the one for $z=240$ to the bottom one). Each plot shows the values for the three replicates for each conditions, so to confront also the variability between replicates, which seems to be quite negligible both in absolute values and trends. The only exception is, maybe, the R-wet sample, in which, closer to the surface, the values for the distance are quite different between the three replicates. One first observation is that the values for all the statistical indicators fluctuate much more from plane to plane in the dry case than in the wet one; the signal remains quite chaotic in all conditions, but it is much more so for dry samples.

Also, we notice that the value of the Fisher-Shannon complexity generally does not exhibits any characteristic tendency along the vertical axis, but the opposite is true for the other indicators. It can be argued that the shape of SEP and distance are often quite similar and somehow reversed in respect to the shape of FIM. Generally, in all the samples, the values of SEP and distance tend to be lower with increasing depth while the values of FIM have the opposite behavior.

In the dry samples without earthworms and in the presence of *A. Rosea* (in this condition all the earthworms died out, making this samples as effectively without earthworms) the curves for SEP and distance exhibits some degree of negative concavity, with a peak at about halfway; while the values of FIM assume a minimum value just before the middle and then start rising again. This behaviour of the statistical quantities is a peculiarity of the dry cases without earthworms, as in all the other samples the curves do not seem to exhibit noticeable convexity.

In the wet cases the tendency of SEP and distance values to diminish and FIM values to rise with depth is maintained, but in all cases this correlation does not start from the beginning. In fact, it seems that all the values remain more or less constant on a plateau for the first half of the way, starting only later to correlate with depth.

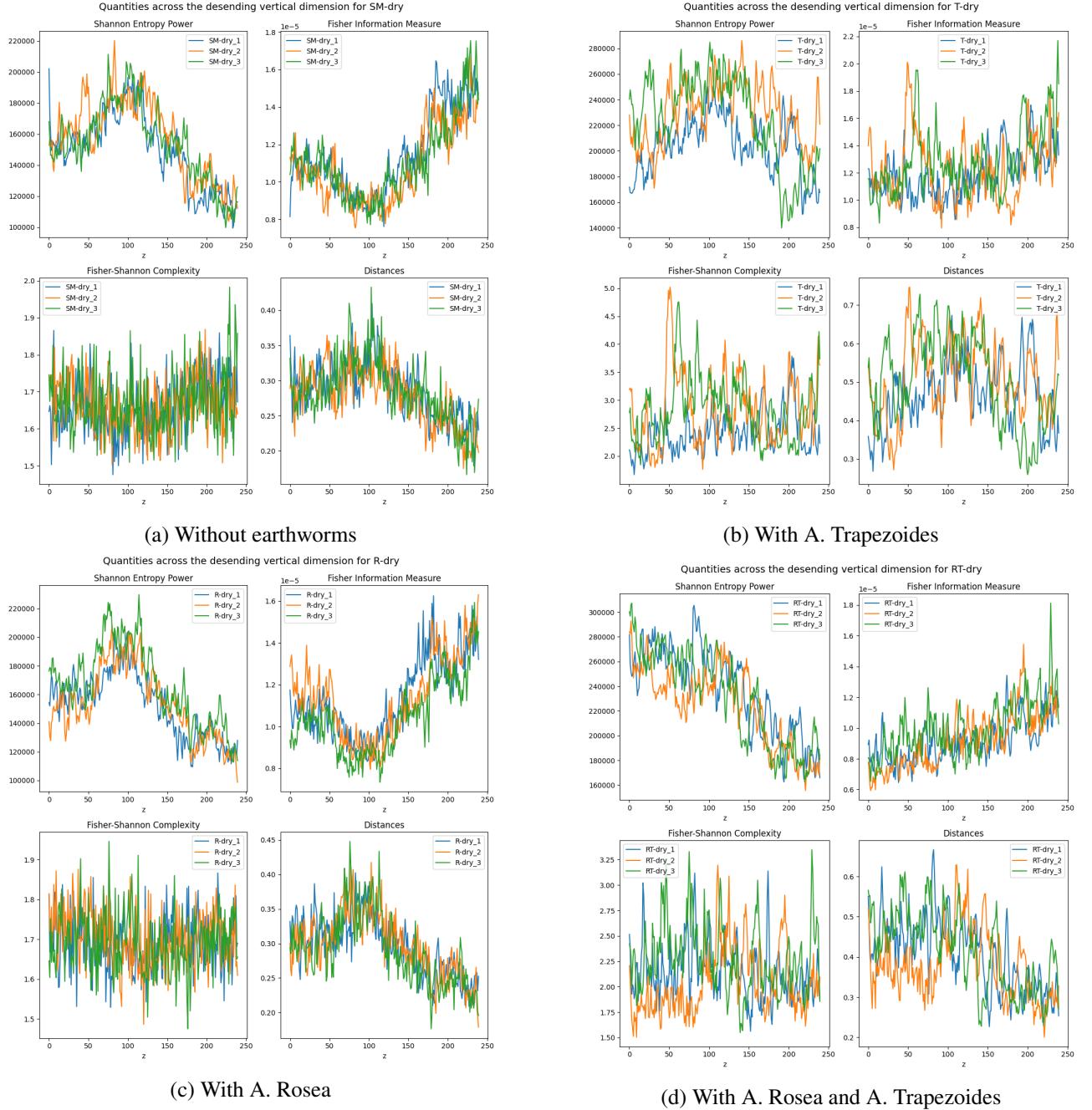


Figure 5: Statistical quantities across the descending vertical axis for all the samples in dry conditions; the four plots of each subfigure represent, in clockwise order starting from the top left, Shannon Entropy Power, Fisher Information Measure, distance from the isocomplexity line in the FSI plane and Fisher-Shannon complexity

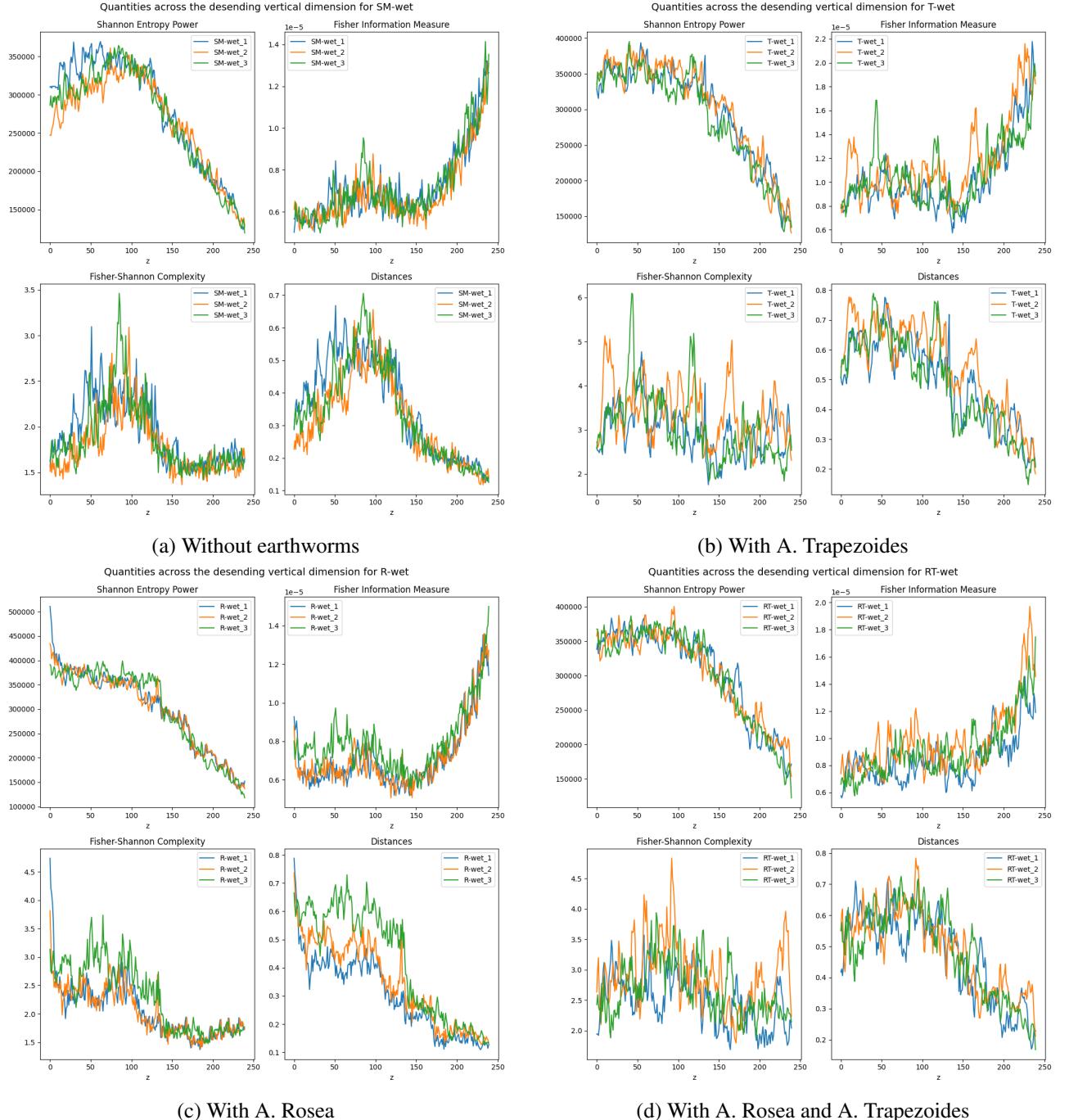


Figure 6: Statistical quantities across the descending vertical axis for all the samples in wet conditions; the four plots of each subfigure represent, in clockwise order starting from the top left, Shannon Entropy Power, Fisher Information Measure, distance from the isocomplexity line in the FSI plane and Fisher-Shannon complexity

5.2 Fisher-Shannon Information analysis

In this section we will go over the results obtained with the tools of information theory, presenting in the values found of Shannon entropy, Fisher Information Measure, Fisher-Shannon complexity and distance from the iso-complexity line. We will also consider the representation of the dataset inside the Fisher-Shannon information plane, to propose classifiers that could be able to discriminate between presence and absence of an ecotype of earthworms.

5.2.1 Fisher-Shannon Information Plane

As a fundamental tool for our analysis, we plotted the samples in the Fisher Shannon information plane, with the value of Shannon Entropy Power on the horizontal axis and the result of Fisher Information Measure on the vertical one. Generally, the samples in different conditions occupy different areas of the plane, remaining quite well separated. In figure 7 the planes for the Columns and Planes partitions are showed; on the left we plotted the average points for each sample with the standard deviations of both SEP and FIM, and on the right side we plotted the normalized quantities, without standard deviations for an easier reading.

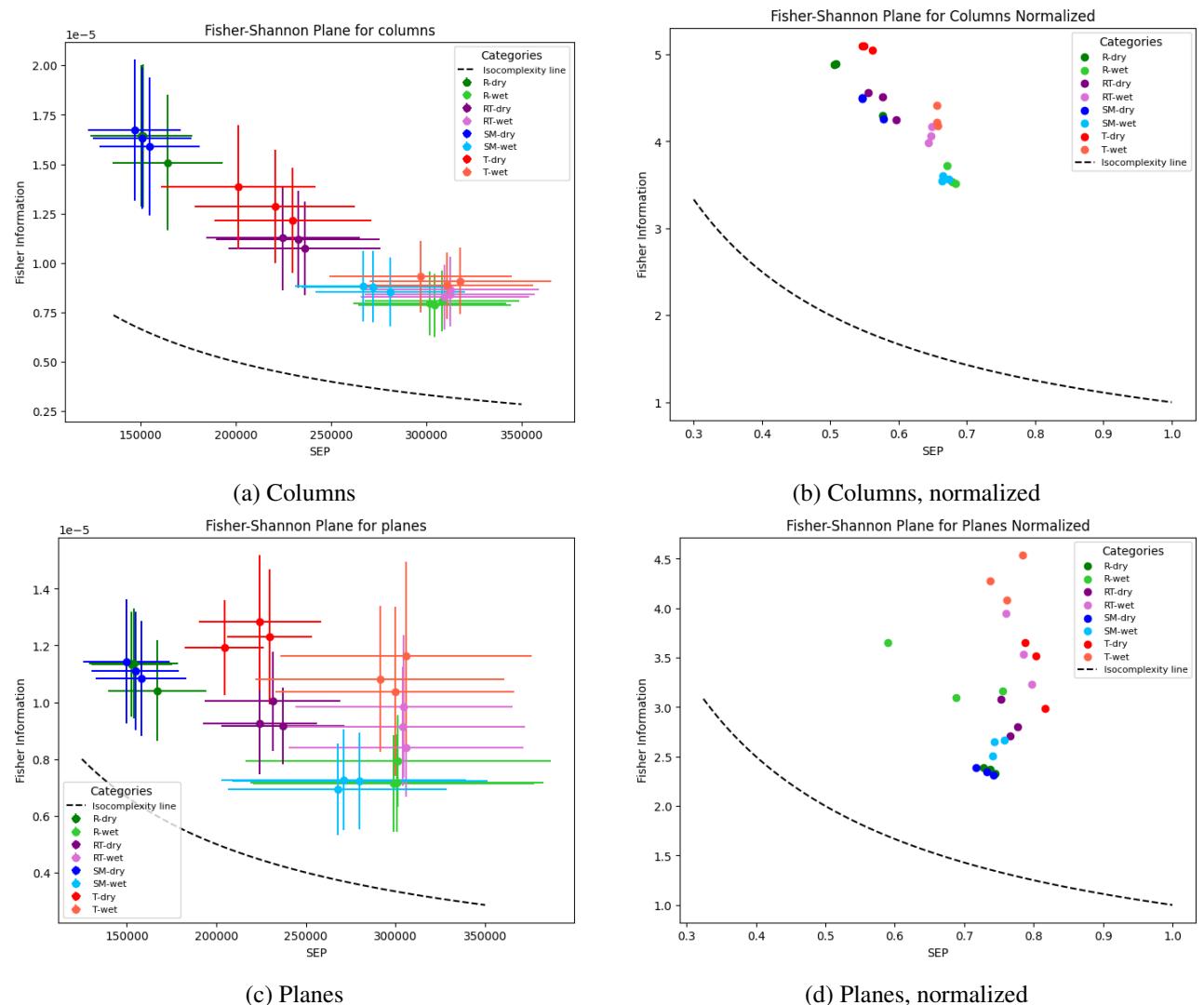


Figure 7: Fisher-Shannon Information plane for Columns and Planes partitions, both with std and normalized without std

In the columns partition the values seem to occupy a linear region, with the dry, earthworm-free samples in the top left corner and the wet samples with earthworms located more towards the bottom right corner. For the planes partition this tendency is not evident and the data appear more mixed and confused. In both examples, the standard deviations associated with the measures are quite big and do not really allow for a clear partition

of the plain between different conditions. Also, we notice that the process of normalization clumps the data points in a small area of the plane and causes considerable overlapping of the data.

In figure 8 we show the same plots for the Walls and Volumes partition. In figure 8c we considered just one average value for each conditions, so to have a value for the standard deviation associated to the statistical measurements with this partition (even though, having just three replicates the sample is not statistically significant). In those two cases, the data are sufficiently separated and the standard deviations small enough to give a pretty clear partition of the space. As was the case for the other partitions, the process of normalization results in less defined separation between samples in different conditions.

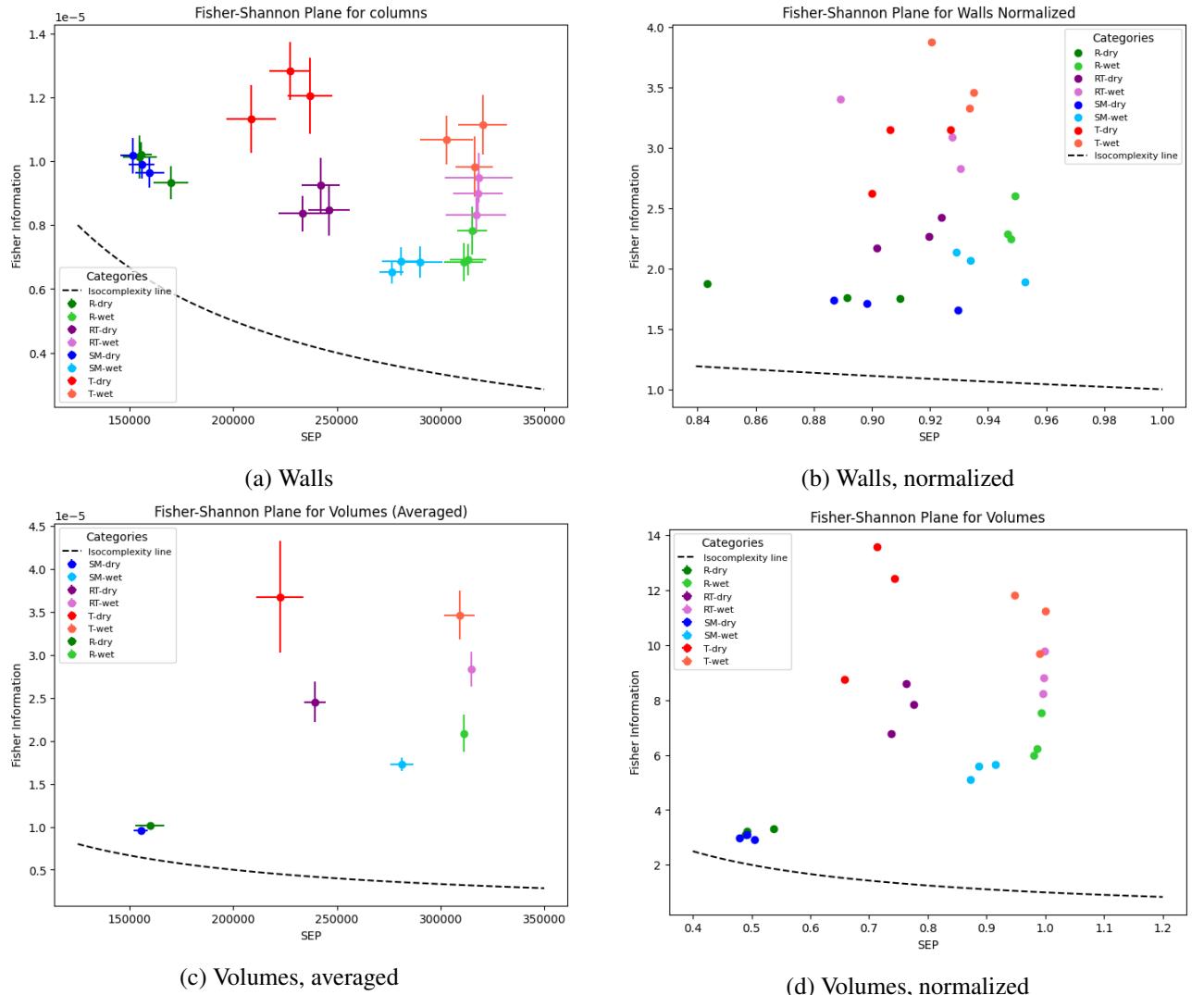


Figure 8: Fisher-Shannon Information plane for Walls and Volumes partition, in (c) there is only one value for each condition that is the average of the three samples in that condition.

5.2.2 Fisher-Shannon Complexity

In figure 9, the mean values of the Fisher-Shannon Complexity ($FSC = SEP * FIM$), with their standard deviation, are plotted for all the samples object of the study and for all the partitions considered. This quantity, in general, varies considerably between the cases with and without earthworms and between the dry and wet conditions. In particular, the samples with animals yield higher values of complexity in respect to the earthworm-free ones. Also, the wet soils consistently have higher values of complexity than the dry ones, and, at least in the absence of earthworms, the standard deviation is much larger in wet conditions too.

In the columns case, the standard deviations associated to the mean are of the same order of magnitude than the dispersion of the data, which means that, with this partition, this parameter is not enough to have a clean differentiation between different conditions. With the planes partition the data are definitely more separated between

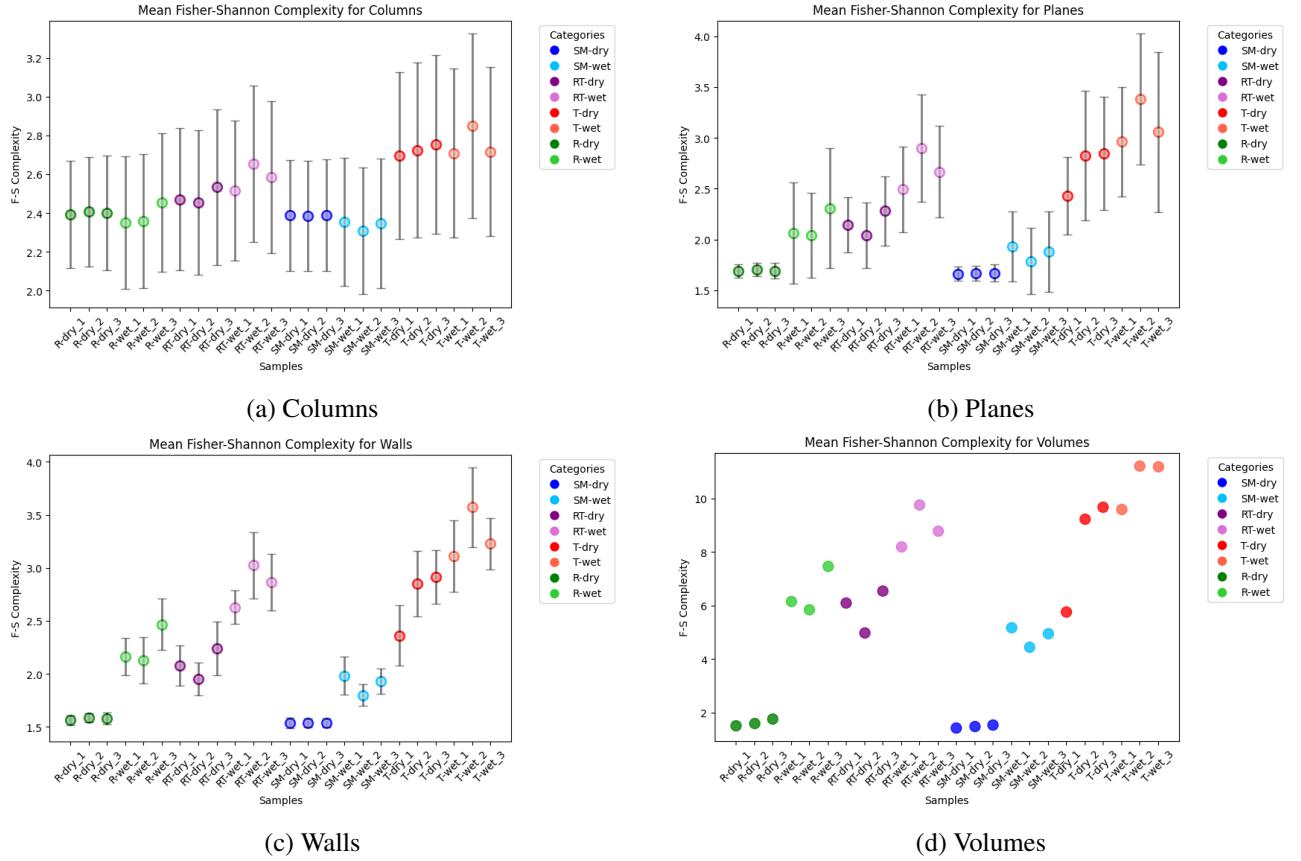


Figure 9: Mean and standard deviation of Fisher-Shannon Complexity value for all the samples, for the volumes partition there is only one value per sample, so we have no std.

them, the uncertainty though is still quite high, especially in wet cases with the presence of earthworms. For the animal-free, dry samples (SM-dry and R-dry, since almost all *A. rosea* died out or stayed in estivation state in this case) the variability is actually very small. The walls partition yields quite strongly separated results, with much smaller variability than the previous cases. For the volumes partition we have only one value for each samples, the single values are plotted in figure 11d and we can here appreciate a sensible division between the samples, with the same characteristics of the walls partition. In figure 10 we calculated one single values of mean and standard deviation of the three samples for each conditions for all the partitions. Here we can appreciate that, with the Volumes partition, the separation of the values of different categories is very clear, the variability is very small, except in the T-dry case.

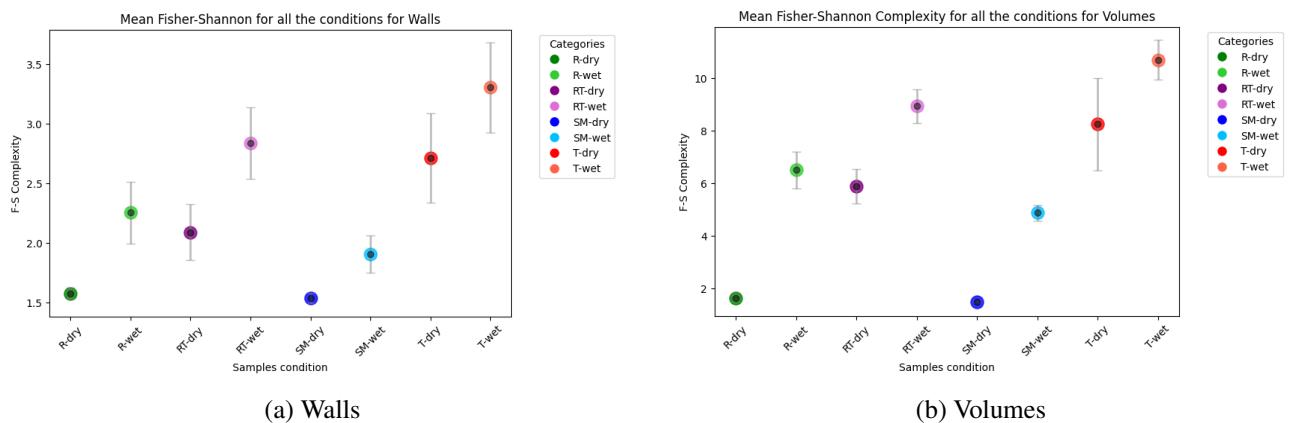


Figure 10: Mean and standard deviation of Fisher-Shannon Complexity value for all the categories for walls and volumes partitions.

5.2.3 Distance from the isocomplexity line

In figure 11 we show the results of the calculation of the distances from the isocomplexity line in the Fisher-Shannon information plane, calculated after the process of normalization, for all the samples and the partitions.

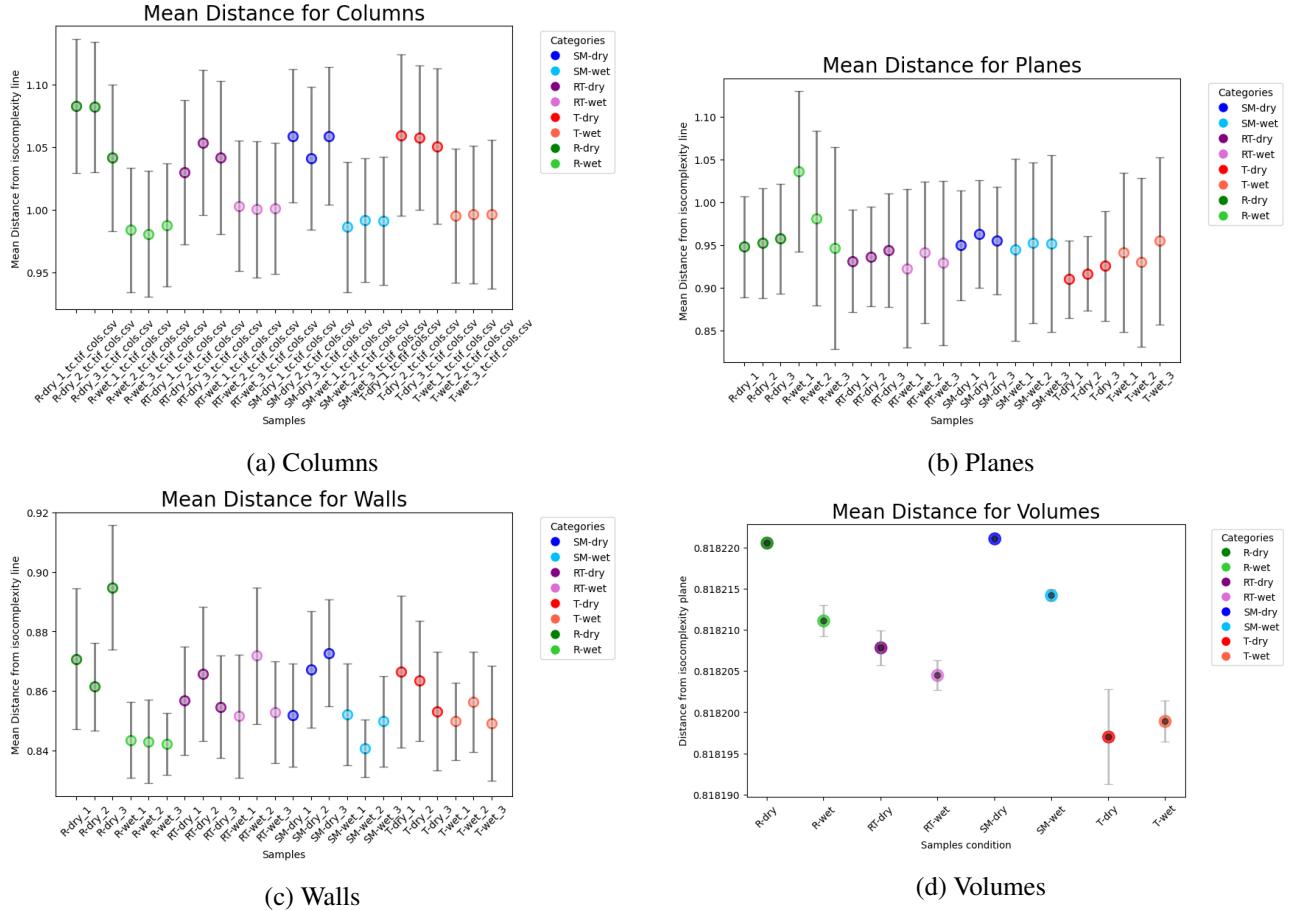


Figure 11: Mean and standard deviation of the distance from the isocomplexity line for all the samples, for the volumes partition there is only one value per sample, so we have no standard deviation.

The distance's value varies considerably across the samples in all the partitions. Specifically, in the columns case, the values of distance seem to be able to discriminate between wet and dry samples, with the former exhibiting a higher value of distance, and therefore of complexity. At the same time it does not seem that there is a marked difference in the values for the samples with different earthworm population. With this partition, though, the standard deviations associated with the mean value of distance for each sample are quite big, often larger than the difference between the means. In the planes partition, the distribution of the means is even less significant, with almost no changes between the different samples. The walls partition also yields poor results, here the mean values fluctuate a lot, without evident trends, and the standard deviations are in general of the order of the differences between the means. In the Volumes partition we plotted just the mean value of the three replicates for every conditions, with the associated standard deviation, the values are much higher for SM-dry and R-dry than for all the other samples.

In figure 12 we plotted the global means for all the conditions (effectively treating the three replicates as one) for the Columns partition, since this was the case in which the distance measure seem more relevant. We can here clearly appreciate the higher value of distance associated with dry samples, and the almost total absence of differences between biological conditions. We point out that, even if the averages seem to be quite well separated, the standard deviations associated to the means are of the same order of magnitude of the difference between the averages.

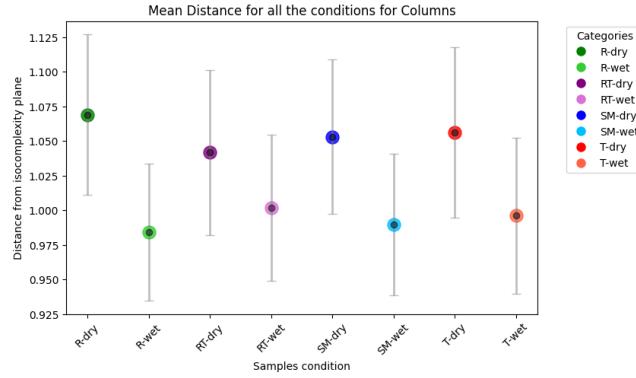


Figure 12: Walls

5.2.4 Moisture level and earthworm presence discrimination

To better assess the discrimination power of our analysis between two different conditions of one parameter, regardless of the variability in the others, we analyzed the dataset aggregating all the samples that share the same value of the parameter in question. Specifically, we created two categories to asses the moisture level: "wet" and "dry", that aggregate all the samples in wet and dry conditions respectively, regardless of earthworm presence. Also, we are interested in the classification of the samples in the two classes "With earthworms" and "Without earthworms", the samples SM-dry, SM-wet and R-dry belong to the former category, and all the others to the latter. We then calculated the mean values for the distance from the isocomplexity line and for the Fisher-Shannon Complexity for the classes of samples described above and produced various boxplots.

In figure 13 we show the most meaningful partition (Columns) to discriminate between wet and dry conditions using the distance from the isocomplexity line.

The plot in figure 11 suggested that this quantity for the columns partition could be able to differentiate between moisture level. We have higher values for the distance in dry conditions than in wet one. But the uncertainty associated with the averages are of the same order of magnitude of the difference between the mean values. This last fact is evident in the large number of outliers present in the boxplot in figure 13

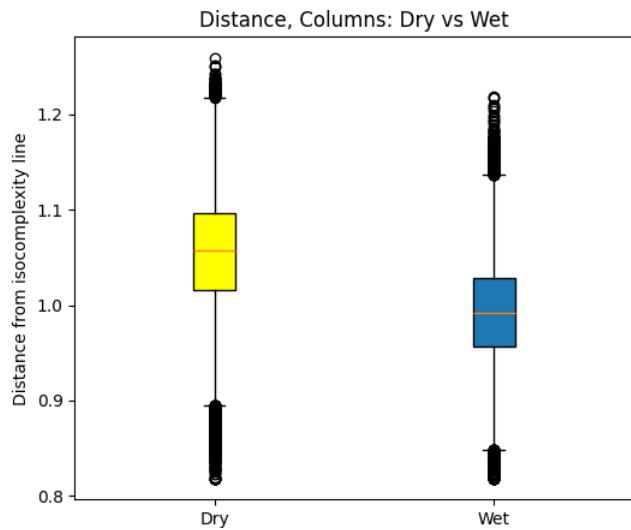


Figure 13: Distance from the isocomplexity line for column partition in wet and dry conditions

We also analyzed the differences in the value of Fisher-Shannon Complexity between wet and dry conditions; the only partition in which this two categories had different enough distributions were the volumes one. In figure 14, we can appreciate how also this quantity assumes higher values for the wet cases than the dry ones, with the mean values more than one standard deviation apart.

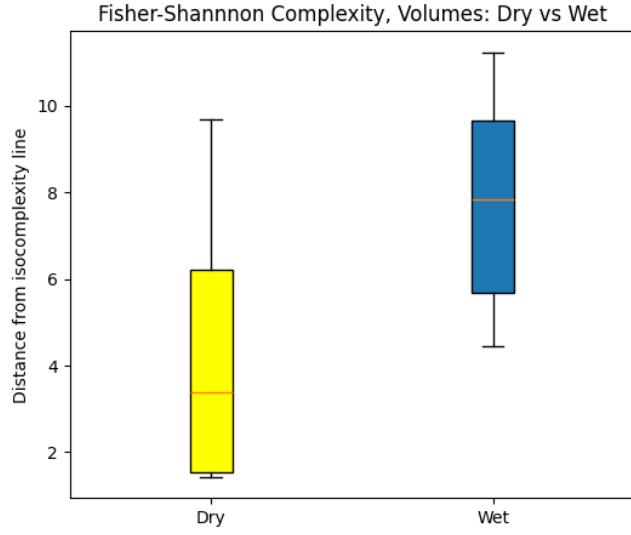


Figure 14: Fisher-Shannon complexity value for the Volumes partition for wet and dry conditions

Lastly, we found that the Shannon Entropy Power alone is able to discriminate very efficiently between different level of moisture in soil. In figure 15 we show the two distributions for the SEP value for the Walls and Volumes partitions in the dry and wet cases. We can appreciate that the values for the different level of moisture are quite different, also the dry values are much more dispersed, with a standard deviation considerably larger than the one for the wet cases.

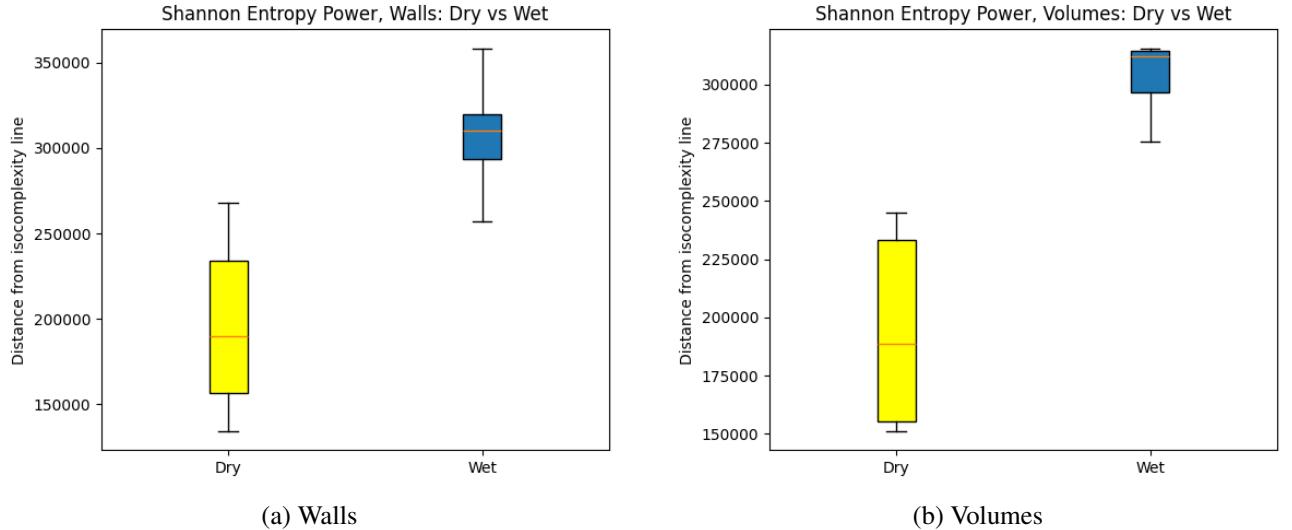


Figure 15: Most meaningful comparisons of Shannon Entropy Power between wet and dry conditions

We are also interested in the discrimination between earthworm presence and absence. In figure 16 we plotted the best partition for the division of earthworm presence according to the distance measure. We can see that the Volumes partition is able to give a clean separation between the cases, when distance is considered. The complexity of the soil with earthworms is significantly higher than the complexity of the ones that do not feature worms.

In figure 17 we show the values of the Fisher-Shannon Complexity, obtained for the samples with and without earthworms, for all the partitions. All partitions exhibit a very high variability, made evident by the numerous outliers in the boxplots. Volumes is the only partition without outliers, and that present a decently differentiated distribution between the two cases, the difference between the means in the two conditions is greater than two standard deviations. Here the values of complexity in the cases in which the animals are present, result much higher than the ones for the earthworm-less samples.

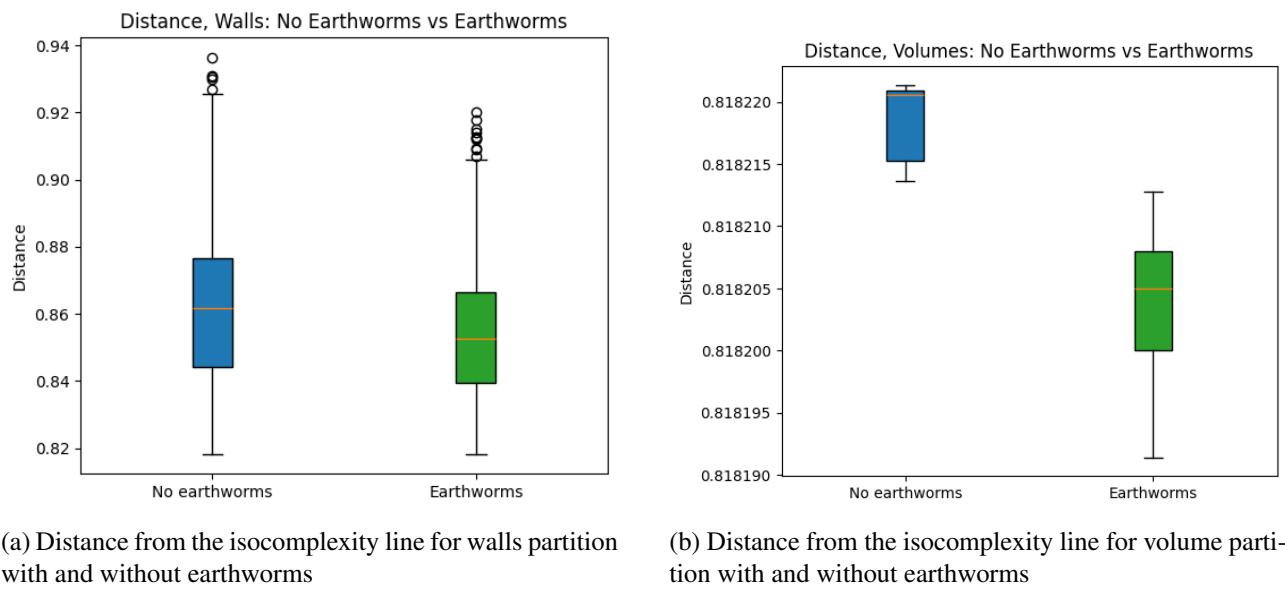


Figure 16: Most meaningful comparisons of distance between presence and absence of earthworms

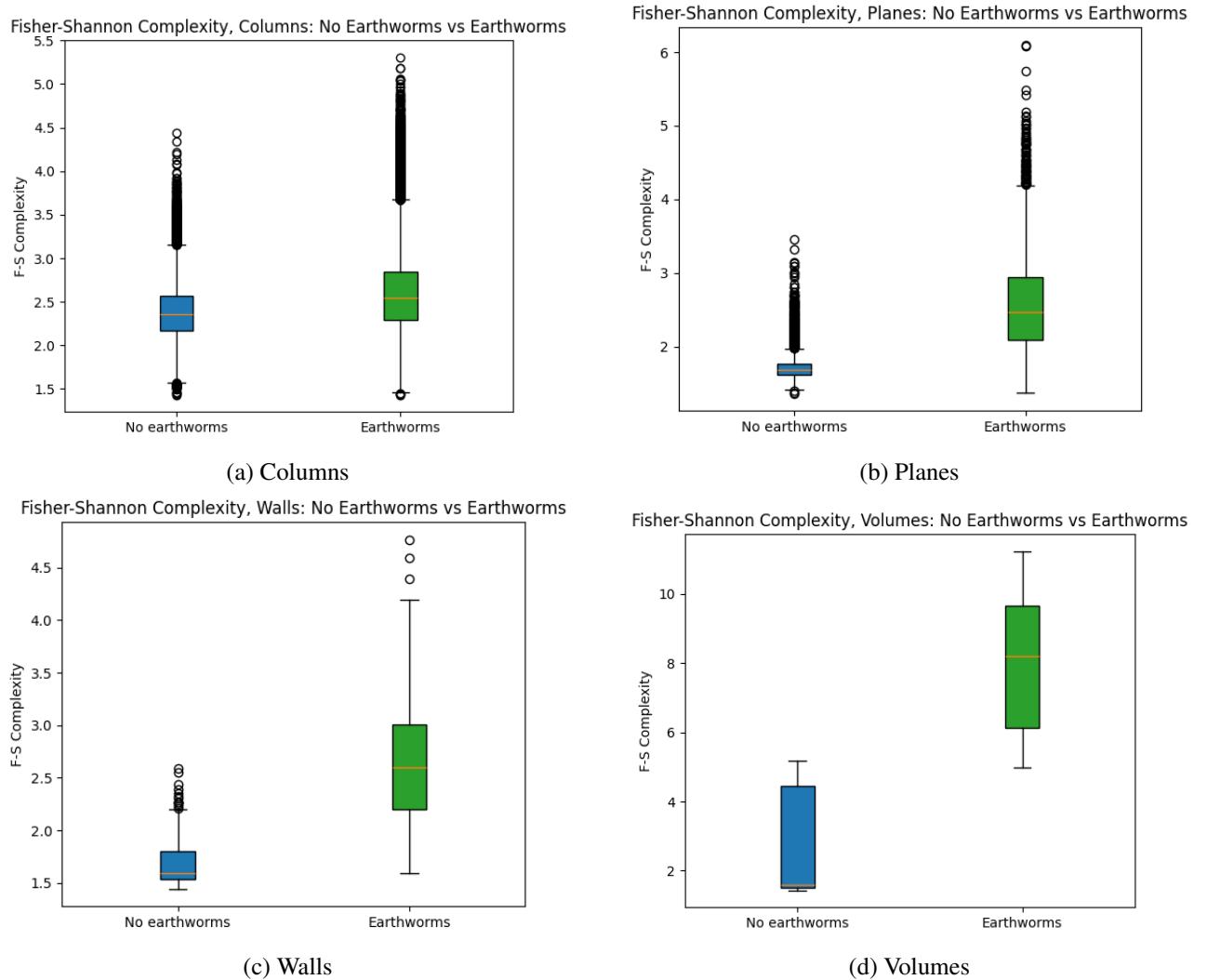


Figure 17: Boxplot of Fisher-Shannon Complexity values for the samples without and with earthworms, for all the partitions

We observed also that the Fisher Information Measure, with the volumes partition, is alone able to produce a very good differentiation between earthworms presence and absence, as can be seen in figure 18. Shannon Entropy Power on the other hand is not able to discriminate between the two conditions.

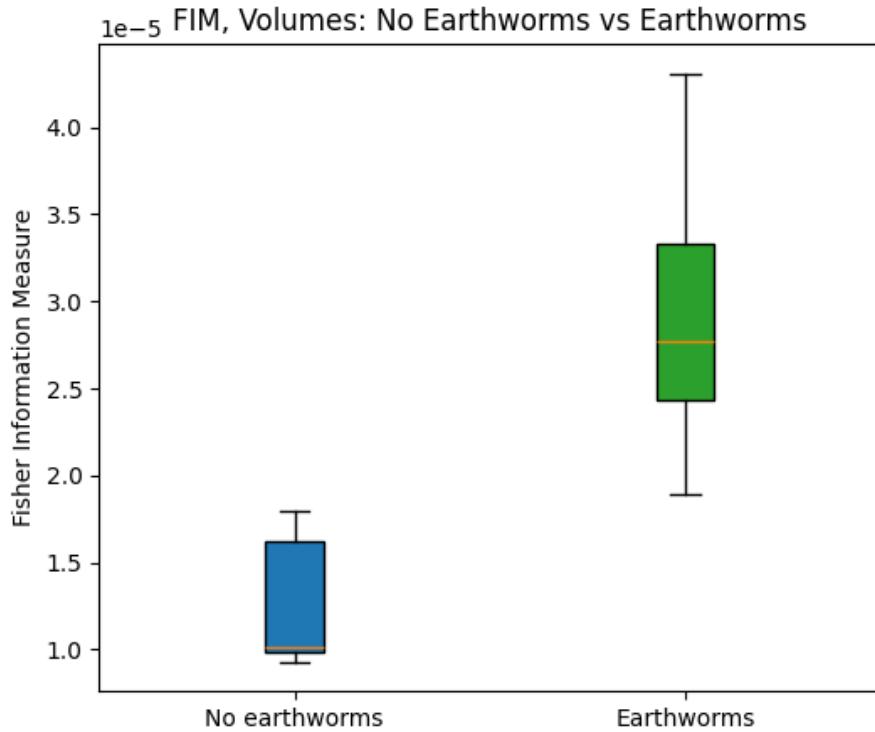


Figure 18: Boxplot of Fisher Information Measure values for the samples without and with earthworms, for the Volumes partition

5.3 Correlation between statistical measures and other soil quantities

In this section, we explore whether the quantities that we calculated with the tools of information theory correlate with the other quantities that have been measured about the system and are summarized in tables 2 and 3. To do so we considered the volumes partition and we created one single dataframe with the results for all the samples, we then added the information about the chemical and biological characterization of the samples. In figure 19 we plotted an heat map for the correlations between all those quantities across all samples.

Ignoring the obvious correlations between the statistical quantities themselves, we notice that the earthworm total mass is the only metric that correlates quite well with the measures of information theory. Specifically, the correlation coefficient of the mass with Fisher Information Measure is 0.9, which means the two quantities are strongly related to each other.

In figure 20 we plotted the values for Fisher Information Measure over the values of the total mass of earthworms measured at the end of the experiment. We can see that the different conditions of the samples result in different positions on this plot. In particular, the presence of anecic earthworms result in a higher value of the total mass and FIM.

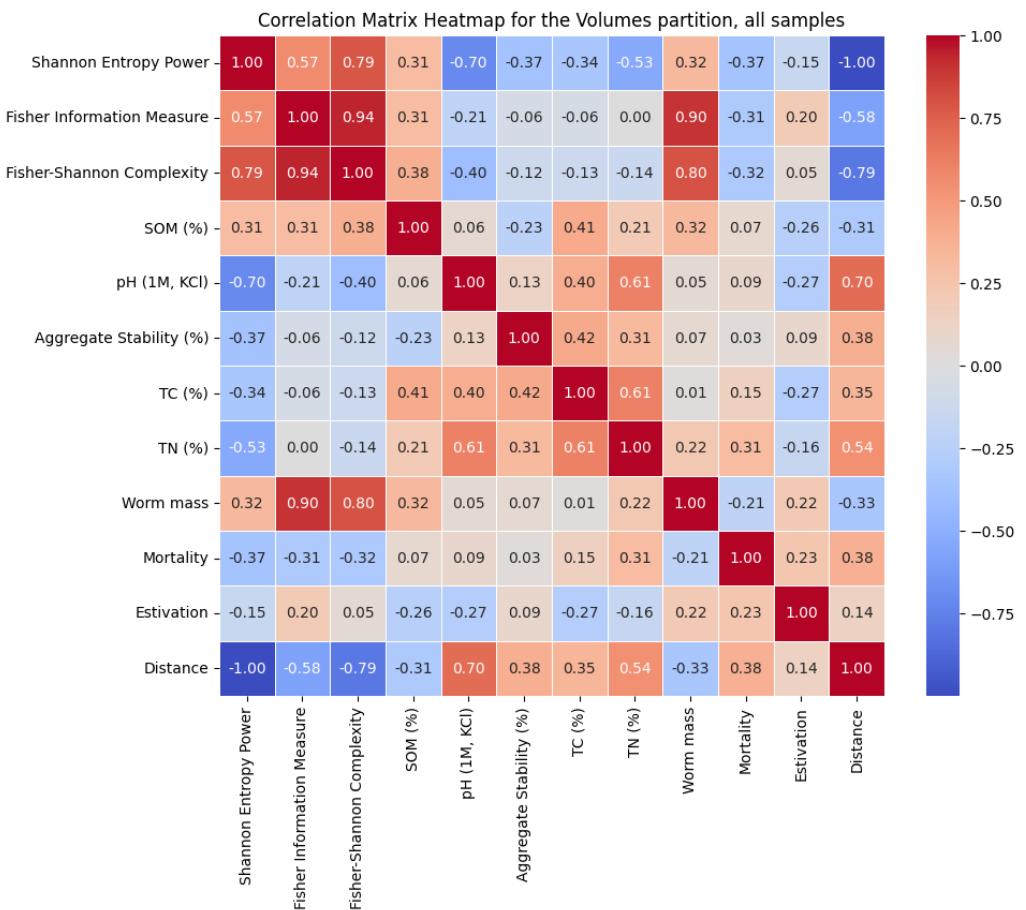


Figure 19: Heat map of the correlations between all the quantities measured for the system, considering the volumes partition of all samples

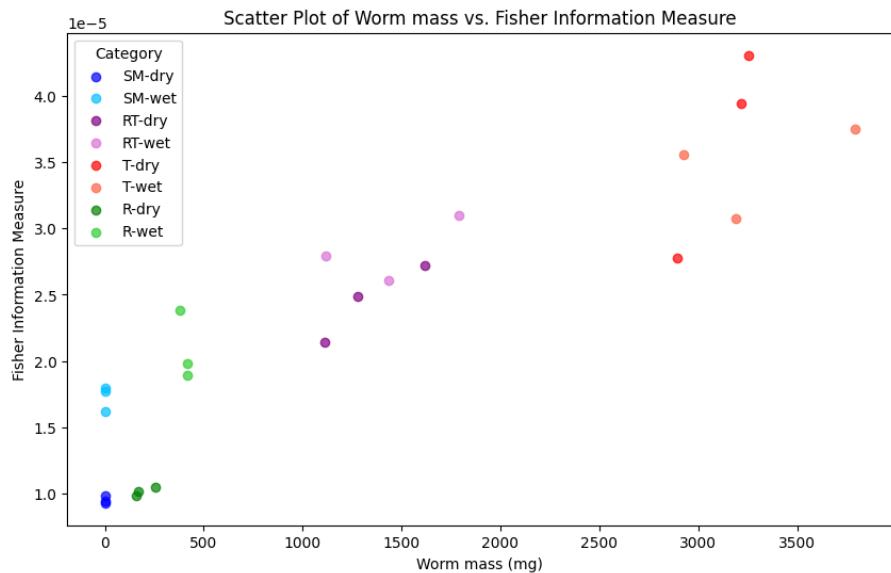


Figure 20: Scatter plot of Earthworm total mass at the end of the experiment and Fisher Information Measure for all the samples, the points are color-coded with the category to which they belong

5.4 Statistical measures and pore space characterization

Here we will present the results concerning the topological measures of the pore space, investigating whether those properties correlate with some of the quantities that we derived from the information-theory based analysis of the system. We found that the correlations are the strongest with the volumes partition, which makes sense since the properties of the pore space refer to the whole volume. First of all, we point out that our segmentation algorithm allows for a clear visualization of both the fragmented pore space in earthworm-less samples, composed of small, isolated voids, mostly resulting from the packing of soil aggregates; and the visualization of the complex and intricate burrow system formed by the animal's activity. This last feature is much more evident in the samples with the presence of anecic earthworms, as they create more stable burrows systems and they tend to not refill those with casts, as opposite to endogeic typical behavior. At our resolution, the samples without anecic earthworms seem to lack any kind of continuous porous system. In figure 21 we show the 3D renderings of the pore space of four samples, all in wet conditions, with different earthworm populations, so to point out qualitatively the differences between the pore space generated by different animal's behavior.

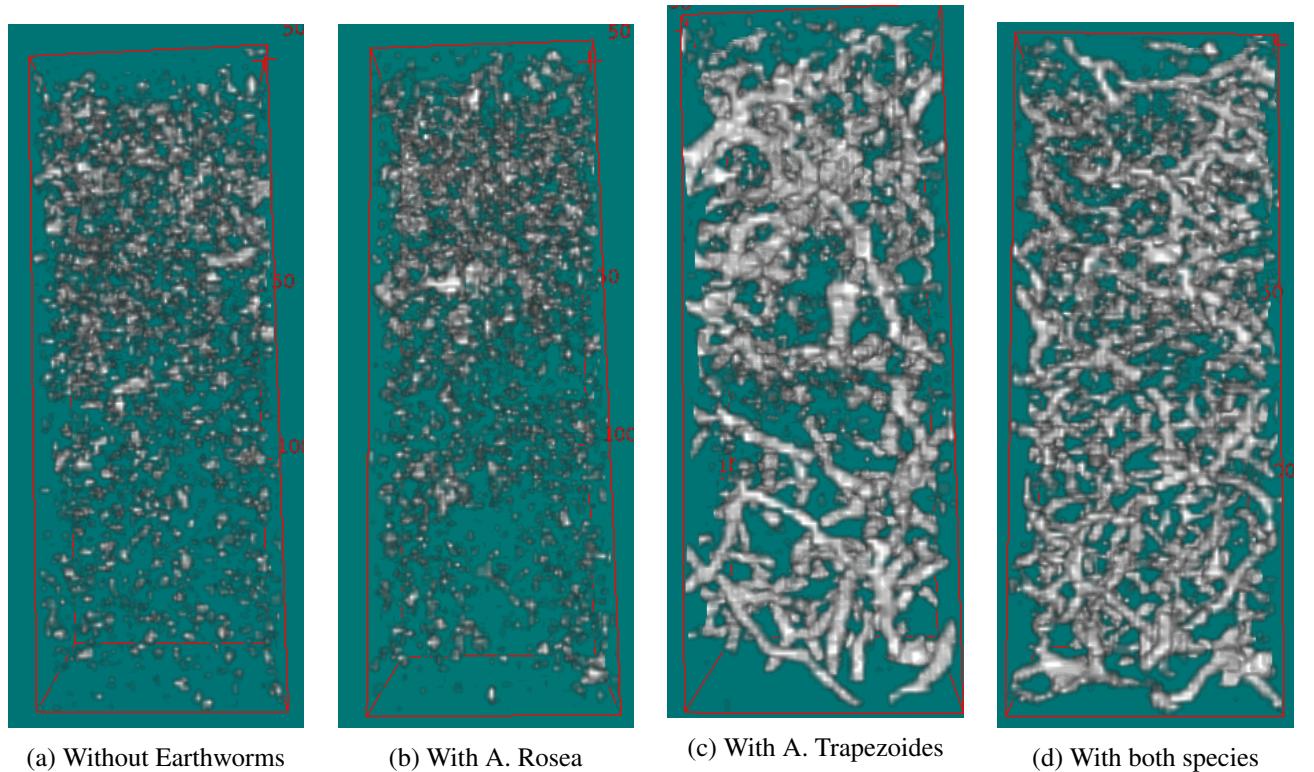


Figure 21: 3D visualization of the pore space of four samples in wet conditions with the presence of different earthworms populations.

In the plots in figure 22 we display the values of the four Minkowsky functional for every sample. We can immediately recognize that M_0 and M_1 have very similar trends across the samples, assuming higher values in presence of earthworms and in wet conditions.

On the contrary, the mean curvature seems to assume lower values for wetter soils, and it is not so strongly influenced by the presence of earthworms. For all the samples the value is positive, meaning that the porous architecture is mainly composed of convex voids.

The Euler characteristic assumes higher values for samples without earthworms or with just the endogeic species, showing how the biological conditions can alter dramatically the number of single, isolated pore structures.

In figure 23 we show the heat map of the correlation matrix for the four Minkowsky Functionals with the four main statistical quantities object of our analysis (Shannon Entropy Power, Fisher Information Measure, Fisher-Shannon Complexity and Distance from the isocomplexity line). The strongest value of correlation is between the pore volume (and the surface, the two topological quantities are very strongly correlated between each other, so they correlate similarly to the other measures) and the Fisher-Shannon complexity, with a value

of 0.97.

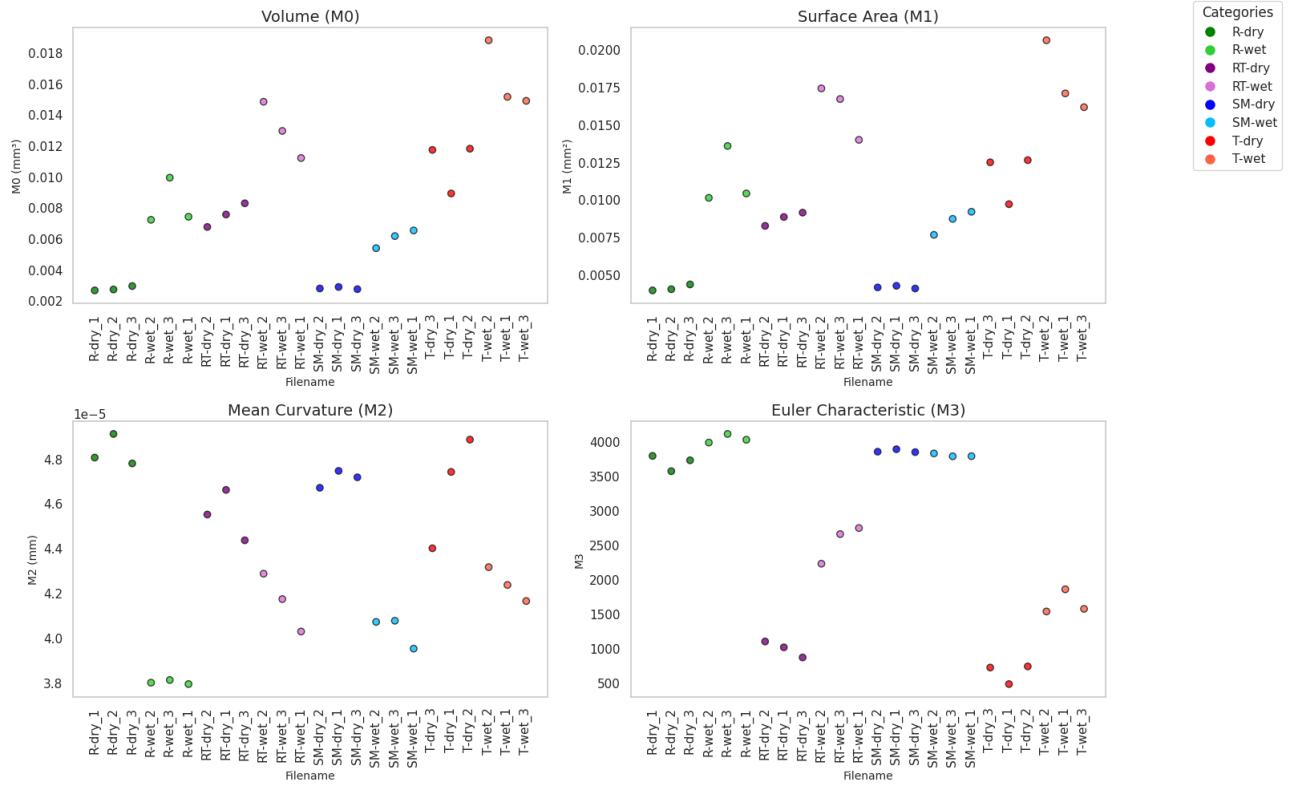


Figure 22: Plots of the values of the four Minkowsky functionals for every sample, color coded by category.

Another high (negative) value for of the correlation matrix is the one between the Shannon Entropy Power and the mean curvature M_2 , suggesting that high entropy is associated to small vales of this Minkowsky functional. A last meaningful correlation is the one between the Euler characteristic (M_3) and Fisher Information Measure; it appears that the higher the number of isolated structure, the lower the value for FIM, as it should be expected from the functional definition and meaning. In figure 24 we show scatter plots for the correlated quantities described above; we can appreciate the evident linear trend in the first figure, the very defined separation between wet and dry cases in the second plot and the demarcate segregation of the samples' categories in the third one.

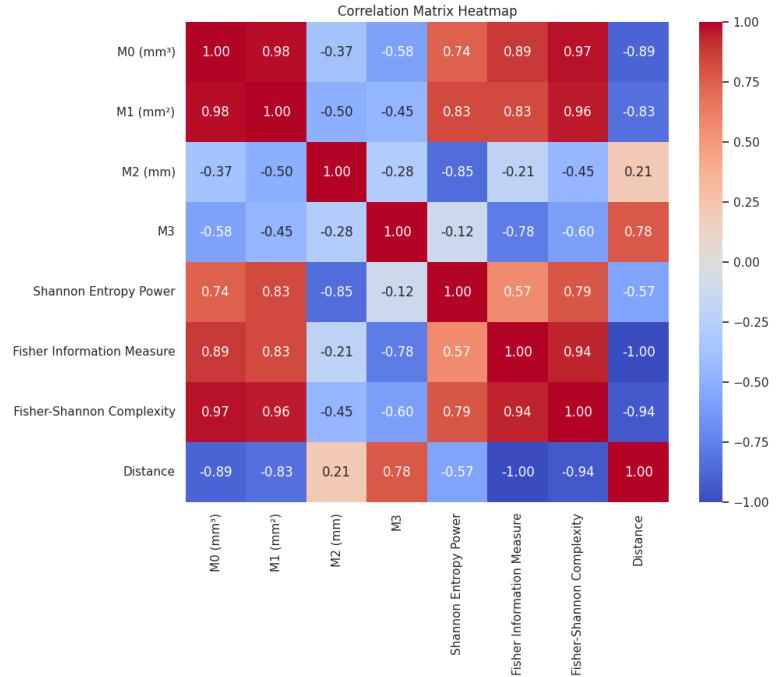
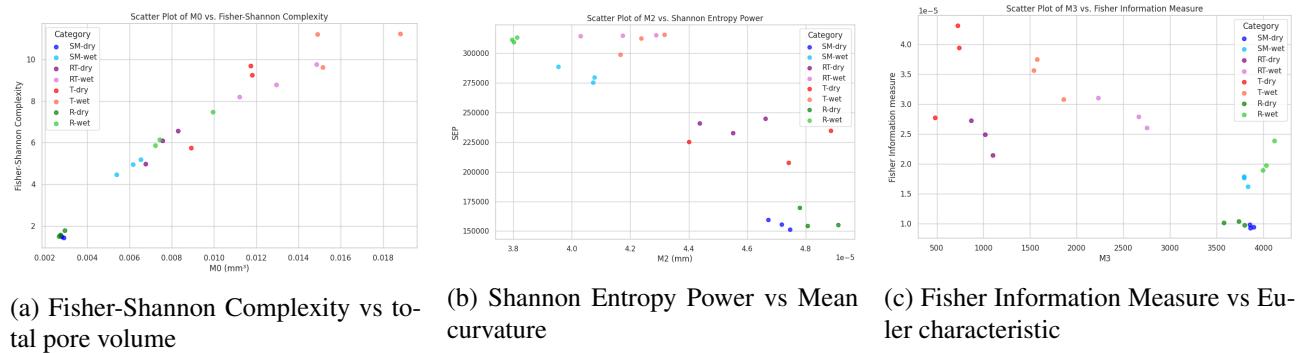


Figure 23: Heat map of the correlations between all the information theory quantities and Minkowsky functionals, considering the volumes partition of all samples



(a) Fisher-Shannon Complexity vs total pore volume (b) Shannon Entropy Power vs Mean curvature (c) Fisher Information Measure vs Euler characteristic

Figure 24: Correlated quantities from information theory and pore space analysis for all samples

6 Discussion

In this chapter we will draw meaningful considerations about the system studied on the basis of our analysis and about the analysis itself. We will try to make sense of our results from a physical and biological prospective, clarifying what are the processes that are likely to influence the metrics we choose to consider. While doing that, we will discuss the opportunity of each one of the passages of our study, considering both the methodologies applied and the results obtained. This will lead to considerations about the best way to study soil systems with this approach.

6.1 Depth influence on information theory measures

The plots in figures 5 and 6 show clearly that the metrics that we choose to consider are very sensitive to depth. This should not come as a surprise but rather as a welcomed confirmation, it is widely known and quite logically obvious that soil structure features are greatly dependent on depth. As an example, in their multifractal analysis of soil porous structure, Martínez and colleagues [65] observed consistent and strong variations of porosity with depth.

The first observation we made was about the fact that dry samples present a higher variability between one plane and the next, resulting in noisier curves, that oscillate much more than the ones derived from wet samples. One possible justification for this is methodological: the act of watering performed on the wet soils could in itself be responsible for a more homogeneous soil between layers, as the repeated percolation of newly added water is cause of material exchange between different layers of the sample. Even without considering this particular interaction, soil structure and functionality is hugely impacted by the moisture level, which influences physical properties of the soil aggregates and metabolic rates of the soil biome. Since water is the main means of transport of material and nutrients in the soil, it is reasonable to think that more humidity leads to more mobility of the finer soil components, ultimately resulting in a mitigation of structural differences on a small scale.

The statistical quantities generally exhibit very well defined trends, with the exception of FSC, which oscillate greatly without showing particular tendencies, except maybe in the SM-wet and R-wet cases, in which it timidly decrease its values with depth. This suggests that the general complexity of soil, as calculated by the FSC, does not correlate with depth, but the kind of complexity most definitely does. The two most significant quantities are SEP and FIM, with the distance that show trends comparable to SEP. The two functionals appear to have an opposite behaviour in response to depth, specifically, the entropy is higher on the surface and decrease at the bottom and the level of structural organization does the opposite. This indicates that the soil layers closer to the surface are characterized by a higher degree of disorder, the soil structure here is more homogeneous; and the deeper layers present a more organized and complex soil architecture; with higher values for pore space. Our findings can be confronted to a visual analysis of the different layers of soil, in figure 25 we show the first and last layer of one of the samples with A. Rosea in wet conditions.

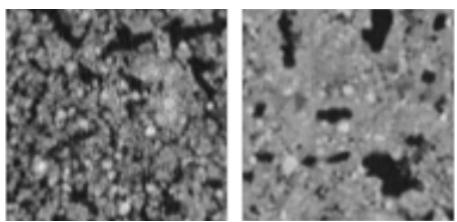


Figure 25: First and last layer of the sample R-wet2

We can here appreciate the two different kinds of complexity of the system: in the top plane it is harder to spot single structures (pores, different macro-aggregates) and the picture appears more chaotic and noisy, this condition is associated with low values for FIM and high entropy. The bottom plane, on the other hand, present a much more well divided image, where pores and solid elements are easily recognisable and the larger part of the solid phase assumes only a narrow range of intensity values; this configuration show a high degree of structural organization, therefore a high value of FIM and low entropy. The reason for this difference is mostly physical: the pressure forces the soil to be less disperse, reducing the macro- and meso-porosity and

compacting the solid phase of the system into a more uniform material. To continue with visual confrontation, the particular shape of SEP and FIM for the dry samples without earthworms, seems to be in qualitative accord with the porosity of the layers.

This differences along the z direction can be appreciated also looking at a 3D picture of the pore space: in figure 26 we can clearly see a region right above the middle that is much less populated by porous structures. This structural feature is reflected in our data by the depression in the curves for FIM and the peak in the curves

for SEP in figure 5c.

This empirical confront suggests that the statistical indicators yield results that are actually in agreement with common sense, and they are deeply influenced by the pore space local characteristic, as they should. This specific tendency is lost in the samples with anecic earthworm presence, probably due to the layer-mixing activity of this kind of earthworms during the excavation of burrows, which homogenizes soil structure, resulting in curves with much more linear trend, less pronounced slopes and no evident concavity.

In wet samples the correlation of the quantities with depth seems to start only after roughly the half of the sample. The uppermost layers of soil present a very uniform distribution of values for all the quantities. Since water permeates easily and uniformly in the first layers of soil we suggest that this tendency could be explained by humidity-induced phenomena. In fact, water is the main responsible for movement of materials in soil and more water means essentially more mobility for the small soil particles, possibly resulting in a more homogeneous distribution of those across the soil structure. Water presence also boost microbial activity, potentially increasing the production of biofilms and other materials which have the potential to create a less variable soil architecture on smaller scales.

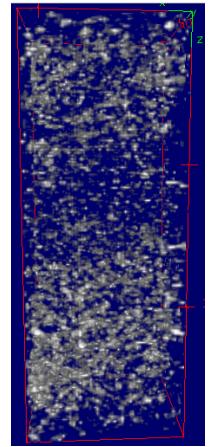


Figure 26: 3D rendering of the pore space for R-dry_2

6.2 Considerations about the different partitions of the volume

The pioneering nature of this research induced us to experiment our analysis on different partitions of the soil volumes. To our knowledge, this kind of soil complexity characterization was performed only once, by Aguiar and colleagues[3], little more than one year ago. In their work, the researchers said that they also performed the analysis on different partitions of the volumes, but they just report the results obtained with the Columns partitions, stating that this was the one yielding the best results for their study. Here we aim to discuss the different performances of the various partitions in identifying the physical and biological conditions of the samples, giving plausible reasons to justify the results.

The Columns partition was, in our case, the one with the worst performances: the average of the statistical quantities calculated are slightly different between the samples in different conditions, but the standard deviation is very large in all cases, making it impossible to attempt an effective classification of the samples. This can clearly be seen in figure 7, where it is evident that this partition is not able to distinguish between the wet cases with earthworms and also without the animals. It appears that this partition is particularly bad at discriminating between conditions with FIM values, especially in the wet cases, for which all the samples assume very similar values, regardless of their biological conditions. We suspect that the reason behind this result is the limited size of the dataset, since the pdf is calculated over only 240 values of intensity for each column, less than one third of the points considered in the work of Aguiar. The small set of data points and the high number of columns per sample result in a huge variability of the pdf shapes, and therefore of the functionals calculated in our analysis. For the reasons expressed above, this partition is not able to discriminate effectively between either moisture or biological conditions. Furthermore, this partition is by far the most computationally intensive, with a required time for calculations that is five to ten times larger than the ones required from the other partitions. This is due to the fact that the Columns case requires to calculate a very large number of pdfs (8.464) for every sample.

The Planes partition present similar features to the Columns one, with a slightly better separation between the categories but still a very high variability. Our hypothesis is that this big standard deviation of the statistical values inside the same sample is due, in this case, more to the inherit differences in the structure of soil at different depth than to the size and number of the subsamples. Even though this partition is not able to proficiently classify the samples in categories, it is useful to draw some considerations about the viability of our analysis in identifying structural features of soil that are typically affected by depth. It is worth noticing that Fisher-Shannon Complexity and distance values for this partition present a much smaller variability when earthworms are not present, suggesting that the biotic influence of those animals is particularly impactful on the variability between subsamples.

The Walls partition is the one that clearly gives us the best results in terms of category segregation of

the values of the pdf functionals considered. In figure 8, we can appreciate the clear division of the various categories inside the Fisher-Shannon Information plane. The values for the wet cases with earthworms are clustered all together since they have very similar values of entropy, but differentiate nicely if considering the Fisher Information Measure. In figures 11 and 9 we can appreciate the limited standard deviation of the distance and FSC values. Again, the variability is considerably smaller in the samples without earthworms, stressing their relevance as a source of noise in the system. We think this is the partition that performed better for at least two reasons: The first is that, with 22.080 voxel in each wall, this is the partitions with the largest subsamples (excluding the Volumes partition); the second, probably more important, is that by considering vertical slice, we are able to construct the pdf on data that already contain the variability in soil structure that comes with depth. In conclusion, this partition seems to be the best balance between quantity of information considered in the single subsample and number of subsamples per sample.

The Volumes partition also yields excellent results, both in terms of segregation of the samples in the Fisher-Shannon Information plane, and of the discrimination efficiency between different physical and biological conditions. This was quite expected, since we were supposing that a probability density function calculated comprising all the information present in the 3D picture would be better in characterizing specifically the single samples. In our case, though, we have only three samples for each condition, making it hard to generalize our conclusions, new experiments with a higher number of replicates should be carried on to assess the consistency of the results obtained in this way. Furthermore, with this choice, we lose every kind of information about the spatial variability of the quantities of interest inside the volume. In our context, this aspect is not that relevant because the system was very well controlled and artificially constructed, but in soils in natural environments, the presence of other elements (rocks, roots, large animals and others) could be a important factor to consider and partitions of the samples a tool to identifying those disturbances, and not get misleading results from the analysis of the whole volume.

6.3 Fisher-Shannon Information Analysis

Here we will discuss the values we obtained for the complexity metrics for the various samples. The representation of the data into the Fisher-Shannon Information plane yielded exciting results for the two partitions that performed best: Walls and Volumes. For the reasons discussed in the paragraph above, in fact, the Columns and Planes partitions have a variability too high to allow for a clear separation of the plane, as can be clearly seen in figure 7. In figure 8, on the other hand, we can appreciate how with both Walls and Volumes, a clear partition of the information plane seem to emerge. In particular, we notice that the samples for SM-dry and R-dry occupy the same left fraction of the plane, which we would expect since the R-dry samples ended up virtually without earthworms. The values for the dry samples with earthworms all fall inside a central region of the plot and the wet samples with earthworms yielded values that occupy the rightmost region. Also the samples with anecic earthworms tend to assume higher values of FIM than the ones with endogeic animals. The sample SM-wet have similar characteristics to the R-wet samples but it remains separated from them.

This compartmentalization of the Fisher-Shannon Information plane suggest that a Nearest Neighborhood Classifier (NNC) could be used to classify soil samples at different moisture level and in different biological conditions. To confirm this hypothesis, more experiment with a much higher volume of replicates and different conditions should be performed. We believe that this should be a very useful instrument to have a fast, qualitative estimation of the physical and ecological conditions of soil, without the need to alter or destroy the samples. In figures 7 and 8 we also plotted the normalized Fisher Shannon Information plane, where Aguiar's process of normalization[3], described in the 'Materials and Methods' section. This transformation result in a much more mixed distribution of the various categories inside the planes. With this operation, the plane is not separable in different areas associated to one conditions, instead the points all occupy the same region. This is the first indication that this process is probably not the most indicated to deal with the measure of distance from the isocomplexity line in our case.

6.3.1 Fisher-Shannon Complexity and Distance from the isocomplexity line

The values of Fisher-Shannon Complexity (figure 9) are generally higher for the samples in wet conditions and the ones with the presence of earthworms. Indicating that the soil structure gains complexity when wet and also when it features biotic structures, like the earthworms' burrows. As before, the columns and lanes

partitions exhibits too large variability to allow a discrimination between the various conditions of the samples. We noticed that in the dry cases without earthworm presence (SM-dry and R-dry), the values not only are lower, but also much more consistent, inside the single sample, and between the different replicates. This observations indicates that in the absence of earthworm activity, and in dry conditions the soil architecture is more spatially consistent and less chaotic. The Walls and Volumes partitions yields better divided results for the different categories, suggesting that this metric could actually be used for classification. Looking at figure 10, we notice that the samples in the same moisture conditions have different results according to the earthworm population presence. Also, wet and dry cases have different values in the same biological conditions. But, for the SM-wet samples, we have similar values to the R-wet and RT-dry cases, so, with no information about the humidity level, it might be hard to discriminate between presence and absence of earthworms, as the wetness of soil seems to strongly influence the worms' soil structure-modifying activity.

The Distance for the isocomplexity line yields controversial results. First of all we want to point out that this value does not seem to be particularly significant to the scope of classify the samples in the categories. As can be seen in figure 11, the values do not exhibit at all any recognizable trend for what concerns earthworm presence. It seems, though, that higher values of this complexity measure are obtained in the dry condition, this can be seen clearly in figure 12. The variability associated to this quantity though does not allow of thing of it to be a good metric to discriminate our samples. We notice that this measure suggests that the soil structure is more complex when the sample is maintained in dry conditions, this result is in contrast to all our other indicators and with our logic. We will try to give some justifications for this in the 'Considerations about Information Theory metrics' section.

6.3.2 Differences between wet and dry conditions

From a theoretical point of view, the level of water abundance impact drastically soil structural organization and its architecture, by means of the complex physical, chemical and biological interactions between water, soil solid elements and the microbial and fungal communities that mainly constitute the biotic part of the soil system.

With our investigation we want to asses if the Information Theory metrics are able to discriminate between wet and dry conditions. We found that The best metric to do so is Shannon Entropy Power: in figure 15, we can clearly see the sharp difference between the distributions of the values of SEP in the two moisture conditions. In particular, the Entropy of soil structure assumes higher and more consistent values in wetter conditions. This can probably be explained by water physical action, that homogenize the structure by facilitating materials' transport, and produce image with more white uncorrelated noise. We also want to point out that FSC and Distance measure show opposite tendencies in this case. For the measure of complexity obtained with FSC, we have that wetter soils are more complex, the opposite is true when we refer to the distance measure. Even though none of the two metrics is actually able to produce a differentiation between the conditions, this controversial result seem to be worth noticing.

We also notice that the plane in figure 24b, in which we plotted SEP vs. Mean curvature, produces an exceptionally clear separation of the two categories; with all the wet samples in the upper left corner and all the dry ones in the bottom right one. Already in figure ?? it was clear that moisture have a main influence on the shape factor M_2 , and the reason is that dry solid aggregates and wet solid aggregates pack themselves differently. In the dry conditions packing voids have a less circular and more fragmented shape.

6.3.3 Earthworm presence and absence

The presence of earthworms is a very important alterative factor for soil macro-structure; different species of earthworms creates burrows systems with very different topological features. We hope, with our analysis, to be able to discriminate between the presence and absence of earthworms in general, and between the presence of animals that belong to different ecotypes and therefore have different behavioral patterns.

For what concern this classifications, for both Distance and Fisher-Shannnon Complexity measure, only the Volumes partition is able to generate distributions different enough for the samples in the two biological conditions (with and without earthworms). As in the wet-dry case, we get an indication of the complexity of the structure that is controversial. The Distance measure seems to indicate that soil unperturbed by earthworm have a more complex structure than earthworm-filled soils. This conclusion goes against the indications that we had

from the other metrics and also against our common sense, discussions about the distance metrics will be carried out in the next section. We also found that the best metric to discriminate between earthworm presence and absence is in fact Fisher Information Measure, as can be seen in figure 20. This can be explained by the fact that this measure quantify the level of structural organization of a signal and the earthworms' impact on soil architecture is, in fact, the creation of organized burrow structures. It should not come as a surprise then, that the FIM value for the samples that featured the animals' presence are much higher, indicating a more organized soil architecture.

6.3.4 Considerations about Information Theory metrics

We want here to draw some considerations about the performances of the various metrics calculated from Fisher Shannon analysis and used in our study.

First, we want to address the Distance from the isocomplexity line as a measure of complexity. This indicator gave us counterintuitive results, suggesting that dry, earthworm-less soils have a higher degree of structural complexity than the wet, populated counterparts. We choose, though, to not completely trust this value though because of the very method that we used to calculate this quantity. We are referring, in particular to the process of normalization described in the 'Materials and methods' section and copied from the work of Aguiar[3]. As can be seen in figures ??, this operation mix together the data points of the different samples in the Fisher-Shannon Information plane, compromising the partition of the plane that naturally occurs with the not normalized measures. Aguiar and colleagues choose to simply multiply or divide the values of FIM and SEP for the maximum value of SEP, so to preserve the value of Fisher-Shannon Complexity and rescale the Fisher-Shannon Information plane. In their case, though, this process resulted in a plane with quantities of the order of the unity for both the axes. In our case, the normalization leaves us with the values of FIM that are roughly one order of magnitude greater than the ones of SEP. It could be argued that a more careful choice of the rescaling parameter could lead us to a situation similar to the one obtained by Aguiar and colleagues. In our opinion, though, this process was done rather arbitrarily, without a solid theoretical justification. For those reasons, we choose to give marginal relevance to the results obtained by the calculation of the Distance from the isocomplexity line.

Fisher-Shannon Complexity appear to be a valid measure of complexity for our system, the results are consistent with the ones obtained from other indicators and with our initial intuitions. This quantity is not very effective in discriminating between moisture conditions, but it is very efficient in doing so according to the biological conditions of the sample.

According to our results, though, the two most informative metrics are the Shannon Entropy Power and the Fisher Information Measure themselves! In particular, it appears that the entropy of the system is very much dependent on the moisture level, while the most impacting parameter for the Fisher Information Measure is the presence of earthworms. This makes all the sense, since SEP is able to capture alterations in the spread-ness of the pdf, making it sensible to factors that alter uniformly the soil structure, as moisture changes. On the other hand, Fisher Information Measure, is an instrument that focuses on identifying structural organization of signals, and is therefore logically very sensible to the introduction in the system of agents that basically create structures where there were not.

6.4 Fisher Information Measure and Earthworm Population

In figure 19 we showed that the Fisher Information Measure is strongly correlated with the value of the total worm mass measured at the end of the experiment. In the plot in figure 20, we can appreciate the linear correlation between the two quantities and the sharp separation of the data points in the different biological conditions. Specifically, the samples with anecic earthworms have much higher values of mass and also higher values of FIM. This confirms that the *A. trapezoides* are considerably larger animals than the *A. rosea* and they produce more distinctive structures inside the soil. It is the result we expected, based on the ecology and behavior patterns of the two species. It seems that indeed the value of the Fisher Information Measure is sensible to the presence and abundance of earthworms and of anecic ones in particular.

The good linear correlation lead us to think that FIM could be an effective, fast and non-disruptive way to estimate the population of anecic earthworms inside a soil sample. To test this hypothesis, a study should be carried on featuring replicates with increasing population of anecic earthworms, so to find out over what range of individuals' number and other environmental conditions the correlation holds. Using this method to

quantify earthworm presence could help soil researchers to have an estimative of not only how many individuals are present in the sample at that time but also how many frequent that patch of soil on average. The structural stability of anecic-constructed burrows system allows to make considerations that refer to a longer period of time.

6.5 Pore Space

We refer now to figure 21 to draw some qualitative observation about the different porous structure of samples in the four biological conditions tested. Here we can appreciate that the anecic earthworm presence is the factor that really changes the features of the pore space, transforming a fragmented collection of isolated amorphous void spaces into a well organized and chaotic network of well defined burrows. The samples that contains only *A. Rosea* present a pore space that have similar feature to the one without earthworms. This is due to the behavior of the endogeic earthworms, that form a more discontinuous and unstable burrow system, often refilling the tunnels with casts. We point out though, that a finer resolution of the X-Ray CT scans could enable the identification of the burrows of also his species. In fact, the endogeic earthworms are considerably (2/3 times) smaller than the anecic ones: their diameter is roughly of the same size as the resolution, making it virtually impossible to effectively identify the burrow system. With higher resolution (of the order of $(1 - 10)\mu\text{m}$) it should also be possible to observe interconnected pore network in the absence of earthworm influence, as the packing voids often form an interconnected and complex pore space at those scales. This consideration got us to the conclusion that the analysis at this resolution could be very effective to identify and possibly quantify the presence of this specific ecotype of earthworms. We can also notice that the pores are more present in the upper layers of soil, this is due to the pressure that, in deeper soil, forces the aggregates closer to each other, eliminating most of the bigger void spaces. For the same reason, in the samples with *A. trapezoides*, even though the burrows are quite homogeneously distributed across the whole volume, near the surface, the soil matrix present a higher incidence of smaller porous structures than at the bottom. Another qualitative observation is that the burrow system in the samples with both species seems to be composed by a larger number of smaller tunnels. Our hypothesis is that, in mixed condition, the *A. trapezoides* individuals remain smaller than when they are alone, resulting in the excavation of burrows with a smaller diameter. This idea is supported by the fact that the average mass of an anecic worm is 546mg in the T-wet samples and only 425mg in the RT-wet samples. The higher number of smaller structures lead to a higher value of entropy and a smaller value of FIM for the samples in mixed biological conditions.

6.5.1 Minkowsky Functionals

From the analysis of the Minkowsky functionals (figure 22) we can draw quantitative considerations about the pore space and its relationship with the Fisher-Shannon analysis of the system. The higher values of M_0 and M_1 in wet conditions with earthworms is, quite obviously, consequence of the burrowing activity of the animals and of the increased porosity of wetter soils. In particular, the anecic species seems to have a deeper impact on soil porous structure, rendering the soil more discontinuous and enlarging noticeably the pore space. The fact that surface area scales with volume in the same way for all samples suggest that the average shape of the pore structures is not altered dramatically in terms of convexity by the different physical and biological conditions. The mean curvature value (M_2) not being strongly influenced by the presence of the animals, also point to the conclusion that, with this particular type of soil, the burrowing activity does not strongly affect the mean shape of the pore space while strongly increasing its extent. We are aware, though, that this could be also the result of the approximations made (marching cube approach) and of the very coarse resolution of our samples, which does not allow to consider smaller packing voids between the soil aggregates. The Euler characteristic (M_3) is considerably lower in than samples with *A. trapezoides*, this is what we would expect, since the anecic individuals tend to create large and interconnected porous structures, effectively diminishing the numbers of individual objects in the pore space. The animals belonging to the *A. rosea* species, even in wet conditions, were not able to produce a significantly lower value for M_3 in respect to the case without earthworms; probably due to the burrowing behaviour of this species, which results in a more fragmented pore space, since the burrows are instable and often refilled with casts.

As can be seen in figure 23, the Minkowsky functionals values correlate quite well with the results that we got through the Fisher-Shannon analysis of the samples. The strong correlation values of M_0 and M_1 indicate

that the quantities that we calculated from information theory are very sensible to the extent of the pore space, as it is easy to qualitatively understand. A less obvious relationship is the one between SEP and M_2 that appear to be strongly inversely correlated, but we can try to understand this relationship. A very high absolute value of the mean curvature would be associated to a pore space where the voids have very similar shapes (curvature, to be precise), and a value near 0 would indicate that there are voids with both positive and negative curvature quite uniformly distributed (so that the average of the curvature is near 0). Considering this, and remembering that all the values for M_2 are positive in our case, it is easier to understand that a higher value of mean curvature would come from a more organized system and, on the contrary, very small values of M_2 would derive from a quite random distribution of voids, therefore explaining the negative correlations between the two quantities. The inverse correlation between M_3 and FIM is easily explained, since this latter quantity is a measure of how structured is a signal, it makes sense that it correlates negatively with the number of individual structures present in the system, as the lesser, bigger structures are present, the more structured the space is. In figure 24, we can appreciate how well the Fisher Shannon Complexity linearly correlates with porosity values, suggesting that the latter quantity could be represented by the former, eliminating the need for the controversial binarization process. We also notice how well the mean curvature separate the samples between wet and dry conditions, indicating, counterintuitively, that moisture level is a much more important factor to determine the shape of voids than the earthworm presence. Euler characteristic in combination with FIM also produces a very clear separation, but this time between samples with and without anecic earthworms, clearly separating the formers in the four categories. This reinforces our suspicion that our analysis at this resolution could represent a useful tool in population dynamic studies of this specific biological group.

6.6 Variability between the samples of the same category

In general, the samples in the same conditions regarding moisture and earthworm presence and type, present similar results between them. This suggest that the analysis implemented is quite robust and not excessively dependent on the specificity of the single sample.

It is worth noticing that the samples without earthworms and in dry conditions are much more homogeneous between them, pointing out the crucial impact of moisture and animal activity. In the systems with earthworms, the variability is much higher, this should not come as a surprise, since biological elements are always a huge source of uncertainty and stochasticity. In fact, even if it could seem not the case to a human eye, every earthworm is different from the next and will behave slightly differently even in the same environmental conditions. Granted that the behavioral pattern are generally the same for all the individuals belonging to the same species, every animal act in response to a complicated elaboration of a wide set of external and internal stimuli, which will inevitably be slightly different for distinct subjects. Increased variability is something that we should always expect when a biological element is introduced to a system, mainly because we are introducing what is effectively a complex system on its own inside a context in which it is going to behave in a complex way; our analysis is able to show this quite clearly.

Our data do not allow to make similar considerations about the moisture level, it is clear that the presence of animals is a much stronger factor that increase variability between samples in the same conditions. However, the metrics obtained for SM-wet are more dispersed than the ones from SM-dry, indicating that a wetter soil can be in a wider range of conditions than a drier one. A possible justification for this observation could lie in the very way water interacts with complex porous systems, where a very small difference in the structure can alter dramatically the liquid behaviour. Another reason could be that, with more water available, the activity of the soil microbiome increases, leading to a stronger biotic influence in soil functionality and so in its structure.

Generally, the categories present consistent rates of mortality and estivation between the samples 3, it would be interesting to understand if variation in those parameters would result in an increase variability between the samples. We have no information, though, about the point in the experiment at which the animals died or entered the estivation state, making it impossible to use this data as real a proxy for earthworm activity. New experiments that feature a more complete survey of population dynamics should be performed, to assess the true variability of those measures in function of earthworm activity.

6.7 Insights into earthworms' influence on water behavior in soil

We have no measurement about the Water Retention Curve or Hydraulic Conductivity in our system, this makes it impossible to draw quantitative conclusions about characterization of hydraulic properties with our approach. Although, we can provide some insights on how earthworm activity would likely affect water dynamic in soil systems. The anecic earthworms' burrowing activity leads to a pore space characterized by large, interconnected porous structures, reducing Euler number (M_3) and increasing the total porosity. This new features of the pore space would likely heavily influence the shape of bot HCC and WRC. In particular, we should expect a much higher value of hydraulic conductivity (K_s) near saturation, though the creation of preferential flow paths. For this same reason, in this conditions, the WRC would probably show a steeper decline near saturation, meaning that a large amount of water would rapidly flow through the system inside the macropore network. This would also significantly increase hysteresis effects in the WRC, thanks to the distinct pathway geometries created by the burrow network. The dramatic reorganization of the pore space by earthworms' activity would likely accentuate the bimodal nature of the Water Retention Curve, creating a clearer separation between matrix and macropore flow. This suggests that, while representing soils with earthworms, it should be better to use a bimodal WRC than the unimodal one from the basic Van Genuchten-Maulem model.

Our results show that earthworm presence brings a higher variability in our metrics, this suggests that hydraulic conductivity would also became more variable inside the sample, with regions that are characterized by very high or moderate conductivity values. These predictions align with the thesis finding that earthworms don't simply increase disorder in soil structure, but rather create organized pathways that would significantly alter water movement and retention patterns. The Fisher Information Measure's strong correlation with earthworm activity suggests these hydraulic impacts would be more related to structural organization than to bulk porosity changes. This considerations also show the limit of using just the total porosity as a parameter for describing soil pore space and its hydraulic influence.

New experiments, that feature direct measurements of hydraulic properties should be performed in order to quantitatively investigate our metrics' ability to inform about water behavior in different conditions.

6.8 Limitations

In this section we will address the limits and criticalities of the study object of the present research work. Some of the limitations have to do with practical constrains, related to the experimental setup, others come from the innovative character of the research and other from methodological choices that cannot find adequate justification. We will try to suggest new experiments and procedures that could solve some of the problems, helping to add context around our findings and perfecting the analysis procedure.

First, we want to stress the fact that only 3 replicates for each condition are not enough to perform a statistically robust analysis; furthermore, the samples come from the same site, and present very similar textural characteristics. This limits the considerations drawn from our work to the specific type of soil considered in the experiment (silty-loamy soil). Since different soil types can have dramatically different textural properties, we should be very careful in generalizing our findings; a series of experiments should be implemented, in order to establish some ground values for different classes of soils, and to assess the variability of the information theory metrics inside the same textural class. We suspect that different textural classes would yield very different results from our analysis. Furthermore, the soil was wet sieved at 5mm and reconstructed in tubes, while maintained in an artificial and very well controlled environment. As before, this is a severe limitation in expanding our conclusions for real-world soils, as those are more complex systems, with various objects of big dimensions (rocks, roots, larger animals' structures) that could influence the analysis and a specific spatial organization that is probably not completely conserved in our samples. Specifically, our considerations about the metrics trends with depth seem to make sense, but we don't know in which measure the repacking process alter the depth-specific structure of soil, and if the two months duration of the experiment is enough for natural-like pore space adaptation to depth. Regarding the X-ray CT images, the coarse resolution (one pixel correspond to 0.67mm) has an ambivalent impact on the analysis validity. It is not possible, at this scale, to make any kind of consideration about microporosity, and also the burrow systems of *A. Rosea* are not well identified in our analysis, for this same reason. On the other hand, this resolution proved to be ideal to identify and even quantify the presence and activity of the earthworms belonging to the *A. Trapezoides* species. We acknowledge here the fact that our analysis is strongly dependent on the resolution of the X-ray CT scans, but, instead of seeing this

as a liability, we can consider it an opportunity to explore different physical and biological characteristics of soil systems. To this end, we suggest that our analysis performed on images at different resolution could help quantify the impact on soil architecture of other ecotypes of earthworms and even of other, smaller animals such as nematodes and insects and also of the microbial and fungal communities. To test this hypothesis, some experiment could be made, or it could be performed a study over a selection of already made experiments that focused on the biological characterization of soils and featured only a qualitative analysis of the X-ray CT scans. Application of our analysis in all those different contexts would surely help to frame better our work and to more efficiently evaluate the results we obtained. This is, though, only the second example of this kind of research, and is therefore natural that our results suffer from a lack of context, we hope with this work to stimulate research in this direction and generally to boot the interest in complexity-oriented analysis of soil structure and functionality.

There are also some methodological concerns about our work. The first liability is in the choice of the bandwidth, our approach was very empirical and lacks a strong theoretical justification. A possible solutions could lie in the work of Botev and colleagues [8], in which an innovative method for bandwidth selection is proposed. We did not, though, investigated this possibility, we suggest it to be tested in future works. The normalization process for the calculation of the distance also seem to be quite arbitrary. Even if the rational behind it is quite comprehensible, the methodology lacks a theoretical context that would relate the newly obtained value of distance from the isocomplexity line with the value that would be obtained without the normalization.

7 Conclusion and future perspectives

Our study successfully implemented a characterization of soil structural complexity with tools from information theory; we demonstrated that Fisher-Shannon analysis of 3D X-ray CT greyscale images is a viable tool to discriminate between soil physical and biological conditions. Specifically, our analysis allowed to quantitatively determine the impact of moisture and earthworm activity on soil architecture's complexity, providing insights into how water and soil biome alter soil structure and, with it, soil functionality.

Our results clearly show that the Shannon Entropy Power is an excellent metric to distinguish between soil wet and dry conditions. The consistently higher values of this quantity for wetter soils, indicate that water abundance increase the disorder of the system, resulting in soils with a more homogeneous and chaotic structural organization. This tendency reflects the role of water as the key facilitator of mineral particles and nutrients' transport across the soil porous matrix, highlighting the impact of this fundamental ecosystemic service. The remarkable correlation of the entropy measure with moisture levels, suggests us that this metric could be used as a proxy for soil humidity assessment, allowing researchers to estimate the water content of a sample with a non-destructive method. To our knowledge, it is the first time that an approach that does not involve direct sample manipulation is proposed to determine the water content of soil. This finding could have impactful repercussions in soil sciences and in the assessment methods used in the field. Further experiments are needed, though, to support our findings, validate them in real world scenarios and for different types of soil and, ultimately, to develop a protocol for the use of this assessment technique.

Our study reveals a very strong correlation ($r = 0.9$) between Fisher Information Measure and the earthworms' total final mass, pointing to a relation between biotic influence and the level of structural organization of soil architecture. We demonstrated that FIM is able to quantify and qualify the level of bioturbation to which a soil sample is subjected, allowing us to discriminate between earthworm presence and absence and even between the abundance of different ecological type of earthworms. We acknowledge the fact that our analysis is resolution dependent, and our results are the reflection of the of the X-ray CT scans' pixel dimension; but instead of seeing this as a liability for our approach, we consider it an opportunity. We believe that similar studies, carried on at different resolution, could help in the quantification of the impact of different groups of soil animals on the spatial organization of soil structures. We point out in our work that our resolution was particularly indicated to identify the presence and abundance of anecic earthworms (*A. trapezoides*), and failed to do the same for the endogeic species (*A. rosea*). Further studies should explore more deeply the relation between Fisher Information Measure and earthworm activity, testing different resolutions and population dynamics, in order to develop a robust, non-destructive tool to quantify bioturbation and animal presence inside a sample of soil, without the need for sample destruction. As before, this method have the potential to revolutionize the process of earthworm population assessment, but more experiments are needed to supplement the results obtained in this work, validating the approach for different soil types and conditions. Since earthworm abundance is often used as a metric to assess soil health, the Fisher Information Measure could represent an important tool in soil science also for quality determination of soil samples.

Our considerations about the depth-dependent patterns of information theory metrics give us important insights in soil structure dynamics and support our understanding of soil physics. The general tendency for entropy to decrease and structural organization to increase with depth aligns with our idea that pressure reduces packing-produced macroporosity. Here, again, the earthworm activity and moisture level were identified by our metrics, that showed less clear trends when the animals were present and in wet conditions, confirming the overall homogenizing activity of both those factors. Of course our considerations on the matter are limited by the soil gathering method used (5mm sieving) and the fact that the soil was not naturally packed, but artificially collected and maintained in laboratory conditions. Further experiments with undisturbed soil samples should be performed in order to validate our findings in natural settings.

The Fisher-Shannon Complexity proved to be a good metric to quantify soil structural complexity: it assumed consistently higher values for the wet and earthworm-filled samples, indicating strongly that biotic activity enhances structural complexity more than disorder in soils. The very clear correlation between FSC and pore volume and surface area ($r \simeq 0.97$), suggests that, for retrieving this quantities, our analysis could substitute the more common analysis of the pore space, eliminating the need for the controversial thresholding procedure. We decided, in fact, to study directly greyscale images, so to take advantage of the whole information contained inside the 3D X-ray CT scans, and to avoid inevitable arbitrary choices in the processing of the images. Our approach enabled us to make the same considerations about the pore space that could be made

calculating the Minkowsky functionals, but with a much faster, reproducible and holistic method. Even though some properties of the pore space cannot be investigated with this approach (topological properties of the network, for example), we stress the relevance of our analysis in the context of the discussion, currently going on in the soil science community, about the opportunity of the binarization process for studying soil structural properties. We demonstrated that greyscale images, if appropriately treated, can yield reliable results about soil pore space properties, without the need of arbitrary processes and loss of information.

Our work also yielded insights with implications for soil hydraulic properties, especially for what concerns earthworm impact on water behaviour in the soil matrix. The burrow system created by anecic earthworms likely increases dramatically hydraulic connectivity by creating paths for preferential flow. Our findings suggest that biotic modifications could accentuate the bimodal nature of the Water Retention Curve, differentiating more between matrix and macropore flow regimes. We want here to point out that our results, about the very different structural organization induced by earthworm activity, highlight the limits of using just the total porosity as a variable in the construction of pedotransfer functions, a new metric that better identify structural complexity and connectivity could be individuated for this kind of calculations. New experiments that feature precise measurements of hydraulic quantities should be performed in order to explore this possibility.

Our approach allows us to make important methodological considerations about the analysis carried out in this work. The most important finding in this sense seem to be the necessity to include vertical variability inside the partitioning of the sample in subsamples, whenever this is necessary for statistical reasons. In fact, we found that the Walls and Volumes partitions are the best ones for the scopes of our investigations, even though just 3 replicates are not really enough for a statistically robust consideration of the Volumes partition.

In conclusion, our research demonstrate that Information Theory provides useful tools for quantifying soil structural complexity and detecting the biological and physical state of soil systems in non-destructive ways. By eliminating the need for direct manipulation of the soil sample to gather information about moisture level and earthworm abundance, our approach have the potential to be revolutionary in soil science. Making use of the whole information contained in the 3D X-ray CT scans, our complexity metrics better captures the multifaceted nature of soil as a complex system where physical, chemical, and biological processes interact across multiple scales. These findings contribute to the ongoing paradigm shift toward viewing soil architecture as a central parameter regulating diverse soil functions, considering the system soil as one, interconnected, multiscale entity rather than a collection of separated processes acting on different levels.

8 Bibliography

References

- [1] Darwin, c. r. 1881. the formation of vegetable mould, through the action of worms, with observations on their habits. london: John murray.
- [2] X-ray imaging of the soil porous architecture.
- [3] Domingos Aguiar, Rômulo Simões Cezar Menezes, Antonio Celso Dantas Antonino, Tatijana Stosic, Ana M. Tarquis, and Borko Stosic. Quantifying soil complexity using fisher shannon method on 3d x-ray computed tomography scans. 25(10):1465.
- [4] Šárka Angst, Carsten W. Mueller, Tomáš Cajthaml, Gerrit Angst, Zuzana Lhotáková, Martin Bartuška, Alexandra Špaldoňová, and Jan Frouz. Stabilization of soil organic matter by earthworms is connected with physical protection rather than with chemical changes of organic matter. 289:29–35.
- [5] S. Barot, J.-P. Rossi, and P. Lavelle. Self-organization in a simple consumer–resource system, the example of earthworms. 39(9):2230–2240.
- [6] E. Blanchart, N. Marilleau, J.-L. Chotte, A. Drogoul, E. Perrier, and Ch. Cambier. SWORM: an agent-based model to simulate the effect of earthworms on soil structure. 60(1):13–21.
- [7] M. Blouin, M. E. Hodson, E. A. Delgado, G. Baker, L. Brussaard, K. R. Butt, and J. J. Brun. A review of earthworm impact on soil function and ecosystem services. *European Journal of Soil Science*, 64(2):161–182, 2013.
- [8] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010.
- [9] Yvan Capowiez, Nicolas Bottinelli, and Pascal Jouquet. Quantitative estimates of burrow construction and destruction, by anecic and endogeic earthworms in repacked soil cores. 74:46–50.
- [10] Yvan Capowiez, Nicolas Bottinelli, Stéphane Sammartino, Eric Michel, and Pascal Jouquet. Morphological and functional characterisation of the burrow systems of six earthworm species (lumbricidae). 51(7):869–877.
- [11] Yvan Capowiez, Stéphane Sammartino, Thomas Keller, and Nicolas Bottinelli. Decreased burrowing activity of endogeic earthworms and effects on water infiltration in response to an increase in soil bulk density. 85-86:150728.
- [12] Yvan Capowiez, Stéphane Sammartino, and Eric Michel. Using x-ray tomography to quantify earthworm bioturbation non-destructively in repacked soil cores. 162(1):124–131.
- [13] Oleg Chertov, Alexander Komarov, Cindy Shaw, Sergey Bykhovets, Pavel Frolov, Vladimir Shanin, Pavel Grabarnik, Irina Priputina, Elena Zubkova, and Maxim Shashkov. Romul_hum—a model of soil organic matter formation coupling with soil biota activity. II. parameterisation of the soil food web biota activity. 345:125–139.
- [14] Oleg Chertov, Cindy Shaw, Maxim Shashkov, Alexander Komarov, Sergey Bykhovets, Vladimir Shanin, Pavel Grabarnik, Pavel Frolov, Olga Kalinina, Irina Priputina, and Elena Zubkova. Romul_hum model of soil organic matter formation coupled with soil biota activity. III. parameterisation of earthworm activity. 345:140–149.
- [15] Enrico Chiesa. Tesi repository, 2024.
- [16] R.E. Creamer, J.M. Barel, G. Bongiorno, and M.J. Zwetsloot. The life of soils: Integrating the who and how of multifunctionality. 166:108561.

- [17] James Curry and Olaf Schmidt. The feeding ecology of earthworms – a review. 50:463–477.
- [18] L.w. de Jonge, P. Moldrup, A.l. Vendelboe, M. Tuller, and D. Wildenschild. Soil architecture and physicochemical functions: An introduction. 11(1). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.2136/vzj2011.0185>.
- [19] Gaby Deckmyn, Omar Flores, Mathias Mayer, Xavier Domene, Andrea Schnepf, Katrin Kuka, Kris Van Looy, Daniel P. Rasse, Maria J.I. Briones, Sébastien Barot, Matty Berg, Elena Vanguelova, Ivika Ostonen, Harry Vereecken, Laura M. Suz, Beat Frey, Aline Frossard, Alexei Tiunov, Jan Frouz, Tine Grebenc, Maarja Öpik, Mathieu Javaux, Alexei Uvarov, Olga Vindušková, Paul Henning Krogh, Oskar Franklin, Juan Jiménez, and Jorge Curiel Yuste. KEYLINK: towards a more integrative soil representation for inclusion in ecosystem scale models. i. review and model concept. 8:e9750.
- [20] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991.
- [21] Carlos R. Estañón, Norberto Aquino, David Puertas-Centeno, and Jesus S. Dehesa. Two-dimensional confined hydrogen: An entropy and complexity approach. 120(11):e26192. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.26192>.
- [22] Bastardie F., Cannavacciuolo M., Capowiez Y., J.-R. De Dreuzy, Bellido A., and Cluzeau D. A new simulation for modelling the topology of earthworm burrow systems and their effects on macropore flow in experimental soils. 36(2):161–169.
- [23] Simone Fatichi, Dani Or, Robert Walko, Harry Vereecken, Michael H. Young, Teamrat A. Ghezzehei, Tomislav Hengl, Stefan Kollet, Nurit Agam, and Roni Avissar. Soil structure is an important omission in earth system models. 11(1):522.
- [24] J. Fernández-Gálvez, J.A.P. Pollacco, L. Lilburne, S. McNeill, S. Carrick, L. Lassabatere, and R. Angulo-Jaramillo. Deriving physical and unique bimodal soil kosugi hydraulic parameters from inverse modelling. 153:103933.
- [25] R. A. Fisher. Theory of statistical estimation. 22(5):700–725.
- [26] Omar Flores, Gaby Deckmyn, Jorge Curiel Yuste, Mathieu Javaux, Alexei Uvarov, Sietse Van Der Linde, Bruno De Vos, Harry Vereecken, Juan Jiménez, Olga Vinduskova, and Andrea Schnepf. KEYLINK: towards a more integrative soil representation for inclusion in ecosystem scale models—II: model description, implementation and testing. 9:e10707.
- [27] B. Roy Frieden. Fisher information, disorder, and the equilibrium distributions of physics. 41(8):4265–4276. Publisher: American Physical Society.
- [28] H. H. Gerke and M. T. Van Genuchten. A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. 29(2):305–319.
- [29] Horst Gerke and Martinus Van Genuchten. A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. 29:305–319.
- [30] Teamrat A. Ghezzehei, Timothy J. Kneafsey, and Grace W. Su. Correspondence of the gardner and van genuchten–mualem relative permeability function parameters. 43(10):2006WR005339.
- [31] Stuart Grandy, Jennifer Fraterrigo, and Sharon Billings. Soil ecosystem resilience and recovery.
- [32] Claire Grosbellet, Laure Vidal-Beaudet, Virginie Caubel, and Sylvain Charpentier. Improvement of soil structure formation by degradation of coarse organic matter. 162(1):27–38.
- [33] Brice Guignard, Chris Button, Keith Davids, and Ludovic Seifert. Education and transfer of water competencies: An ecological dynamics approach. *European Physical Education Review*, 26(4):938–953, 2020.

- [34] R. M. L. Guimarães, B. C. Ball, and C. A. Tormena. Improvements in the visual evaluation of soil structure. *Soil Use and Management*, 27(3):395–403, 2011.
- [35] Paul D. Hallett, Kamal H. Karim, A. Glyn Bengough, and Wilfred Otten. Biophysics of the vadose zone: From reality to model systems and back again. 12(4):1–17.
- [36] R. J. Hanks and S. A. Bowers. Numerical solution of the moisture flow equation for infiltration into layered soils. 26(6):530–534.
- [37] Nils-Bastian Heidenreich, Anja Schindler, and Stefan Sperlich. Bandwidth selection for kernel density estimation: a review of fully automatic selectors. 97(4):403–433.
- [38] A.N. Houston, W. Otten, R. Falconer, O. Monga, P.C. Baveye, and S.M. Hapca. Quantification of the pore size distribution of soils: Assessment of existing software using tomographic and synthetic 3d images. 299:73–82.
- [39] Don Kirkham and W. V. Bartholomew. Equations for following nutrient transformations in soil, utilizing tracer data: II. 19(2):189–192.
- [40] Alexander Komarov, Oleg Chertov, Sergey Bykhovets, Cindy Shaw, Marina Nadporozhskaya, Pavel Frolov, Maxim Shashkov, Vladimir Shanin, Pavel Grabarnik, Irina Priputina, and Elena Zubkova. Romul_hum model of soil organic matter formation coupled with soil biota activity. i. problem formulation, model description, and testing. 345:113–124.
- [41] K. Kosugi. General model for unsaturated hydraulic conductivity for soils with lognormal pore-size distribution. 63(2):270–277.
- [42] M. Larsbo, J. Koestel, and N. Jarvis. Relations between macropore network characteristics and the degree of preferential solute transport. 18(12):5255–5269.
- [43] Julia Le Noë, Stefano Manzoni, Rose Abramoff, Tobias Bölscher, Elisa Bruni, Rémi Cardinael, Philippe Ciais, Claire Chenu, Hugues Clivot, Delphine Derrien, Fabien Ferchaud, Patricia Garnier, Daniel Goll, Gwenaëlle Lashermes, Manuel Martin, Daniel Rasse, Frédéric Rees, Julien Sainte-Marie, Elodie Salmon, Marcus Schiedung, Josh Schimel, William Wieder, Samuel Abiven, Pierre Barré, Lauric Cécillon, and Bertrand Guenet. Soil organic carbon models need independent time-series validation for reliable prediction. 4(1):158.
- [44] Xudong Li, Luciano Telesca, Michele Lovallo, Xuan Xu, Jun Zhang, and W.G. Song. Spectral and informational analysis of pedestrian contact force in simulated overcrowding conditions. *Physica A: Statistical Mechanics and its Applications*, 555:124614, 04 2020.
- [45] M.T Martin, F Pennini, and A Plastino. Fisher's information and the analysis of complex signals. 256(2):173–180.
- [46] Serge Martin and Patrick Lavelle. A simulation model of vertical movements of an earthworm population (*millsonia anomala omodeo*, *megascolecidae*) in an african savanna (lamto, ivory coast). 24(12):1419–1424.
- [47] Katharina Meurer, Jennie Barron, Claire Chenu, Elsa Coucheney, Matthew Fielding, Paul Hallett, Anke M. Herrmann, Thomas Keller, John Koestel, Mats Larsbo, Elisabet Lewan, Dani Or, David Parsons, Nargish Parvin, Astrid Taylor, Harry Vereecken, and Nicholas Jarvis. A framework for modelling soil structure dynamics induced by biological activity. 26(10):5382–5403.
- [48] R. Milleret, R. C. Le Bayon, F. Lamy, J. M. Gobat, and P. Boivin. Impact of roots, mycorrhizas and earthworms on soil physical properties as assessed by shrinkage analysis. 373(3):499–507.
- [49] Yechezkel Mualem. A new model for predicting the hydraulic conductivity of unsaturated porous media. 12(3):513–522. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/WR012i003p00513>.

- [50] M. Naveed, E. Arthur, L. W. de Jonge, M. Tuller, and P. Moldrup. Pore structure of natural and regenerated soil aggregates: An x-ray computed tomography analysis. *Soil Science Society of America Journal*, 78(2):377–386, 2014.
- [51] Muhammad Naveed, Emmanuel Arthur, Lis Wollesen De Jonge, Markus Tuller, and Per Moldrup. Pore structure of natural and regenerated soil aggregates: An x-ray computed tomography analysis. 78(2):377–386.
- [52] J R Nimmo and Menlo Park. Porosity and pore size distribution. 3.
- [53] Xi Pang, Ulla Mörtberg, Ola Sallnäs, Renats Trubins, Eva-Maria Nordström, and Hannes Böttcher. Habitat network assessment of forest bioenergy options using the landscape simulator LandSim – a case study of kronoberg, southern sweden. 345:99–112.
- [54] A. Papadopoulos, N. R. A. Bird, A. P. Whitmore, and S. J. Mooney. Investigating the effects of organic and conventional management on soil aggregate stability using x-ray computed tomography. *European Journal of Soil Science*, 60(3):360–368, 2009.
- [55] Bernard C. Patten. A simulation of the shortgrass prairie ecosystem. 19(6):177–186.
- [56] Edith Perrier, Nigel Bird, and Michel Rieu. Generalizing the fractal model of soil structure: the pore–solid fractal approach. 88(3):137–164.
- [57] A. R. Plastino, M. Casas, and A. Plastino. Fisher’s information, kullback’s measure, and h-theorems. 246(6):498–504.
- [58] A. Poulovassilis and W. M. El-Ghamry. THE DEPENDENT DOMAIN THEORY APPLIED TO SCANNING CURVES OF ANY ORDER IN HYSTERETIC SOIL WATER RELATIONSHIPS:. 126(1):1–8.
- [59] Peter A. C. Raats and John H. Knight. The contributions of lewis fry richardson to drainage theory, soil physics, and the soil-plant-atmosphere continuum. 6. Publisher: Frontiers.
- [60] E. Rabot, M. Wiesmeier, S. Schlüter, and H. J. Vogel. Soil structure as an indicator of soil functions: A review. 314:122–137.
- [61] Inge C. Regelink, Cathelijne R. Stoof, Svetla Rousseva, Liping Weng, Georg J. Lair, Pavel Kram, Nikolaos P. Nikolaidis, Milena Kercheva, Steve Banwart, and Rob N.J. Comans. Linkages between aggregate formation, porosity and soil chemical properties. 247-248:24–37.
- [62] Riccardo Rigon and Concetta D’Amato. The rosetta stone of the darcy-buckingham law 3.
- [63] J. S. Russell. Mathematical expression of seasonal changes in soil organic matter. 204(4954):161–162.
- [64] M. Samec, A. Santiago, J. P. Cárdenas, R. M. Benito, A. M. Tarquis, S. J. Mooney, and D. Korošak. Quantifying soil complexity using network models of soil porous structure. 20(1):41–45.
- [65] F. San José Martínez, M. A. Martín, F. J. Caniego, M. Tuller, A. Guber, Y. Pachepsky, and C. García-Gutiérrez. Multifractal analysis of discretized x-ray CT images for the characterization of soil macropore structures. 156(1):32–42.
- [66] Ophélie Sauzet, Roxane Kohler-Milleret, François Fülleman, Yvan Capowiez, and Pascal Boivin. Nicodrilus nocturnus and allolobophora icterica drill compacted soils but do not decrease their bulk density – a laboratory experiment using two contrasted soils at two different compaction levels. 402:115164.
- [67] S. Schlüter, H. J. Vogel, O. Ippisch, and J. Vanderborght. Combined impact of soil heterogeneity and vegetation type on the annual water balance at the field scale. *Vadose Zone Journal*, 12(4), 2013.
- [68] Andreas Schomburg and Pascal Turberg. *EARTHWORMS TYPES, ROLES AND RESEARCH*. Nova Science Publishers Inc., 2017.
- [69] Simon Sheather. Density estimation. 19.

- [70] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. CRC Press, Boca Raton, FL, 2018.
- [71] Sarah Smet, Eléonore Beckers, Erwan Plougonven, Angélique Léonard, and Aurore Degré. Can the pore scale geometry explain soil sample scale hydrodynamic properties? 6:20.
- [72] Brigitta Szabó, Gábor Szatmári, Katalin Takács, Annamária Laborczi, András Makó, Kálmán Rajkai, and László Pásztor. Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. 23(6):2615–2635. Publisher: Copernicus GmbH.
- [73] A. M. Tarquis, R. J. Heck, D. Andina, A. Alvarez, and J. M. Antón. Pore network complexity and thresholding of 3d soil images. 6(3):230–239.
- [74] I.G. Torre, Juan J. Martín-Sotoca, J.C. Losada, Pilar López, and A.M. Tarquis. Scaling properties of binary and greyscale images in the context of x-ray soil tomography. 365:114205.
- [75] Iván G. Torre, Juan C. Losada, and Ana M. Tarquis. Multiscaling properties of soil images. 168:133–141.
- [76] Berwin Turlach. Bandwidth selection in kernel density estimation: A review.
- [77] Benjamin Turner. Soil as an archetype of complexity: A systems approach to improve insights, learning, and management of coupled biogeochemical processes and environmental externalities. 5.
- [78] H. Vereecken, A. Schnepf, J.W. Hopmans, M. Javaux, D. Or, T. Roose, J. Vanderborght, M.H. Young, W. Amelung, M. Aitkenhead, S.D. Allison, S. Assouline, P. Baveye, M. Berli, N. Brüggemann, P. Finke, M. Flury, T. Gaiser, G. Govers, T. Ghezzehei, P. Hallett, H.J. Hendricks Franssen, J. Heppell, R. Horn, J.A. Huisman, D. Jacques, F. Jonard, S. Kollet, F. Lafolie, K. Lamorski, D. Leitner, A. McBratney, B. Minasny, C. Montzka, W. Nowak, Y. Pachepsky, J. Padarian, N. Romano, K. Roth, Y. Rothfuss, E.C. Rowe, A. Schwen, J. Šimůnek, A. Tiktak, J. Van Dam, S.E.A.T.M. Van Der Zee, H.J. Vogel, J.A. Vrugt, T. Wöhling, and I.M. Young. Modeling soil processes: Review, key challenges, and new perspectives. 15(5):1–57.
- [79] Harry Vereecken, Melanie Weynants, Mathieu Javaux, Yakov Pachepsky, M. Schaap, and Martinus Van Genuchten. Using pedotransfer functions to estimate the van genuchten– mualem soil hydraulic properties: A review. 9:795–820.
- [80] C. Vignat and J. F. Bercher. Analysis of signals in the fisher–shannon information plane. 312(1):27–33.
- [81] H.-J. Vogel, W. Amelung, C. Baum, M. Bonkowski, S. Blagodatsky, R. Grosch, M. Herbst, R. Kiese, S. Koch, M. Kuhwald, S. König, P. Leinweber, B. Lennartz, C. W. Müller, H. Pagel, M. C. Rillig, J. Rüschhoff, D. Russell, A. Schnepf, S. Schulz, N. Siebers, D. Vetterlein, C. Wachendorf, U. Weller, and U. Wollschläger. How to adequately represent biological processes in modeling multifunctionality of arable soils. 60(3):263–306.
- [82] Hans-Jörg Vogel, Maria Balseiro-Romero, Alexandra Kravchenko, Wilfred Otten, Valérie Pot, Steffen Schlüter, Ulrich Weller, and Philippe C. Baveye. A holistic perspective on soil architecture is needed as a key to soil functions. 73(1):e13152. _eprint: <https://onlinelibrary.wiley.com/doi/10.1111/ejss.13152>.
- [83] Hans-Jörg Vogel, Stephan Bartke, Katrin Daedlow, Katharina Helming, Ingrid Kögel-Knabner, Birgit Lang, Eva Rabot, David Russell, Bastian Stöbel, Ulrich Weller, Martin Wiesmeier, and Ute Wollschläger. A systemic approach for modeling soil functions. 4(1):83–92.
- [84] A. W. Warrick. *Soil water dynamics*. Oxford university press, 2003.
- [85] Tobias Karl David Weber, Lutz Weihermüller, Attila Nemes, Michel Bechtold, Aurore Degré, Efstatios Diamantopoulos, Simone Faticchi, Vilim Filipović, Surya Gupta, Tobias L. Hohenbrink, Daniel R. Hirmas, Conrad Jackisch, Quirijn De Jong Van Lier, John Koestel, Peter Lehmann, Toby R. Marthews, Budiman Minasny, Holger Pagel, Martine Van Der Ploeg, Shahab Aldin Shojaeezadeh, Simon Fiil Svane, Brigitta Szabó, Harry Vereecken, Anne Verhoef, Michael Young, Yijian Zeng, Yonggen Zhang, and Sara Bonetti. Hydro-pedotransfer functions: a roadmap for future development. 28(14):3391–3433.

- [86] Ulrich Weller, Lukas Albrecht, Steffen Schlüter, and Hans-Jörg Vogel. An open *Soil Structure Library* based on x-ray CT data. 8(2):507–515.
- [87] Laosheng Wu. Relationship between pore size, particle size, aggregate size and water characteristics.
- [88] Erwin Zehe, Ralf Loritz, Conrad Jackisch, Martijn Westhoff, Axel Kleidon, Theresa Blume, Sibylle K. Hassler, and Hubert H. Savenije. Energy states of soil water – a thermodynamic perspective on soil water dynamics and storage-controlled streamflow generation in different landscapes. 23(2):971–987.

9 Appendix A

Here a table of the acronyms used in the thesis and their meaning.

Acronym	Definition
SOM	Soil Organic Matter
PDE	Partial Differential Equation
WRC	Water Retention Curve
RRE	Richard Richardson's Equation
HCC	Hydraulic Conductivity Curve
SHP	Soil Hydraulic Properties
PTF	Pedotransfer Functions
MWC	Maximum Water Content
SEP	Shannon Entropy Power
FIM	Fisher Information Measure
FSC	Fisher-Shannon Complexity
CT	Computed Tomography (often as X-ray CT)
μ CT	Micro Computed Tomography
IQR	Interquartile Range
pdf	Probability Density Function
KDE	Kernel Density Estimator
EPI	Extreme Physical Information
TC	Total Carbon
TN	Total Nitrogen

Table 4: List of acronyms and their definitions