Portraying Urban Diversity Patterns Through Exploratory Data Analysis

Lorena Salazar Llano

Supervisors:

Martí Rosas Casals
María Isabel Ortego Martínez

Institute for Sustainability Science and Technology

Universitat Politècnica de Catalunya

Ph.D. Thesis

December 2019

Acknowledgements

In the first place, I would like to thank my supervisors, professors Martí Rosas and Maribel Ortego, for their support and guidance. To Martí, thanks for motivating me into the field of complex systems: you outlined the path that exact moment when you suggested me the book "Diversity and complexity" (Page, 2011), which in a certain way gave rise to this thesis. To Maribel, thanks for explaining me the multivariate analysis field: for all your patient and detailed statistical explanations acknowledging my previous background on architecture.

I especially want to thank Camilo, my partner in life and my fundamental support during all these years. Thank you for having believed in this project, and for having supported me not only scientifically but in the deeper sense of life. To Emilia, my beloved daughter who has grown up in parallel with this thesis, you have been my biggest motivation: this work is dedicated to you. Thanks to you and your boundless love I haven't lost myself in the way. You both bring a lot of happiness and meaning to my life.

To our family, infinite thanks for having supported us in the distance, and now that we are back home.

To my friends, with whom I have shared this process, thanks for the conversations at lunchtime, for the time we spent in the library, but above all, for the experiences we shared not in the university. You made this time much more enjoyable. To Barcelona —my case study—, and to everyone whom I took part with there, thanks. All those years were outstanding and personally very fulfilling.

Finally, I would like to acknowledge the doctoral scholarship that I had received from Fundación CEIBA, and the big support that the Sustainability Measurement and Modeling Lab, and the Universitat Politecnica de Catalunya, gave to this research.

Abstract

This thesis analyzes the complexity of the urban system, being described with multiple variables that represent the environmental, economic, and social characters of the city. The portrayal of the urban diversity and its relationship with a better response of the city to disturbances, hence to its sustainability, is the main motivation of the study. Certainly, this thesis aims to provide theoretical knowledge through the application of statistical and computational methodologies that are developed progressively in its chapters. Beginning with the introduction, which draws the city as an abstract urban system and reviews the concepts and measures of diversity within the theoretical frameworks of sustainability, urban ecology, and complex systems theory. Afterward, the city of Barcelona is introduced as the case study: it is constituted by a set of districts and represented by an information system that contains temporal measurements of multiple environmental, economic, and social variables. A first approach to the sustainability of the city is made with the entropy of information as a measure of the urban system's diversity. But the fundamental contribution of the thesis focuses on the application of Exploratory Multivariate Analysis (EMA) to the urban system: Principal Component Analysis (PCA), Multiple Factorial Analysis (MFA), and Hierarchical Cluster Analysis (HCA). From this EMA approach, diversity is analyzed by identifying the similarity — or dissimilarity — between the different parts that make up the urban system. Some other techniques based on computer science and physics are proposed to evaluate the temporal transformation of the urban system, understood as a three-dimensional data cloud that gradually deforms. Differentiated characters and distinctive functions of districts are identifiable in the EMA application to the case study. Moreover, the temporal dependency of the dataset reveals information about the district's differentiation or homogenization trends. Finally, the conclusions of the most relevant results are presented and some future lines of research are proposed.

Resumen

Esta tesis analiza la complejidad del sistema urbano, descrito con múltiples variables que representan las características ambientales, económicas y sociales de la ciudad. La motivación fundamental para emprender este estudio consiste en describir la diversidad de la ciudad y su relación con una mejor respuesta a perturbaciones y amenazas, y por lo tanto, a su sostenibilidad. La tesis plantea aportar conocimiento teórico mediante la aplicación de metodologías estadísticas y computacionales que se desarrollan progresivamente en sus capítulos. En la introducción se presenta la abstracción de la ciudad como un sistema urbano, y se hace una revisión de los conceptos y medidas de la diversidad dentro de los marcos teóricos de la sostenibilidad, la ecología urbana y la teoría de los sistemas complejos. Posteriormente, se introduce el sistema urbano de la ciudad de Barcelona, constituido por un conjunto de distritos y representado mediante un sistema de información que contiene mediciones temporales de múltiples variables ambientales, económicas y sociales. Se hace una primera aproximación a la sostenibilidad de la ciudad empleando la entropía de la información como medida de diversidad del sistema urbano. Pero el aporte fundamental de la tesis se centra en la aplicación del Análisis Exploratorio Multivariante (EMA) en el sistema urbano: Análisis de Componentes Principales (PCA), Análisis Factorial Múltiple (MFA) y Análisis de Agrupamiento Jerárquico (HCA). Desde dicho enfoque se analiza la diversidad identificando la similaridad — o disimilaridad— entre las distintas partes que componen el sistema urbano. Se plantean también algunas de las técnicas de las ciencias de la computación y la física para evaluar la transformación temporal del sistema urbano, entendido como una nube de datos tridimensionales que se deforma gradualmente. En el análisis del estudio de caso se identifican características diferenciadas y funciones distintivas de los distritos. Además, la dependencia temporal del conjunto de datos revela información sobre las tendencias de diferenciación u homogeneización de los distritos. Finalmente, se exponen las conclusiones de los resultados más relevantes y se enuncian algunas líneas futuras de investigación.

Contents

List of Figures

Chapter 1

Introduction

1.1 Urban sustainability assessment

Cities have been recognized by humanity as centers for economic development. This acknowledgment has been granted to such a degree that, by 2050, it is expected that 66% of the total population will be urban [1]. Cities are also identified as centres for productivity, creativity, innovation, and cultural transformation [2–4]. However, disruptions in the environmental, as well as socio-economical areas are some direct consequences of the increasing urbanization. In this sense, the exhaustive analysis of the cities, and the quantification of the urban assets play a key role in facing the regional and global urbanization challenges [5–7].

Many concepts and definitions of urban sustainability have emerged since the 90s [7–10]. Up to date, there is not a widely shared definition of the concept of urban sustainability. However, one of the most recent and accepted is the following: "urban sustainability is defined as an adaptive process of facilitating and maintaining a virtuous cycle between ecosystem services and human well-being through concerted ecological, economic, and social actions in response to changes within and beyond the urban landscape" [7]. The concept of urban sustainability is related to the ability of the city to respond to hazards or disturbances without failing, where the most prevailing hazards have historically been the economic and societal collapses, civil conflicts, pandemics, human-enhanced natural catastrophes, and environmental problems. Traditionally, the concept of a failing city has been exclusively associated with the concept of a loss in its population. However, this is a reductionist approach: it does not even explain the process in which the city is not able to recover to a functional state after the event of a disturbance.

Several frameworks have been proposed for studying and achieving urban sustainability: for example, the urban ecology framework [6, 11–13], the industrial ecology framework [5, 14], the urban resilience framework [15–18], and the landscape ecology framework [19, 20].

The framework of urban ecology [13] offers interdisciplinarity: it adopts methods both from natural and social sciences, and link together researches and stakeholders —from different disciplines— in the urban planning. There are two main perspectives to analyze cities in urban ecology; one is the "ecology of cities" that consider the whole city as an ecosystem [6], and the other is the sustainability of cities. Urban ecology studies Spatio-temporal patterns, environmental impacts, and sustainability of urbanization, with emphasis on the ecosystem's services and processes [7].

Within the framework of industrial ecology, some concepts (i.e. urban metabolism) and methods (i.e. Material Flow Analysis (MFA) and Life Cycle Assessment (LCA)) have been applied to understand urban systems' flows and patterns. Precisely, the concept of urban metabolism can be defined as the entire technical and socio-economic processes that take place in cities, resulting in growth, energy production, and waste disposal [5]. The importance of the metabolism analogy relies on its applicability: it becomes a method of analysis that quantifies the matter, energy, and information flows, which are inherent to urban systems. This concept was first introduced by Wolman (1965) [21] to analyze the metabolism of a hypothetical US city of 1 million inhabitants. Since then, urban metabolism has been widely used in the analysis of different urban systems [5, 22–27]. Most of these studies have measured the metabolism of cities focusing on the four fundamental flows of urban metabolism: water, materials, energy, and nutrients [5]. Particularly, at the neighborhood scale, Codoban & Kennedy (2008) [25] analyzed the metabolism of four representatives Toronto neighborhoods through a material flow analysis of energy, water, food, and waste flows. More recently, some studies have linked the analysis of urban metabolism with urban sustainability indicators [28–30].

The urban resilience framework relies on the resilience theory applied to socio-ecological systems [31–34]. Within this framework, it has been argued that urban systems must improve their resilience capacity in order to recover from disturbances and to achieve sustainability [15, 35]. Although the concept of resilience was introduced in ecology since the 70s [31], and in the last decade it has been used in the conceptualization of urban systems, the concept of urban resilience was built only in recent years, mainly due to the difficulty of delimiting the concept of urban. Urban resilience has been defined by [16] as "the ability of an urban system and all its constituent socio-ecological and socio-technical networks across temporal and spatial scales to maintain or rapidly return to desired functions in the face of a disturbance, to adapt to change, and to quickly transform systems that limit current or future adaptive capacity". Some other scholars [36, 37] have built definitions of urban resilience, however, they have not addressed some key aspects of the ecological concept of resilience, such as non-equilibrium paradigm, a general adaptive capacity, and different scales —temporal and spatial.

Interestingly, most of these frameworks seem to converge in some fundamental concepts and theories: ecosystems services, resilience, complex adaptive systems, and socio-ecological systems (SES). In this context, cities —or, in general, urban systems— can be defined as complex adaptive systems which are characterized by the high diversity of their components and spatial structures, nonlinear feedbacks, across-scales interactions, and the ability to self-adjust in response to changes [7, 38]. This conceptualization of the city is aligned with the theoretical frameworks of complex systems and science of cities [39]. Indeed, from the perspective of complexity science the city is commonly studied as a diverse, adaptive, and socio-ecological system that displays multiple interactions between its components and scales [4, 40]: urban systems exhibit, among other features, emergence and adaptive phenomena, and are a prime example of complex systems, behaving between order and randomness. Within this framework, the city is represented as a system with multiple components, among which, its infrastructure, resources, and population stand over. This set of components is singular in the sense that they cannot be easily described either predicted and that their relations are very sensitive to their current state.

Several scholars have linked the sustainability of socio-ecological systems with the concepts of resilience and/or robustness [41–44]. On one hand, the concept of robustness has been

widely investigated within the theory of complex systems [45]. Robustness, in general, can be defined as the ability of a system to maintain functionality in the face of some disturbance, which could be internal or external to the system [45–47]. Resilience, on the other hand, refers to the ability of a system to adapt to a modifying process while retaining its functionality and not necessarily returning to the previous equilibrium state [34, 43, 48]. The definition of resilience comes from the non-equilibrium paradigm of ecology [43], and fits the complexity and uncertainty inherent to cities. Some studies used the terms robustness and resilience interchangeably since the meanings of both concepts can be made equivalent [49]: both refer to the capability of a system to continue to perform in diverse circumstances, not necessarily returning to the previous equilibrium state, but resilience includes adaptation as a key process. Moreover, resilience provides a wider scientific framework for understanding complex systems and how their temporal and spatial scales interact [44]. A great number of studies have studied the resilience concept in urban systems [16, 18, 24, 50–52]. In this thesis, we are interested in the concept of resilience —instead of robustness— in cities, and its relation with diversity and sustainability. In a certain sense, measuring resilience is tightly related to inspecting the system's property to remain functional over time, hence, to sustain.

Some principles and strategies for building resilience in socio-ecological systems [53], and particularly in urban systems [15], have been proposed in the last years. Biggs and colleagues [53, 54] identified seven generic policy-relevant principles for enhancing resilience: (1) maintain diversity and redundancy, (2) manage connectivity, (3) manage slow variables and feedbacks, (4) fostering and understanding SES as Complex Adaptive Systems (CAS), (5) encourage learning and experimentation, (6) broaden participation, and (7) promote polycentric governance systems. In the urban context, Ahern (2011) proposed five urban planning and design strategies for building urban resilience: (1) multi-functionality, (2) redundancy and modularization, (3) bio and social diversity, (4) multi-scale networks and connectivity, and (5) adaptive planning and design. Remarkably, two of the principles are centered on improving diversity.

Accurately, in this thesis, we are concerned about diversity and its relationship to resilience. Regarding this relation, other scholars have argued that diversity enhances the resilience of complex systems (i.e., its ability to maintain functionality) [32, 41, 55–57]. However, it should be noted that since diversity is a scale-dependent concept [56], the diversity-stability debate [47, 58, 59] remains open. In the following section, we will deepen in the concept of diversity and stability, with an original approach to this debate.

1.2 Deepening on the Diversity-Stability(-Resilience) debate

The aforementioned relation between diversity and resilience relies on the argument that a highly diverse system possesses many different types, and therefore, a bunch of individuals belonging to these different types that are able to perform similar functions at a wider range of conditions [15, 55]. From an urban sociology perspective, diversity is a fundamental property that ensures the city's functionality in the face of disturbances. Since 1961, Jane Jacobs [2] mentioned that diversity is an essential factor for economic growth, attractiveness, and liveability of cities. Some researchers [60, 61] have supported this thesis arguing that diversity fosters productivity, innovation, and therefore, economic development in cities. However,

declaring the direct relationship between diversity and resilience is insufficient: a formal characterization of diversity and its relationship with the resilience of the system is required in order to make quantitative and qualitative assessments of the city's sustainability.

Regarding diversity, several definitions and concepts can be found in the literature. Specifically, in the urban literature, diversity has been conceptualized and analyzed mainly from social and cultural approaches [62–65]. From that perspective, diversity has been understood as a mixture of social (i.e. income, race, ethnicity, age, family types), and physical (i.e. land uses, building types) components within a place. Most of the studies in urban planning arguing for diversity rely on Jacob's work [61, 63, 66–68]. The concept of diversity is central in Jacobs's discourse [2]. She claimed that it is essential for the vitality and the success of cities, and stated four conditions to foster urban diversity: mixed land uses, small blocks, the mixture of building ages, and population density [2].

Since we aim at studying the relationship between diversity and resilience, and we understand the city as a complex urban system, we are interested in the prime definitions and characterizations of diversity arising from ecology and complex systems theory. In ecology, diversity applies to populations or collections of entities like ecosystems having multiple types of flora and fauna; or like cities, with different types of people, organizations, buildings, infrastructural systems, etc. Two main parameters introduced in [69, 70] have been applied to described diversity in ecology; the alpha (α) diversity and the beta (β) diversity. The former refers to the diversity within a particular sample or system and is usually expressed by the number of species (i.e., species richness) in that system. The latter refers to the variation in the identities of species among samples: it portrays the degree of community differentiation (i.e., species diversity between systems) [71]. More recently, [72] outlined a framework for interdisciplinary analysis of diversity, and identified three basic properties of diversity: variety, balance, and disparity. Variety is the number of categories in which the elements of the system are distributed, and responds to the question "how many types of things do we have?". The balance represents the pattern of distribution of elements across categories, it answers the question: "how much of each kind of thing do we have?". Disparity refers to the degree of differentiation between the elements, and answers the question: "how different from each other are the types of thing that we have?". From a complexity approach, three types of diversity apply for populations or collections of entities [47]: (1) diversity within a type, or variation in some attribute or characteristic (such as differences in the length of finches' beaks), (2) diversity of types and kinds within a system, or species in biological systems (such as different types of stores in a mall), and (3) diversity of composition, that refers to differences in configuration —in how the types are arranged— (such as different connections between atoms in a molecule). These multiple definitions of diversity can also be applied to the (complex) urban system. Our approach in the diversity stability debate will appropriate some of them: we assess the alpha (α) diversity, and the (β) diversity —or what is the same, the disparity [72] or diversity of composition [47]— in an urban system. We describe all these definitions since the main topics of this thesis will explore the patterns of diversity in an urban system.

We rely on classical physics to portray the stability-resilience of a diverse system. In the following paragraphs, we expose another approach to add to the debate [58, 59]: one from the classical mechanics, where diversity can be defined as multiple individuals linked with springs and dampers, and its stability the dynamic displacement of the system that contains them. It should be noted that stability and resilience (i.e. robustness) are not the same [47].

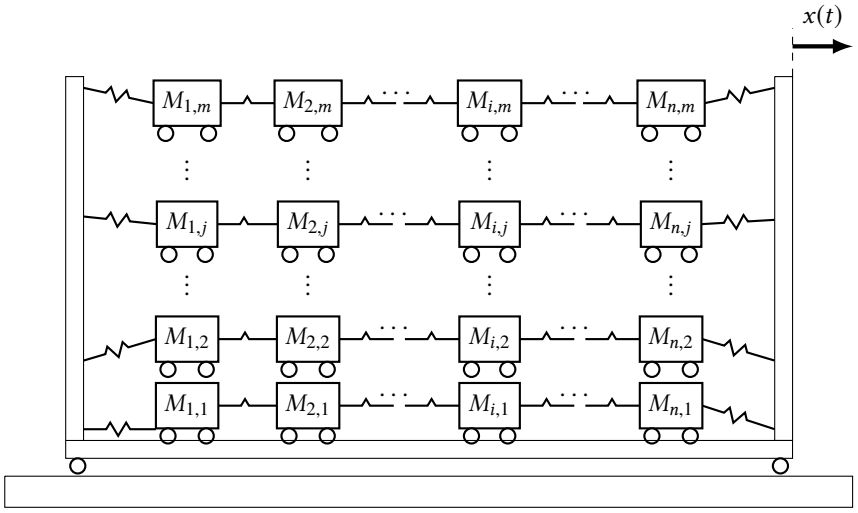

Figure 1.1: Mechanical system composed by $n \times m$ translational components $M_{i,j}$ connected by springs, and a non-inertial translational component M that supports them.

Its difference lies in the fact that the former refers to the tendency of a system to return to an equilibrium given a dynamic process, and the later supports multiple equilibrium states. Since we acknowledge this conceptual difference (i.e. diversity-stability-resilience), we will examine whether the system continues to function and not whether it returns to the same equilibrium. In order to conceptualize the diversity-stability(-resilience) debate, we pose the following problem: suppose the mechanical translational system depicted in Figure 1.1, which is composed by $n \times m$ translational components $M_{i,j}$ interconnected by springs inside a translational container component M. This simple mechanical system models the response —or displacement $x(t)$— of the overall system when a perturbation is applied to the elements. Let us make the supposition that the springs are linear and elastic with spring constant k being homogeneous for all springs. We aim to study the stability of the system when it faces perturbations, where the perturbation can be modeled as a given time-dependent force $F_{i,j}(t)$ acting over any translational component $M_{i,j}$ in the system. The governing differential equation for the translational motion of each component $M_{i,j}$ is given by the relation

$$k\left[x_{i-1,j}(t) - 2x_{i,j}(t) + x_{i+1,j}(t)\right] + F_{i,j}(t) = M_{i,j}\ddot{x}_{i,j}(t), \tag{1.1}$$

where $x_{i,j}(t)$ is the displacement and $\ddot{x}_{i,j}(t)$ the aceleration of $M_{i,j}$. This previous relation applies for all components $M_{i,j}$ inside the container, with indices running $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Also note that $x_{0,j}(t) = x_{n+1,j}(t) = x(t)$. For the container component M, we suppose that this component represents the complete (global) system by implying that $M = \sum_{i=1,j=i}^{n,m} M_{i,j}$. The governing equation for its translational movement is

$$k\Big[x_{1,1}(t) + x_{1,2}(t) + \cdots + x_{1,j}(t) + \cdots + x_{1,m}(t) + \\ -x_{n,1}(t) - x_{n,2}(t) - \cdots - x_{n,j}(t) - \cdots - x_{n,m}(t)\Big] = M\ddot{x}(t). \tag{1.2}$$

A system of $(n \times m) + 1$ differential equations results from using equations (1.1) and (1.2) for all the elements in the problem. Introducing the vector of unknowns $\boldsymbol{X} =$

$\left[x(t)\, x_{1,1}(t)\, x_{2,1}(t)\ \ldots\ x_{i,j}(t)\ \ldots\ x_{n,m}(t)\right]^{\mathrm{T}}$, the system of differential equations can be written in vector form as

$$\boldsymbol{A}\ddot{\boldsymbol{X}} = \boldsymbol{K}\boldsymbol{X} + \boldsymbol{F}, \tag{1.3}$$

where $\boldsymbol{A}$ and $\boldsymbol{K}$ are the inertial and elasticity matrices, and $\boldsymbol{F}$ is the vector accounting for the external forces —or perturbations. Equation (1.3) is a second-order differential equation that needs to be integrated in time, including the initial and forcing conditions of the problem. Our approach is to use the fourth-order explicit Runge-Kutta method as the temporal integration scheme that guarantees an accurate numerical solution [73]. Nevertheless, analytic approaches like the Laplace and Fourier transform can be applied as well.

As described before, diversity can be defined as:

- the multiple relations between individuals (Balance or diversity of composition),
- or the multiple attributes of individuals (Variety, or $\alpha-$diversity and $\beta-$diversity).

We characterize diversity by varying these two conditions in the mechanical system. Regarding the first diversity condition, we evaluate multiple n and m components, such that the system differs in the amount and type of relations. Increasing n leads to multiple components connected in series by springs, while $m > 1$ leads to the parallel connection of those components —attached to the non-inertial container component with springs. In the case of the second diversity condition, we suppose that the components vary in magnitude according to relations $M_{i+1,j} = \alpha M_{i,j}$ and $M_{i,j+1} = \beta M_{i,j}$, where $\alpha, \beta \in \mathbb{R}^+$ are algebraic coefficients. We run several simulations varying n, m, α, and β parameters, but fix a constant type of perturbation $F_{i,j}(t)$. Concretely, we set the perturbation to only affect component $M_{1,1}$ and to be a single period of 2π of a sinusoidal function of amplitude M.

The numerical results for one first set of simulations are presented in Figure 1.2. In this first set of simulations we vary n and m, but keep $\alpha = \beta = 1$. The displacements of $M_{1,1}$ and M and their phase diagrams of momentum against displacement are presented in that figure. We observe that a single component $n = m = 1$ setting is unstable: the amplitude of the non-inertial container's displacement $x(t)$ grows unbounded, and its phase function diverges from an initial state. Adding components to the system, on the other hand, stabilizes the translational movement of the container. In those cases, the container's position is displaced from its original state, and even though it is dynamic, it remains bounded with a certain amplitude. This dynamical system is fully conservative —without the presence of dissipative mechanical elements—, but the bounded asymptotic behavior of displacements is only appreciated when there are several constitutive elements: the complex but never divergent response of M is portrayed. Even if the perturbation coincides exactly with the resonance mode of the single component, we observe that including elements to the system displaces the natural resonance modes. Another observation that can be extracted from the present model results is that the dynamic response of the system is complex and cannot be easily predicted. More importantly, it can be noticed that this kind of diversity fosters the stability-resilience of this complex system. Resilience being understood as a bounded response, and diversity as the multiplicity of relationships between constitutive individuals.

In the second set of simulations we fix the $n \times m$ number of constituent elements of the system, but evaluate a change in an attribute of the element. This is characterized by varying the α

and β coefficients in a range of values close to one. We simulate the time span $0 \leq t \leq t^f = 100$ and process the numerical results by calculating the temporal integration of the displacement $x(t)$. Hence, the integral result can be used as a measure of resilience: a small value refers to a stable state, while a high value means an uncontrolled state that tends to diverge. The contour plots of the integral value of displacement are presented in Figure 1.3. We observe that cases $n = m = 2, 3, 5, 10$ differ in the integral value result, which can be explained by the different average displacement $x(t)$ resulting from $F_{1,1}(t)$: the higher the number of elements, the higher the amplitude of the perturbation force and therefore the mean displacement. Nevertheless, the important result is to locate the high results of the displacement's integration for each $n = m$ case. We observe that increasing α –which has to do with the variation of the strongly related components to each other— damages the stability of the system. We also detect some control given by the β parameter for cases $n = m = 5$ and $n = m = 10$, such that it can improve the stability for systems with several constitutive elements. Consequently, we can not conclusively declare that all types of diversity have a positive relationship with enhancing the stability of the system. We can only state that when α is increased, global stability is reduced compared to having homogeneous constituent elements $\alpha = 1$. But if diversity is characterized by the $\beta-$diversity —which has to do with the variation of the fewer related components to each other— this seems to promote the resilience of the complete system.

With the aid of this mechanical model, we obtain results for a wide range of possible configurations. The most important conclusions have to do with the direct relationship that diversity —defined as the variation of the relations between the constituent elements of a system, and partially, of the attributes of elements— has with the improvement of the complex system's resilience. However, a completely positive relationship between the two concepts cannot be declared. Future investigations using this mechanical model can detail the causes. In particular, the displacement of the system can be solved by analytical techniques in order to detail the configurations where diversity has a positive relationship with stability and/or resilience.

1.3 Assessing urban sustainability and diversity from a multivariate approach

The goal from the sustainability approach, anyway, is to improve the resilience of cities [17, 74]. An important question is how is it going to be measured and achieved. Certainly, the path towards sustainability needs active decisions which must be addressed with scientific knowledge. The departure step is to understand cities' trends towards sustainability —or unsustainability—, with the necessary task of quantifying the environmental, social and economic capitals of the city. Yet, measuring these components is difficult because arising with a single measure of those can be done in several different ways.

1.3.1 Urban sustainability indicators

The most widely used approach to assess sustainability is to use an information system of indicators [10, 75–80]. Therein, the sustainability concept can be incorporated by accounting for its most relevant aspects. Most systems of indicators ease the processing of information

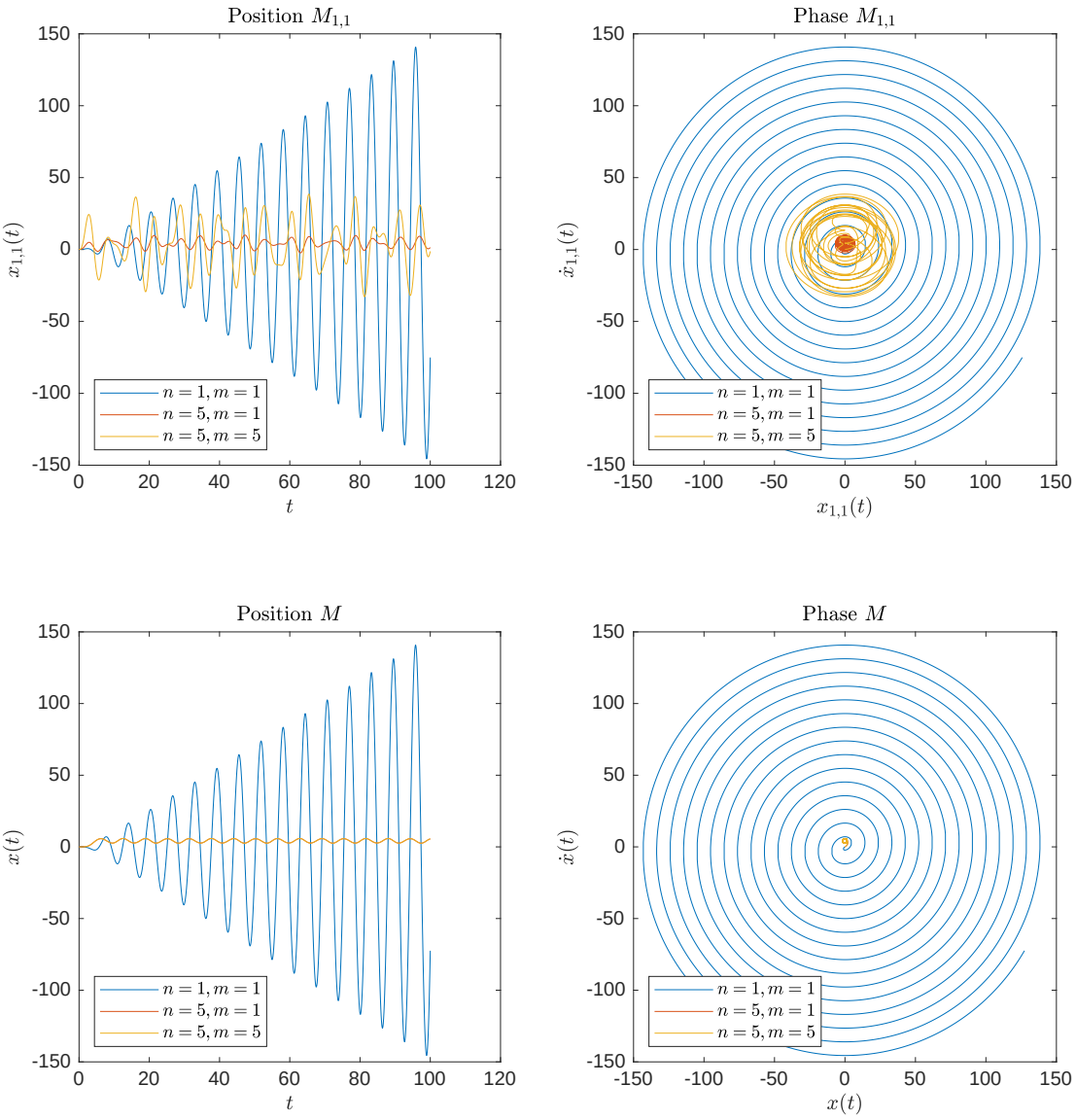

Figure 1.2: Stability of the mechanical system varying n and m, and fixing $\alpha = \beta = 1$.

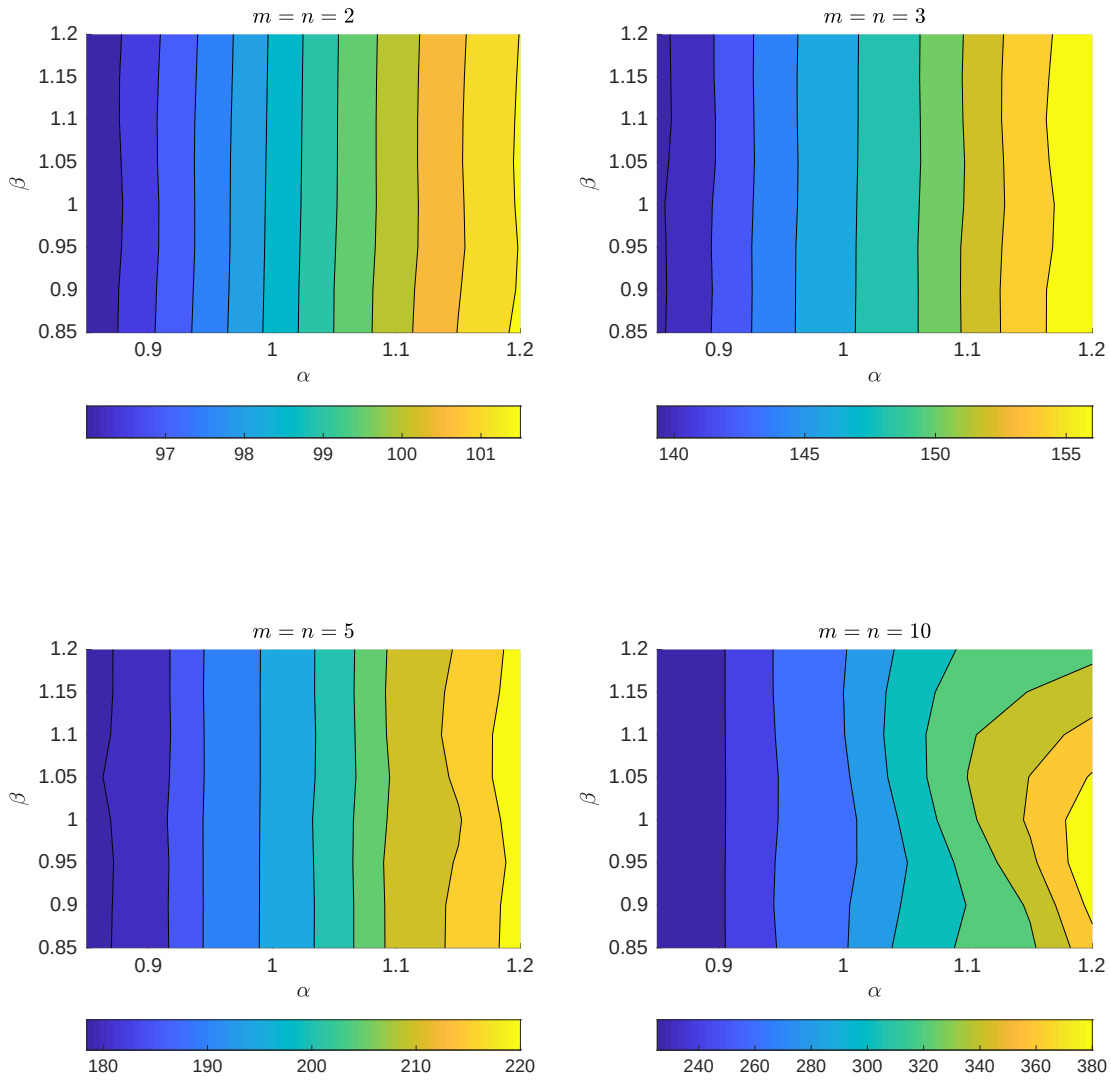

Figure 1.3: Stability of the mechanical model varying α and β.

and therefore, help to understand the urban system's performance. Urban indicators in particular play an important role in advancing the science and practice of sustainability. In practice, systems of indicators provide the most reliable tool to support urban planning: they can be used to track the city's trends and to evaluate the effectiveness of urban policies and strategies [76]. However, up to date, there is not a common framework for building an information system of indicators to evaluate urban systems [81]. Each one encompasses a different level of abstraction depending on its final goal, and on the scale and site of application. Therefore, the selection process of the indicators and the design of the information system represents a research topic in itself.

Some basic concepts need to be introduced in order to build an information system of indicators. A variable is any characteristic of the system, whereas an indicator represents a certain phenomenon that cannot be measured directly [79]. What is more, indicators are often the combination of several variables. Indices, or composite indicators, on the other hand, are the aggregation of variables or indicators that are typically aggregated with weighting methods [79, 80]. In this regard, some indices have been proposed within the framework of urban sustainability, for example, Ecological Footprint (EF), City Development Index (CDI), Green City Index (GCI), Environmental Sustainability Index (ESI) [75, 80]. However, urban systems —as well as other complex systems— are better represented using separated indicators instead of combined or aggregated attributes. Gathering multivariate and temporal information into single indices is usually not convenient for the following reasons. Firstly, most of the methods for aggregating information into indices produce a disconnection between the more intuitively original variables and the quantities resulting from these indices [82]. Secondly, aggregation also requires some form of human judgment, so that, it relies on potentially distorting assumptions [82, 83]. Lastly, aggregate indices barely capture interactions within complex systems [83]. Nonetheless, the direct interpretation of disaggregated indicators is difficult, and some level of aggregation is needed to portray an integrated picture of the system: aggregated measures surely reduce the amount of information, but they also highlight patterns and enhance the communication of results.

Within the urban sustainability framework several systems of indicators have been proposed, some of which adopted the sustainable development's dimensions as a hierarchical categorization [76, 79]. Systems of indicators that describe the environmental, economic, and social components of the city. Each one of these components (i.e. sustainable development dimensions) is further divided into categories (e.g. energy, transport, air quality, housing, education, businesses) that group indicators. Moreover, in the industrial ecology framework, some sets of indicators have been proposed: most belong to Material Flow Analysis (MFA) [84], but also to the urban metabolism approach [24, 85, 86]. With regard to this last approach, based on the dissipative structure theory [87], Zhang et al. (2006) [24] developed a system of indicators that characterizes the metabolism function of the urban ecosystem, grouped in four metabolic categories: sustaining input, imposed output, regenerative metabolism, and destructive metabolism. Later in this thesis, we will consider —and extend— the description made by these categories as a meaningful classifying hierarchy. Certainly, we will include some additional indicators that characterize social interactions, diversity of land-uses, services, and liveability.

In the present research, we aim to seek the relation between diversity and sustainability in urban systems using an information system's approach. As commented before, arising with

a single indicator of sustainability is not straightforward: the way to measure diversity and resilience, and the analysis of their correlation, are the main scientific motivation. The most problematic part is the need for reference values to provide a reasoned judgment of sustainability. In this sense, we decided to start by studying the city of Barcelona as a multivariate system that incorporates measurements of environmental and socio-economic indicators between the years 2003 and 2015.

1.3.2 Diversity measures

With respect to diversity measures, most of them are designed to compress data and to transform differentiated entities into single values. The cost of analyzing the system's diversity with the compressed information requires to be acknowledged: in the process of measuring diversity, one may treat economic, social, and environmental variables indistinctly. This can lead to inaccurate readings about the real system's diversity.

Page [47] classified diversity measures within five different types: (1) variation, (2) entropy measures, (3) distance measures, (4) attribute measures, and (5) disjoint population measures. Moreover, Anderson [71] stated two classes of $\beta-$diversity measures: (1) classic metrics (e.g. Shannon, Whitaker), and (2) multivariate measures, which are based on pairwise dissimilarities or distances among sample units. Even though the classifications of diversity measures given by [47, 71, 72] are relevant, we will use the following categories: variation measures, entropy measures, and distance measures.

The first category, variation measures, comprises statistical indicators of variation, such as variance (i.e. statistical variance [88]), the coefficient of variation (i.e. standard deviation), and the species richness. This type of measures fails to portray diversity when the data is distributed homogeneously at the extremes [47]. Within the entropy measures category, the following measures have been widely applied (see [89] for a deep explanation on this topic): Shannon's entropy [90], Shannon-Wiener's index [91], Simpson's index [92], and the Rao's quadratic entropy [93]. In general, entropy measures include relative abundance information: it considers the number of types and the distribution across types [47]. Information entropy indicators (e.g. Shannon's entropy or Simpson's diversity index), account for widespread types of distributions which are truly related to diversity but fail to take into account the extent of differences between types. Distance measures, on the other hand, are used to capture differences between types in a sample —disparity—, and differences between samples (i.e. disjoint populations [47]). Some of the statistical techniques that associate distance measures with diversity are the following: Weitzman [94], Phylogenic Distances, Multi-dimensional scaling (MDS), Discriminant Analysis, some types of Cluster Analysis (CA), and ANalysis Of VAriance (ANOVA). Some of these are multivariate measures built over pairwise dissimilarities (e.g. Euclidean distances) that give an accurate indication of diversity in a dataset.

In this respect, Multi-Dimensional Scaling (MDS) or Principal Coordinates Analysis (PCoA) have been used recently in biology to examine and visualize diversity patterns in multivariate datasets [95–98]. The potential relationship between observations and variables are explored in these methods by superimposing labels on observations (i.e. identifying groups) [71]. Since PCoA is a distance-based method (dissimilarities), it does not provide a direct link between the components and the original variables which makes it hard the interpretation of components and variables' contributions [99]. Instead, the components and variables can be

easily analyzed using Principal Component Analysis (PCA) since it is based on correlations. Regarding diversity, PCA (conceptually similar to PCoA) have been used in genetic diversity studies [100, 101] and in urban typologies [102–104].

Our approach is to explore the diversity of urban systems through multivariate statistical methods. Specifically, we opted for PCA and Multiple Factor Analysis (MFA). Given that multivariate data systems are complex and hold much information, these methods provide a balanced trade-off between the large volume of initial data and its final aggregation. The way to measure diversity with multivariate statistical methods is explained by their capacity to reduce the dimensionality of the dataset, transforming the original data of correlated variables into a set of a few representative variables —known as principal components— that extract the most relevant information [105, 106]. Under this approach, diversity can be seen as equivalent to dispersion in the multivariate dataset. The objective is to visualize and measure the relations between the system's components, such that, the more differentiated —distant— entities, the more diversity indication. MFA is an extension of PCA that reveals the relationship between observations and groups of variables (if there are any) [107, 108]. Additionally, we apply the Hierarchical Clustering Analysis (HCA) as a multivariate technique that groups observations based on similarities.

Our concern, in this thesis, it to apply multivariate statistical methods to understand the diversity-resilience relationship in a certain information system that describes a city. The contribution relies on the dimensionality reduction given by multivariate statistical methods, and therefore, the possibility of the human reading of the great amount of information comprised in the information system. From a formal perspective, the characterization of urban systems has been increasingly performed by means of a multiplicity of variables instead of aggregated attributes. This has been motivated by the accessibility to open data sets and the proliferation of novel telecommunication technologies, mainly of those designed to perform measurements in time and space through remote sensors. In point of fact, sensors in a single city can generate more than 10M observations per year (between the localized and temporal observations), which agrees with the definition of big amount of data or Big Data. Reading this amount of information surpasses human capabilities, and even worst, too many information can affect negatively the conscious decision making.

1.4 Computer-based urban analytics

The processing of a big amount of information can only be achieved by means of computers. That is exactly one of the issues in the computational science field: analyzing the big amount of information comprised in the multivariate information system requires the development of algorithms that can be executed efficiently by computers. Especially, the algorithms that reduce the dimensionality of information systems: those methods that aid the system's readability. In our case, we are interested in the development of a computer-based methodology to reveal the diversity of the information system by applying the PCA as the dimensionality reduction technique. By applying PCA, the description offers the possibility of revealing a different representation of system's relations, as (a great percentage of) the information contained in the original variables are encoded in the new set of uncorrelated variables. The first uncorrelated components can be selected in order to illustrate graphically the relationships be-

tween indicators, system's elements, and uncorrelated components, which in the case of the raw information system could not be possible to describe. Since distanced positions of the observations in the PCA indicate a diverse behavior, exposing those differentiated results should be done somehow: one basic visualization method is the plot of variables and the system's elements in relation with the uncorrelated components —called biplot—, and the analysis of the differentiated values. Complimentary, a descriptive analysis of the principal components can be performed: the comprised information in those principal components can be described and interpreted by considering uncorrelated components' loadings and biplots [105, 109], and thereby giving a conceptual label to each of them. Although multivariate information systems are widely used in urban studies, to the best of our knowledge, a PCA-based methodology to assess diversity in an urban sustainability framework has not been done so far. There are no previous applications of PCA or MFA in urban systems that present diversity results in synthetic and meaningful graphs, neither that summarizes the whole output relating observations and variables. At most, it has been used in references [102–104] to identify neighborhood typologies using a multivariate description of urban systems, where several variables describing race and ethnicity, age, family structure, education, employment, income, and housing, were measured.

It is also remarkable that the dimensionally reduced representation reveals the temporal evolution of the system in the measured time period: the time-dependent results can be visualized by notating distinctly the initial from the final description in the biplots. One naif approach to understanding the temporal change of the system is to use *Confidence Ellipses* in the dimensionally reduced representation [110–112]. This is a technical approach that quantifies the dispersion of the time-dependent results. But the multidimensional (reduced) description that evolves in time is analogous to a n-dimensional cloud of points that deforms gradually with time. This is exactly the field of study of morphology change, where research is mainly devoted to interpreting the temporal movements of the face in the recognition of expressions and feelings [113]. Several biomechanical problems of movement are linked with the temporal tracking of a deforming cloud of points which represent moving limbs [114–118]. For these movements, classical numerical methods have been proposed to recover displacement or velocity fields. Research in the medical area has also developed computer algorithms that facilitate the human readability of images —or videos— of moving media inside the human body, like the four-dimensional obstetrical ultrasound imaging [119–121]. It is common nowadays that the acquisition of the time-dependent and three-dimensional data is done through non-invasive tests in the medical imaging applications, or by computational vision, e.g. using RGBD cameras, such that, big amounts of data need to be stored and processed.

After collecting the information of the three-dimensional location of points that gradually displace from one capture to another, the deformation information can be stored in several forms. Note, anyway, that the amount of information can be large —of the order of Megabytes—, where computational methods to process large data quantities become imperative. Powerful algorithms have been developed in this sense. Most aim to represent the abstract information in a human advantageous way [122–125]. Diagnostic indices can be one approach, but these are not useful in most Big Data applications [126]. A recent approach is to use machine learning to build the deformation register, where deformation models are built over a mathematical basis: basically, algorithms that are designed to approximate the discrete displacement —or velocity— fields by using optimization techniques, which try to

fit the coefficients of a polynomial function by solving a least square problem over the discrete dataset [127]. The data information of the movement —or four-dimensional morphology change model— can be compressed, using machine learning, in a few coefficients of the order of bits.

Our idea in this thesis is to use geometry and physics-based approach in the analysis of the temporal change of multidimensional data. In this line, the data cloud can be understood as a deforming material which is subjected to the physical laws of motion. Specifically, to the continuum kinematics of deformable bodies, where a profound mathematical framework can be used to quantify the deformation of the data cloud [128]. Actually, the mathematical framework of tensors —which is usually not well spread among the scientific community— can determine exactly the strain condition of a given material that is typically subject to the imposition of displacements or loads. But the problem remains to be similar: the deforming data cloud constitutes a discrete description of the material, while the kinematic theory relies upon a continuum approach. Hence, several methods have made an effort to approximate the discrete displacement fields. The use of spline interpolations of point-wise data to obtain continuous representations of the strain-rate have been extensively used in earth sciences and medical imaging analysis. Indeed, the calculation of a strain-rate state has been used for studying point-wise observations of displacements made by geodetic networks and strain evaluation for geophysical information. In this sense, the earthquake focal mechanisms have been examined by using strain analysis: locally maximizing the strain-rate along with any reference system (mostly given by satellite measurements). Furthermore, some medical imaging analysis relies on the localization and reconstruction of continuous strains obtained from discrete tomographic methods [129, 130]. This type of applications is mostly related to the quantification of structural integrity by means of inverse methods.

Since a quantitative methodology that describes the temporal change of the urban system —represented by a multivariate data-set— has not been carried out before, several chapters of this thesis are devoted to proposing a novel methodology in this field by including the strain-rate tensor as the fundamental metric. The only similar work has been the calculation of the strain-rate condition of geodetic observations in [131, 132], which gave accurate conclusions about the stress concentration rates of tectonic plates, and therefore, of possible seismic hazard potentials. Still, we think that the inclusion of this mathematical object can be a good starting point to estimate the time-dependent variation of discrete data-sets. For example, the presence of localized strain-rate patterns has a direct relationship with the observed system's trends. An analysis of the extension and contraction rates can indicate a strong influence of transient behavior. High extension rates are associated with the separation of the system's elements with respect to the multivariate description. This is a differentiated behavior of each element in the system's description. In contrast, the contraction rates indicate clustering of the elements of the system with time. One chapter of this thesis raises a computational science technique that facilitates the identification and visualization of the strain-rate patterns: the technique adopts the principal components and principal orientations of the strain-rate tensor, which are responsible for describing the maximum —extension— and minimum —contraction— strain-rate values and their orientations. Post-processing those principal components allow to visualize the orientations of extension and contraction rates inside the data cloud, and the plots of these results aid to localize patterns of diversification or convergence of the elements of the system. If the city's description based on the information system relies on accurate data,

and an adequate visualization and communication strategy (like the strain-rate visualization technique) is implemented, then the diversity state of the urban system can be read.

Nevertheless, an up-to-date and permanent self-assessment framework is required in order to deliver precise directions based on the reading of the results. Building a methodology in this sense —not only the sustainability framework— can be a tool for decision making in urban governance. The dimensionality reduction given by PCA enables the possibility of live prediction of urban phenomena: computer science and computational methods on computer vision and machine learning use dimensionality reduction and pattern identification to foretell the future states of the system [133]. As described before, the displacement field of a cloud of points can be represented by a few parameters of a polynomial function. Theoretically, one can predict the future state of the displacement field by evaluating this function in future times. These algorithms, which are commonly referred to as Artificial Intelligence (AI), have been used in Risk Analysis [134] and to predict failure of systems [135]. The logic behind these techniques rely on the pattern identification in few-dimensions, and the evaluation —on live— of the system's future states. Nevertheless, those mathematical methods are subjected to probability, since non-physical models can be low-accurate in reality [136].

Our approach in this thesis, again, relies on the classical physics of movement in order to foretell the future states of the system. When considering the cloud of points that defines the state of the system as a deformable medium subjected to the laws of motion, one can go a step further and consider that the medium is subjected to forcing constraints that make it deforms in that way. If those forces are known —or supposed— one can obtain the mechanical properties of the medium, like the elasticity tensor, by solving the inverse kinetics problem of elasticity [137, 138]. Hence, the mechanical properties of the deformable medium mean that the stress field can be calculated, complimentary with the strain field. More importantly, by solving the elasticity problem one can predict the future states of the system. This has another profound impact, which has not been explored up to this date: there is a direct link between the resilience of the urban system and the resilience of the deformable media, which is the energy that the medium is capable to dissipate in the event of a deforming process [139]. This energy is related precisely with its stress-strain product [128].

1.5 Thesis outline

The work developed in the framework of this thesis can be enclosed in the main objective of proposing a methodology for analyzing complex urban systems, described as multivariate information systems that portray the environmental, economic, and social characters of cities. In particular, the methodology has to be capable of describing the diversity and its relation with a better response to perturbations or hazards, and therefore, sustainability through time.

The specific content of this work is divided into several topics, which are developed progressively as displayed in Figure 1.4, and that will be presented in the document as follows.

Chapter 2 opens the methodology with the application of an information entropy measure to the information system. The analysis of diversity with this simple index is intended to introduce the theoretical and conceptual description of the city's behavior and how do we model it. We use the case study of four districts of Barcelona, with a temporal span of 10 years from 2003 to 2012.

Chapter 3 is devoted to the PCA of the information system describing four districts of the city in order to reduce its dimensionality. The exhibition and measurement of diversity among the elements of the urban system is the main objective of this chapter and specifically, it is extensively covered with respect to four districts of Barcelona in the temporal span of 13 years, from 2003 to 2015.

Chapter 4 evaluates some other multivariate statistical approaches to the measurement of diversity and the temporal-change analysis of the urban system. Specifically, we apply PCA, MFA, Cluster Analysis, and Confidence Ellipsoids as the statistical techniques. In that chapter, we also expand the case study and extend it to the ten districts of the city of Barcelona in the time span between the year 2003 and 2015.

Chapter 5 quantifies and visualizes the temporal change of the urban system by using some techniques in computational sciences. Concretely, this is achieved with a computational mechanics approach based on the continuum mechanics theory. Precisely, we apply the kinematic ideas of continuum deformable media in order to measure the deformation rates (or strain-rates) of the dimensionally-reduced description of Barcelona from 2003 to 2015.

Finally, Chapter 6 closes the thesis with some conclusions and the summary of further possible research lines. Especially, it leaves behind the ideas to predict future states of the system and physically relate the concepts of strain-stress of the historical data and resilience of the urban system.

1.6 Research dissemination

The research work contained in each chapter is quite self-contained even if this implies the need for repeating some information, especially the definition of the case study that grows in data dimensionality through the document. The notation is gradually introduced as it is required and may vary (slightly) from one chapter to another. This is due to the fact that each chapter of this thesis has been disseminated in the form of a book chapter, and in the format of articles in peer-reviewed scientific journals, as follows:

1. Chapter 2:
   L. Salazar-Llano, and M. Rosas-Casals, ”Assessing sustainability in urban systems based on information entropy. The case of four districts of Barcelona between 2003 and 2012.”, in *Urban Resilience: Methodologies, Tools, and Evaluation*. Resilient Cities: Re-thinking Urban Transformation (Book Series). Springer. Accepted.

2. Chapter 3:
   L. Salazar-Llano, M. Rosas-Casals, Ortego, M.I. ”An Exploratory Multivariate Statistical Analysis to Assess Urban Diversity”. *Sustainability*, 2019, 11, 3812; https://doi.org/10.3390/su11143812.

3. Chapter 5:
   L. Salazar-Llano, and C. Bayona-Roa, ”Visualization of the strain-rate state of a data-cloud: Analysis of the temporal change of an urban multivariate description”, *Applied Sciences*, 2019, 9, 2920; https://doi.org/10.3390/app9142920.

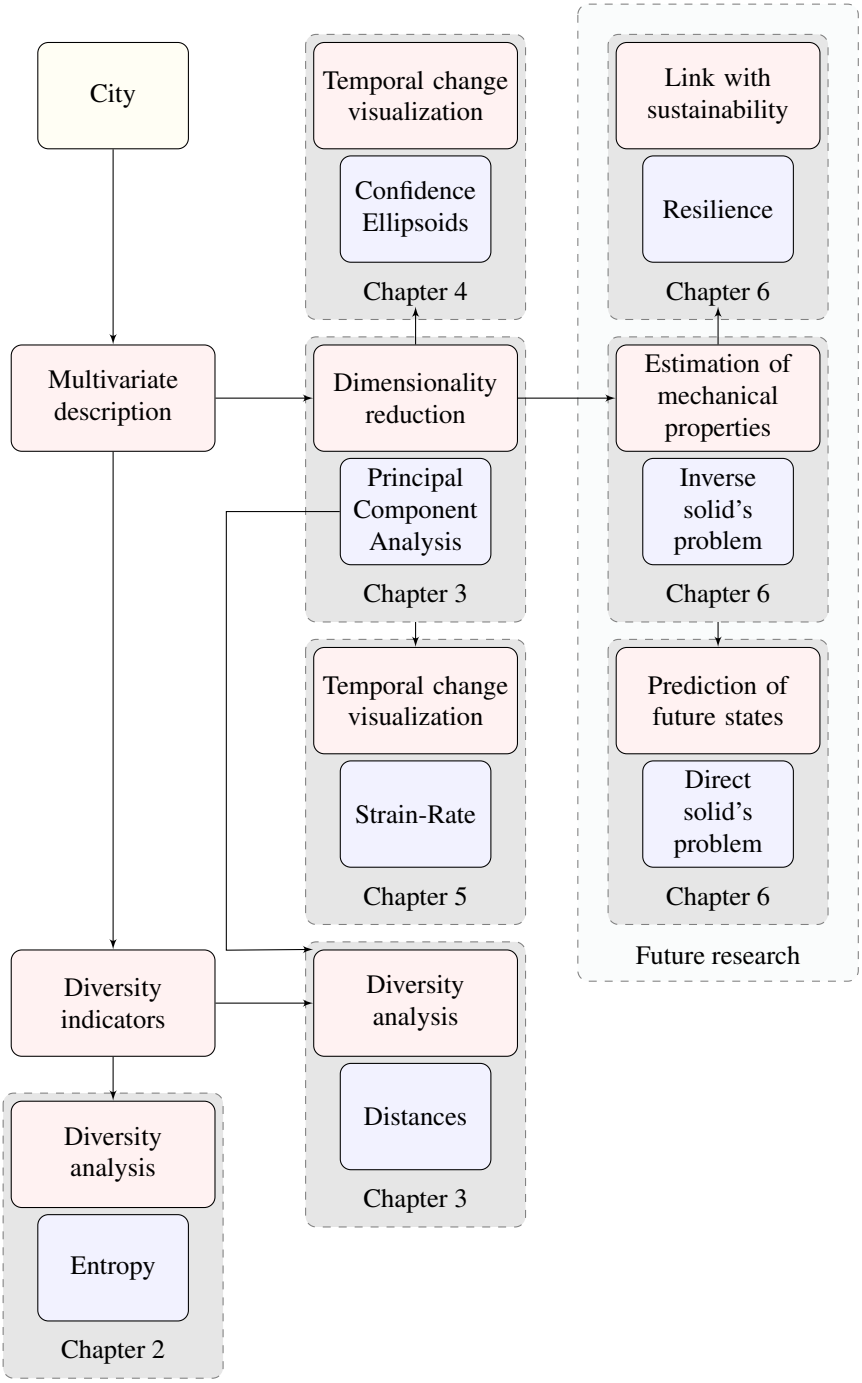

Figure 1.4: Flow chart of the thesis outline.

Chapter 2

Assessing an urban system based on information entropy. The case of four districts of Barcelona between 2003 and 2012.

The recent progress in urban ecology comes with several indicators and concepts to analyze urban systems. Cities are systems that can be considered as complex, adaptive, dynamic, and heterogeneous, also characterized by highly diverse components and multi-scale interactions. In this context, the ecological concepts of the urban ecosystem and urban metabolism can be introduced in the city's diagnosis and management. Moreover, diversity has been defined as a fundamental property of complex systems, but also as a descriptive indicator of ecosystems in ecology. Researchers have mainly used the information entropy as a measure of diversity that relates the science of complexity and ecology. Particularly, in the urban context, information entropy has been used as a measure of morphological diversity, functional diversity, and metabolic diversity. In this chapter, we propose to use an information entropy based methodology to describe the metabolism of the urban ecosystem and to assess its compliance with sustainable goals. This methodology comprises a temporal analysis of the urban metabolism of a city. Specifically, we apply it in the case study of four districts of Barcelona: Ciutat Vella, Eixample, Les Corts, and Gràcia, between 2003 and 2012. Results demonstrate that the methodology is able to track the temporal trends of the urban metabolism at the district's scale.

2.1 Introduction

In recent decades, the world has experienced rapid growth in the urbanization process. This has been the cause of important social, economic and environmental impacts, which have not been sufficiently understood. It has also contributed to global-scale problems such as climate change, and it has also been related to the local crisis related to the availability of resources and services [6]. In this sense, adequate urban management can bring opportunities for economic growth, cultural transformation, creativity, productivity, and innovation to the society

[18]. Certainly, urban policies are considered to potentially aid to mitigate climate change and satisfy the basic needs of the population. But this applies only if adequate directions, based on the deep knowledge of the socio-economic and environmental processes in the city, are implemented. Following these ideas, it has become necessary to understand and predict the relations and dynamics that govern the urbanization process. That is, to systematically study the interactions between society and environment in the urban land from the urban system perspective.

The recent concepts of urban ecology aid in the analysis, design, and management of cities [11]. The urban ecology field integrates theories and methods of both natural and social sciences to the study of urban systems [12, 13]. Essentially, the urban ecology science comprises the study of the multiple interactions and processes among the urban elements, and enables the description of spatio-temporal trends, socio-economic transactions, environmental impacts, and urban sustainability. This, having an emphasis on biodiversity, ecosystem processes, and ecosystem services [7]. Among urban ecology, cities are studied as heterogeneous, dynamic landscapes and as complex, adaptive, socio-ecological systems, in which ecosystem services relates the society and the environment at the different scales [7].

In both urban sustainability and urban ecology literature, the concept of the city as an ecosystem is widely used to analyze and identify environmental issues. As described by the ecology, environmental problems are an imbalance between the matter and energy flows in the ecosystem. Also, the concept of urban metabolism has been developed as a method of analysis that quantifies the flows of matter, energy, and information, inherent of urban systems. That method allows for identifying the level of order and development of the urban ecosystem within the urban ecology framework. In this sense, the ecosystem approach also allows to design urban policies, mostly in the aspects of natural processes reintegration, increasing resources efficiency, waste recycling, energy conservation, and production goals [5, 22]. Indeed, several authors have used the concept of urban metabolism to analyze the environmental issues associated with the different metabolic fluxes [5, 14, 22–24, 26, 140]. For example, Wolman (1965) [21] used this concept in his urban studies in order to analyze the metabolism of a hypothetical US city of 1 million inhabitants.

From the viewpoint of urban metabolism, it has been further associated as the entire technical and socio-economic processes that take place in cities and urban systems [5]. This concept directly relates growth, energy production, and waste disposal in the city: it is based on the analogy between the city and a living organism that demands resources from and releases wastes into the environment. This analogy has been useful for quantifying consumption and waste generation trends in growing cities [85]. But the first step of this methodology is necessarily to quantify matter and energy flows. The main advantage of the urban metabolism concept is that accurate temporal descriptions can be made by continuously evaluating the city, such that the development direction can be traced and control planning can be achieved. Indeed, being able to detect the change —and especially the metabolic change— of cities can be very helpful in the understanding of city's sustainability since the dynamic path of cities cannot be fully described due to the inherent uncertainty of urban systems [4]. Following the principles of urban ecology, the metabolic state of the city can be guided towards sustainable states through urban planning and policies' design [15, 18, 35].

More recently, the study of cities has been revisited from the perspective of complexity science. They are the principal example of a complex system: cities are diverse and exhibit

emergence and adaptive phenomena, among other features. Within complexity, theories have demonstrated the scaling of some properties and indicators related to the city's development and size [3, 141–144]. Also, Batty (2008, 2012) [4, 145] has studied urban growth and its relation to the form, the scale, and the transportation patterns of urban systems. One of the fundamental properties of complex systems has been diversity, that has been surprisingly used as a descriptive indicator of ecosystems in ecology. Some studies in socio-ecological systems argue that a certain increase in diversity improves responsiveness to disturbances and thus stability. In this way, it has been considered that diversity is related to the system's complexity, and therefore, to its robustness and resilience. However, the diversity-stability debate remains open due to the broadness of diversity's concept and the variety of measures that quantify it, as explained in the first chapter. Particularly, the information entropy concept has been used in urban systems as a measure of morphological diversity [146–150], functional diversity or land use diversity [151], and metabolic diversity [24, 152]. But all those previous works have described separated urban system's properties. Let us at this point explain in more detail the entropy of information as a measure of diversity.

Claude Shannon introduced the information entropy equation in 1940, within the information theory [90]. The information entropy measures the uncertainty of an information source, and can be considered as the average amount of information contained in a particular arrange of symbols. Also, it is a measure of the order of a system, meaning by order the state in which is easy to distinguish a differentiated structure. The information entropy, S, is mathematically described as follows: it is defined as the expected information content of a random variable. In an uncertain system, in which X is a random variable that expresses the system's state $X = \{x_1, x_2, \ldots, x_n\}$ for $n \geq 2$, the probability of occurrence of an event P is the frequency of the event with respect to the total number of occurrences of all events, $P = \{p_1, p_2, \ldots, p_n\}$ $(0 \leq Pi \leq 1; i = 1, 2, \ldots, n; \sum P_i = 1)$. One can compose a mathematical expression with the probability distribution, such that an overall quantity

$$S(x) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i), \tag{2.1}$$

is exactly zero when only one of the probabilities is equal to one and the others are zero. On the contrary, that quantity is maximum when all events are equally likely to occur. The previous equation is the entropy of the probability distribution, and quantifies the expected information in the distribution. In that equation, b indicates the base of the logarithm, such that when the base is $b = 2$, the units of entropy are bits of information, and when the base is Euler's number $b = e$, the units are called *nat*, which are typically used in the calculation of real variables. Another base that can be used is $b = 10$, called *dut*, which are also used for real variables.

From an analytical perspective of the urban ecosystem, Zhang (2006) [24] developed an information entropy-based methodology to measure the energy and matter flows of an urban system. It was based on the idea of maximizing the occurrences of optimal measured values. In that work, the information entropy was introduced as a measure of the system's optimality, and could be applied into a system of multiple indicators that describe the urban metabolism of a city. Using that methodology, the direction of change in urban development could be assessed in the urbanization process of the city of Ningbo between 1996 and 2003, since yearly

calculations were possible. Lin (2013) [152] replicated the methodology and made some contributions in order to evaluate the sustainable development of Guangzhou between 2004 and 2010. In the present chapter we evaluate that information entropy-based methodology in [24, 152] over an original elaboration of a multivariate system of indicators representing an urban system. The methodology aims to analyze the urban system's metabolic state and its evolution as described by four different metabolism types. This methodology is detailed in the next Section, together with the multivariate system of indicators for the case study of the four districts of Barcelona (Spain). Generally, the above methodology includes: first, the definition of the system of indicators comprising four metabolic categories (or types). Then, the calculation of the information entropy for the four types of metabolic categories, and finally, the composition of an overall appraisal score. In Section 2.3, the results of applying the methodology to the case study are presented. Finally, some conclusions close this chapter in Section 2.4.

2.2 Materials and methods

In this section, we first introduce the case study of four districts of Barcelona between 2003 and 2012. Then, we present the system of indicators and the entropy concept in the analysis of urban systems.

2.2.1 Case study

We study the case of four districts of Barcelona; Ciutat Vella, Eixample, Les Corts, and Gràcia. These districts are depicted in Figure 2.1 and are chosen considering four main criteria: (1) urban morphology or structure, (2) social characteristics of the population, (3) economic features and (4) the kind of development; planned or unplanned.

Barcelona had a surface of 10216 ha and a population of 1602386 inhabitants in 2014, therefore a density of 157 inhabitants/ha. Since 1987, the city has been divided into 10 administrative districts, which are subdivided into a total of 73 neighborhoods. The districts of Ciutat Vella and Eixample constitute the historical city center. Ciutat Vella has an area of 437 ha that comprises four neighborhoods: el Raval (1), el Gótic (2), la Barceloneta (3) and Sant Pere, Santa Caterina i la Ribera (4), and it has a total population of 100685 inhabitants. Eixample has an area of 748 ha which comprises six neighbourhoods: Fort Pienc (5), la Sagrada Familia (6), la Dreta de l'Eixample (7), l'Antiga Esquerra de l'Eixample (8), la Nova Esquerra de l'Eixample (9) and Sant Antoni (10), and it has a population of 263565 inhabitants. The district of Eixample is the densest of Barcelona, with a population density of 353 inhabitants/ha [153].

The districts of Les Corts and Gràcia, on the ohter hand, were annexed to the city in the last decade of the XIX century. Previously, they were former independent municipalities — Les Corts of Sarrià and Vila de Gràcia—, which after a conurbation process were included in Barcelona's municipality. The district of Les Corts has an area of 602 ha that comprises three neighborhoods: Les Corts (19), la Maternitat i Sant Ramon (20) and Pedralbes (21), and it has a total population of 81200 inhabitants. Finally, Gràcia has an area of 419 ha that comprises five neighborhoods: Vallcarca i els Penitents (28), el Coll (29), la Salut (30), la Vila de Gràcia (31) and el Camp d'en Grassot i Gràcia Nova (32), and it has a total of 120273 inhabitants

Figure 2.1: Barcelona area and location map of the four districts. Adapted map from Urban Guide City Map [154].

[153].

Thus, the case study covers an area of 2206 ha, being the 22% of the total city surface. With respect to population, the case study comprises 565723 inhabitants, accounting for 35% of the city's population. We believe that these districts are a representative sample of both the urban structure and the social network of Barcelona. In this sense, these structures are relevant actors of the energy, material and informational flows in the urban ecosystem.

2.2.2 Information system of indicators

We aim to represent the urban system by constructing an information system. The main idea is to collect time-dependent measurements of several social, economic, and environmental indicators at the district's scale. In order to build the system of indicators we take the following criterion into account. We define the information indicators in correspondence with previously benchmarked indicators, some of which belong to the *input supportive*, *output pressure*, *destructive* and *regenerative* metabolic categories defined in [24, 152]. We also consider the

availability of temporally disaggregated data by district. And this has been also the argument to select the four districts of the case study: these provide the most information that can be collected at the time of this investigation. Also, the potential information provided by each indicator in assessing the urban ecosystem evolution towards sustainability or unsustainability is included in the selection process. In this sense, we have examined the annual goals and the sustainable cities indicators in [22, 76, 79] and considered benchmarked indicators for composing our information system.

We consider a total of 43 indicators that are presented in Table 2.1 for composing the set of indicators, 18 of which belongs to the *Input Supportive* category, 10 belongs to the *Output Pressure* category, 6 belongs to the *Destructive Metabolism* category, and 6 belongs to the *Regenerative Metabolism* category of indicators. The Input Supportive category describes the natural capital of the city, but also the livability and the complexity of urban services. In that category, we include indicators that describe the diversity of legal entities, economical activities, and institutions associated with each district. We refer to the land use areas (see indicators A1-A12): housing, parking, commerce, industry, corporations, education, sanitation, sports, religion, and entertainment, among others. As a remark, Barcelona is considered as the support system of the four districts in some indicators of the Input category.

In the *Output Pressure* category we consider 10 indicators that describe the pressure exerted by the economic and social subsystems over the natural subsystem. That category is mostly related to the population size (social subsystem) and its consumption demand. In the *Destructive Metabolism* category, we include 6 indicators that reveal the environmental pollution within the urban ecosystem, especially the ones related to atmospheric pollution and waste generation. In order to measure the non-liveability of the urban ecosystem, we also include in the Destructive category the number of traffic accidents. The *Regenerative Metabolism* category comprises 6 indicators that are related to the ecological construction upon the regenerative infrastructure of the urban ecosystem and environment care.

We compile the data measurements for the system of indicators from the open databases of [155] and [153]. Indeed, the data is compiled for a period of time between 2003 and 2012 based on the data availability. The compiled datasets for the four districts of the case study are available online at `https://summlabbd.upc.edu/SalazarLlano/PhDThesis/`.

2.2.3 System analysis using information entropy

Once the information system is set, we apply the information concept in order to analyze the compliance of the measured data to the type of metabolisms. Let us introduce the following mathematical notation: we denote as $\boldsymbol{X} = \left(x_{ij}\right)_{n\times m}$ the dataset matrix, where x_{ij} is the measured value in the $j-$th event for the $i-$th indicator of the system. Let $i = 1, 2, \ldots, n$ be the counter of the n indicators in a certain metabolism category of indicators, and $j = 1, 2, \ldots, m$ the counter of the m events — or years. Each dataset matrix comprises the temporal data for the district in the period of time between 2003 and 2012, and the measured values for the indicators in each metabolism category,

Table 2.1: System of indicators.

Type		Id	Indicator	Unit
INPUT	Supportive	A1	Residential area	m^2
		A2	Parking area	m^2
		A3	Commercial area	m^2
		A4	Industrial area	m^2
		A5	Office buildings	m^2
		A6	Educational buildings	m^2
		A7	Health service buildings	m^2
		A8	Hotel buildings	m^2
		A9	Sport area	m^2
		A10	Religious buildings	m^2
		A11	Entertainment buildings	m^2
		A12	Other land uses	m^2
		A13	Cultural centres	Unit
		A14	Sports centres	Unit
		A15	Security level in the neighbourhood	Points
		A16	Public water fountains	Unit
		A17	Playgrounds	Unit
		A18	Active market stalls	Unit
		A19	Street trees (unit = 5.6 m^2)	Unit
		A20	Urban parks	ha
		A21	Urban green area besides parks	ha
OUTPUT	Pressure	B1	Natural growth rate	-
		B2	Population density	people/ha
		B3	Total number of motor vehicles	Unit
		B4	Total volume of water supplied to the commerce	m^3
		B5	Total volume of water supplied to domestic consumption	m^3
		B6	Total volume of water supplied to the industry	m^3
		B7	Total volume of water supplied to other usages	m^3
		B8	Available household income per capita	Index
		B9	Accommodation places	Unit
		B10	Final energy consumption	Mwh/year
DEST-RUCTIVE		C1	Average levels of PM_{10}	$\mu g/m^3$
		C2	Maximum value of average daily levels SO_2	$\mu g/m^3$
		C3	Average levels of NO_2	$\mu g/m^3$
		C4	Average levels of CO	mg/m^3
		C5	Collected volume of non-recyclable waste	ton
		C6	Traffic accidents	Unit
REGENE-RATIVE		D1	Collected volume of paper and cardboard	ton
		D2	Collected volume of glass	ton
		D3	Collected volume of containers	ton
		D4	Maintenance cost	Miles of €
		D5	Streets and zones with pedestrian priority	ha
		D6	Treated small containers	ton

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}. \tag{2.2}$$

Using one given dataset matrix for a metabolism category and a district in the time period, in the following lines we demonstrate the calculation of the information entropy quantity associated to that specific matrix. Understandably, this same procedure is performed for the other categories and districts of the case study.

The first step is to indicate the system's optimality; Ideal indicator's values are established by analyzing the type of data values that the category should have. An assessment criterion —such as the ideal values for each indicator of the system – is set formally by introducing an assessment matrix: the assessment criteria matrix $\boldsymbol{V}$ is defined as $\boldsymbol{V} = (v_i)_n$, where v_i is the value assessment criteria for the $i-$th indicator.

In our case, we define ideal values as: $v_i = \max\{x_{ij}\}$ for $j = 1, 2, \dots, m$ for the *Input supportive* and *Regenerative metabolism* categories —positive indicators—, and $v_i = \min\{x_{ij}\}$ for $j = 1, 2, \dots, m$, for the indicators within *Output pressure* and *Destructive metabolism* categories —negative indicators. The closeness between x_{ij} and its ideal value of the assessment criteria matrix describes the state of the system relative to the ideal reference value.

Naturally, the proximity relationship to the ideal state can be defined with a normalization matrix $\boldsymbol{Q} = \left(q_{ij}\right)_{n\times m}$, where q_{ij} is the $j-$th data of the $i-$th indicator, and $q_{ij} \in [0, 1]$. Based on the previously defined assessment criteria, for positive indicators the normalization is $q_{ij} = \frac{x_{ij}}{v_i}$ for $j = 1, 2, \dots, m$, and for negative indicators is $q_{ij} = \frac{v_i}{x_{ij}}$ for $j = 1, 2, \dots, m$.

The probability matrix $\boldsymbol{P}$ is defined as $\boldsymbol{P} = \left(p_{ij}\right)_{n\times m}$, where each component is calculated as $p_{ij} = \frac{q_{ij}}{\sum_{j=1}^{m} q_{ij}}$ for $j = 1, 2, \dots, m$ and $i = 1, 2, \dots, n$. This probability measures the appraisal for each indicator with respect to the set of events. Also, the probability fulfils the properties $p_{ij} \in [0, 1]$ and $\sum_{j=1}^{m} p_{ij} = 1$.

The information entropy S_i, given the probability distributions for the system, is calculated for the indicators as $S_i = -\kappa \sum_{j=1}^{m} p_{ij} \ln\left(p_{ij}\right)$ for $i = 1, 2, \dots, n$. Here κ represents a constant $\kappa = \frac{1}{\ln m}$ which guarantees that $S_i \in [0, 1]$. When the probability is zero, $p_{ij} = 0$, the supposition that $p_{ij} \ln\left(p_{ij}\right) = 0$ is implemented.

We can also weight the indicators using the information entropy results. We use a weighting value into each indicator to reveal its importance in the description of the system and obtain relevant information about its evolution. The smaller the value of the information entropy, the better the indicator is to providing compliant information along the set of events, and therefore, describe the system changes in the time span. On the contrary, the bigger the value of the information entropy, the lesser compliant information of the system is provided by the indicator. This evince a lack capability of the information given by the indicator to describe the system's optimality and therefore, it is neglected. Formally, the entropy-based weight W_i is defined for each $i-$th indicator as

$$W_i = \frac{1 - S_i}{\sum_{i=1}^{n}(1 - S_i)}, \tag{2.3}$$

where $W_i \in [0, 1]$ and $\sum_{i=1}^{n} W_i = 1$.

A new normalization matrix $\boldsymbol{Q}^* = \left(q_{ij}^*\right)_{n\times m}$ can be calculated using the previously obtained weights: each component of the weighted normalization matrix is given by $q_{ij}^* = W_i q_{ij}$ for each $j = 1, 2, \ldots, m$ and $i = 1, 2, \ldots, n$. This matrix is used both for the yearly information entropy calculation —to be explained next—, and for the Appraisal Score calculation.

Accordingly, one can calculate the information entropy *for each each year of the study*. This approach estimates the response of the information given by the indicators *in a specific year*. In that case, the information entropy can be calculated using a *yearly probability* for each $j = 1, 2, \ldots, m$, as,

$$S_j = -\frac{1}{\ln n} \sum_{i=1}^{n} p_{ij}^* \ln\left(p_{ij}^*\right), \tag{2.4}$$

with $\frac{1}{\ln n}$ being the scaling constant with respect to the entropy of all n indicators. Note that the probabilities of events are directly calculated using the weighted normalized data

$$p_{ij}^* = \frac{q_{ij}^*}{\sum_{i=1}^{n} q_{ij}^*} \qquad \text{for } j = 1, 2, \ldots, m \text{ and } i = 1, 2, \ldots, n. \tag{2.5}$$

The newly defined probability also fulfils the properties $p_{ij}^* \in [0, 1]$ and $\sum_{j=1}^{m} p_{ij}^* = 1$.

The bigger the value of information entropy for each $j-$th year, the greater the number of occurrences (compliances) of the information given by all indicators in that year. This means that in a given year the data of the group of indicators is behaving as expected in the assessment criteria. The instantaneous metabolic state and the evolution of the system can therefore be described using this calculated quantity.

Finally, an overall appraisal score (G) can be defined based on information entropy [24]. This score is calculated with the information entropy-based weights and the weighted normalized entropy, and it is used to integrate multidimensional information of the system. For which, the bigger the value, the better the behavior of the system with respect to the assessment criteria set in the system of indicators. It can be calculated as

$$G_j = \sum_{i=1}^{n} W_i q_{ij} \qquad \text{for } j = 1, 2, \ldots, m. \tag{2.6}$$

2.3 Results

In this section we present the results of applying the information entropy as a measure of the urban ecosystem of the case study, as introduced in [24, 152]. First, we discuss the information entropy results obtained for each metabolism type of the districts: Input Supportive (ΔeS1), Output Pressure (ΔeS2), Destructive Metabolism (ΔiS2) and Regenerative Metabolism (ΔiS1). Afterwards, we compare and discuss the Appraisal Score (G) evolution for the districts in the time span.

The results obtained for each metabolism category (grouped indicators) express the behavior of the district with respect to that metabolism. As explained before, we use an assessment criterion to normalize the dataset using an optimal measured value for each indicator. That is, we define the maximum positive measured value of each indicator as the optimal value for the *Input supportive* and *Regenerative metabolism* types, since the highest value in these indicators represents fitness and regeneration capacity, whereas, the minimum positive value is used as optimal value for the *Output pressure* and *Destructive metabolism* categories because it represents a lesser negative load over the ecosystem.

Annual information entropy results can be read as follows: equally-like occurrences of the normalized dataset in a certain category is reflected as an equal-probability overall indicators, thus, a maximum information entropy result. That is, for that year the system behaves in accordance with the optimum defined criterion for the category of indicators. On the contrary, a single appearance of the normalized dataset within an indicator gives zero information entropy. The system, in this case, does not behave as optimally defined for the group of indicators.

In the following paragraphs we read the case study's behavior as described by the information entropy measure and the metabolic categories.

2.3.1 Input supportive metabolism

One crucial metabolic category is the Input supportive category. Figure 2.2 shows the comparative results of *Input Supportive* category as described by the information entropy measure ($\Delta eS1$). Results demonstrate a generalized growing tendency, which is referred to as the evolution of the *Input Supportive* category towards the optimum state. In order words, the tendency demonstrates the growth of the load capacity of the system. There can also be seen as a turning point at 2008. At that moment, the information entropy tends to stabilize into its maximum value. We can associate that moment to the Spanish economic crisis when the country ceased its growth. Most of the indicators in this metabolic type are referred to the diversity in urban functionalities, equipment, and services, which are also related to the liveability of the urban ecosystem. In this sense, the results in Figure 2.2 shows stabilization in the *Input supportive* entropy and therefore, a homogeneous distribution of urban services among the different districts.

2.3.2 Output pressure metabolism

The second metabolism type is the Output pressure category. Figure 2.3 shows the comparative results of *Output pressure* metabolism as given by the information entropy ($\Delta eS2$). In this figure, we can observe a general decay of the *Output pressure* metabolism during the whole time span. The selected indicators in this category describe the system's consumption. We can observe that information entropy is departing from an optimal state, and as the optimal criterion for this type of indicators is defined to be a minimum value, it can be inferred that the pressure imposed by the economic and social subsystems is increasing during the study. That is, the system is facing a temporal evolution in the degenerating direction. The *Output pressure* results for Eixample and Les Corts districts remain in a constant plateau, with both districts behaving similarly. The districts of Ciutat Vella and Gràcia, on the other hand, exhibit

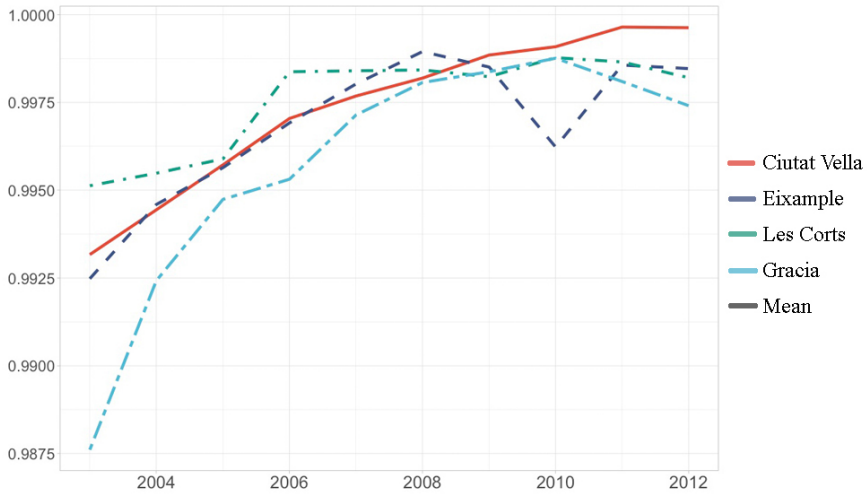

Figure 2.2: Input supportive metabolism as described by the information entropy measure (ΔeS1).

a similar downward trend: exposing a several decrease between the years 2007 and 2008, but recovering at the year 2011.

2.3.3 Destructive metabolism

The third metabolic category that we evaluate is the Destructive type. Figure 2.4 shows the comparative results of the *Destructive Metabolism* information entropy measure (ΔiS2). The results exhibit an overall growth tendency with some overshoots for certain districts. The indicators in this category describe the destruction of the ecosystem and its exposure to pollution. A close to one entropy result indicates the fulfilment of a minimum destructive trend. Therefore, we can consider that the ecosystem is, in general, tending to diminish its destructive metabolism during the time period. Isolated fluctuations are only described for the Gràcia district in the 2005 – 2009 period, being the district with the least favorable results.

2.3.4 Regenerative metabolism

In the case of the Regeneratuive metabolism, Figure 2.5 shows the comparative results for the information entropy measure. It can be observed a similar trend for all districts: the entropy measure tends to grow during the first years of the study, there is an inflection point between 2008 and 2009 when the ecological construction reaches its maximum level, and from these years on the entropy diminishes. On one side, Les Corts exhibits the best regenerative capacity between 2005 and 2009, but it also presents the most significant contraction after 2009. Ciutat Vella shows a reduced regenerative capacity during 2003 and 2008, being the lowest among the districts, but from 2009 this district improves its regenerative metabolism and reaches the best record in the case study. In this sense, Ciutat Vella presents the closest to the optimal behavior between 2003 and 2012 and exhibits a great resilience capacity in the renerative metabolism type.

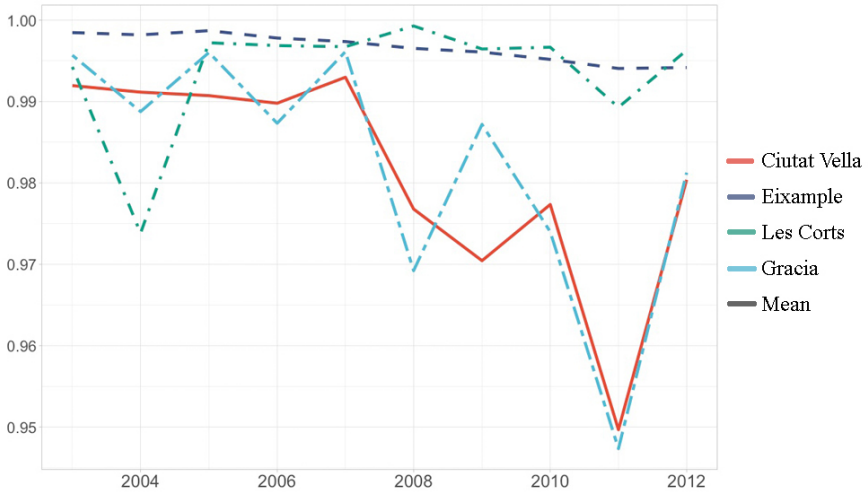

Figure 2.3: Output pressure metabolism as described by the information entropy measure (ΔeS2).

Figure 2.4: Destructive metabolism as described by the information entropy measure (ΔiS2).

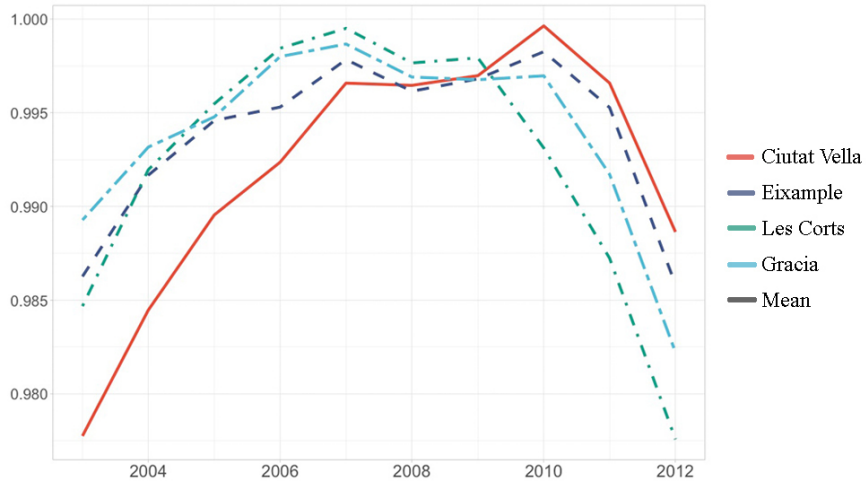

Figure 2.5: Regenerative metabolism as described by the information entropy measure (ΔiS1).

Figure 2.6: Appraisal Score (G).

2.3.5 Appraisal Score

Finally, we calculate the Appraisal score as another measure of the overall district's behavior. Figure 2.6 shows the comparative results of the Appraisal Score (G) for the districts in the case study. This indicator uses the calculated weights for each indicator; it is intended to estimate the evolution of the ecosystem towards the sustainable direction. Results show a growing tendency for all districts: the positive slope demonstrates that the system is evolving towards an optimum direction described by the selected indicators and reveals that during 2003 and 2012, the system (4 districts of Barcelona) has displayed its resilience capacity against disturbances.

2.4 Conclusions

In this chapter, we have demonstrated the application of an information entropy-based methodology that can be used to measure the urban ecosystem's metabolism. Our approach has been to use the information entropy as a measure of the state of an information system describing the city. As information entropy is a widely used concept, we can also apply in the multidimensional description of the urban system. The information entropy measure has been intended to report the state of the city in relation to four metabolic categories of an urban ecology framework that represent energy, water, matter, and information flows.

The proximity relationship of the observations to the ideal state of the system has been defined using a normalization process that correlates the measured and the optimal values. A probability of occurrence matrix has been constructed over the normalized quantities, to which the information entropy measure has been applied. The measured output of the information entropy is zero when exactly one of the probabilities of occurrence of the observations is one and the others are zero. The maximum output of the measure occurs when all events are equally likely to occur. In the case of the information system, this meant that the entropy value was maximum if the system fitted the expected optimal behavior. In the opposite situation, the information entropy gave minimum results if only one of the measurements had an occurrence and all others were zero. Some temporary readings of the behavioral change of the case study could be made based on the yearly calculation of the information entropy to the set of measured variables. The drawback is that the optimal behavior should be previously known —or at least understood.

We have applied the information entropy-based methodology in the case of study of four districts of Barcelona: Ciutat Vella; Eixample; Les Corts and Gràcia, between 2003 and 2012. First, we have constructed an information system of indicators for the case study: temporal measurements for several indicators in the metabolic categories have been collected from open data bases. We obtained the temporal description of the metabolism of these districts by applying the information entropy measure. Indeed, results for the *Input supportive* metabolism reveals a homogenized growing trend in the urban services among the districts. These results demonstrate optimal trends of the supportive category during the range of the study. Complimentary, entropy results for the *Output pressure* metabolism show that the system is also facing a temporal evolution in a degenerative direction. Nevertheless, the entropy for the *Destructive metabolism* indicates that the urban ecosystem, in general, is tending to a decrease in its destructive condition during the studied period. Regarding the *Regenerative metabolism*, there is a growing tendency of regeneration before 2008, from which the ecological construction decreases. The four districts show similar performance in that metabolism type, but Ciutat Vella exhibits the closest to the optimal behavior, and therefore, the greatest resilience capacity of the case study. Finally, we have computed an Appraisal Score that revealed an evolution of the districts towards the optimal direction during 2003 and 2012.

Hence, the methodology has been able to exhibit global and local events such as economic crisis, public policies, and local disturbances in the urban system. In particular, we have noticed that the environmental problems of urban ecosystems do not entirely rely on the degradation of nature, but that they are particularly linked with the lack or poor management of energy and mass flows, and with the city's model as a highly dissipative system. In this sense, descriptions made by urban metabolism may contribute to solve environmental problems: those quantifies

the flow of resources and the urban pressure exerted over the environment [24].

Chapter 3

An exploratory multivariate statistical analysis to assess urban diversity

Understanding diversity in complex urban systems is fundamental to face current and future sustainability challenges. In this chapter, we apply an exploratory multivariate statistical analysis (i.e., Principal Component Analysis (PCA) and Multiple Factor Analysis (MFA)) to an urban system's abstraction of the city's functioning. Specifically, we relate the environmental, economical, and social characters of the city in a multivariate system of indicators by collecting measurements of those variables at the district's scale. The statistical methods are applied to reduce the dimensionality of the multivariate dataset, such that, hidden relationships between the districts of the city are exposed. The methodology has been mainly designed to display diversity, being understood as differentiated attributes of the districts in their dimensionally-reduced description, and to measure it with Euclidean distances. Differentiated characters and distinctive functions of districts are identifiable in the exploratory analysis of a case study of Barcelona (Spain). The distances allow to identify clustered districts, as well as those that are separated apart exemplifying dissimilarity. Moreover, the temporal dependency of the dataset reveals information about the district's differentiation or homogenization trends between 2003 and 2015.

3.1 Introduction

It is expected that 66% of the world population will be urban by 2050 [1]. This on-going urbanization process entails major environmental and socio-economic impacts that compromise the function and stability of cities. As complex systems, cities generate interesting patterns and trends that are not easily described nor predicted [145]. Systems that produce complexity consist of diverse rule-following entities whose behaviors are interdependent [156, 157]: those entities interact over a contact structure or network, and they often adapt (i.e., by learning in a social system, or by natural selection in an ecological one) [47]. Due to their adaptive behavior and complexity, cities have been also identified as centers for economic development, productivity, creativity, innovation, and cultural transformation [2–4]. In this scenario, understanding cities' complexity will play a key role in facing current and future sustainability challenges at regional and global scales [5–7].

One of these challenges will be to improve the resilience of cities [15, 16, 35, 41, 51] and this is not straightforward since one action may not give the expected transformation of the complex system. Resilience refers to the ability of a system to adapt to a modifying process while retaining its functionality and not necessarily returning to the previous equilibrium state [34, 43, 48]. This definition comes from the non-equilibrium paradigm of ecology [43] and fits the complexity and uncertainty inherent to cities, as it includes adaptation as a key process that cannot be easily predicted. Specifically, urban resilience has been defined by Meerow (2016) as "the ability of an urban system-and all its constituent socio-ecological and socio-technical networks across temporal and spatial scales to maintain or rapidly return to desired functions in the face of a disturbance, to adapt to change, and to quickly transform systems that limit current or future adaptive capacity". This definition comprises the three pathways to a resilient state: persistence, transition, and transformation [158]. To some extent, measuring resilience is tightly related to inspecting the property of the system to remain functional over time —from persistence to adaptation—, and hence, to sustain itself. In any case, there are five questions that need to be address during the planning of effective resilience policy in a city: who? Resilience for whom, who benefits with that strategy; what? Which part of the city is going to be more resilient; when? Is the policy oriented to face short-term disruptions or long-term stress; where? Is the policy limited to a spatial scale, is portraying cross-scalar interactions, why? Which is the goal of the program [50].

We are interested in the relationship between diversity and resilience in the urban system, and in the effects of diversity fostering resilience. Basically, in the development of a statistical framework to evaluate urban systems, which may be helpful in the formulation of urban policy and in answering these questions. When we speak about diversity, we refer to any of the three characteristics of a population [47]: (1) diversity within a type, or variation in some attribute (such as differences in the length of finches' beaks), (2) diversity of types and kinds, or species in biological systems (such as different types of stores in a mall), and (3) diversity of composition, which refers to differences in configuration (such as different connections between atoms in a molecule). The diversity concept has characterized populations or collections of entities like ecosystems, with multiple types of flora and fauna, but it is also suitable to be applied in cities, having different types of people, organizations, infrastructural systems, etc. Regardless that the influence of diversity in the stability of a system is still a controversial issue [47, 58, 59], a number of studies in socio-ecological systems have pointed out a positive correlation between diversity and resilience [32, 41, 55–57]. From an urban ecology perspective, it is argued that cities with higher levels of environmental, economic, and social diversity have a higher resilience capacity [15, 32, 35]. This argument relies on the assumption that a highly diverse system possesses many different entity types, and therefore, a bunch of individuals belonging to these different types that are able to perform similar functions at a wider range of conditions —response diversity— [15, 55]. Furthermore, it has been shown that diversity is a fundamental property that ensures the city's functionality in the face of disturbances; it's an essential factor for economic growth, attractiveness and liveability of cities [2]. In the last decade, some scholars have supported this thesis by arguing that diversity fosters productivity, innovation, and therefore, economic development in cities [60, 61, 159].

From a formal perspective, the assessment and characterization of a system's diversity have been increasingly performed by means of a multiplicity of variables instead of aggregated attributes. This has been motivated by the enormous increase of computer power and

accessibility to open datasets in recent years. However, direct interpretation of multivariate information is not easy and some level of aggregation is required to display an integrated picture of the system: aggregated measures surely reduce the amount of information, highlight patterns, and enhance the communication of results. Nevertheless, gathering all the information into one single diversity index is usually not convenient for the following reasons. Firstly, most of the methods for aggregating indicators into indices produce a disconnection between the more intuitively original variables and the quantities resulting from these indices [82]. Secondly, aggregation also requires some form of human judgment, so that, it relies on potentially distorting assumptions [82, 83]. Lastly, aggregate indices (e.g.,variance, information entropy, or phylogenetic distances [47]) barely capture interactions within complex systems [83].

Here we explore diversity in urban complex systems through multivariate statistical methods (i.e., Principal Component Analysis (PCA) and Multiple Factor Analysis (MFA)). These techniques are robust methodology in the sense that they provide a moderate trade-off between the large volume of initial data and its final aggregation. In this regard, PCA [105, 107] is probably the most widely used multivariate statistical technique. It has been applied in urban geography since the 1970s to analyze the landscape change [160–162], but it has lately emerged as a meaningful statistical tool to understand the geographical complexity of globalization processes [163], sustainable development [83, 164–166], and socio-economic resilience trends [167]. At the urban scale, it has been used to identify urban typologies —of neighborhoods or cities— in metropolitan areas. By identifying typologies, these studies have exhibited —to some extent— the degree of diversity of the urban systems, being described by several socio-economic variables (i.e., race, ethnicity, age, family structure, education, employment, income...) [102–104]. MFA [168, 169], on the other hand, is a multivariate data analysis technique –an extension of PCA- for summarizing and visualizing complex data, where variables are organized into groups of variables: it reveals an integrated picture displaying the relationship between observations and groups of variables [107, 108, 168]. MFA has been applied mostly in sensory analysis; see for example the recent work of [170–173]. In the urban context, few studies have applied MFA [174, 175].

The way to measure diversity with these methods is explained by their capacity to reduce the dimensionality of the multivariate dataset by transforming the original data of correlated variables into a set of a few representative variables —known as principal components— that extract the most relevant information [105, 106]. The objective is to visualize the relations between the system's components, such that, the more differentiated entities, the more diversity indication. Since distanced results in the dimensionally-reduced description indicate diverse attributes of districts as described by the multivariate system of indicators, we may seek for exposing those differentiated results: one basic visualization method is the plot of variables and individuals in relation with the principal components (i.e., biplot). Another, is the quantitative measurement of such differences with Euclidean distances.

In this chapter, we apply a descriptive multivariate statistical analysis (i.e., PCA and MFA) to an urban system's abstraction of the city's functioning. Specifically, we relate the environmental, economical, and social characters of the city of Barcelona (Spain) in a multivariate system of indicators by collecting measurements of those variables at the district's scale. We analyze diversity qualitatively and quantitatively by means of the dimensionality reduction of the dataset and the recognition of interactions and patterns among districts. The main purpose of this work is to exhibit the diversity of the urban system since —at a broader sense— we

are interested in its relationship with resilience. In this sense, the temporal span of the dataset —from the year 2003 to the year 2015— not only brings a wider and reliable perspective to the city's description, but it also allows for a temporal analysis: it constitutes the main tool to understand diversity trends and to identify key points towards the achievement of resilience. We also apply MFA to expose relations between urban metabolism groups of variables and the principal components.

This chapter is organized as follows. In Section 3.2, we present our case study, the multivariate dataset and the statistical techniques that we apply to expose diversity. Results of the multivariate statistical methods are presented in Section 3.3. Discussion is given next in Section 3.4. The chapter ends with some conclusions and future research inquiries in Section 3.5.

3.2 Materials and methods

In this section, we present the statistical methodology for exposing diversity in the urban system. For doing so, the case study is described in the first part of the section. Then, we introduce the main hypothesis and transform the city into an abstraction of multiple variables that allows us to quantify diversity. Finally, the implemented statistical techniques are presented in the last part of this section.

3.2.1 Case study

Barcelona city —our case study— can be studied at its largest scale, or as a composition of smaller divisions. The choice of the scale influences the overall interpretation of diversity: cities may seem diverse on one scale, but when looking at another scale it may appear as homogeneous. At its largest scale, Barcelona has a population of approximately 1.6 million inhabitants living in 102.16 km^2 (15865 inhabitants per km^2) [153]. We choose to represent the city at the district's scale with the aim of revealing particular functions and patterns of these urban subsystems while covering efficiently most of the city's territory.

Since 1987, Barcelona has been divided into 10 administrative districts which can be compared with neighborhoods in a common metropolitan area. Our selection of districts —and therefore, of the city's representation— ends up with Ciutat Vella, Eixample, Les Corts, and Gràcia as our case study. These districts are chosen to characterize Barcelona since the surface covered by the case study comprises a total area of 22.06 km^2 and represents the 22% of the total city's surface, as shown in Figure 2.1. With respect to population, the case study comprises 571183 inhabitants accounting for 35% of the city's population.

We also consider the following criteria in our selection: urban form, socio-economic characteristics of the population, age of urban development —year or period of establishment—, proximity to the city center —distance to historic center—, and type of urban planning — organic or planned.

The urban form refers to the physical layout of the city: it is set by the distribution of urban elements, such as streets, blocks, plots, and buildings. There are different types or urban forms: grid or orthogonal, linear, radial, and irregular. Representative samples of the urban form for each one of the districts are shown in Figure 3.1, where plans are drawn at the same scale, blocks are shown in black, streets in white and pedestrian streets in light brown. Ciutat Vella

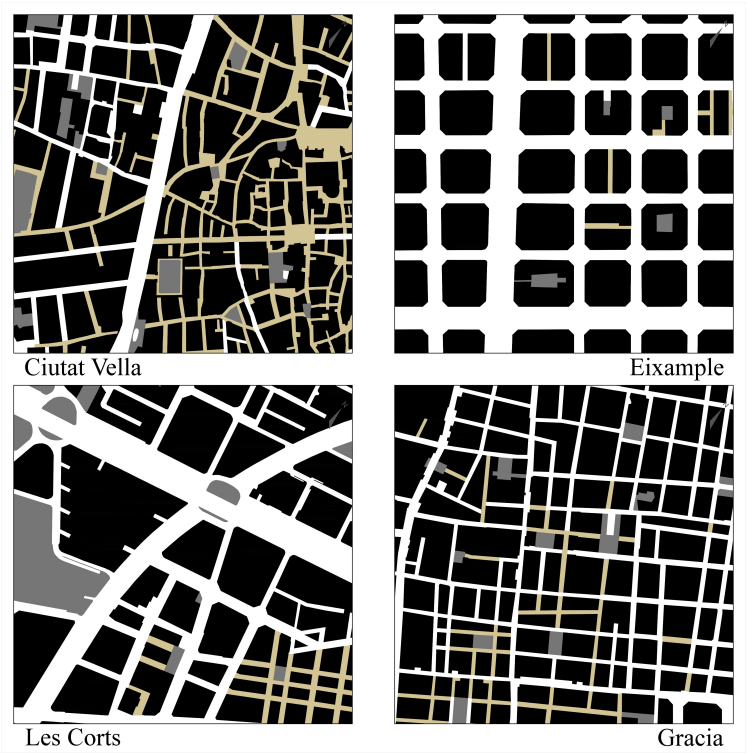

Figure 3.1: Urban form of Ciutat Vella, Eixample, Les Corts and Gràcia.

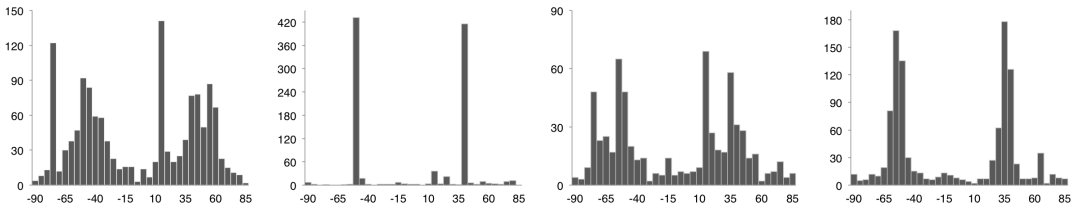

Figure 3.2: Frequency histograms of the case study streets' orientation. From left to right: Ciutat Vella, Eixample, Les Corts and Gràcia.

Table 3.1: Urban structure's metrics using Shannon's entropy.

Metric	Ciutat Vella	Eixample	Les Corts	Gràcia
Streets' lenght	2.963	2.910	3.746	3.345
Streets' orientation	4.648	2.156	4.615	4.046
Buildings height	2.905	3.137	3.398	3.207
Average entropy	3.506	2.734	3.920	3.533

exhibits an organic and irregular structure —distinctive of Middle Age—, Eixample a planned grid structure, Les Corts recalls a radial plan, and Gràcia is based on an orthogonal structure of different size plots. Hence, a fine-grain urban fabric characterizes Ciutat Vella and Gràcia, and a coarse-grain urban fabric distinguishes Eixample and Les Corts. Based on previous measurements of the urban street networks [146–150, 176], we compute Shannon's entropy of the urban structure to provide quantitative information that is useful in the diversity analysis. In this methodology, the street orientation, length of the built tract of the street, and height of the adjacent buildings are measured in the four districts' area. Then, the values for each variable and district are arranged —by absolute frequencies— in fixed bins $j = 1, \ldots, m$: histograms illustrate this distribution (e.g., histograms for streets' orientation) in Figure 3.2. Once the measures are binned, Shannon's entropy S_i is computed for each $i-$th distribution (e.g. the streets' orientation in a district) following $S_i = -\kappa \sum_{j=1}^{m} p_{ij} \ln\left(p_{ij}\right)$ for $i = 1, 2, \ldots, n$. Where κ represents a positive constant, and p_{ij} is the probability of the variable falling in the $j-$th bin. When the probability is zero, $p_{ij} = 0$, the supposition $p_{ij} \ln\left(p_{ij}\right) = 0$ is implemented. Entropy values for each one of the $i-$th variables and districts are shown in Table 3.1. This information entropy metric reveals the degree of order of the urban structure: the low entropy exhibited by Eixample corresponds to its grid structure with a high degree of order. Les Corts and Ciutat Vella on the other hand, present a high entropy that displays their heterogeneous urban structure, with neighborhoods having a dissimilar urban layout, not only in formal aspects, but also in the size of their entities (streets, plots, blocks, and empty urban spaces, etc.).

Regarding the socio-economic component, we consider characteristics such as employment rate, income, education, population age, and percentage of foreigners (see Table 3.2 for relevant demographic data of the case study, including percentages for each district with respect to the city). The district of Ciutat Vella is the oldest one, constituting itself the historical city center. The district of Eixample was designed by Ildefons Cerdà in 1859 as part of the widely known urbanism plan called "Plan Cerdà" [177–179]. Eixample district is also the densest district of Barcelona with a population density of 35617 inhabitants per km^2. In the case of the districts of Les Corts and Gràcia, those were former independent municipalities annexed to the city in the last decade of the 19th-century after a major conurbation process. As commented before, we select these districts to be representative and to constitute the urban essence of Barcelona.

3.2.2 Multivariate description of the case study

The urban system description of the case study is achieved through a multivariate dataset that incorporates measurements of environmental and socio-economic variables from the year 2003 to 2015. This dataset is a matrix containing 40 indicators (variables) and 52 individ-

Table 3.2: Case study demographic description.

City/District	Population		Surface	Pop. density	Foreigners		
			km^2	(inhab/km^2)	Population	of district	of city
Ciutat Vella	101387	6%	4.37	23201	45357	45%	16%
Eixample	266416	16%	7.48	35617	53393	20%	19%
Les Corts	82033	5%	6.02	13627	9796	12%	3%
Gràcia	121347	7%	4.19	28961	19574	16%	7%
Barcelona	1620809	100%	102.16	15865	284907	18%	100%

Data for 2017 from the Index Statistical yearbooks of Barcelona [153].

uals: each individual represents a district in a given year (e.g. "CiutatVella 2003", "Ciutat Vella 2004", "Eixample 2003",...). To construct the multivariate description of the case study we considered the availability of time-dependent disaggregated data in the case of the four districts. Our dataset collects the information in the open databases of Departament d' Estadistica (2017) [153] and Agencia de Salut Publica de Barcelona (2016) [155] and it is shown in Table 3.3 for the district of Ciutat Vella between 2003 and 2015. The compiled datasets for all the districts (Ciutat Vella, Eixample, Les Corts, and Gràcia) are available online at `https://summlabbd.upc.edu/SalazarLlano/PhDThesis/`. This particular time-span —from 2003 to 2015— has been chosen regarding the data availability at the district's scale: before 2003 only few indicators were reported at this scale.

Most of the variables have been measured on a *per capita* basis in order to neutralize the effect of districts' population size. This fact allows us to make a comparison focused on services and consumptions for each inhabitant of a given district. However, the discussion on analyzing urban systems based on aggregated or disaggregated data, normalized either by surface or population, remains an open debate. Particularly, we consider raw values for general or disaggregated indicators: Security level in the neighborhood (A12), Natural growth rate (B1), Population density (B2), Available household income per capita (B8), Average levels of PM_{10} (C1), Average levels of NO_2 (C2), and Average values of CO (C3).

Within the framework of urban ecology, we use the concept of urban metabolism in order to describe and classify the interactions between the different components that take place in the city. One of the main ideas of this methodology is to integrate social, environmental and economic variables into categories that characterize the ecological structure, diversity of amenities, population pressures, and consumption of the urban system: it is intended to portray the flows of matter, information, and energy in an ecosystem. We arrange our system of variables based on some previous indexes developed within the frameworks of urban metabolism and urban sustainability [22, 24, 76, 79], but we include some additional variables that characterize social interactions, diversity of land uses, and liveability. Following Zhang (2006) [24], we group the variables (presented in Table 3.3) into four metabolic categories:

- Input Supportive. This category comprises 18 indicators and portrays the natural capital of the city, but also the liveability and the complexity of urban services. Some indicators in this category describe the diversity of economic activities and institutions associated

with each district. We refer to the land use areas and buildings: housing, commerce, offices, education, health services, sport, religion, and entertainment, among others.

- Output Pressure. It includes 12 indicators that describe the pressure exerted by the economic and social subsystems over the natural subsystem. This category is mostly related to the population size (social subsystem) and its consumption demand.
- Destructive Metabolism. It includes 5 indicators that reveal the environmental pollution within the urban system, particularly described by air pollution and waste generation indicators. We also include the number of traffic accidents in order to measure one particular aspect of the non-liveability character of an urban system.
- Regenerative Metabolism. It comprises 5 indicators related to ecological construction upon regenerative infrastructure and environmental care.

3.2.3 Multivariate statistical analysis of the dataset

Since there is a large amount of information comprised in the multivariate dataset its examination is complicated, thus, statistical methods are implemented to ease the analysis of the urban system. In the first place, we display in Figure 3.3 the correlation matrix among the 40 variables (that has been produced in R [180] using the corrplot package [181]), which is colored according to the correlation coefficient. We realize that the interactions that take place in the city between districts and indicators are far too complex, and therefore, the dimensionality reduction given by the PCA is justified. Specifically, we rely on this mathematical procedure to transform the original $x-$dimensional dataset space into a reduced $y-$dimensional component space —being the cardinality of y lesser or equal than x— by solving the eigenvalues problem [106]. For the sake of conciseness, we omit the mathematical formulation of the eigenvalues problem of the correlation matrix (see [105] for a thorough explanation on this topic). Here we are mainly interested in the transformation given by the solution of the eigenvalue problem.

The PCA transformation leads to the description of the original dataset by means of a set of linearly independent variables, which are the so-called principal component [105, 182]. The principal components are uncorrelated and ordered so that the first component describes most of the variance of the dataset, and each successive component describes the largest possible variance after the previous components [105]. Therefore, if only the subset of the first y principal components is selected, $y << x$, the dimensionality can be reduced whilst a great percentage of the variance of the original correlation matrix is maintained. The main advantage of PCA relays in its ability to reduce the dimensionality of the multivariate system while extracting its main characteristics in the dimensionally-reduced description. In the case study, we reduce the 40-dimensions description to only 3-dimensions per individual. The PCA calculation is conducted in R Software [180] using the FactoMineR package [183]; the results are displayed using the ggplot2 [184] and factoextra [185] packages.

Since the variables are measured in different units, and the measurements for each variable can vary up to two orders of magnitude, we perform a scaling of the raw data matrix prior to the application of the PCA. The data preparation is intended to make comparable data of different units and scales, such that the standardized data is scaled to have standard deviation one and mean zero.

Table 3.3: Dataset for the district of Ciutat Vella between 2003 and 2015.

Type	Id	Indicators	Unit	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
INPUT Supportive	A1	Residential area	m^2/person	37.0	36.7	35.5	34.9	35.5	35.4	36.9	37.9	38.2	38.0	38.4	39.3	39.5
	A2	Commercial area	m^2/person	5.6	5.6	5.4	5.4	5.5	5.6	5.9	6.1	6.2	6.2	6.3	6.5	6.5
	A3	Office buildings	m^2/person	6.6	6.6	6.4	6.3	6.4	6.4	6.6	7.1	7.0	6.9	7.2	7.5	7.4
	A4	Educational buildings	m^2/person	3.0	3.0	2.9	2.9	3.0	3.0	3.2	3.3	3.4	3.5	3.5	3.6	3.7
	A5	Health service buildings	m^2/person	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.7
	A6	Sport area	m^2/person	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4
	A7	Religious buildings	m^2/person	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	A8	Entertainment buildings	m^2/person	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.8	0.8	0.8
	A9	Other land uses	m^2/person	2.5	2.5	2.5	2.4	2.5	2.5	2.6	2.8	2.9	2.7	2.7	2.8	2.8
	A10	Cultural centres	unit/person ($\times 10^{-4}$)	2.6	2.0	2.2	2.2	2.1	2.1	2.4	2.8	2.7	2.6	2.6	2.8	2.8
	A11	Sport centres	unit/person ($\times 10^{-3}$)	1.1	1.1	1.0	1.1	1.1	1.1	1.4	1.5	1.4	1.4	1.5	1.7	1.7
	A12	Security level in the neighbourhood	Points	4.8	4.7	4.9	5.0	5.3	5.2	4.9	4.9	4.8	5.2	5.2	5.4	5.2
	A13	Public water fountains	unit/person ($\times 10^{-4}$)	9.9	9.6	9.2	9.1	9.0	9.1	9.4	9.6	9.8	9.6	9.5	9.4	10.1
	A14	Playgrounds	unit/person ($\times 10^{-4}$)	2.1	2.0	1.8	1.9	2.2	2.4	2.4	2.8	2.7	2.8	3.1	3.2	3.2
	A15	Active market stalls	unit/person ($\times 10^{-3}$)	9.8	9.5	8.7	8.6	8.6	8.5	8.8	9.1	9.2	9.1	9.2	9.4	9.4
	A16	Street trees (Unit = 5.6 m^2)	unit/person ($\times 10^{-2}$)	6.4	6.3	5.8	5.8	6.0	5.8	6.5	6.5	6.7	6.8	6.7	6.9	6.9
	A17	Urban parks	ha/person ($\times 10^{-4}$)	3.5	3.5	3.3	3.3	3.3	3.3	3.5	3.6	3.6	3.6	3.6	3.7	3.7
	A18	Urban green area besides parks	ha/person ($\times 10^{-4}$)	2.4	2.4	2.3	2.2	2.3	2.3	2.4	2.5	2.5	2.5	2.5	2.5	2.5
OUTPUT Pressure	B1	Natural growth rate	-	-4.7	-1.5	-1.2	-1.1	-1.4	0.1	0.4	0	0.9	-0.3	-0.5	-0.5	-0.3
	B2	Population density	people/ha	244	246	254	258	255	255	245	239	236	239	237	231	229
	B3	Total number of motor vehicles	unit/person	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.4	0.4	0.4
	B4	Total volume of water supplied to the commerce	m^3/person	36.3	35.6	34.0	32.7	34.2	32.7	32.7	33.5	33.8	33.1	33.5	9.6	9.7
	B5	Total volume of water supplied to domestic consumption	m^3/person	41.7	40.8	39.0	37.7	38.1	37.1	38.1	38.7	38.9	37.9	37.1	37.6	38.0
	B6	Total volume of water supplied to the industry	m^3/person	5.3	5.0	4.6	4.4	4.3	3.6	3.4	3.3	4.2	3.0	2.7	26.6	27.5
	B7	Total volume of water supplied to other usages	m^3/person	10.6	10.3	9.7	10.8	10.0	7.0	8.4	8.6	9.1	9.5	9.3	9.8	12.6
	B8	Available household income per capita	Index	71.0	70.8	72.0	72.6	73.5	71.2	74.2	75.2	76.4	76.6	77.2	79.7	85.5
	B9	Accommodation places	unit/person ($\times 10^{-1}$)	1.1	1.2	1.2	1.2	1.2	1.3	1.6	1.8	1.9	1.9	1.9	2.0	2.1
	B10	Final energy consumption	MWh/year.person	2.9	3.1	3.2	2.9	2.8	2.9	3.1	3.3	2.9	3.1	2.9	3.2	3.2
	B11	Parking area	m^2/person	3.2	3.3	3.3	3.4	3.6	3.7	3.9	4.4	4.5	4.5	4.6	4.8	4.8
	B12	Industrial area	m^2/person	5.1	5.0	4.7	4.5	4.5	4.4	4.5	4.5	4.5	4.3	4.3	4.4	4.4
DEST-RUCTIVE	C1	Average levels of PM_{10}	$\mu g/m^3$	51	46	47	52	43	43	33	29	31.6	31	26	27	31
	C2	Average levels of NO_2	$\mu g/m^3$	57	43	48	47	46	42	46	46	40	42	35	37	42
	C3	Average levels of CO	mg/m^3	0.5	0.5	0.5	0.3	0.4	0.3	0.4	0.4	0.3	0.3	0.3	0.3	0.3
	C4	Collected volume of non-recyclable waste	ton/person ($\times 10^{-1}$)	2.4	2.2	2.1	2.1	2.2	2.0	2.0	2.2	2.2	1.9	1.9	2.0	2.0
	C5	Traffic accidents	unit/person ($\times 10^{-3}$)	7.8	7.0	6.8	6.2	6.9	6.0	5.9	5.3	5.5	5.5	5.5	5.3	5.9
REGENE-RATIVE	D1	Collected volume of paper and cardboard	ton/person ($\times 10^{-2}$)	2.2	2.5	2.9	3.0	3.4	3.5	3.2	3.9	2.9	2.4	2.1	2.3	2.3
	D2	Collected volume of glass	ton/person ($\times 10^{-3}$)	8.0	8.4	8.8	9.1	10.8	11.4	12.1	15.1	14.9	14.1	14.2	15.1	15.0
	D3	Collected volume of containers	ton/person ($\times 10^{-3}$)	4.2	4.5	4.7	4.9	6.1	6.6	7.0	9.4	9.0	8.3	8.0	8.4	8.4
	D4	Maintenance cost	Miles of €/person ($\times 10^{-3}$)	6.9	6.8	5.6	6.9	7.2	7.4	7.8	8.3	8.4	8.1	8.2	8.4	8.4
	D5	Streets and zones with pedestrian priority	ha/person ($\times 10^{-4}$)	2.0	2.1	2.0	1.9	2.0	1.9	2.1	2.2	2.3	2.3	2.3	2.7	2.8

In order to visualize the outcome of the analysis, we display the relationship among the variables (indicators) and individuals (districts) using covariance biplots [105, 109, 186]. In biplots, we make profit of the new low-dimensionality of the system by plotting individuals and variables against the three selected principal components, which are placed at the axes of the plots [109]. Using biplots and loadings, we identify correlations (direct, inverse or null) between variables and principal components, such that a conceptual meaning is given to each principal component in a process that we call "labelling". Individuals can be analyzed in their relative positions to the labeled principal components. This will be further exemplified in the next sections.

Also, distance measures can be used to quantify the differences —dissimilarities— between individuals. We perform some statistics of the districts' performance, where districts' positions are accounted as the events: each district's location is accounted as a single observation, such that, the average value of the districts' positions —or centroids— can be used as a representative location. We propose to calculate the component-wise and Euclidean distances between districts' centroids, respectively, in order to quantify diversity. Thus, the next step of analyzing diversity fostering resilience can be done based on quantitative results.

Furthermore, we implement MFA [168] for summarizing and visualizing urban metabolism categories and their relation with the principal components' description of districts. Particularly, MFA is an extension of PCA that takes advantage of the categorical groups of variables by weighting —or balancing— the contribution of the variables in each group [108]. Conclusions on the metabolism of the city can be addressed by understanding the contribution of variables in each urban metabolism category to the principal component's description of districts, and hence, of the city.

3.3 Results

In this section we present the results of applying the multivariate statistical methodology to our case study. Results are presented in the following order: firstly, the dimensionality reduction given by the PCA application, the description of each district as given by the PCA and the variables' loadings for each principal component. Secondly, we explain the labeling of the selected principal components. Lastly, the MFA results of the urban metabolism categories.

3.3.1 Principal Component Analysis

As described in the methodology of this chapter, we use R Software [180] and FactoMineR package [183] to perform the standardized PCA. Since FactoMineR uses a singular value decomposition algorithm, the PCA is calculated over the standardized correlation matrix, where a matrix of 40 uncorrelated components is obtained. Table 3.4 shows the percentage of variance and the eigenvalues for the first 10 components of this matrix. The remaining (30) components correspond to a residual amount of variance. By selecting only the first three principal components we reduce the dimensionality of the multivariate description, so that, the graphical representation and its subsequent interpretation are simplified. The first three principal components describe 87.7% of the total variance: the first component describes 42.7% of the variance, the second one describes 29.3% of the variance, and the third component describes

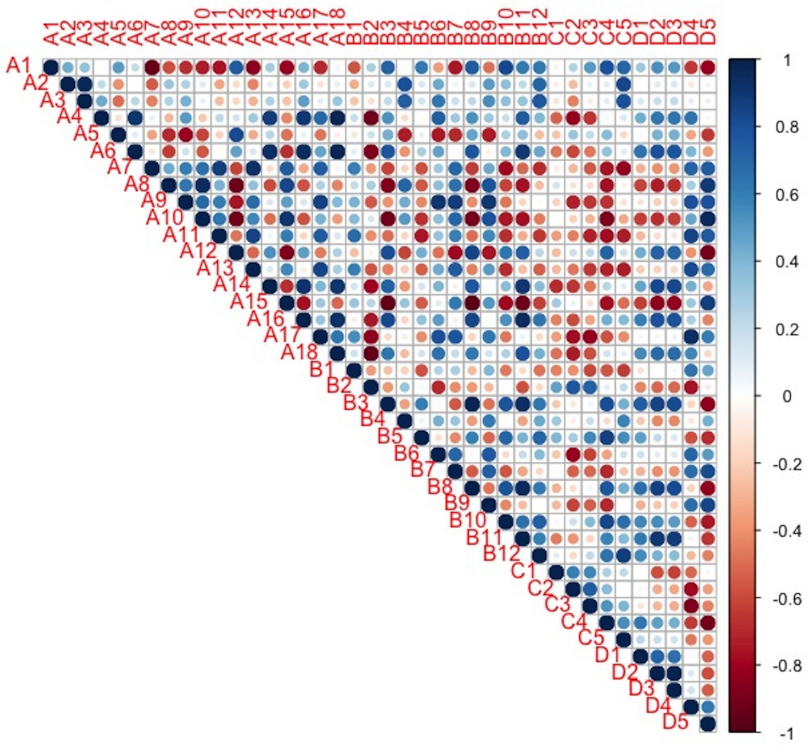

Figure 3.3: Correlogram of the multivariate description of the case study.

15.8% of the variance. In the case of the goodness of fit, we rely on the following metrics to verify the choice of the three first components: the Root Mean Square of the Residuals (RMSR) is 0.05 and the fit based upon off-diagonal values is 0.99. We label these three principal components as: "Social Background" (PC1), "Ecological Background" (PC2), and "Urban morphology and architectural typology" (PC3) according to the urban description that they make (to be explained next).

Correlations between indicators and principal components are described by the components' loadings. Table 3.5 shows the (positive, negative and null) loads for each one of the three principal components. A high absolute load indicates an important contribution of that variable to the percentage of variation covered by that particular component. Its sign implies a direct (+) or an inverse (-) correlation.

Figures 3.4 to 3.6 show two-dimensional biplots, each one showing the combination of two of the three principal components. Points (individuals) represent the districts in each year of the time period under analysis. Individuals are shown with a specific color and shape for each district. The temporal evolution of districts is depicted with dashed arrows connecting the different coordinates, from the year 2003 to the year 2015.

We also calculate the district's displacement from the initial to the final year of the time span. This calculation is given in Table 3.6 for each principal component, and it is useful to describe the district's variation related to the principal component. Averaged positions are presented in Table 3.7 with the notation for each district. Finally, in Tables 3.8 and 3.9 we present the component-wise and Euclidean distances between districts' centroids, respectively.

Table 3.4: Decomposition of variance per component of the PCA.

Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
1	17.07	42.69	42.69
2	11.70	29.25	71.94
3	6.31	15.78	87.73
4	2.14	5.36	93.10
5	0.82	2.05	95.15
6	0.48	1.20	96.36
7	0.35	0.87	97.24
8	0.27	0.68	97.92
9	0.19	0.48	98.40
10	0.14	0.37	98.78

3.3.2 Labeling the Principal Components

Vectors in biplots of Figures 3.4 to 3.6, portray the correlations of the different variables with the principal components: the vector length represents the influence —or loading— of the variable into that component. In order to give a conceptual label to each principal component, we select some of the variables that have large loadings in the component. Then, we identify the type of correlation between the highlighted variables (direct or inverse), but also classify the variables showing no relation with the component. A conceptual meaning is given to each component by recognizing the type of variables that characterizes it. Each labeling of the principal components is detailed next.

3.3.2.1 First Principal Component PC1 – "Social background"

As presented in Table 3.4, the First Principal Component (PC1) explains 42.7% of the total variance. It can be observed in Figures 3.4 and 3.6, and in Table 3.5, that the variables with higher loadings on this component are, among others: A1, A7, A8, A10, A11, A12, A14, A15, A16, B3, B7, B8, B11, B12, C4, D1, D2, D3, D5. The positively correlated variables are related to available income, the number of vehicles, and housing. Instead, the negatively correlated variables are related to urban amenities such as market stalls, pedestrian zones, and cultural, religious, and sports centers. This component is, therefore, displaying the diversity of land uses and confronts two different types of neighborhoods: on the one side, housing and predominant car-usage zones, and on the other, diverse and pedestrian zones. It also reveals the social interactions that take place in the city, which are directly related to diversity, and thus, with economic development, productivity, and innovation [2, 47]. We also observe that variables that are not represented by this component are: population density, commercial areas, offices, and average levels of PM_{10} and NO_2. Hence, we label this component as "Social background".

3.3.2.2 Second Principal Component (PC2) – "Ecological background"

In the case of PC2, it accounts for 29.3% of the total variance. As shown in Figures 3.4 and 3.5, and in Table 3.5, variables with noteworthy loadings on this component are related to air

Table 3.5: Component loadings (correlations between variables and components) for the three principal components. Values above the 0.4 threshold are highlighted in boldface font.

Type		Id	Indicators	PC1	PC2	PC3
INPUT	Supportive	A1	Residential area	**0.88**	-0.23	0.23
		A2	Commercial area	0.16	0.11	**0.91**
		A3	Office buildings	0.17	0.33	**0.90**
		A4	Educational buildings	0.24	**0.95**	0.10
		A5	Health service buildings	**0.54**	-0.23	**-0.66**
		A6	Sport area	**0.53**	**0.82**	-0.10
		A7	Religious buildings	**-0.82**	0.31	-0.39
		A8	Entertainment buildings	**-0.86**	-0.12	**0.46**
		A9	Other land uses	**-0.65**	**0.65**	0.38
		A10	Cultural centres	**-0.95**	-0.01	0.23
		A11	Sport centres	**-0.72**	**0.46**	-0.37
		A12	Security level in the neighbourhood	**0.89**	-0.01	-0.40
		A13	Public water fountains	**-0.69**	**0.52**	-0.35
		A14	Playgrounds	**0.51**	**0.81**	-0.13
		A15	Active market stalls	**-0.96**	-0.22	0.00
		A16	Street trees (Unit = 5.6 m^2)	**0.62**	**0.75**	0.15
		A17	Urban parks	**-0.50**	**0.83**	-0.07
		A18	Urban green area besides parks	0.34	**0.92**	0.00
OUTPUT	Pressure	B1	Natural growth rate	**-0.47**	**0.42**	-0.38
		B2	Population density	-0.05	**-0.95**	0.15
		B3	Total number of motor vehicles	**0.93**	0.33	-0.04
		B4	Total volume of water supplied to the commerce	-0.29	-0.16	**0.92**
		B5	Total volume of water supplied to domestic consumption	**0.70**	-0.19	0.26
		B6	Total volume of water supplied to the industry	-0.28	**0.78**	**0.51**
		B7	Total volume of water supplied to other usages	**-0.74**	**0.44**	0.24
		B8	Available household income per capita	**0.93**	0.33	-0.05
		B9	Accommodation places	**-0.66**	**0.47**	**0.50**
		B10	Final energy consumption	**0.86**	-0.04	0.27
		B11	Parking area	**0.81**	**0.57**	0.00
		B12	Industrial area	**0.63**	0.20	**0.72**
DESTRUC-	TIVE	C1	Average levels of PM_{10}	-0.08	**-0.62**	0.25
		C2	Average levels of NO_2	0.16	**-0.88**	-0.17
		C3	Average levels of CO	0.34	**-0.81**	0.19
		C4	Collected volume of non-recyclable waste	**0.91**	-0.21	0.08
		C5	Traffic accidents	**0.54**	-0.11	**0.81**
REGENE-	RATIVE	D1	Collected volume of paper and cardboard	**0.62**	0.35	0.00
		D2	Collected volume of glass	**0.73**	**0.55**	-0.17
		D3	Collected volume of containers	**0.68**	**0.56**	-0.18
		D4	Maintenance cost	**-0.51**	**0.82**	-0.14
		D5	Streets and zones with pedestrian priority	**-0.95**	0.19	0.18

Table 3.6: Displacements of districts in the principal components.

District	PC1	PC2	PC3
Ciutat Vella	-0.526	3.263	-0.933
Eixample	-0.417	1.985	-1.752
Gràcia	-0.401	2.547	-2.266
Les Corts	0.233	2.492	-2.627

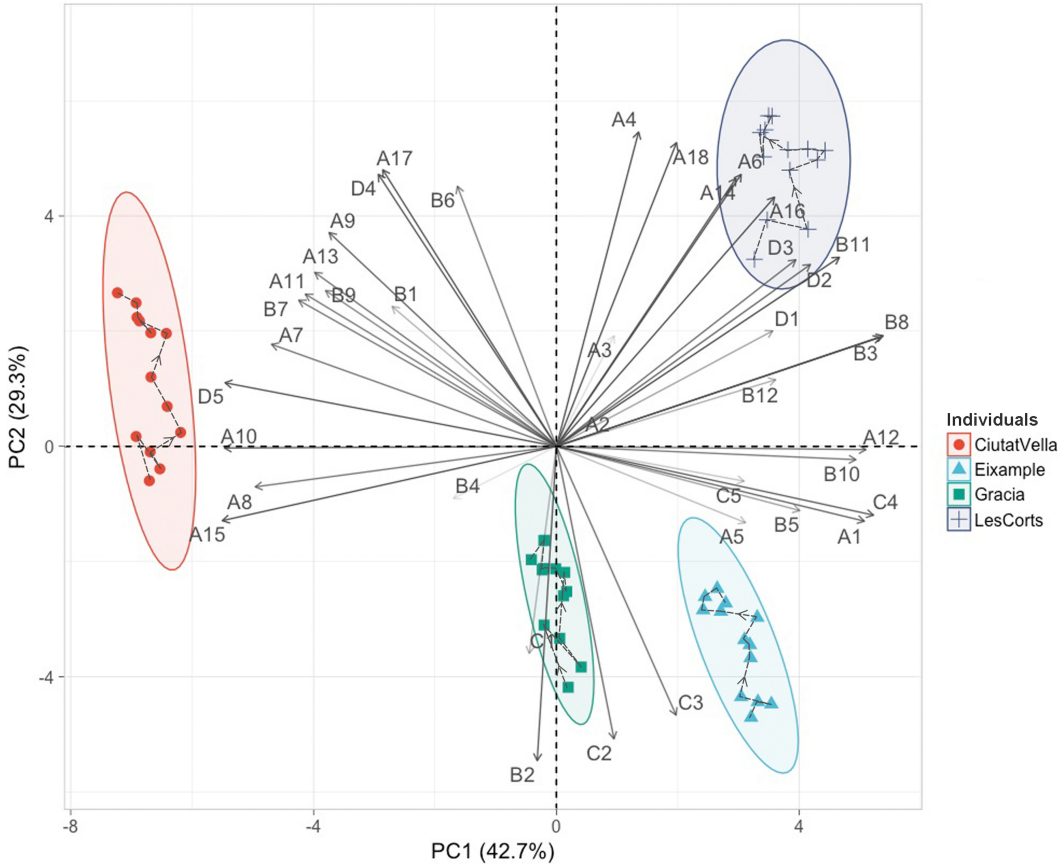

Figure 3.4: Biplot of individuals and variables for first and second principal components.

Table 3.7: District's centroids.

District	Notation	PC1	PC2	PC3
Ciutat Vella	$\overline{C}$	-6.705	1.133	1.227
Eixample	$\overline{E}$	2.993	-3.457	2.632
Gràcia	$\overline{G}$	-0.03	-2.569	-3.828
Les Corts	$\overline{L}$	3.743	4.893	-0.031

Table 3.8: Component-wise distances between districts' centroids.

PC1					PC2					PC3				
	$\overline{C}$	$\overline{E}$	$\overline{G}$	$\overline{L}$		$\overline{C}$	$\overline{E}$	$\overline{G}$	$\overline{L}$		$\overline{C}$	$\overline{E}$	$\overline{G}$	$\overline{L}$
$\overline{C}$	0	9.698	6.675	10.448	$\overline{C}$	0	4.59	3.702	3.761	$\overline{C}$	0	1.405	5.055	1.257
$\overline{E}$	9.698	0	3.023	0.75	$\overline{E}$	4.59	0	0.888	8.351	$\overline{E}$	1.405	0	6.461	2.663
$\overline{G}$	6.675	3.023	0	3.773	$\overline{G}$	3.702	0.888	0	7.462	$\overline{G}$	5.055	6.461	0	3.798
$\overline{L}$	10.448	0.75	3.773	0	$\overline{L}$	3.761	8.351	7.462	0	$\overline{L}$	1.257	2.663	3.798	0

pollution, population density, street trees, urban parks, and green areas. Moreover, there are some variables that are not described by this component at all, such as the number of cultural centers, security level, and energy consumption. Consequently, it is related to the ecological structure of each district, and we label this component as "Ecological Background".

3.3.2.3 Third Principal Component (PC3) – "Urban form and architectural typology"

PC3 explains 15.8% of the total variance. We observe in Figures 3.5 and 3.6, and in Table 3.4, that variables with high loadings on this component are related to commercial, corporative and industrial areas, as well as to traffic accidents. There are some variables, such as green areas, collected paper and cardboard, market stalls, and parking areas that are not described by this component. Although the indicators related to this principal component are describing large-surface activities such as: commerce, offices, and industries, we label it on the fact that it also suggests a distinctive neighbourhood character, with a particular urban form and architectural typology. Consequently, we label this component as "Urban and architectural typology".

3.3.3 Multiple Factor Analysis

We complement the PCA description by reading the contribution of the variables in each of the four metabolic categories. In addition to the districts' coordinates (Figures 3.4 to 3.6),

Table 3.9: Euclidean distances between districts' centroids.

	$\overline{C}$	$\overline{E}$	$\overline{G}$	$\overline{L}$
$\overline{C}$	0	10.821	9.155	11.175
$\overline{E}$	10.821	0	7.188	8.797
$\overline{G}$	9.155	7.188	0	9.184
$\overline{L}$	11.175	8.797	9.184	0

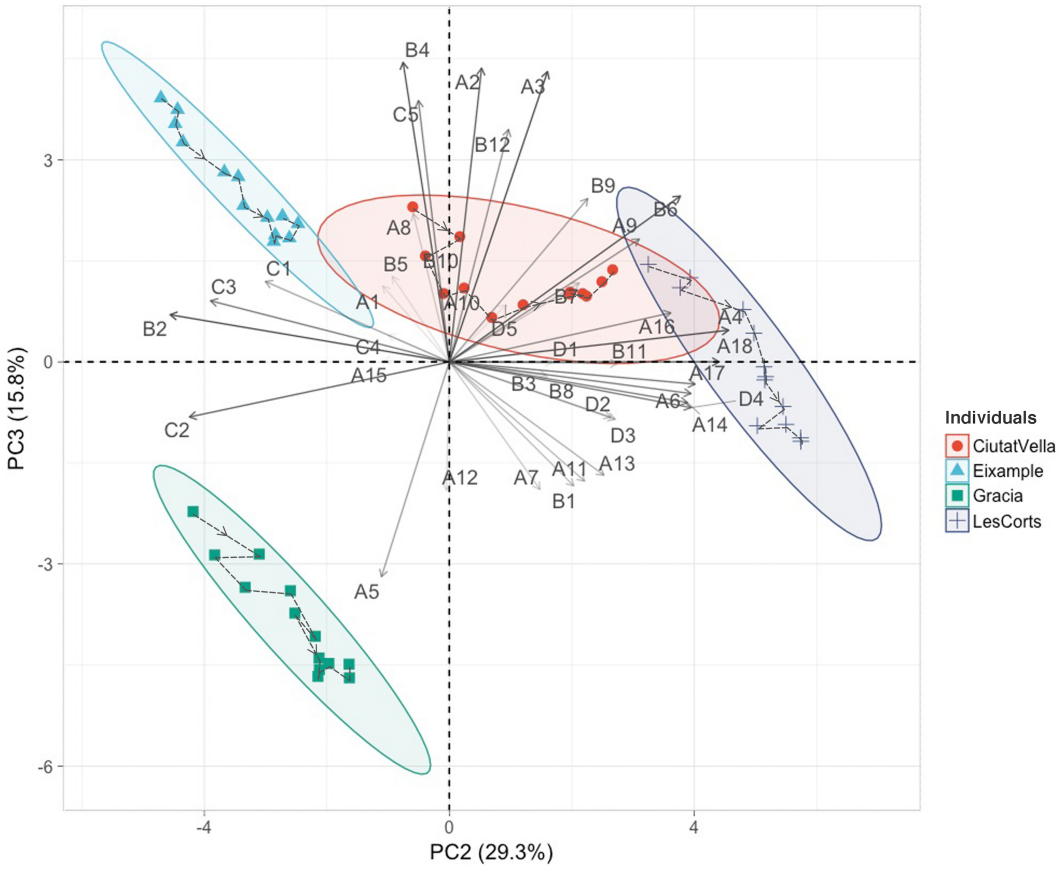

Figure 3.5: Biplot of individuals and variables for second and third principal components.

Table 3.10: MFA results for the groups of variables.

Metabolism Categories (Groups)	Coordinates			Contribution (%)		
	PC1	PC2	PC3	PC1	PC2	PC3
Input supportive	0.95	0.66	0.38	29.12	27.77	34.49
Output pressure	0.97	0.48	0.42	29.87	20.05	38.07
Destructive	0.51	0.70	0.27	15.72	29.58	24.37
Regenerative	0.82	0.53	0.03	25.19	22.58	3.05

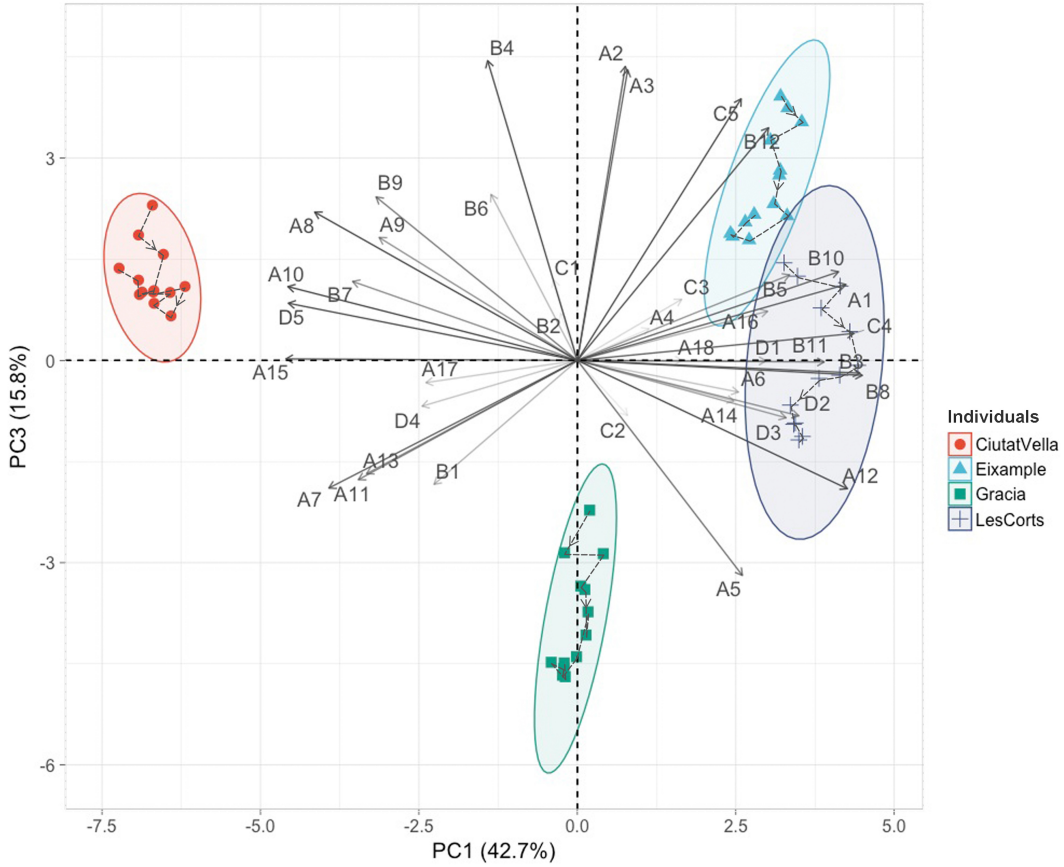

Figure 3.6: Biplot of individuals and variables for first and third principal components.

new relations arise between metabolic types of variables when those categories are described using the principal components. In particular, Table 3.10 shows the relation between groups and principal components; each group of variables —a metabolism type— contributes to the construction of the principal components. As shown in that table, Output Pressure and Input Supportive types of variables contribute the most to the construction of the first component (29.9% and 29.2%, respectively); the minor contribution, in this case, is from the Destructive Metabolism type of variables (15.7%). Instead, the four groups of variables are very close together in the second component, only the Destructive Metabolism category has a differentiated contribution in this component with the 29.6%. Regarding the third component, it is related to both Output Pressure and Input Supportive types of variables —whose contributions are 38.1%, and 34.5%, respectively— but it is not linked with the Regenerative Metabolism category (contributing only with 3.0%). These facts suggest that the Output Pressure and the Input Supportive groups of variables are defining the first and third components, and in the case of the second component, although all metabolism groups contribute similarly, the Destructive Metabolism group mostly characterizes it.

3.4 Discussion

Now we interpret the comprised information in the dimensionally-reduced description by considering the districts' performance. Since the principal components have been labeled, in the first part of this section we read the description of each district: urban diversity can be recognized and quantified using the results of the previous section. Then, we evaluate the metabolic performance of districts by correlating individuals, principal components, and metabolic groups of variables, as given by the MFA results. We further discuss the link between urban diversity and resilience. Finally, we analyze the time-dependent evolution of the districts and the possible increase —or decrease— of urban diversity.

3.4.1 Urban diversity

We follow the previous studies in [187–192], which have taken advantage of biplot graphs to display diversity in multiple systems. Using the PCA representation (i.e., biplots), we perform a descriptive analysis of diversity with respect to principal components. To achieve this reading we determine each district's performance [186] (positions in biplots of Figures 3.4 to 3.6) and quantify diversity as the Euclidean distance [193] between districts' centroids.

The first notable characteristic of the districts' performance in the PC1-Social Background, is that this component separates the districts of Ciutat Vella and Les Corts roughly by 10.5 units of distance (see Table 3.8), one unit more than the distance between Ciutat Vella and Eixample (9.7), and four units more than distance between Ciutat Vella and Gràcia (6.7). The districts of Eixample and Les Corts are closely together (i.e., the distance between them is of 0.75), an relatively near to Gràcia (being 3.0 units apart from Eixample and 3.8 from Les Corts). Certainly, differentiated social interactions for Ciutat Vella are identifiable from the raw data: this district owns community amenities for the whole city, and, in this sense, it exhibits high loads on the indicators related to religious and entertainment buildings, exclusive pedestrian areas, market stalls, and cultural and sport centers, among others. This can be explained because in the last

decade an intensive culture-based program of urban regeneration has been developed in the district —mainly in The Raval, one of the neighborhoods of Ciutat Vella— in order to face urban marginality and social stigmatization: constituting itself a center for culture and creativity [194]. On the other hand, Les Corts —and Eixample, to a lesser extent— are mainly residential districts, which are characterized, among other features, by higher income, higher security level, a large number of vehicles, and large parking areas. Therefore, PCA results agree with the actual nature of the districts, where districts are differentiated by the quantity and quality of social interactions associated with the diversity of land uses and functional typologies.

Figure 3.5 offers a remarkable characteristic, it clearly shows the emergence of "two Barcelonas", with PC2-Ecological Background drawing a division between two groups of districts: Eixample and Gràcia on one hand (Figure 3.5, left), and Les Corts on the other (Figure 3.5, right), with an average distance of eight units between them. Les Corts has a good performance in the indicators related to the quality of public spaces, and with green and recreational areas. Particularly, this district is defined by the following indicators: urban parks (A17), urban green areas (A18), educational buildings (A4), sports area (A6), playgrounds (A14), and street trees (A16). We observe that these indicators are inversely correlated with air pollution indicators (see Figure 3.3). In contrast, the group of districts concerning Eixample and Gràcia is described mainly by the average levels of PM_{10}, NO_2, and CO. Therefore, Eixample and Gràcia exhibit a great flow of pollutants —compared to all the districts of the city—, linked to high population density (B2), and a lack of green and recreational spaces (see the correlation in Fig. 3.3). According to Barcelona City Council (2016), the road transport activity in 2013 was responsible for the 67.6% and 68.9% of the NO_2 emissions in Eixample and Gràcia, respectively. Furthermore, because of its high traffic, Eixample experiences the highest noise levels (over 70 dB(A)) in the city [195]. In order to improve the environmental quality of these districts and move forward to positive performances in PC2-Ecological Background, some planning strategies should be applied. Since the compact form and the mixed use development of these districts encourages walkability [196], actions should focus on reducing road activity and promoting cycling. Moreover, taking into account the high density of these districts and the scarcity of green areas, green roofs and green walls represent some alternative solutions.

Regarding PC3-Urban Form and Architectural Typology, this principal component segregates Eixample and Gràcia (see Figures 3.5 and 3.6), with a gap of around seven units of distance (6.5). This can be explained since Eixample is mainly associated with a larger amount of offices and greater commercial and industrial areas than those in Gràcia. In fact, Gràcia has the lowest records in these indicators for the whole case study. On the one side, Eixample displays a metropolitan scale due to its size, compactness, high density and proximity to the city center (see Figure 2.1). Eixample, being designed in 1859, has a compact urban form (see Figure 3.1) consisting of a strict grid pattern with large straight streets, square blocks with chamfered corners (133 meters side), and wide crossing avenues [177–179]. In terms of architectural typology, Eixample is characterized by large-surface and high-rise buildings (up to 7 floors) compared with the average height of buildings in Barcelona. Hence, it allows a mixed-used development. Gràcia, on the other hand, is more related to a human scale of interactions since it was originally an independent and smaller town. This district was established originally in 1626 as a village —out of Barcelona's boundaries— and displays a distinct urban layout: it has an orthogonal grid, with small and medium size blocks, crossed by narrow streets

(see Figures 3.1 and 3.2, and Table 3.1). It is defined by low and medium-rise buildings of medium-size plots.

The indicator of traffic accidents also unveils the urban form and the amount and scale of interactions: because of its straight streets, Eixample's urban layout encourages car mobility with higher levels of street traffic and higher car speeds [195], hence, traffic accidents are more frequent. Instead, human scale and car-restricted streets of Gràcia possess lower traffic, and consequently less road and car accidents. In summary, both districts exhibit differentiated neighborhood characters and scales which are captured in the multivariate analysis.

The PCA applied on the multivariate system description reveals correlations among indicators and districts, which are difficult to identify only with the raw data (as it that has been shown in Figure 3.3 or with the urban structure metrics in Table 3.1), and allows us to understand the city at a macro level. It exhibits distinctive attributes of the four districts of the case study, as demonstrated by each one of the principal component descriptions presented before. In the case of the first component, it exposes an "inner diversity" of land uses, contrasting Ciutat Vella and Les Corts: that own different functional typologies and are the most distant, or what is the same, the most different of the case study. Our findings portray the different types of "Barcelonas", and thus, the re-evaluation of the *a-priori* homogeneity of the city. This is more evident when looking at the second principal component; the description made by this component draws a clear division between two distinguishable groups or types of districts: Les Corts on one side, characterized by large public spaces and green and recreational areas, and Eixample and Gràcia on the other, identified as dense and highly polluted.

Moreover, the measurement of the relative distances between the districts as described by PCA gives the main instrument for quantifying diversity at the city scale. The Euclidean distance among districts (i.e., three-dimensional distance between districts' centroids) determine that Ciutat Vella and Les Corts are the most distant —different— districts of the case study, with an average distance of 11.17 units; Ciutat Vella is also widely separated —10.82 units— from Eixample; and Eixample and Gràcia are the closest —similar— ones, with an average distance among them of 7.19 units (as presented in Table 3.9). These distances in a non-diverse case study would give close to zero values, and imply districts' stacks in specific locations. This has not been the case of the present results, in which significant distances among the districts indicate a diverse behavior of the city.

This diverse behavior among districts can be explained by factors such as building construction type, urban development period, economic prosperity, urban form, geographic setting, prevailing land use, and other biophysical and economic characteristics of the neighborhoods. The amount and type of activities within districts are also related to the previous factors. This has been illustrated by the third principal component, where there is a clear differentiation in the types of land uses between Eixample and Gràcia, mainly explained by their distinctive building type, urban form, and neighborhood characters.

3.4.2 Urban trend analysis

The multivariate system description of the current case study is a time-dependent data: it is, therefore, possible to analyze districts' evolution between 2003 and 2015. Trends are clearly drawn by biplot graphs [197] and districts' displacements with respect to principal components are shown in Table 3.6. For example, with regard to PC1, we observe that none of the districts

move —change— in the time span. Contrarily, the displacements of districts are substantial in PC2: although there is no relative divergence or convergence movement between the districts, all of them move simultaneously towards positive values. In the case of Gràcia, Eixample and Les Corts, they advance about two units of distance during the 12 years (2.55, 1.99, and 2.49, respectively). Specifically, Eixample and Gràcia (districts at the bottom of Figure 3.4) move forward to a reduction of the indicators related to air pollution (C1, C2, C3). This trend agrees with the findings from Barcelona City Council (2016) and Baro (2014): the measures from the municipal monitoring stations show a steady trend for NO_2 values, and a minor decrease for PM_{10} since 2006. In the case of Les Corts (see top of Figure 3.4), there is an increase in units and areas of urban services such as education (A4), sport (A6), playgrounds (A14), street trees (A16) and urban green (A18), upgrading its carrying capacity —Input supportive category. Ciutat Vella, on the other hand, is the district with the highest displacement in the time span: it moves more than three units (3.26) in the time period.

The displacement of the districts along PC3 demonstrates a particular divergence behavior of Ciutat Vella compared to the other districts, and therefore, the urban diversity increases. While Ciutat Vella stagnates, the other districts move backward in this component about two units of distance in the time span (Eixample: -1.75, Gràcia: -2.27, and Les Corts: -2.63).

3.4.3 Contribution of the metabolism categories to the PCA description

Although we acknowledge that the urban metabolism approach can be developed in greater depth (e.g. by applying other urban metabolism measures [198]), here we compare the contribution of the metabolic groups of variables against the districts' performance to have some insights on the metabolic character of districts. Figure 3.6, for example, detail the location of Les Corts and Eixample against the first and third principal components. And if we superpose this figure with the results given in Table 3.10, we observe that both districts are characterized by the Input and Output metabolic groups of variables. Ciutat Vella is also segregated when described by PC1 and therefore, it can be said that the Input and Output categories of variables are contributing the most to its description. As indicated previously, this district is related to the urban amenities of the city. Instead, it is not correlated with industry and pollution.

In the case of the third principal component, for which the contribution from all metabolism categories is similar, except from the Regenerative Metabolism type, we can see a gap between Gràcia and the rest. Indeed, Gràcia is not described by PC1, and therefore, not characterized by the Input and Output metabolisms. Instead, it is mainly described by PC2 and PC3, which are contributed the most by the Destructive type and few by the Regenerative type. This district displays correlation with industry and infrastructure, such as industrial areas (B12), water supplied to domestic consumption (B2), traffic accidents (C5), and air pollution indicators (C1, C2, C3), and can be labeled as a destructive type of district. Thereby, planning actions must be taken in order to reduce its destructive metabolism and enhance its regenerative metabolism (e.g. plans to reduce air pollution, mitigate industrial activity and encourage recycling).

3.4.4 A link with resilience

Although the concept of differentiated characteristics has been reported in other urban studies at the neighborhood scale (see for example [25]), the relation with resilience was not consid-

ered. Given the definition of resilience as the ability of a system to adapt to a modifying process while retaining its functionality and not necessarily returning to the previous equilibrium state [55, 57], it can be said that the case study of 4 districts of Barcelona has been resilient between 2003 and 2015. The system, or the subsystems (if we consider the four districts as isolated systems), have adapted and responded to disturbances (i.e. 2008 - to present Spanish financial crisis), performing in diverse circumstances while retaining their basic functions. Particularly, if we look at the system of indicators, and specifically at those included in the input supportive category which increases in the time period (e.g. buildings of commerce, offices, education, health services, playgrounds, sport, and cultural centers), it can be stated that urban planning has built adaptable social and economic infrastructures, thus strengthen urban resilience.

Despite the above, an extensive analysis of urban resilience is required for each district. For example, it has also been reported in other studies [194] that after some urban renewal programs Ciutat Vella improved its cultural infrastructure. Since those programs were mainly focused on physical aspects, gentrification processes have taken place [199] threatening the district's social resilience. Precisely this is one of the challenges within the framework of urban resilience: cross-scale trade-offs among the different strategies that aim at fostering resilience [50, 74, 158]. Hence, the application of the PCA to the multivariate dataset representing the districts, and the visualization and measurement of the individuals, can support effective urban planning.

In order to strengthen system's resilience, planners must identify the processes and disturbances that the urban system is likely to face. Yet, to reach accurate conclusions about urban factors fostering diversity and influencing urban resilience, complementary types of analysis are needed. Recent studies have suggested that urban resilience must be evaluated at multiple scales, including the scale of the city's subsystems [74, 158]. This is exactly the purpose of the present approach, where the analysis is undertaken at district's scale. Certainly, resilience is understood as a property of the system that is not subject to a specific scale (e.g. the city's scale) [74]. Focusing only on one scale could enhance resilience in a single district without considering the effects in the other districts of the city. On the other hand, a general perspective may neglect subsystems that can help to further improve resilience in the whole city.

An insight we can provide is that there are "many types of Barcelonas", which apparently should strength the city's resilience as a whole. As remarked by [55], the functions and services provided by those types can be sustained over a wider range of conditions, and the system will have a greater capacity to recover from disturbance. Nonetheless, an immediate consequence of this fact is a reductionist approach to properly assess resilience at the lower levels of a system, in our case at the neighborhood scale, for example. At the city scale, diversity improves the resilience capacity of the system by having several types of districts that are capable of performing different and specialized urban functions. For example, Ciutat Vella has been identified as a diverse and pedestrian zone that owns amenities for the whole city. In contrast, Les Corts emerges as a predominant housing and car-usage zone. The answer to how this trade-off between global diversity and local specialized function at the district's level favors resilience remains elusive and, at the same time, it raises serious ethical concerns for the governance of a city.

MFA results can also contribute to the understanding of the resilience and sustainability level of the urban system. Actually, the metabolism of the city can be seen as a living organism that struggles towards life —or sustainability: the Input and Regenerative categories must

exceed (in quantity) the Output and Destructive ones for life to be possible and sustain. In this sense, we raise the question on how to quantify these categories for each district and possibly conclude about the metabolism of the city itself. The differences between the contributions of the metabolism categories in the present results are somehow faint. Nevertheless, we believe that including these categories is important as a departure step to link this exploratory multivariate statistical approach with the urban metabolism framework. In this regard, metabolism types —categories— can highlight trends and patterns, and provide measures, which can be easily identified by decision makers for assessing the degree of sustainability in the urban development [200]. Therefore, urban planning strategies could and should be applied to improve the Input and Regenerative capacity of districts and thus the sustainability of the whole system.

3.5 Conclusions

In this chapter we have applied exploratory multivariate statistical techniques to expose the diversity between the districts of a city. Specifically, we applied PCA to a multivariate system of indicators that collects measurements of environmental, economical, and social indicators of the city of Barcelona (Spain) between 2003 and 2015. The assembly of the dataset for the particular case study of four districts of the city allowed us to evaluate the methodology in a complex urban system, which also constitutes itself a reliable dataset for future research. After applying PCA to the dataset, three principal components have been found to account for 88% of the total variance: "Social Background" (PC1), "Ecological Background" (PC2), and "Urban and architectural typology" (PC3). One of the main findings, is that many types of "Barcelonas" emerge in the dimensionally-reduced description: districts exhibit differentiated neighborhood characters and distinctive functions at the city's scale. We could also quantify diversity as the Euclidean distances between separated districts. The distances allow to identify clustered districts, as well as those that are separated apart. At least two groups of districts are clearly identifiable, one that is characterized by large public spaces and green and recreational areas, and another one that is dense and highly polluted.

Results demonstrated that Ciutat Vella and Les Corts are the most distant —different— districts, with an average distance of 11.17 units; Ciutat Vella is also widely separated —10.82 units— from Eixample; and Eixample and Gràcia are the closest —similar— ones, with an average distance of 7.19 units. Moreover, the temporal-dependency of the dataset reveals information about urban diversity trends, where the district of Ciutat Vella seems to diverge from the rest increasing the diversity of the city. Temporal results also demonstrate that, for example, some districts have moved forward to an increase in urban services, while some others have decreased air pollution throughout the years.

The potential applications of this methodology are not only restricted to the development of urban policy, for which this framework constitutes another way to evaluate urban systems. It can also be exploited in the understanding of cities from the perspective of the science of cities [39], where the aim is to quantify the complex relationships between the different elements that compose the urban system. Since this work is based on novel algorithmic developments in the field of exploratory multivariate statistical analysis (i.e, MFA), it can also be used by researchers to contrast and validate their own methods.

Some open questions left by the elaboration of this chapter are the following. Should di-

versity be assessed at the city scale or at the district scale? Why robustness, a system level property, emerges in urban systems as well as in natural ecosystems? How to relate all the temporal trends and the direction of the city towards the increase of sustainability? Future studies on the science of cities and on urban metabolism will help to unveil these open questions and will allow us to more fully understand the links between diversity, urban resilience, and their trade-offs with sustainability [74]. Nevertheless, our multivariate statistical approach provides a framework for further mapping of Barcelona's diversity. The present methodology is feasible to be replicated at other scales and urban systems, for which particular indicators to each case study should be selected. One future work is focused on reviewing accurate indicators that describe urban sustainability in terms of diversity at the city scale, as measured with the present methodology. Cross-city profiles could be developed in future multivariate statistical analyses, in order to find diversity patterns and causes influencing urban resilience, and the scaling of those particular patterns in different cities. Another future work is to contrast the exploratory multivariate analysis with some recent statistical methods such as Self-Organising Maps (SOMs) [201, 202], compositional data techniques [203], and cluster analysis [204]. This can be done by including all the 10 districts of Barcelona, similarly as it will be presented in the next chapter.

Chapter 4

Unveiling patterns and trends in a complex urban system

In this chapter, we apply multivariate statistical methods to understand the complexity of the city of Barcelona and its temporal change. First, we build an abstraction of the case study by means of a multivariate dataset for the 10 districts of the city in the time span between 2003 and 2015. The multivariate dataset comprises measurements of 40 indicators grouped in the economic, environmental, and social dimensions of sustainable development. We explore the evolution of Barcelona by applying Exploratory Multivariate Analysis (EMA) over its multivariate description: Principal Component Analysis (PCA), Multiple Factor Analysis (MFA), and Hierarchical Cluster Analysis (HCA). Furthermore, some multivariate visualization techniques (i.e., correlograms, biplots, dissimilarity matrices, and Confidence Ellipsoids) are implemented in order to understand urban complexity patterns.

4.1 Introduction

Following the main objective of this thesis, we seek to expose the relation between diversity and sustainability in urban systems using an information system's approach. In this chapter, we apply different multivariate statistical methods to understand the complex relationships of the elements in an information system that describes the city as a composite of subsystems. Concretely, the case study now incorporates measurements of environmental and socio-economic indicators for the 10 districts of Barcelona between the years 2003 and 2015. Besides, in order to give some insights about the sustainability state of the system the multivariate system for the expanded case study is arranged following the three main pillars of sustainable development. Given that the amount of raw information comprised in the increased multivariate system, and therefore its complexity is huge, here, we explore some other multivariate statistical methods —besides from Principal Component Analysis (PCA). Specifically, we expand on Multiple Factor Analysis (MFA), apply Hierarchical Cluster Analysis (HCA) and explore other multivariate visualization techniques (i.e., biplots, correlograms, dendrograms) in order to understand the urban complexity given by the multivariate description of the city. Because the case study is now extended to the 10 districts, we apply HCA to find groups [205] among the districts, and validate those groups with the PCA description.

The processing of a big amount of information can better be achieved by means of computers. Multivariate statistical analysis methods are intended to build up a reduced description of the system and to, among other goals, select new representative variables, model data, and recognize patterns. The process of dimensionality reduction requires the development of algorithms that can be efficiently executed in computers. Indeed, a large number of studies rely on the dimensionality reduction given by the PCA to build up a synthetic abstraction of a large amount of information [206]. As examined in previous chapters, PCA can also reveal patterns of similarity (or dissimilarity) that are difficult to see directly in the raw data.

MFA is based on PCA but extends its applicability to categories —current sustainable development's dimensions. It reveals an integrated picture of the system by displaying the relationship between observations and groups of variables [107, 108, 168]. Hence, it offers an alternative method to picture the diversity of urban systems and to relate it with the dimensions of sustainability. Furthermore, HCA exhibits hierarchical relations between the different types of entities: it uses the pairwise distance matrix between observations as the clustering criteria. There are two types of hierarchical clustering methods: agglomerative and divisive. Agglomerative methods start by placing each observation in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Divisive methods do just the opposite, all observations start in one cluster, and splits down the hierarchy [207].

Regarding measuring and displaying diversity with these methods; PCA, sometimes in combination with clustering, has been applied to evaluate diversity in urban contexts [102, 104, 208–210]. Moreover, HCA has been widely used in microbiology and genetic studies to portray similarity and dissimilarity among individuals [96, 211–213]. With respect to MFA, it has been applied in a few studies assessing diversity [214] and urban systems [175, 215]. To the best of our knowledge, an exploratory multivariate analysis to assess diversity within urban sustainability frameworks has not been done so far.

Above all the previously mentioned features, multivariate statistical methods potentially reveal the temporal evolution of the system. If the dataset observations have been measured throughout time, then the time-dependent results can be visualized by highlighting the path of individuals —their initial and final points. Analyzing the temporal change of the data comprised in the reduced-dimensional plots still can be hard. Even in the bi-dimensional description, the data points gradually displace from one year to another. The amount of information about these consecutive displacements is large, and the complete analysis of time-span becomes exhaustive. Hence, algorithms have been developed to process the displacements information: most aim to represent the abstract information of displacements in a human advantageous way [122–125]. One approach is to use machine learning to build the deformation register, where deformation models are built over a mathematical basis: basically, algorithms that are designed to approximate the discrete displacement —or velocity— fields by using optimization techniques that fit the coefficients of a polynomial function. With those methods the temporal change can be compressed in a few coefficients of the order of bits: the so-called four-dimensional morphology change models.

Since a quantitative methodology that describes the temporal change of the urban system —represented by a multivariate data-set— has not been carried out before, this chapter presents a first approximation in this respect. Our idea is to use Confidence Ellipsoids in the dimensionally-reduced representation of the system to understand its temporal change. We think that it is a technical approach that quantifies the dispersion of the time-dependent results,

Table 4.1: Socio-economic indicators for the case study.

		Ciutat Vella	Eixample	Sants Montjuïc	Les Corts	Sarrià-Sant Gervasi	Gràcia	Horta	Nou Barris	Sant Andreu	Sant Martí	Barcelona
Population		101387	266416	181910	82033	149279	121347	168751	166579	147594	235513	1620809
		6.3%	16.4%	11.2%	5.1%	9.2%	7.5%	10.4%	10.3%	9.1%	14.5%	100%
Surface (km^2)		4.37	7.48	22.94	6.02	20.10	4.19	11.94	8.04	6.56	10.52	102.16
		4%	7%	22%	6%	20%	4%	12%	8%	6%	10%	100%
Pop. density (inhabitant/km^2)		23201	35617	7930	13627	7427	28961	14133	20719	22499	22387	15865
Foreigners	Population	45357	53393	35507	9796	17792	19574	21015	26272	17604	38597	284907
	% of district	45%	20%	20%	12%	12%	16%	12%	16%	12%	16%	18%
	% of Bcn.	16%	19%	12%	3%	6%	7%	7%	9%	6%	14%	
Household income per capita		86.9	119.3	79.1	136	182.4	105.4	79.2	55	74.5	87.1	100
Education	WE	2.4%	1.7%	3.1%	1.7%	0.7%	2.0%	4.0%	5.3%	3.1%	3.3%	2.8%
	ES	24.3%	12.2%	22.4%	11.2%	6.2%	12.4%	20.1%	27.9%	21.3%	20.3%	18.0%
	HE	30.6%	41.4%	25.2%	42.1%	49.8%	40.8%	23.8%	13.8%	21.8%	25.7%	30.9%
Population Age	Children(0-14)	10.8%	11.0%	11.8%	12.6%	15.9%	12.1%	12.4%	13.3%	13.5%	13.6%	12.7%
	Youth(15-24)	9.6%	8.5%	8.8%	8.8%	11.0%	7.9%	8.7%	9.2%	8.7%	8.7%	8.9%
	Adults(25-65)	65.9%	58.4%	59.3%	53.1%	51.6%	58.2%	55.1%	54.0%	56.0%	57.4%	56.9%
	Seniors (65<)	13.7%	22.1%	20.2%	25.5%	21.4%	21.7%	23.9%	23.5%	21.8%	20.4%	21.5%

WE: Whitout education, ES: Elemantary school, HE:Higher Education, Bachelor's Degree, Graduate, Master's, PhD.

and allows to identify patterns of diversification or convergence of the elements of the system.

This chapter is organized as follows: In Section 4.2 we shall introduce the multivariate dataset for the case study and the statistical techniques to display the complex relations of the urban system. In Section 4.3 we demonstrate the application of these statistical methods and discuss some particular findings for the case study. Finally, some conclusions close the chapter in Section 4.4.

4.2 Materials and methods

In the first part of this section, we describe the case study and and its abstraction as a multivariate system. We build upon the multivariate dataset —with 40 indicators— presented in Chapter 3 but construct a larger set that comprehends the measurements for all the 10 districts of Barcelona. Then, we describe the exploratory multivariate statistical approaches and the temporal change analysis that we apply to the urban system.

4.2.1 Case study

We study the Mediterranean city of Barcelona (Spain) between 2003 and 2015 as a composite of all its districts, as depicted in Figure4.1. Barcelona is the capital and the largest city of Catalonia —and the second-largest city of Spain. It has a population of 1.6 million inhabitants and a surface of 102.16 km^2 —within the city's boundaries. The city and its 36 neighboring municipalities constitute the Barcelona metropolitan area (àrea Metropolitana de Barcelona (AMB)) with a surface of 636 km^2 and more than 3.2 million inhabitants [153].

Barcelona is divided into 10 administrative districts (i.e. boroughs), which are subdivided into 73 neighborhoods. We consider all the 10 districts of the city; Ciutat Vella (1), Eixample (2), Sants-Montjuïc (3), Les Corts (4), Sarrià-Sant Gervasi (5), Gràcia (6), Horta-Guinardó (7), Nou Barris (8), Sant Andreu (9), and Sant Martí (10). All these districts comprise the case study. Table 4.1 presents some relevant socio-economic data for the case study.

Figure 4.1: Barcelona area and location map. Adapted map from Urban Guide City Map [154].

Table 4.2: Information System

SD	Category	Id	Indicator	Unit
Economic	Businesses	E1	Commercial area	m^2/person
		E2	Office buildings	m^2/person
		E3	Industrial area	m^2/person
		E4	Active market stalls	unit/person
		E5	Accommodation places	unit/person
	Income and Expenses	E6	Available household income per capita	Index
Environmental	Green space, ecosystems and heritage	En1	Street trees (Unit = 5.6 m^2)	unit/person
		En2	Urban parks	ha/person
		En3	Urban green area besides parks	ha/person
	Water	En4	Total volume of water supplied to the commerce	m^3/person
		En5	Total volume of water supplied to domestic consumption	m^3/person
		En6	Total volume of water supplied to the industry	m^3/person
		En7	Total volume of water supplied to other usages	m^3/person
	Energy	En8	Residential energy consumption	MWh/year.person
	Air quality	En9	Average levels of PM_{10}	$\mu g/m^3$
		En10	Average levels of NO_2	$\mu g/m^3$
		En11	Average levels of CO	mg/m^3
	Waste	En12	Collected volume of non-recyclable waste	ton/person
		En13	Collected volume of paper and cardboard	ton/person
		En14	Collected volume of glass	ton/person
		En15	Collected volume of containers	ton/person
	Transport	En16	Total number of motor vehicles	unit/person
		En17	Traffic accidents	unit/person
		En18	Parking area	m^2/person
Social	Demographics	S1	Natural growth rate	-
		S2	Population density	people/ha
	Housing	S3	Residential area	m^2/person
	Education	S4	Educational buildings	m^2/person
	Health	S5	Health service buildings	m^2/person
	Social and community services	S6	Sport area	m^2/person
		S7	Religious buildings	m^2/person
		S8	Entertainment buildings	m^2/person
		S9	Other land uses	m^2/person
		S10	Cultural centres	unit/person
		S11	Sport centres	unit/person
		S12	Public water fountains	unit/person
		S13	Playgrounds	unit/person
	Security	S14	Security level in the neighbourhood	Points
	Expenses and public administration	S15	Maintenance cost	Miles of €/person
	Well being	S16	Streets and zones with pedestrian priority	ha/person

4.2.2 City's abstraction: multivariate system of indicators

With the aim of identifying urban interactions and patterns towards sustainability, we construct a multivariate description of the 10 districts in a time-dependent —from 2003 to 2015— dataset comprising 40 indicators grouped in the three sustainable development dimensions: Economic, Environmental, and Social. This dataset is a matrix that collects the measurements for the 40 indicators and the 10 districts in each year of the time-span (13 years); thus, a total of 130 individuals are considered (i.e., "CiutatVella 2003", "Eixample 2003"… to "SantMartí 2015"). In the present study, we suppose the city as a complete composite of smaller subsystems (districts). Accordingly, we collect data exclusively from disaggregate indicators at the district's scale to enable a further comparison among the districts: we build our multivariate dataset upon open databases from [153, 155] taking into account the availability of disaggregated data at this scale.

Following previous studies on sustainability indicators [76, 79, 80], the index system is arranged over the three Sustainable Dimensions (SD) and its respective categories (see Table 4.2). The 40 indicators' classification into these categories follows the indications given by the focus group of experts in The Sustainability Measurement and Modeling Lab (SUMMLab). Thus, our index system is organized as follows: the Economic SD dimension groups 2 categories; Businesses, and Income and Expenses, and includes 6 indicators. The Environmental SD dimension comprises 6 categories; Green space, ecosystems and heritage, Water, Energy, Air Quality, Waste, and Transport, including 18 indicators. The Social SD dimension groups 8 categories; Demographics, Housing, Education, Health, Social and community services, Security, Expenses and public administration, an Wellbeing, and comprises 16 indicators. Because the dataset for the case study is too large (40 variables and 132 individuals) to be displayed here, we arrange separated tables for each district. For example, we present the dataset for the district of Sants-Montjuïc in Table 4.3. All the compiled datasets are available online at `https://summlabbd.upc.edu/SalazarLlano/PhDThesis/`.

The ratio of variables to data individuals is 0.3, guaranteeing the occurrence of reliable correlations. Also, the information was taken in nearby but not contiguous places (e.g. the yearly averaged values of PM_{10}, which were measured by distinct air quality stations across the districts of the city). The hypothetical dependence between the raw data has been a frequent issue in urban studies (see [216, 217] for a detailed analysis of this problem). We are also aware of the possible common external influences for some of the measured variables (e.g. the yearly weather fluctuations). Nevertheless, we think that the dataset is not entirely subject to these occurrences and, therefore, future studies are greatly encouraged following this information system.

4.2.3 Exploratory Multivariate Analysis

The exploratory multivariate statistical analysis that we apply in this chapter is organized in 6 stages, which are shown in Figure 4.2 together with their main outputs. The first stage of the methodology refers to the city's abstraction into a multivariate system of indicators grouped into the three SD dimensions (as described in the preceding subsection). The subsequent three stages constitute the exploratory multivariate analysis itself: in stage 2 we perform PCA, in stage 3 we compute MFA, and in stage 4 we apply an HCA. Based on the PCA results, in stage 5 we construct a Distance Matrix to explore diversity. Finally, in stage 6, we use the statistical tool of Confidence Ellipsoids to understand temporal trends of the PCA data points.

Before applying the three stages of the exploratory multivariate analysis, we construct the correlation matrix by computing the Pearson's correlation coefficient between the variables of the dataset. This correlation matrix is depicted with a Correlogram in Figure 4.3, that will be later used in the multivariate analysis to explore links among indicators. The Correlogram highlights the most correlated variables in the dataset: direct (+) correlations are colored in blue, and inverse (-) correlations are colored in red. Again, we realize that the interactions that take place in the city between districts and indicators are far too complex, and therefore, the exploratory multivariate statistical is justified. In the remainder of this section, we will explain the statistical methods, their utility, and main outcomes.

Table 4.3: Dataset for the district of Sants-Montjuïc between 2003 and 2015.

Id	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
E1	6.47	6.45	6.37	6.26	6.29	6.17	6.13	6.10	5.98	6.02	5.99	6.09	6.03
E2	3.23	3.26	3.27	3.26	3.32	3.30	3.28	3.29	3.38	3.50	3.53	3.52	3.44
E3	23.14	22.55	21.76	20.88	20.48	19.57	18.91	18.23	18.13	16.38	16.60	16.30	15.61
E4	2.8E-03	2.9E-03	2.7E-03	2.6E-03	2.5E-03	2.3E-03	2.0E-03	1.8E-03	1.8E-03	1.7E-03	1.6E-03	1.4E-03	1.3E-03
E5	2.7E-02	2.8E-02	2.8E-02	2.8E-02	2.9E-02	3.1E-02	3.6E-02	3.6E-02	3.6E-02	3.6E-02	3.9E-02	3.8E-02	3.9E-02
E6	80.86	80.69	80.70	81.50	82.50	80.70	78.40	76.10	76.30	76.30	75.30	75.80	78.10
En1	8.8E-02	8.9E-02	8.7E-02	8.4E-02	8.5E-02	8.3E-02	8.3E-02	8.8E-02	9.0E-02	9.0E-02	9.0E-02	9.1E-02	9.1E-02
En2	1.5E-03	1.5E-03	1.5E-03	1.4E-03	1.5E-03	1.4E-03	1.4E-03	1.4E-03	1.4E-03	1.4E-03	1.4E-03	1.4E-03	1.5E-03
En3	1.8E-04	1.8E-04	1.8E-04	1.9E-04	2.2E-04	2.2E-04	2.3E-04	2.7E-04	3.1E-04	3.0E-04	3.0E-04	3.3E-04	3.1E-04
En4	4.92	4.88	4.74	4.51	4.57	4.40	4.16	4.03	4.00	3.94	3.91	3.64	3.68
En5	44.40	44.25	42.81	41.21	40.65	38.92	38.11	37.94	37.88	37.23	37.74	35.74	35.84
En6	30.44	29.75	29.96	29.78	29.08	26.64	22.47	23.20	23.07	23.17	21.34	21.58	22.19
En7	9.82	8.67	7.01	6.39	6.59	3.66	4.24	4.25	4.00	3.98	3.96	4.01	4.04
En8	2.79	2.96	3.12	2.89	2.73	2.74	2.87	2.96	2.56	2.61	2.53	2.78	2.78
En9	46.00	52.00	49.00	62.00	54.00	44.00	48.00	34.00	39.00	38.00	25.00	26.00	29.00
En10	37.00	37.00	49.00	31.00	47.00	45.00	41.00	41.00	40.00	37.00	33.00	32.00	36.00
En11	0.70	0.50	0.40	0.30	0.40	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
En12	5.1E-01	5.3E-01	5.2E-01	5.1E-01	5.2E-01	5.1E-01	4.8E-01	4.8E-01	4.5E-01	4.4E-01	4.4E-01	4.5E-01	4.5E-01
En13	3.6E-02	4.0E-02	4.9E-02	5.1E-02	5.9E-02	6.0E-02	5.4E-02	4.6E-02	3.7E-02	3.2E-02	2.7E-02	2.9E-02	2.9E-02
En14	1.2E-02	1.3E-02	1.4E-02	1.5E-02	1.8E-02	1.9E-02	1.9E-02	1.7E-02	1.8E-02	1.8E-02	1.8E-02	1.9E-02	1.9E-02
En15	6.7E-03	7.2E-03	8.0E-03	8.5E-03	1.1E-02	1.1E-02	1.1E-02	1.1E-02	1.1E-02	1.1E-02	1.1E-02	1.1E-02	1.1E-02
En16	0.61	0.63	0.66	0.63	0.63	0.60	0.58	0.58	0.58	0.57	0.58	0.58	0.59
En17	6.7E-03	6.5E-03	7.0E-03	6.8E-03	7.2E-03	5.2E-03	5.3E-03	5.1E-03	4.9E-03	5.5E-03	5.7E-03	5.8E-03	5.7E-03
En18	5.77	6.54	7.24	7.89	8.72	9.32	10.51	10.80	11.05	10.71	14.04	14.26	14.35
S1	-1.70	-0.50	-0.30	-0.50	-1.30	0.40	0.20	0.30	-0.20	-1.10	-0.90	-0.30	-1.10
S2	76.76	76.73	77.44	78.48	77.80	79.09	79.51	79.74	79.69	80.24	79.64	78.82	78.80
S3	33.60	33.69	33.46	33.09	33.45	32.98	32.81	32.87	32.97	32.92	33.18	33.48	33.58
S4	1.44	1.46	1.47	1.48	1.52	1.51	1.49	1.57	1.60	1.62	1.63	1.67	1.66
S5	0.08	0.08	0.08	0.08	0.13	0.18	0.23	0.22	0.34	0.41	0.42	0.49	0.49
S6	1.30	1.34	1.37	1.40	1.45	1.46	1.50	1.51	1.50	1.68	1.70	1.72	1.69
S7	0.13	0.13	0.13	0.12	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
S8	0.71	0.71	0.70	0.69	0.69	0.68	0.13	1.59	0.36	0.63	0.64	0.66	0.66
S9	2.13	2.01	1.87	1.73	1.63	1.48	1.96	0.84	0.87	0.85	0.83	0.84	0.90
S10	8.5E-05	1.3E-04	1.1E-04	1.1E-04	1.1E-04	9.9E-05	1.2E-04	1.3E-04	9.8E-05	9.8E-05	9.9E-05	1.1E-04	1.1E-04
S11	1.2E-03	1.2E-03	1.2E-03	8.7E-04	8.8E-04	8.7E-04	1.0E-03	1.1E-03	1.0E-03	1.0E-03	1.0E-03	1.1E-03	1.1E-03
S12	1.1E-03	1.1E-03	1.1E-03	1.0E-03	1.1E-03	1.1E-03	1.0E-03	1.0E-03	1.1E-03	1.1E-03	1.0E-03	1.0E-03	1.1E-03
S13	3.1E-04	3.2E-04	3.3E-04	3.3E-04	3.5E-04	3.4E-04	3.5E-04	3.6E-04	3.6E-04	3.5E-04	3.6E-04	3.8E-04	3.9E-04
S14	5.60	5.50	5.30	5.50	5.70	6.00	5.60	5.60	6.00	5.80	5.80	5.90	6.00
S15	4.6E-03	4.6E-03	3.9E-03	4.7E-03	5.0E-03	5.1E-03	5.2E-03	5.3E-03	5.3E-03	5.1E-03	5.1E-03	5.2E-03	5.2E-03
S16	1.7E-05	1.7E-05	1.7E-05	1.9E-05	2.0E-05	2.0E-05	2.0E-05	2.0E-05	2.1E-05	2.2E-05	2.2E-05	7.0E-05	7.5E-05

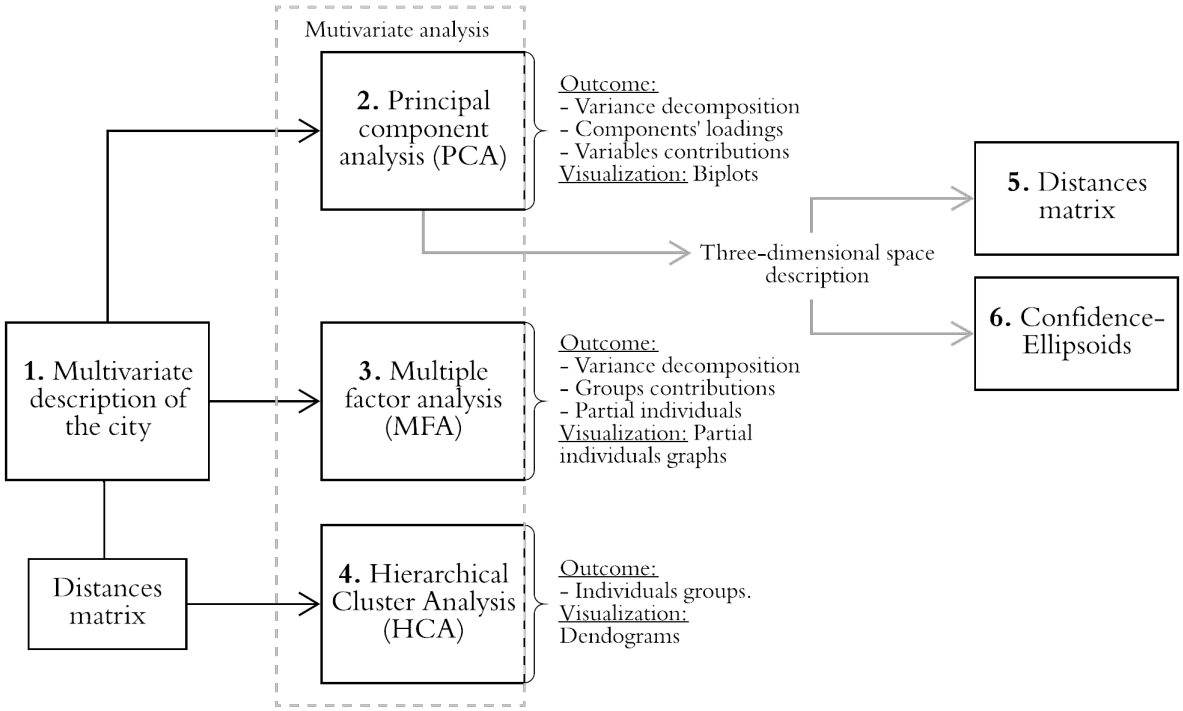

Figure 4.2: Flow chart of the exploratory multivariate statistical analysis methodology.

4.2.3.1 Principal Component Analysis

The application of PCA is intended to reduce the number of original variables in the dataset into a set of new uncorrelated variables —know as principal components— that account for most of the original variance [105]. In stage 2 of the methodology, we apply PCA to the standardized version of the dataset that includes the measurements for all the 10 districts. The data preparation is intended to make comparable data of different units and scales, such that the standardized data is scaled to have standard deviation one and mean zero.

With the application of PCA, we can reduce the 40-dimensions' description to only 4-dimensions per individual. We portray the outcome of the reduced-dimensional description using Biplots. This type of graphs offers a meaningful visualization tool that synthesizes interactions among variables, individuals, and principal components [109]. In Biplots, we make a profit of the new low-dimensionality of the system by plotting individuals and variables against two of the four selected principal components, which are placed at the axes of the plots. Nevertheless, selecting only the first three principal components constitute the starting point for PCA-based measures: Distance Matrices and Confidence Ellipsoids.

4.2.3.2 Multiple Factor Analysis

In stage 3, we apply MFA [108, 168] to the standardized dataset. MFA is an extension of PCA that takes into account the groups of variables in the statistical analysis. In the present case study, this means that the outcome of the MFA relates the three SD dimensions —groups of variables— with individuals and principal components. The principal results from MFA are the decomposition of variance for each component and the contribution of the groups of variables

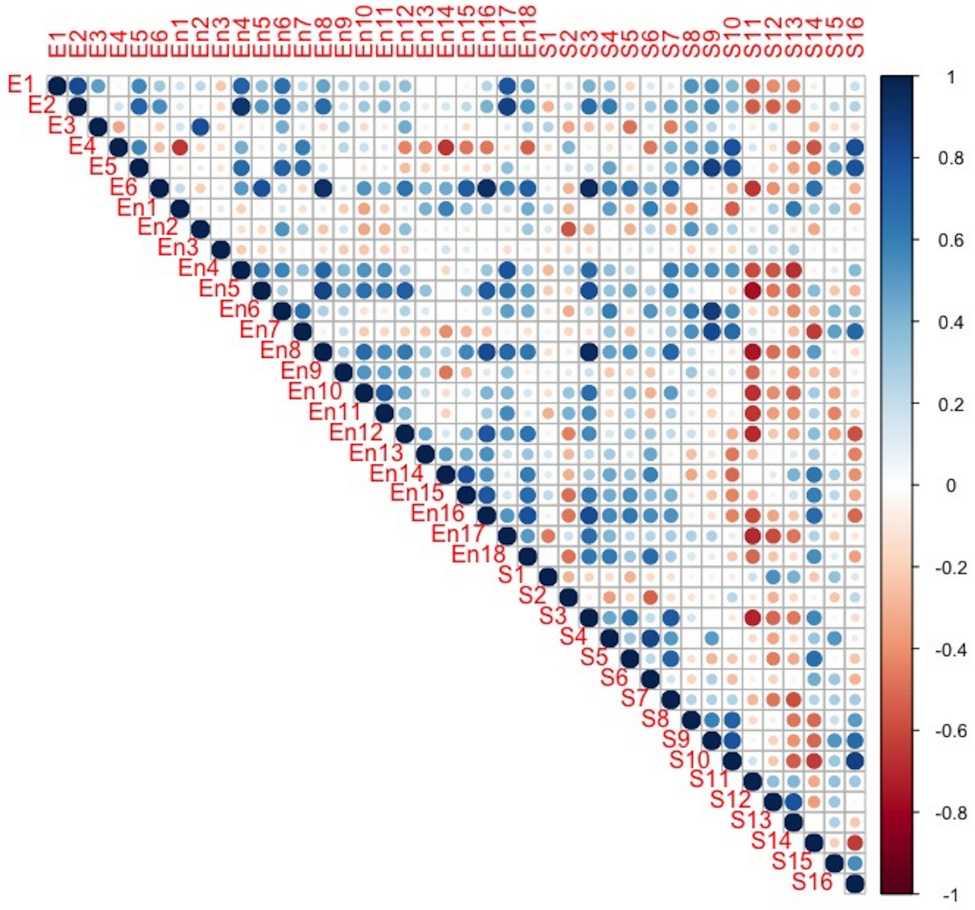

Figure 4.3: Correlogram of the dataset.

—SD dimensions— to the construction of the principal components.

4.2.3.3 Hierarchical Cluster Analysis

In stage 4, we apply HCA to the dataset in order to find similarities between individuals. Hierarchical Clustering (HC) or Hierarchical Cluster Analysis (HCA) uses the pairwise distance matrix among observations as the clustering criteria. Thus, it is useful for identifying groups of similar observations in a dataset [205]. First, a Euclidean distance matrix is computed for all the observations: 130 individuals —district per year— described by the 40 indicators in the dataset. Then, we use Ward's hierarchical clustering method [218, 219] to advance in the construction of the cluster tree. Results from hierarchical clustering are shown in Dendrograms, which are a tree-based representation of the relationships between elements.

4.2.3.4 PCA-based measures

As described before, we take advantage of the uncorrelated description given by PCA and select the first three principal components to construct a three-dimensional space. Indeed, the description given by the first three principal components can be understood as a specific location in the three-dimensional space. The three-dimensional coordinates of the individuals are the input for the methods described below.

In stage 5 we assess the system's diversity by computing distances among individuals. Initially, Euclidean distances can be computed between the 130 individuals –district per year– and portrayed in a distances' matrix. Some patterns can be observed in this large matrix that is commonly referred to as the dissimilarity matrix. However, since we aim at analyzing similarity and dissimilarity between districts, we compute Euclidean distances only among the ten districts' centroids and construct a smaller dissimilarity matrix. Districts' centroids refer to the mean temporal positions of districts' individuals in the time span. Hence, by doing this process we reduce the dimensionality of the dissimilarity matrix to the single relationships between the 10 districts.

In the final stage, we use the statistical method of Confidence Ellipsoids over the time-dependent and three-dimensional coordinates of the districts. The 95% confidence intervals of the spatial distribution of the coordinates in each dimension compose a three-dimensional ellipsoid. The main idea is to compare the shape of the ellipsoid at different years in order to analyze the system's temporal change. To calculate the confidence intervals —and the dispersion matrix—, we first compute the variance-covariance matrix of the set of three-dimensional coordinates (distribution). The way to select which coordinates compose the observations of the statistical distribution is done by two different approaches.

One approach is to define the data points of all the 10 districts in a given year. The aim of this approach is to analyze the temporal change of the overall city from one year to another. This can be done by plotting the confidence ellipsoids for each yearly distribution and calculating the volumetric change of the generated ellipsoids between consecutive years.

The other approach is to use the data points of each district in the time span. With this type of distribution one can calculate the dispersion —and the ellipsoid— accounting for the temporal distribution of this given district, and to compare it with the other districts' distributions. But, an analysis of the temporal change of each district can not be easily done in this regard.

Instead, one can compose a subset constituted by the points per district in the first years (e.g. the first five years of the study). The confidence ellipse can be calculated for this initial distribution. From this initial set, we add only the consecutive observation (coordinate at the next year) and recalculate the confidence ellipsoid. This procedure is repeated until the whole time span is covered, such that bringing conclusions about the district's temporal changes becomes possible.

We implement both approaches. In both of them, the magnitude and direction of extension and contraction of the confidence ellipsoids bring clues about its temporal trends: larger confidence interval indicates a wider variety in the observations and hence, diversification. On the contrary, smaller ellipsoids indicate a stack in a specific location of the description and hence, homogenization. Also, in the first approach, the direction of the confidence intervals can be related to the principal components, and therefore, to the description that they provide to the system.

4.3 Results and discussion

In this section, we present the main results of the exploratory multivariate analysis applied to an urban system. Concretely, we analyze the exploratory statistical methodology in the case study of Barcelona, being understood as a composition of districts (subsystems). First, we present the components' description given by the PCA and the MFA, followed by the districts' performance in that reduced description. Then, results about the hierarchical relationship between districts are presented. Finally, we exhibit some results of the temporal trends given by the Confidence Ellipsoids calculation.

4.3.1 Components' description as given by the PCA and MFA

As described in the previous section, we calculate the PCA and MFA over the standardized correlation matrix. Table 4.4 shows the eigenvalues —which measure the amount of variation— and the decomposition of variance for the first 10 components of PCA and MFA. The four principal components of PCA account for 29.1%, 21.8%, 14.2%, and 9.9% of the variance, respectively. In the case of the MFA, the four principal components account for 28.9%, 23.5%, 13.4%, and 11.0% of the variance. Nonetheless, there is not a meaningful difference between the two analyses; comparatively, MFA explains a slightly greater variance with the first four principal components (76.8% against 74.9%).

As it was expected, the first principal components from PCA and MFA applied in the 10 districts explain a lower percentage of variance with respect to the analysis of the 4 districts in the previous chapter, where the first three principal components explain 87.7% of the total variance. This is by the fact that the dataset has increased its size and thus, its variance: the measurements now include 130 individuals accounting for all the districts.

In the case of the components' description provided by PCA and MFA results, the main findings are presented as follows. Table 4.5 shows the correlations between variables and components (loadings) given by PCA. Figure 4.4 displays the contribution of the most contributing variables —expressed in percentage— to the construction of each principal component. The larger the value of the contribution, the more the variable contributes to a component (defini-

tion). In that figure, the red dashed line indicates the expected average contribution. Hence, all the variables (larger than this cutoff) are considered to importantly contribute to the component. We also take into account the contribution of the groups of variables —SD dimensions— in the construction of each principal component. The group of variables' contribution is the main result of MFA, and it is presented in Table 4.6.

In the case of the First Principal Component (PC1), it explains 29.1% of the total variance. It is observed that the variables with high loadings on this component are in order of relevance —regarding of its contribution: En8, S3, E6, En16, En5, En15, S11, E2, En18, En4, En12, S7, En10, S12, S4. As a remark, four of the variables with the higher loadings in this component are linked to the residential subsystem of the city: En8-Residential energy consumption (0.94), S3-Residential area (0.93), E6-Available household income *per capita* (0.93), and En5-Total volume of water supplied to domestic consumption (0.86). This principal component is also contributed by indicators related to road activity: En16-Total number of motor vehicles (0.87), En17-Traffic accidents (0.79), and En18-Parking area (0.74). Also —to a lesser extent— this component is correlated with 5 indicators from the Social and community services category (S7, S6, S11, S12, and S13). Therefore, it can be said that the indicators in the Environmental and Social dimensions are contributing the most to the construction of this component. This interpretation is validated with the output from MFA, that take into account the groups from the information system. As shown in Table 4.6, the three SD dimensions have a similar contribution to the construction of PC1. Nevertheless, the Environmental, and Social dimensions have a slightly greater contribution (34.7% and 33.9%, respectively) than the Economic dimension (31.4%).

The Second Principal Component (PC2) accounts for 21.8% of the total variance. It is observed in Table 4.5 that the variables with high loadings on this component are in order of relevance: S10, S16, S9, E5, E4, En7, S8, S14, E4, and En14. Inspecting those variables, it is clear that the social and economic groups of variables mainly explain this component. This fact is also observed in the MFA output. Results in Table 4.6 demonstrate that the social dimension is contributing the most to the construction of the PC2 component (44.5%), followed by the economic dimension (32%). With respect to social indicators, this component is highly and positively correlated with; S10-Cultural centers (0.95), S16-Streets and zones with pedestrian priority (0.83), S9-Other land uses (0.82)(e.g., institutional and administrative buildings), and S8-Entertainment buildings (0.72). As a remark, the security level in the neighborhood (S14) is negatively correlated with this component (-0.68), and thus, with the previously mentioned indicators. On the other hand, the economic indicators highly correlated with this component are E5-Accommodation places (0.82), E4-Active market stalls (0.81), E2-Office buildings (0.48), and E1-Commercial area (0.47), all of them corresponding to the Businesses category. Therefore, this component is exposing a clear picture of a certain "type of city": it is displaying a cultural, administrative, touristic, commercial, and walkable Barcelona.

The Third Principal Component (PC3) explains 14.2% of the total variance. As shown in Table 4.5, variables with high loadings in this component in order of contribution are S6, S15, En10, S2, En11, S4, En1, En6, En14, and S9. It is observed that none of the indicators from the economic dimension have a meaningful correlation with this component. The SD contributions to this component are the following: the Social dimension accounts for 41.3%, and the Environmental dimension for 35.7%. On one side, some positively correlated variables with this component are: S6-Sport area (0.70), S15-Maintenance cost (0.68), S4-Educational build-

Table 4.4: Decomposition of variance per component of the PCA and MFA.

Component	Eigenvalue		Percentage of variance (%)		Cumulative percentage of variance (%)	
	PCA	MFA	PCA	MFA	PCA	MFA
1	11.63	2.51	29.0	28.94	29.07	28.94
2	8.70	2.04	21.76	23.45	50.82	52.39
3	5.68	1.16	14.20	13.39	65.02	65.78
4	3.95	0.95	9.87	10.95	74.90	76.73
5	2.32	0.51	5.79	5.91	80.69	82.64
6	1.95	0.41	4.87	4.70	85.56	87.34
7	1.45	0.28	3.62	3.19	89.18	90.53
8	1.00	0.18	2.51	2.04	91.69	92.57
9	0.71	0.12	1.78	1.39	93.46	93.96
10	0.56	0.11	1.39	1.29	94.86	95.25

ings (0.63), E1-Street trees (0.60), S9-Other land uses (0.82)(e.g., institutional and administrative buildings), and En2-Urban parks (0.49). On the other, negatively correlated variables are En10-Average levels of NO_2 (-0.65), S2-Population density (-0.64), En11-Average levels of CO (-0.64), and in a lesser extent, En9-Average levels of PM_{10} (-0.35). Thus, this component on one side describes; green spaces, and some social services and buildings; and on the other side, air pollution and population density —both directly correlated.

The Fourth Principal Component (PC4) accounts barely for 9.9% of the variance explained, thus it is difficult to relate it with the variables. As shown from correlations in Table 4.5 and contributions in Figure 4.4, some of the variables with meaningful contributions to this component are: E3-Industrial area (18.5%), En2-Urban parks (11%), En12-Collected volume of non-recyclable waste (8%), En9-Average levels of PM_{10} (7.5%), S7-Religious buildings (7%), and S5-Health service buildings (6.5%).

4.3.2 Districts' performances

Figures 4.5 to 4.7 show two-dimensional biplots, each one showing the individuals as described by the principal components. Specifically, we place at the axes of biplots the combination of two of the four principal components. Points (individuals) represent the districts in each year of the time period under analysis. The points are shown with a specific color and shape for each district. With the aid of the graphical display given by biplots, in the following, we detail the districts' performances.

Regarding PC1, this component is confronting two groups of districts: Eixample, Les Corts and Sarrià from one side, and Nou Barris, Sant Martí, Horta and Sant Andreu in the other, drawing a clear division between high-income and low-income-working-class districts. Eixample, Les Corts, and Sarrià have an available household income per capita index of 119.3, 136, and 182.4, respectively (see the E6 indicator in the raw datasets), and an average of 149.9. In contrast, Nou Barris, Sant Martí, Horta and Sant Andreu, districts on the east side of the city, have an available household income per capita index of 55, 87.1, 79.2 and 74.5, respectively, and a mean of 73.9 —being less than half the income of the high-income group of districts.

The high-income group —in the right side of Figure 4.5— is, on one hand, directly (+) correlated with available household income per capita (E6) and with housing-related indicators:

Table 4.5: Component loadings for the four principal components.

SD	Id	Indicator	PC1	PC2	PC3	PC4
Economic	E1	Commercial area	0.55	0.47	0.18	-0.37
	E2	Office buildings	0.75	0.48	0.18	0.07
	E3	Industrial area	-0.04	0.10	0.33	-0.87
	E4	Active market stalls	-0.16	0.81	-0.29	0.34
	E5	Accommodation places	0.20	0.81	0.36	0.26
	E6	Available household income per capita	0.92	-0.24	0.04	0.16
Environmental	En1	Street trees (Unit = 5.6 m^2)	0.12	-0.46	0.60	-0.06
	En2	Urban parks	-0.09	0.18	0.49	-0.67
	En3	Urban green area besides parks	-0.15	-0.20	0.15	0.21
	En4	Total volume of water supplied to the commerce	0.71	0.63	-0.12	0.06
	En5	Total volume of water supplied to domestic consumption	0.86	-0.01	-0.23	-0.13
	En6	Total volume of water supplied to the industry	0.41	0.60	0.59	-0.22
	En7	Total volume of water supplied to other usages	-0.04	0.76	0.36	0.07
	En8	Residential energy consumption	0.94	-0.01	-0.11	0.09
	En9	Average levels of PM_{10}	0.27	0.29	-0.35	-0.56
	En10	Average levels of NO_2	0.59	0.06	-0.65	-0.05
	En11	Average levels of CO	0.53	0.07	-0.64	-0.16
	En12	Collected volume of non-recyclable waste	0.70	-0.25	-0.02	-0.58
	En13	Collected volume of paper and cardboard	0.37	-0.41	0.23	-0.15
	En14	Collected volume of glass	0.30	-0.61	0.50	0.24
	En15	Collected volume of containers	0.55	-0.51	0.33	0.24
	En16	Total number of motor vehicles	0.87	-0.40	0.15	-0.02
	En17	Traffic accidents	0.79	0.23	-0.07	-0.18
	En18	Parking area	0.74	-0.25	0.46	-0.16
Social	S1	Natural growth rate	-0.26	-0.14	0.30	-0.12
	S2	Population density	-0.15	0.28	-0.64	0.20
	S3	Residential area	0.93	-0.04	-0.15	0.20
	S4	Educational buildings	0.58	0.06	0.63	0.37
	S5	Health service buildings	0.56	-0.27	-0.19	0.50
	S6	Sport area	0.42	-0.26	0.70	0.07
	S7	Religious buildings	0.66	0.27	-0.07	0.52
	S8	Entertainment buildings	0.19	0.72	0.14	-0.30
	S9	Other land uses	0.14	0.82	0.50	0.00
	S10	Cultural centres	-0.08	0.95	0.06	0.07
	S11	Sport centres	-0.76	0.02	0.29	0.33
	S12	Public water fountains	-0.59	-0.25	0.30	-0.16
	S13	Playgrounds	-0.53	-0.57	0.34	0.05
	S14	Security level in the neighbourhood	0.54	-0.68	-0.10	0.25
	S15	Maintenance cost	-0.06	0.27	0.68	0.46
	S16	Streets and zones with pedestrian priority	-0.22	0.83	0.14	0.37

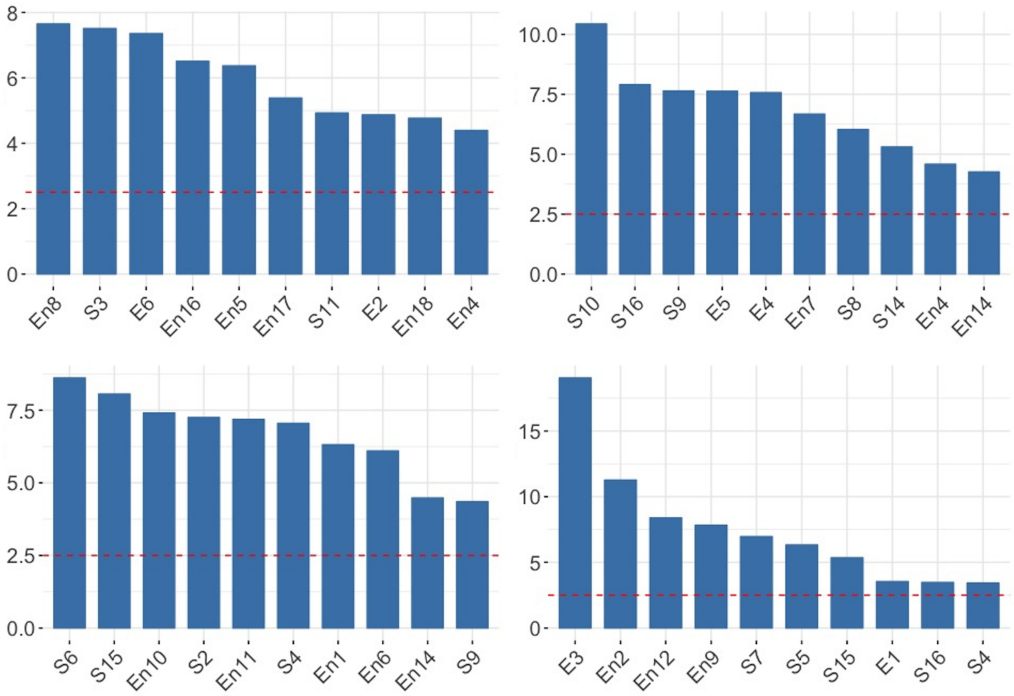

Figure 4.4: Contribution of variables to Principal Components. PC1 (top-left), PC2 (top-right), PC3 (bottom-left), and PC4 (bottom-right).

Table 4.6: Contributions of groups of variables to Principal Components.

SD Dimension (Groups)	Contribution (%)			
	PC1	PC2	PC3	PC4
Social	33.88	44.87	41.29	39.71
Economic	31.40	32.03	23.03	34.73
Environmental	34.72	23.09	35.68	25.55

residential energy consumption (En8), residential area (S3), and residential water supplied (En5). Besides, the high-income group is positively correlated with car-related indicators: total number of motor vehicles (En16), traffic accidents (En17), and parking area (En18). These indicators are highly and positively correlated between them. Notably, available income is directly correlated with residential area and number of vehicles, and therefore, with residential water and energy consumption, traffic accidents, and parking area. Given that indicators are measured in per capita values, this means that a person from one of these neighborhoods has a higher income with a larger dwelling —a greater surface to live in—, and a higher water and energy consumption. Positively correlations between high income, large housing areas, and domestic electricity consumption have been found in previous urban studies [220, 221].

Moreover, districts in the low-income group are inversely (-) correlated with indicators describing high-income districts, and directly (+) correlated with three indicators in the social and community services' category: sport centers (S11), public water fountains (S12), and playgrounds (S13). This trend can be explained because some urban policies have been implemented in those neighborhoods, which were intended to improve community services — social SD dimension. Other relationships can be expanded by looking at the raw values in the datasets. For example, inspecting the districts at the extreme years of the study: in 2015 a person from Sarrià - Sant Gervasi had an income index of 188, lived in 57.4 m^2, consumed 4.61 MWh/year of energy and 45.2 m^3 of water. In contrast, a person from Nou Barris had an income index of 53.8, lived in 30.4 m^2, consumed 2.41 MWh/year of energy and 32.5 m^3 of water. Therefore, results from the exploratory multivariate analysis reveal that there are many types of "cities" with many types of citizens. Despite urban planning actions to increase social and environmental resilience, economic dimension reveals —to some extent— urban inequalities.

Furthermore, results clearly show that Ciutat Vella exhibits a distinctive performance in the time-span. This behavior is observed more accurately along with the PC2 description. As shown in Figures 4.5, 4.6 and 4.7, Ciutat Vella is notably separated from the 9 remaining districts. This district shows a particular behavior and performs as an outlier of the urban system. In particular, Ciutat Vella is mainly described by the following indicators: pedestrian streets and zones (S16), cultural centers (S10), active market stalls (E4), other land uses (S9) (e.g., institutional and administrative), accommodation places (E5), and entertainment buildings (S8). These indicators coincide with the ones largely contributing to the construction of PC2, which displays a certain "type of city". In fact, Ciutat Vella is displaying a clear functional typology: it is a cultural, touristic, administrative and walkable district, providing services for the whole city. Two facts can be outlined in this regard. The first one is that the specific urban plan of cultural enhancement in the Raval neighborhood has been effective [194]. Yet, the effect of this urban planning has failed at improving its social component [199]. The second is that the promotion of Barcelona as a tourist destination has focused mostly in this district [199]. The question is whether this district offers its cultural services for the citizens or for the tourists.

Regarding PC3, another differentiation is being displayed: Les Corts on one side, and Gràcia and Eixample on another. These groups exhibit an opposite behavior, agreeing with previous results from the PCA applied to the four districts presented in Chapter 3. Eixample and Gràcia are directly (+) correlated with air pollution and population density: average levels of NO_2 (En10), average levels of CO (En11), and population density (S2). As seen in Figure 4.3, these indicators are positively correlated among them. Some previous studies

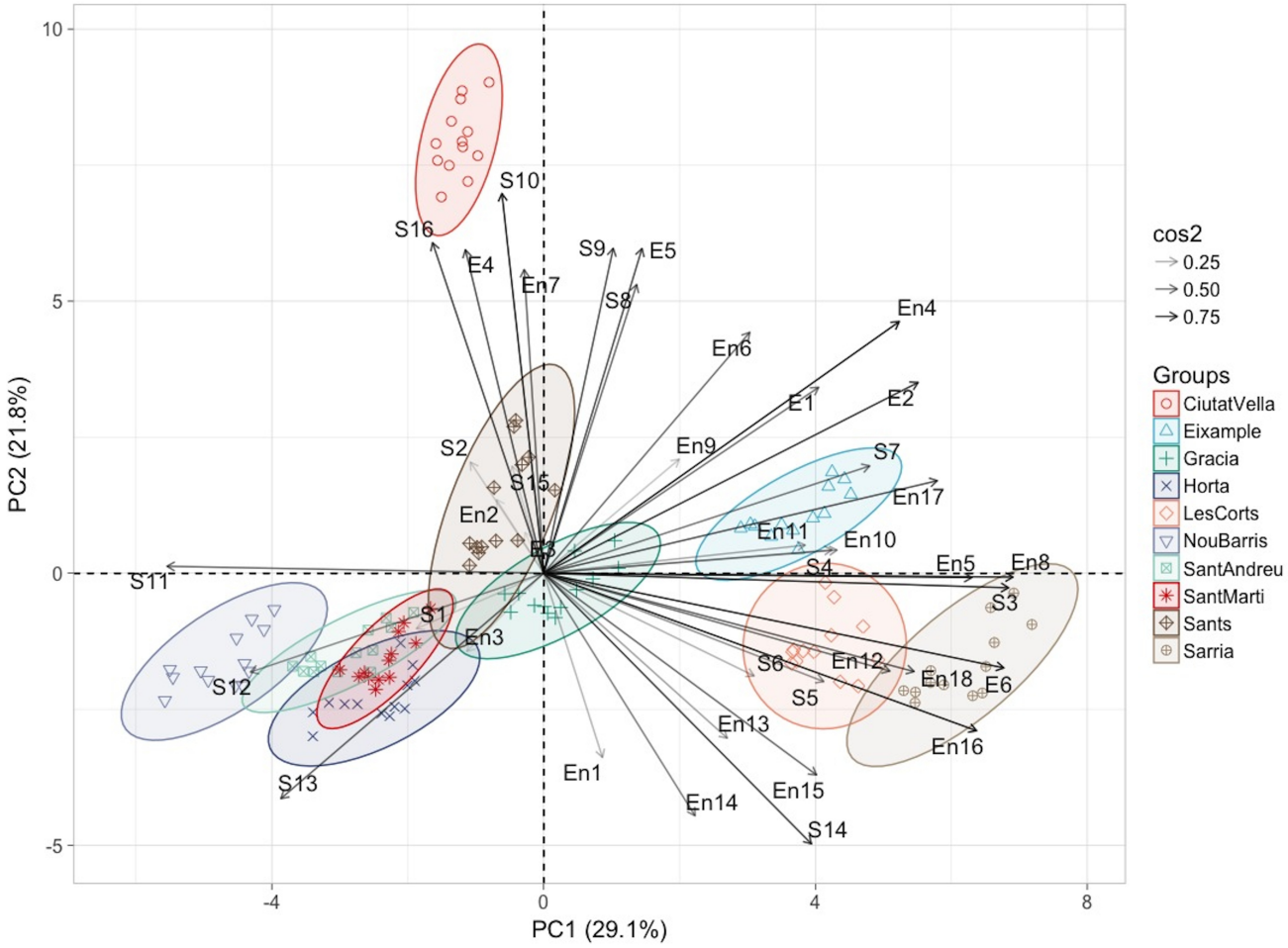

Figure 4.5: Biplot of individuals and variables using the first and second principal components.

have found also a positive correlation between high population density and air pollution [222–224]. In this sense, it must be highlighted that Eixample has a large population density (35617 people/km^2) —even greater than Manhattan's population density (27544 people/km^2) [225]. Thus, air pollution is expected to be high. Moreover, Les Corts is inversely (-) correlated with the indicators describing Eixample and Gràcia, and directly correlated with social indicators such as sports area (S6), educational buildings (S4), maintenance cost (S15), and street trees (En1). A positive performance in S4 is explained by the fact that Les Corts comprise within its territory two of the largest university campuses in the city: the University of Barcelona and the UPC-BarcelonaTech.

4.3.3 Urban diversity

Following the previous explanation, districts are found to exhibit particular typologies and exhibit —to some extent— urban diversity. To complement the exploratory multivariate methodology regarding the system's diversity, we also calculate a hierarchical clustering analysis and the distance matrix between PCA individuals.

First, we present the HCA results in Dendrograms (Figures 4.9 and 4.8) that display hier-

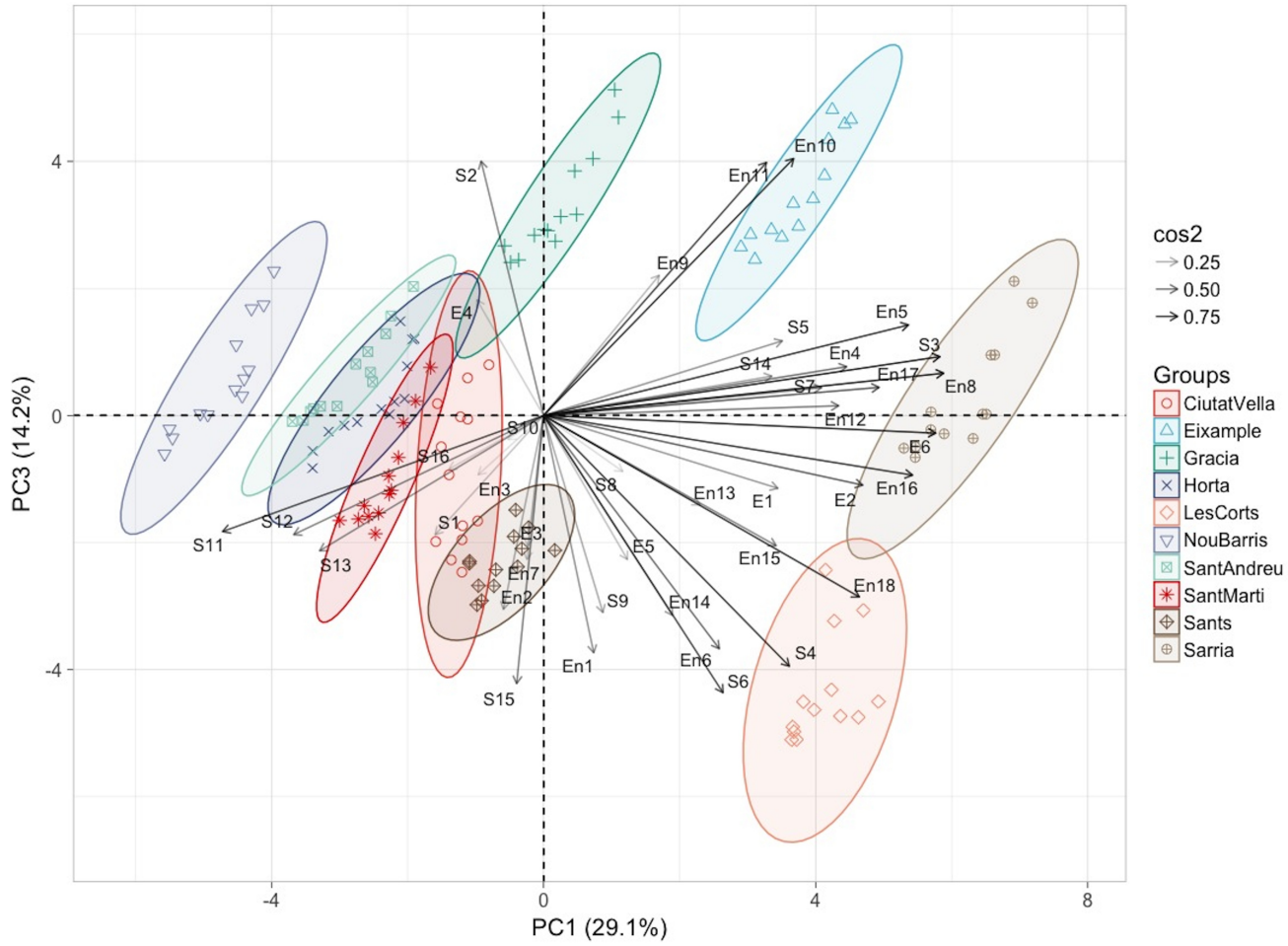

Figure 4.6: Biplot of individuals and variables using the second and third principal components.

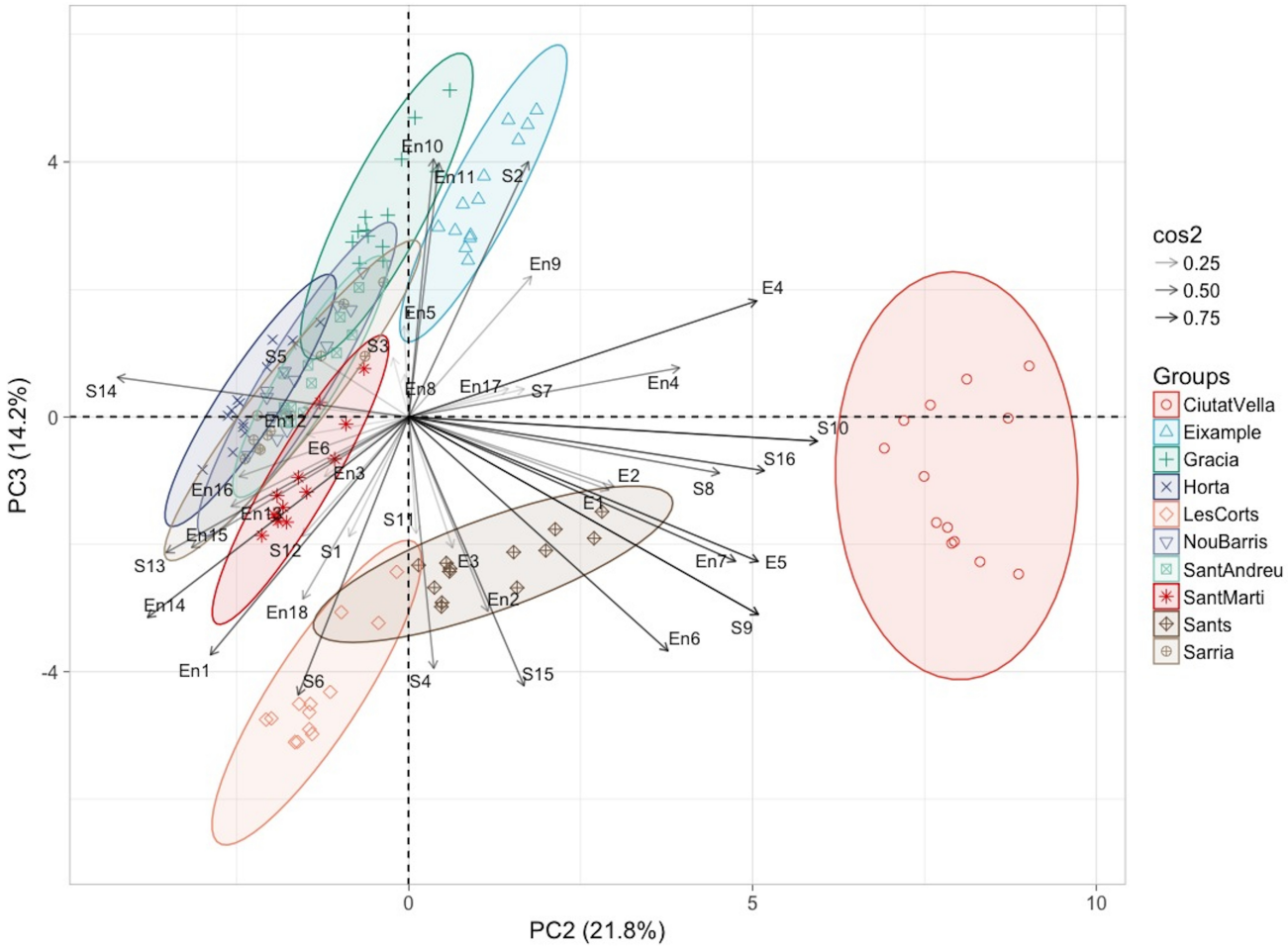

Figure 4.7: Biplot of individuals and variables using the second and third principal components.

archical links between the individuals (districts per year). Even though the HCA is applied to the standardized multivariate dataset (composed by the 130 observations —districts in a given year— of the 40 indicators), the results validate the diversity patterns displayed by PCA and MFA. We introduce the following notation for the districts in the plots: C is Ciutat Vella, E is Eixample, G is Gràcia, H is Horta, L is Les Corts, N is Nou Barris, A is Sant Andreu, M is Sant Martí, S is Sants-Montjuïc, and R is Sarrià. In the first analysis, the HCA is set to display five major clusters —groups of districts—, which gives: (1) Gràcia and Eixample, (2) Sarrià-Sant Gervasi and Les Corts, (3) Ciutat Vella, (4) Sants-Montjuïc, and (5) Horta, Sant Martí, Nou Barri, and Sant Andreu. The graphical representation of these five clusters —from left to right— is displayed in Figure 4.8. Although PCA accounts for a reduced variance than HCA, both findings agree well: the first group —Gràcia and Eixample— has been pictured by PC3; the third group —outlying Ciutat Vella— has been displayed by PC2, and the fifth group —low-income districts— has been portrayed by PC1. The first group —Gràcia and Eixample— has also been found in the exploratory multivariate statistical analysis of the reduced —four districts— case study in Chapter 3.

When the HCA is set to display only three major clusters, Ciutat Vella remains an isolated cluster. The second cluster comprises four districts: Gràcia, Eixample, Sarrià, and Les Corts. And the third cluster includes five districts: Sants-Montjuïc, Horta, Sant Martí, Nou Barris, and Sant Andreu. These results are displayed in Figure 4.9, where a circular Dendrogram that uses the graphic space more efficiently to display the 130 individuals is depicted. Agreeing with PCA results, HCA groups high-income districts from one side, an atypical district in between, and low-income districts on the other side. Despite from the differences portrayed by the PC1 in Figures 4.5 and 4.6 (income, residential area, and flow consumptions), there are two main socio-economic features that distinguish the two segregated groups: the available household income per capita and the percentage of people with higher education. In the high-income group, the average income per capita is 136, and 43.5% of the population has a higher education diploma —bachelor, master, or Ph.D. In the low-income group, the average income index is 75, and just 22.1% of the population has a higher education diploma. A positive correlation between higher education levels and higher incomes has been reported since the 1970s in [226], and as a direct consequence, well being [227]. The remaining sociodemographic variables do not segregate both groups: the percentage of foreigners is reported to be the same (15%), and the composition of the population by groups of age is also similar.

Moreover, the HCA is complemented by calculating the distance matrix between the districts' centroids given by PCA. This matrix of Euclidean distances is displayed in Figure 4.10 and portrays similarities and dissimilarities between each pair of districts. Particularly, the distance matrix shows similarities —in blue— and dissimilarities —in red— among districts' centroids. As shown in that figure, the most similar —closest— districts are Horta and Sant Andreu, with a distance among their centroids of 1.0 unit. This distance is followed by; Horta and Sant Martí (1.5 units); Sant Andreu and Sant Martí (1.7 units); Sant Andreu and Nou Barris (1.9); Nou Barris and Horta (2.3); Sant Martí and Nou Barris (2.9); Sants-Montjuïc and Sant Martí (3.6); and Gràcia and Eixample (3.8). In the case of Gràcia and Eixample districts' centroids, those appear closer than in the 4-districts' exploratory analysis (with 7.19 units). On the other side, the most dissimilar —furthest— districts are Ciutat Vella and Sarrià, separated by 12.3 units of distance, followed by; Les Corts and Ciutat Vella separated by 11.3 units —and matching the distance obtained in the four-districts case study—; Nou Barris and Sarrià

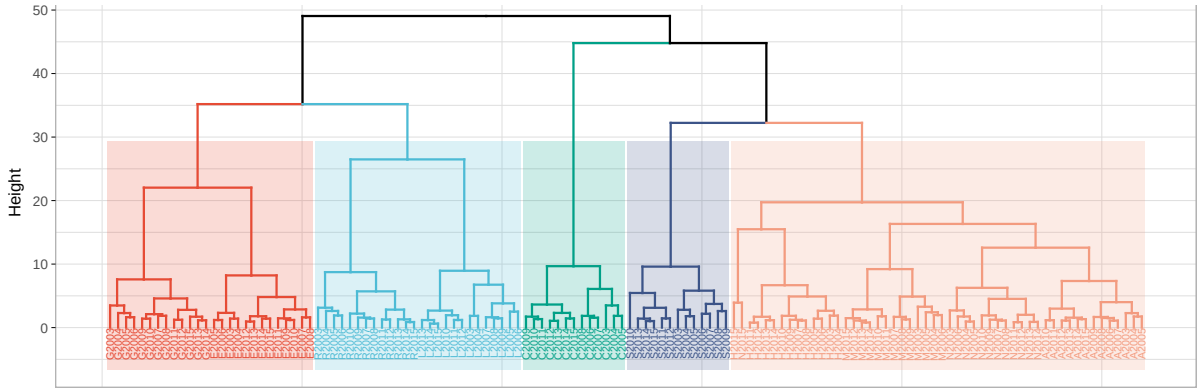

Figure 4.8: Rectangular dendogram for the five clusters given by HCA.

(10.9 units apart); and Horta and Ciutat Vella (10.4 units apart). Even though HCA reveals some clustered districts, unquestionably, Nou Barris, Sant Martí, Horta and Sant Andreu are exceptionally clustered, making up a homogeneous continuum in the city. This is explained by the fact that the four districts in this cluster perform quite similar in the multivariate description throughout the time span.

4.3.4 Temporal change analysis of the urban system using Confidence Ellipsoids

Since PCA and MFA results are temporal distributions of data points for each district, analyzing the changes of these points is hard. As explained in the previous section, our first approach here is to use Confidence Ellipsoids to quantify the dispersion of the time-dependent results. The results of applying these methods to the case study are explained in the following lines.

4.3.4.1 Temporal change analysis of the city

In the case of the temporal change analysis of the overall city, we calculate the 95% confidence intervals of the dispersion matrix for all the districts of the city in a given year. Those confidence intervals in each spatial direction define the ellipsoid's axes. The confidence ellipsoids for the initial and final years of the time span are depicted in Figure 4.11. This graph, in brief, provides us a tool to evaluate the change of the city with respect to the coordinate's axes (principal components), including the deformation trend of the distribution and its direction of change.

One analysis that can be derived from Figure 4.11 is the indication of the magnitude and direction of extension and contraction in the district's distribution: larger confidence interval indicates a wider variety in the observations. For example, the increment in the ellipsoid's size can be associated with a diversification pattern. Certainly, the shape of the ellipsoid vary from the first to the last time-step: we observe extension in both the PC1 and PC2 axes, mostly located between clustered districts. We can also detail an overall displacement of the initial ellipsoid towards negative values of PC1 and PC2, simultaneously with positive displacement towards PC3. These directions of change can be linked with the description of the system made by each principal component.

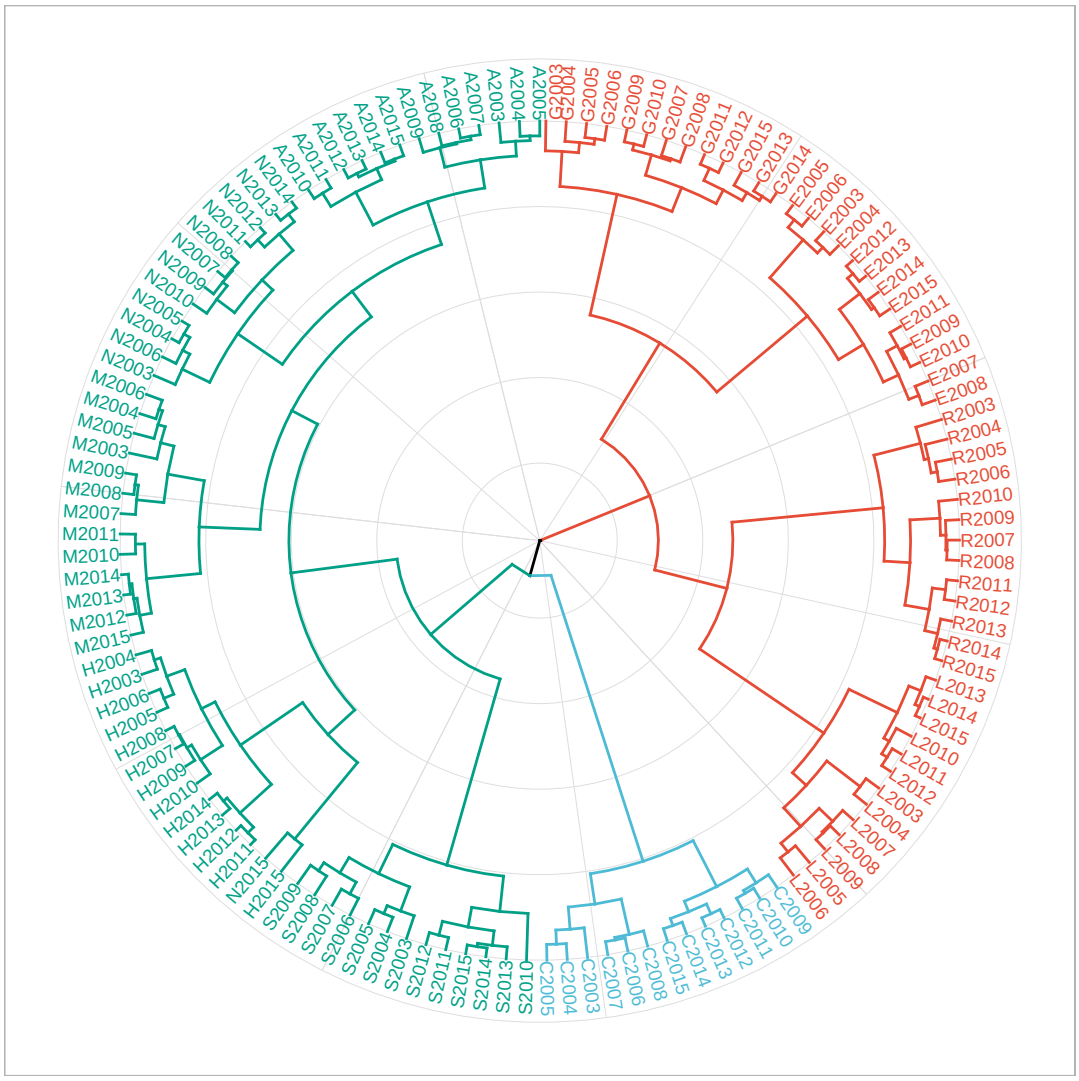

Figure 4.9: Circular dendogram for the three clusters given by HCA.

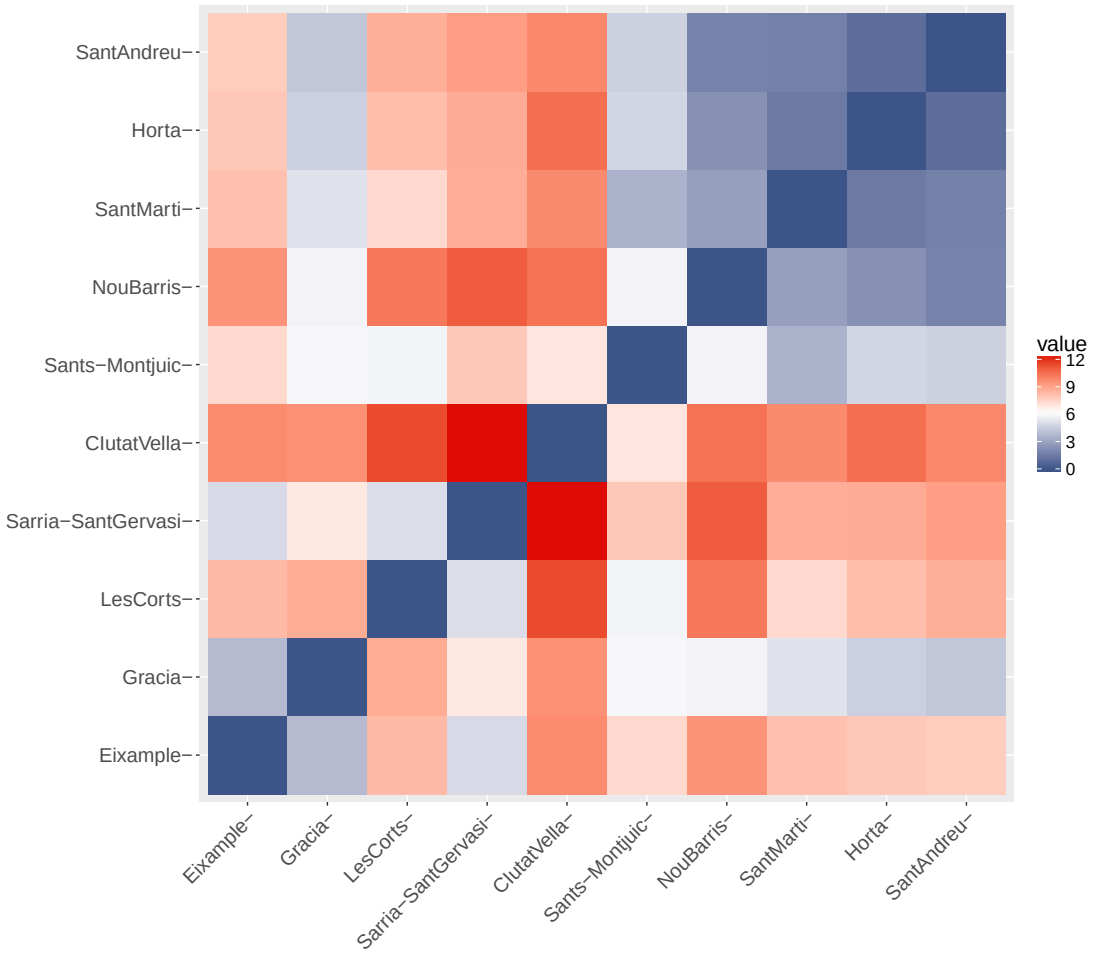

Figure 4.10: Distance Matrix.

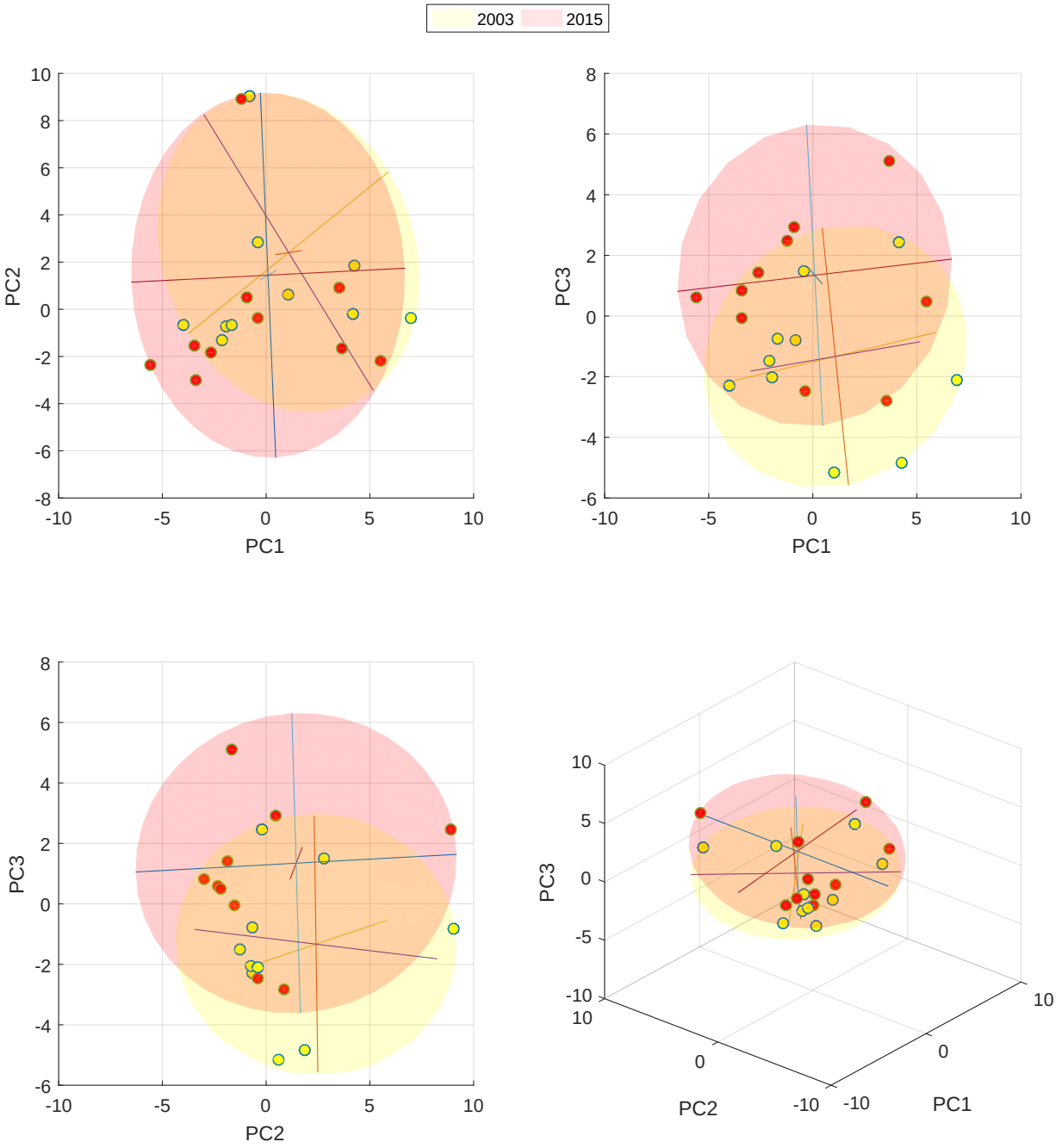

Figure 4.11: Confidence ellipsoids of the overall city between the initial and final years of the time span.

4.3.4.2 Temporal change analysis of the districts

Another option to analyze the temporal change of the urban system is to calculate the 95% confidence intervals of the dispersion matrix for the temporal distribution of each district. It has been explained before that calculating the confidence ellipsoid using all the years in the distribution can be unproductive. In that case, only a few comparisons between the ellipsoids' sizes of districts can be made. Instead, one can define a basic subset that is constituted by the distribution of points for each district during the first five years —from 2003 to 2008. The confidence ellipsoids over this initial distribution are depicted in Figure 4.12. The remaining points of the following years are also presented in the figure, where certainly they stand aside from the confidence intervals (ellipsoids). From the initial set, we add only the next observation (district at the next year) and recalculate the confidence ellipsoid. This procedure is repeated until the whole time span is covered.

In Figures 4.13 and 4.14 we show the resulting confidence ellipsoids for each district and each incremental year. In those figures, we have organized the ellipsoids in such a way that they are aligned over their r1, r2, and r3 confidence intervals (axes). The first observation that we do is that the sizes of the ellipsoids vary greatly between the districts. There are some districts with small sizes of ellipsoids such as Horta, Nou Barris, and San Martí, that indicate a homogeneity of the data for the district and hence, a negligible variation. On the other hand, there are districts with large sizes of ellipsoids such as Ciutat Vella, Les Corts, Sants-Montjuïc, and Sarrià, that indicate a great dispersion of the data in the study period. Also, meaning large temporal variations of those districts.

We can deepen the analysis of temporal change through the recognition of the sizes' change of the ellipsoids from one year to another. We observe that Horta, Nou Barris, and Eixample, are the districts with the most abrupt change in the eight consecutive years —from 2008 to 2015. The districts of Gràcia and San Martí involve an appreciable temporal change. The remaining districts, on the other hand, show a stable behavior without great variations during the study. Thanks to the graphical arrange of the ellipsoids, we can verify that the major temporal changes occur with respect to their main axis (r1), and smaller changes occur in r2 and r3 axes. However, this was expected since the r1 axis represents the greatest variance in the distribution. The drawback of this approach is that we can not interpret what is the direction of change regarding the principal components of the PCA since we have arranged the ellipsoids based on their axes. Note the difference between the arrangement of the original ellipsoids in Figure 4.12 and the organized arrangement of Figures 4.13 and 4.14. Also, the way to recalculate the ellipsoids per year could consider —or not— the observations from initial years. In this regard, several techniques can be adopted as well.

4.4 Conclusions

In this chapter, we have applied exploratory multivariate statistical techniques to a multivariate dataset that represents the urban system of Barcelona. First, we have built an abstraction of the city of Barcelona by collecting measurements of 40 indicators for the 10 districts of the city during the time span between 2003 and 2015. The indicators have been grouped in the economic, environmental, and social dimensions of sustainable development. Our approach has been to illustrate the complex relationships between the different elements that compose the ur-

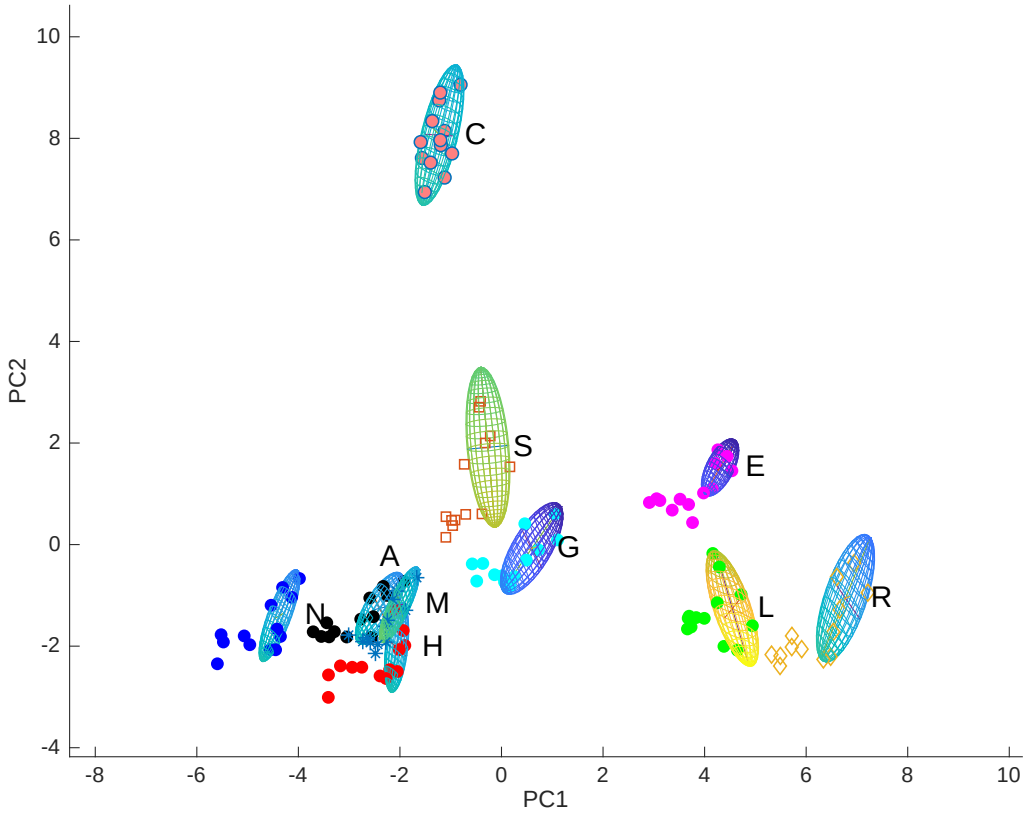

Figure 4.12: Confidence ellipsoids of the districts' initial subset.

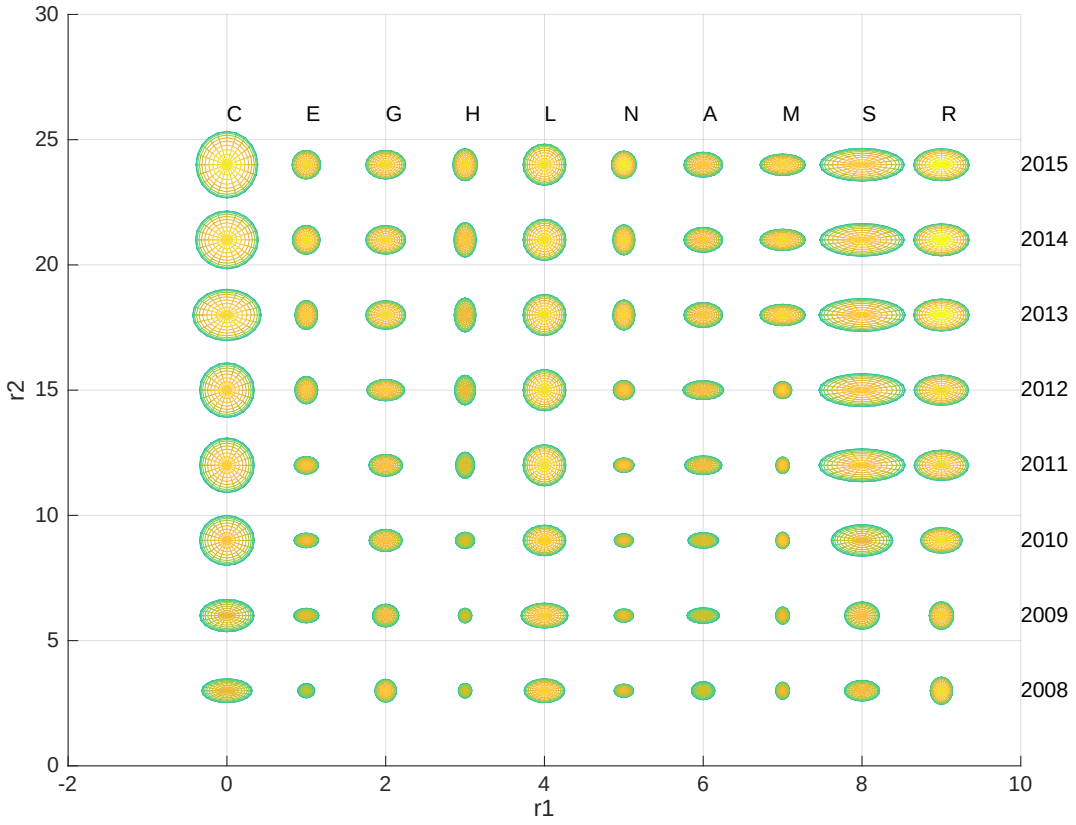

Figure 4.13: Confidence ellipsoids of the districts' subsets between 2008 and 2015. Ellipsoids are arranged and displayed with respect to r1 and r2 ellipsoids' radii.

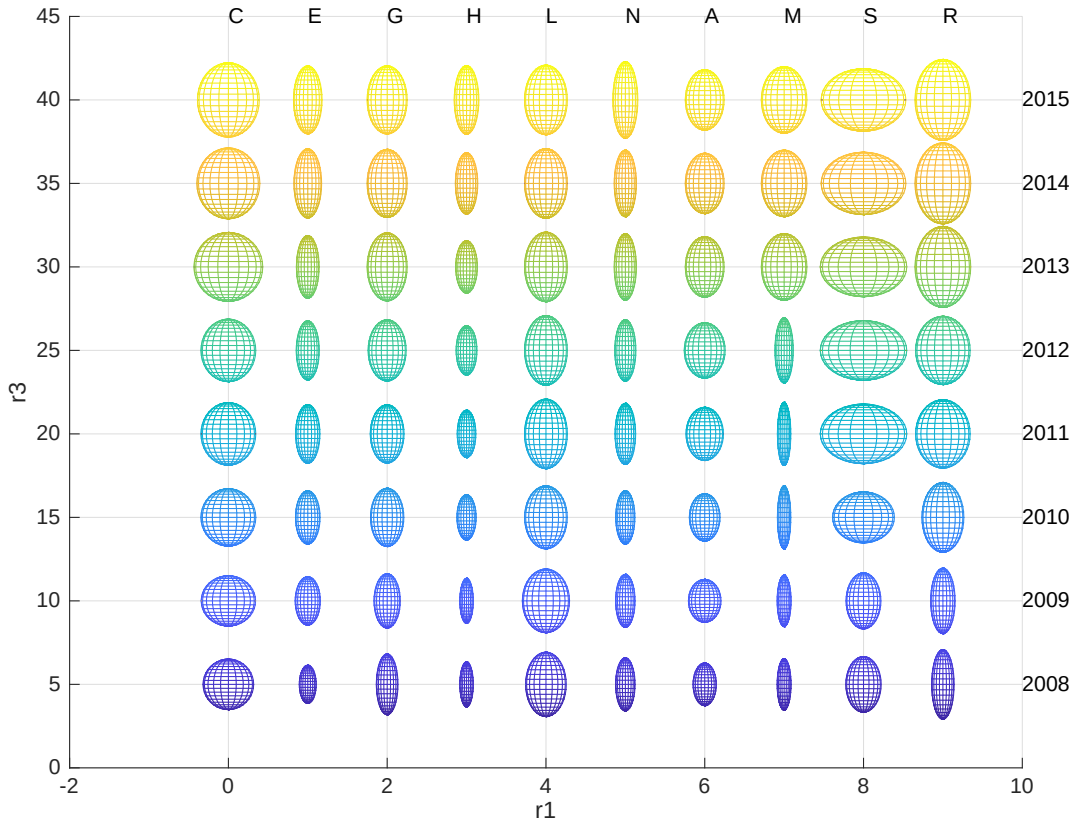

Figure 4.14: Confidence ellipsoids of the districts' subsets between 2008 and 2015. Ellipsoids are arranged and displayed with respect to r1 and r3 ellipsoids' radii.

ban system. This has been attempted based on novel developments in the field of exploratory multivariate statistical analysis (i.e, PCA, MFA and HCA). Using the reduced-dimensional description given by PCA and MFA, together with the clustering patterns of HCA, we have identified at least three distinct groups of districts: one that is characterized by high income and large housing surfaces, one atypical district (Ciutat Vella), and another with low income and small housing surfaces. The correlations presented by these groups of districts with the environmental and socio-economic variables is not simple. On the contrary, describing such correlations requires statistical support and a verification process with the external causes.

The case of the Eixample, clustered in the high-income group, exemplifies this complex relationship. Eixample was originally designed to respond to the historic center's hyper-densification, but, nowadays, it has a high density, which is not bad: it is the effect of good urban planning and a compact urban form. However, this district suffers from a lack of green areas and pollution and traffic problems associated with transportation. One possible solution to these problems can be the re-conversion of the blocks' centers into green areas. In addition, due to the lack of space, the planting of green roofs has proved to be efficient in reducing air pollution in urban environments [195].

The opposite type of districts, Nou Barris, Sant Martí, Horta, and Sant Andreu, are characterized as low to middle-class neighborhoods where the working population settles. This predominant character of the districts makes them group when all the three different exploratory methodologies are applied. Also, the correlation between income and education levels has been depicted, suggesting that the improvement of public education would decrease income-inequities. This indeed, would improve the social dimension of the city, but also raises some other ethical concerns. For example, having identified that segregation between districts, not so clearly highlighted when looking at the raw data, there is evidence that the behavior of the citizen is different depending on the district where he lives, and so his privileges.

Ciutat Vella has been found to be an isolated type of district, being at the same time the historic core of the city and the most important cultural, recreational, and touristic part of the city. We have emphasized the inequality that exists inside the district with respect to the urban policies, where the effects have been both positive and negative [199].

In this chapter, we have also attempted some approaches to the analysis of the temporal change of the city. To do so, we have computed the confidence intervals for the three-dimensional distribution of data points from PCA, and have drawn ellipsoids using these intervals. The analysis lies in the evaluation of the geometrical characteristics of the ellipsoids, especially their volume, to which we can associate the variance of the distribution. One first approach has been to calculate the ellipsoid for the distribution of all the districts of the city in a certain year. Hence, we have read the temporal change between the initial and the final year of the study mainly by inspecting the ellipsoids' size and direction of change. It is precisely this characteristic, to be able to relate the changes in the city directly with the main components of the PCA, one of its advantages. We have found extension patterns in both the PC1 and PC2 axes, mostly located between clustered districts, for the case study. The second approach has been to compute the confidence ellipsoids for the distributions per district. Specifically, we have proposed to compute several consecutive ellipsoids from subsets of the temporal distribution. In this sense, we have been able to evaluate the change of the size of the districts' ellipsoids in consecutive years, as well as the differences between the districts. But the main drawback of this second approach is the impossibility of linking the direction of change to

the main components of the PCA since we have proposed a rearrangement of the consecutive confidence ellipsoids in the graphs. Also, that ether approach can not locate the most dynamic variations between the districts. Even so, we believe that these approaches are technical and systematic, and serve as a starting point for the development of characterization techniques for large volumes of time-dependent data.

Chapter 5

Visualization of the strain-rate state of a data cloud: Analysis of the temporal change of an urban multivariate description

One challenging problem is the representation of three-dimensional datasets that vary with time. These datasets can be though as a cloud of points that gradually deforms. But point-wise variations lack of information about the overall deformation pattern, and more importantly, about the extreme deformation locations inside the cloud. The present chapter applies a technique in computational mechanics to derive the strain-rate state of a time-dependent and three-dimensional data distribution, by which one can characterize its main trends of shift. Indeed, the tensorial analysis methodology is able to determine the global deformation rates in the entire dataset. With the use of this technique, one can characterize the significant fluctuations in a reduced multivariate description of an urban system and identify the possible causes of those changes: calculating the strain-rate state of a PCA-based multivariate description of an urban system, we are able to describe the clustering and divergence patterns between the districts of the city and to characterize the temporal rate in which those variations happen.

5.1 Introduction

One challenging problem in the data analysis is the representation of three-dimensional discrete data [121]. This analysis becomes harder when a time-dependent change of the data takes place, introducing a new temporal dimension. In the present chapter, we explore formal approaches to quantify the temporal change of discrete three-dimensional data. Specifically, we build a methodology to assess the transformation of a data cloud that is derived from a *Principal Component Analysis* (PCA): a 13-years span multivariate description given in Chapter 4 that provides a reduced description of an urban system given only by the first three principal components. Since the points represent an abstraction of an urban system, one main goal is to understand the temporal variation of the multivariate description of the districts in order to analyze the behavior of the overall city in the time-span. Our main hypothesis is that these

three-dimensional datasets can be though as a cloud of points that gradually deforms.

Still, the challenging issue is that deformation between consecutive times cannot be visualized accurately. There are some methods to overcome this difficulty. One is the vector plot of three-dimensional displacements or velocities, that is typically used to visualize results in Computational Mechanics applications [228–230]. In example, these are used in the Kinematic Visualization of Motion [114–118], but they are restricted to display the relative motion among the data, and are not able to identify the most dynamic regions of the dataset. Another is the Parallel Coordinate Technique that successfully exhibits the temporal change of highly-dimensional statistical and information datasets [231]. Yet, multi-dimensional data is mostly segmented into two-dimensional subsets, which are easier to deal with. Like the Computer Tomographic scans of medical imaging [119–121]. But all these methods are not suitable for understanding the patterns of diversification or conformation, which are closely related to the temporal change of the differences inbetween join data values: the maximum and minimum magnitudes of variation and the evaluation of their direction can be significantly helpful to identify differentiation patterns in the data [232]. Or the opposite, to locate uniformity for a dataset which was previously differentiated.

One common approach to understanding the temporal change of the dataset is to use *Confidence Ellipses* in dimensionally-reduced representations (i.e. PCA) [233–235]. This type of approach has been applied to several time-dependent problems, such as the ones in [110, 111, 236, 237]. The Confidence Ellipses quantify the dispersion of the results: the main directions and intervals of variance accounting for the spatio-temporal distribution are calculated first, and then, the Ellipses (or Ellipsoids in 3D) are drawn using those intervals as the main axes. A temporal analysis can be achieved by evaluating the geometrical characteristics of the ellipses, especially their area, to which the change of the distribution can be associated. Its main drawback lies in the failure to locate the most dynamic areas in the dataset [238], as demonstrated in the previous chapter.

The field of continuum mechanics provides a measure of the temporal variation of the distance in between points: *the Strain-Rate tensor* (see, for instance, [128]). The continuum mechanics theory —which arises from the classical Newtonian mechanics— analyzes the causes and effects of motion for a deformable media composed by an infinite group of particles. When a continuous media is being deformed in various directions at different rates, the strain-rate of a certain position in the medium cannot be expressed by a scalar value solely. It cannot even be expressed by using a single vector. Instead, the rate of deformation must be expressed by the rank-two strain-rate tensor with its components determined by the positional derivatives along each spatial dimension. Hence, the mathematical framework of tensors can determine exactly the deformation that is accumulated in a location inside the medium —that is typically subject to the imposition of displacements or loads. This tensor is commonly used to detail the amount of elastic energy in the physical descriptions of multiple materials, like solids or fluids. See [239] for a complete mathematical exposition. Most of those models are formulated as the product of a constitutive tensor and the strain-rate tensor, giving the stress condition of the material that is balanced in the kinetic equations. In the present study, the calculation of the strain-rate tensor is not related to the kinetics of any material, and thus, it can only be a mathematical tool that supports the examination of the deformation rates in the discrete statistical data.

Still, the strain-rate tensor arises from the continuum assumption, and discrete displace-

ments of points rather than continuous distributions take place in the variation of the data cloud. Typically, the issue of applying derivatives to discrete displacements of points is solved by several different approaches. Few statistical techniques use co-variance functions to represent directly the strain-rate field (see e.g. [240]). The common approach is to compute a continuous version of the displacement —or velocity— field, so that, derivatives can be applied to the continuous displacements. Some methods, in this line, have interpolated the discrete displacements by minimizing the residual —or distance— between the continuous interpolation and the discrete version [241]. Other interpolation techniques weight the distance between an interpolated piece-wise continuous field and the discrete displacement field, as in [242]. This method results in a minimization technique where a continuous strain-rate field can be derived. In example, the piece-wise continuous field can be defined as to be splines, or as the widely used linear polynomials in variational formulations [243]. These techniques have been applied in earth science and medical imaging works [129, 130, 244], but also in the strain-rate calculation of geodetic observations in [131, 132].

Another fundamental issue is the representation of the strain-rate state. One of the possible techniques that can help to visualize the deformation rate of the dataset is to plot the main components of the tensor using *Strain-rate diagrams*, where concentrations of strain-rate patterns can be displayed as vector fields (see for example the ones in geodetical observations of the earth's mantle [245–247]). The main drawback of strain-rate diagrams is that the strain-rate components are visualized as the projection of three-dimensional vector fields into the two-dimensional framework, and therefore, the third-dimension component has to be necessarily neglected. Another method, more suitable to two-dimensional plots, is the contour graph of principal stresses, where the stress patterns in structural elements [248, 249] and tectonics [250] are visualized using continuous lines that depend on the stress magnitude. That method overcomes the three-dimensional issue, but it does not give insights about the orientation of the principal stresses. A dual form of the contour graph is to calculate the family of curves that are instantaneously tangent to the *extension* and *contraction* components of the strain-rate tensor: the so-called *trajectory* curves in the continuum mechanics field [128]. The trajectories of each principal component of the stress can be depicted in a separated plot with the stress magnitude colored along the trajectory line such that stress patterns are completely shown in a two-dimensional framework.

Since a robust methodology that describes the temporal change of the urban system —represented by a multivariate dataset— has not been carried out before, we choose to perform a quantitative analysis by including the strain-rate tensor as the fundamental metric. In this work, we calculate the strain-rate state of the discrete dataset without *a priori* assuming the mechanisms by which the system experiences transformation. In order to apply the continuum mechanics principles into the discrete dataset, we use interpolation methods, such as the ones applied in discrete variational formulations (i.e. *Finite Element Methods* (FEM), Particle Methods, Collocation Methods, Mesh-less methods, etc.). Specifically, we derive the three-dimensional strain-rate tensor from a FEM interpolation of the discrete velocity field, as introduced in [251–253]. We include the trajectory lines of the principal strain-rate components as the methodology that can be used in a computational (two-dimensional) framework for visualizing the main patterns of change in any time-dependent data cloud. This approach overcomes the three-dimensional representation problem by separating the strain-rate state of the cloud in several two-dimensional plots —one for each principal component of the strain-rate—, which

exhibit the magnitudes and orientations of the deformation patterns.

The remaining parts of this chapter are organized as follows. In Section 5.2 the methodology to compute the discrete version of the strain-rate tensor is presented. Since the main problem involves the calculation of the derivatives of discontinuous —discrete— velocities and the visualization of the strain-rate patterns, we extensively detail the numerical techniques that are adopted in the present approach. Next, in Section 5.3, we present the application of the methodology to the case study —the urban system of Barcelona— by visualizing its main strain-rate patterns, meaning the city's environmental, social and economic change. Finally, in Section 5.4 some conclusions of the proposed methodology close this chapter.

5.2 Methods

We begin this section with a review of the strain-rate tensor calculation provided a discrete three-dimensional data cloud. For doing so, the formal problem of the time-dependent dataset is introduced first. Then, we explain the numerical techniques that transform the discrete dataset into a mathematical framework by which the strain-rate tensor can be computed. Most of the ideas rely on the geometrical analysis of the discrete dataset by computing the spatial discretization of the dataset into geometric elements through a *Delaunay Triangulation*. After doing that, the computation of the strain-rate is performed with a FEM interpolation of the velocity field. Finally, we address the *eigen-problem* for the strain-rate tensor, such that the solution of the eigenvalues/eigenvectors gives the extrema strain-rates at each geometric element. The flow chart diagram of this methodology is abstracted in Fig. 5.1, including the main outcomes of each step. The extended explanation of the methodology is developed along the present section.

5.2.1 Time-dependent three-dimensional dataset

Since the main objective of this work is to reveal the temporal transformation of a three-dimensional and time-dependent dataset, let us first introduce some notation in order to clarify the mathematical ideas to be used. Let us define the discrete time-dependent data to be the set of points $\mathcal{P} = \{p_i\}$, with $i = 1, 2, \ldots, m$, being m the total data. The values in each one of the three dimensions can be seen as scalar coefficients for a set of basis vectors. These tuple of components compose the vector that we call the *position* or *coordinate* $\boldsymbol{x}_i = \begin{bmatrix} x_{i,1} & x_{i,2} & x_{i,3} \end{bmatrix}^\top$, with the superscript $\top$ denoting the transpose operation, the first subscript referring to the point i and the second to the dimension. Hence, we call $\mathcal{X}_n(t) = \{\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \ldots, \boldsymbol{x}_i(t), \ldots, \boldsymbol{x}_n(t)\} \in \mathbb{R}^3$ the *positions* of the points in a certain time t. Let us consider a uniform partition of the time span $t \in [t^d, t^g]$ in a sequence of discrete time-steps $t^d = t^0 < t^1 < \ldots < t^n < \ldots < t^N = t^g$, with $\delta t > 0$ the time-step-size defining $t^{n+1} = t^n + \delta t$ for $n = 0, 1, 2, ..., N$. Hence, our dataset is $\{\mathcal{X}_n(t)\}$, $t \in [t^d, t^g]$. Thereby, we use the superscripts to denote the discrete time-steps, with the only exception of denoting the transpose operation with the superscript $\top$.

Since the time-dependent dataset of the case study comes from a PCA reduction of a higher-dimensional multivariate dataset $\mathcal{Y}_n(t) = \{\boldsymbol{y}_1(t), \boldsymbol{y}_2(t), \ldots, \boldsymbol{y}_i(t), \ldots, \boldsymbol{y}_n(t)\} \in \mathbb{R}^k$, with $k >> 3$ and $t \in [t^d, t^g]$, into a lower-dimensional one $\mathcal{X}_n(t)$, $t \in [t^d, t^g]$, that pos-

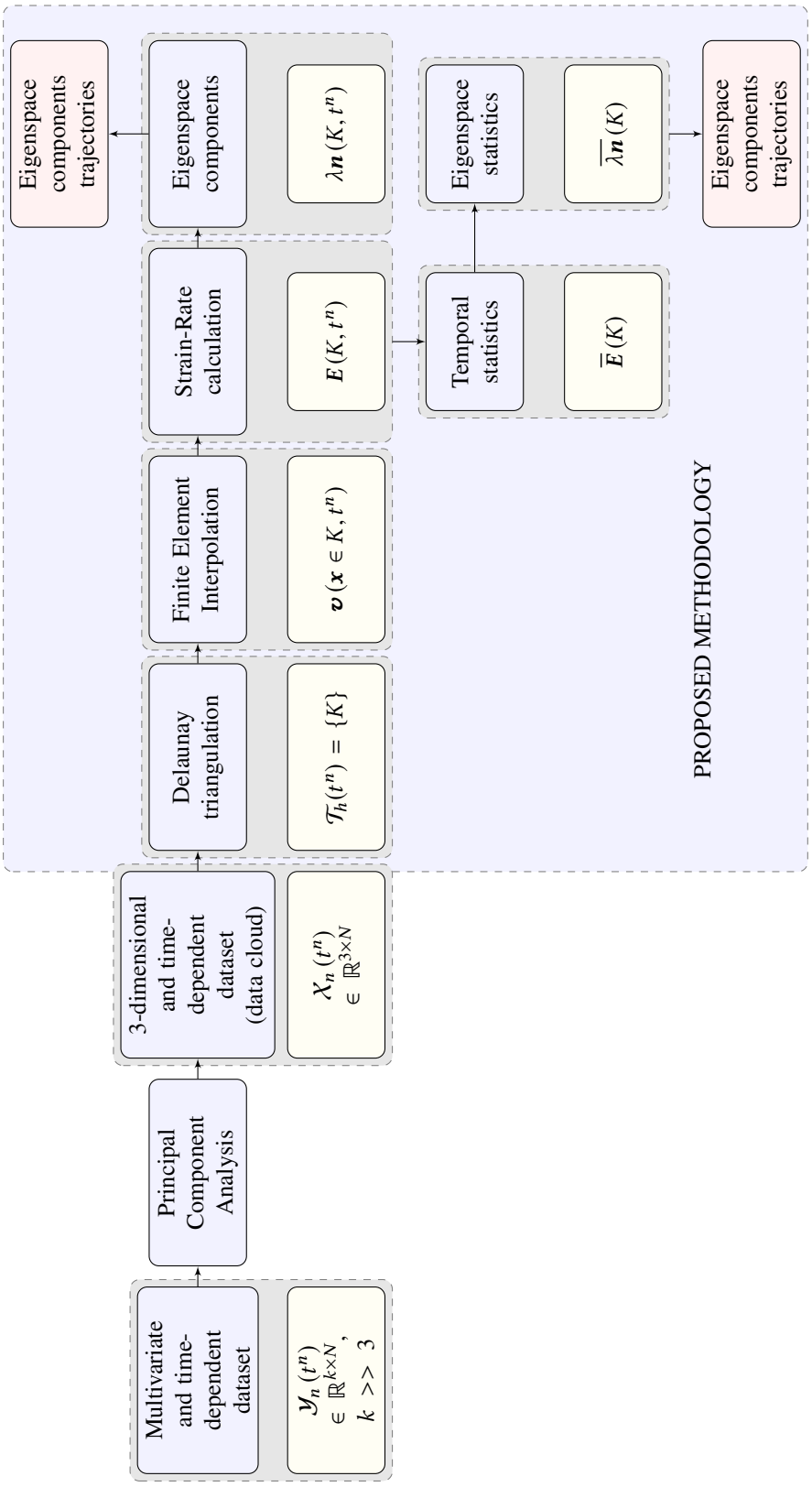

Figure 5.1: Flow chart of the visualization methodology.

sesses only three independent dimensions: Principal Component 1 (PC1), Principal Component 2 (PC2), and Principal Component 3 (PC3), we use the Cartesian coordinate system straightforwardly with each principal component being a dimension. This is, $\boldsymbol{x}_i\,(t^n) = \left[x_{i,\mathrm{PC1}}\,(t^n) \quad x_{i,\mathrm{PC2}}\,(t^n) \quad x_{i,\mathrm{PC3}}\,(t^n)\right]^\top$. Hence, the discrete time-dependent data-set can be thought as a cloud of points in the three-dimensional space that deforms gradually throughout time.

5.2.2 Finite Element Method interpolation

The main idea of the present approach is to transform the discrete cloud of points into a mathematical framework —similar to a deformable medium— by which the strain-rate tensor can be computed. To do so, we generate a mesh $\mathcal{T}_h(t) = \{K\}$ from the set of points $\mathcal{P}$ that is composed by non-overlapping and conforming geometrical elements K of diameter h. There are several methods to generate a mesh from a set of points, all which are studied in the computational geometry field [254]. Here, we apply the Delaunay Triangulation $DT\,(\mathcal{P})$ because of several reasons. The first is that the aspect ratio of the triangulated elements produce a high-quality mesh. The second is because fast Delaunay triangulation algorithms have been developed recently (see for example the one in [255]).

The result of applying the Delaunay triangulation over the set of points is a discrete mesh $\mathcal{T}_h := DT\,(\mathcal{P})$ which possess the following characteristics: it covers exactly the convex hull Ω of the point set, no point p_i is isolated from the triangulation, and all the elements $\{K\}$ are 4-points tetrahedron, which are completely defined by the position of their four corner points $K := \left\{\boldsymbol{x}_j\right\}$, with $j = 1, 2, 3, 4$. The generated mesh $\mathcal{T}_h = DT\,(\mathcal{P})$ can be seen as a —material— domain Ω that suffers deformations from the displacements of the points between consecutive time-steps. Since only discrete displacements between consecutive time-steps are known for the set of points, we now explain how the continuous velocity field inside the mesh is calculated.

Even though the FEM has been used to perform interpolation using the point-wise data (see [252, 253] for applications in one and two-dimensional data), in this work we apply this well-known method in a three-dimensional setting. In FEM, the finite interpolating space $\boldsymbol{\mathcal{V}}_h$ is defined as made of continuous piece-wise polynomials $N(\boldsymbol{x})$ in the mesh $\mathcal{T}_h$, where the discrete approximation $\boldsymbol{F}_h(\boldsymbol{x}, t) \in \boldsymbol{\mathcal{V}}_h$ of any multi-dimensional function $\boldsymbol{F}(\boldsymbol{x}, t)$, $\boldsymbol{x} \in \Omega$, can be written as

$$\boldsymbol{F}(\boldsymbol{x}, t) \approx \boldsymbol{F}_h(\boldsymbol{x}, t) := \sum_{i=1}^{n} N(\boldsymbol{x}_i)\boldsymbol{F}_i(t), \qquad \boldsymbol{x} \in \Omega. \tag{5.1}$$

We use the simplest three-dimensional finite element: the tetrahedron with *linear polynomials* and four nodes. Again, some notation is required to define the polynomials inside the element. The set of normalized coordinates $\eta_1, \eta_2, \eta_3, \eta_4$ in each tetrahedron K are such that the value of η_i is one at the point $p_i \in K$, zero at the other three corner points, and varies linearly from that point to the opposite edges. This set of coordinates has the property that the sum of the four coordinates (each belonging to one tetrahedron point) in any location inside the tetrahedron is identically one: $\eta_1(\boldsymbol{x}_i) + \eta_2(\boldsymbol{x}_i) + \eta_3(\boldsymbol{x}_i) + \eta_4(\boldsymbol{x}_i) = 1$, with $\boldsymbol{x}_i \in K$. Hence, the shape functions inside each linear tetrahedron are defined to be these coordinates:

$N_i(\boldsymbol{x}_i) = \eta_i(\boldsymbol{x}_i)$, with $i = 1, 2, 3, 4$ denoting the corner points. The FEM interpolation (5.1) of a *three-dimensional vector function*, say $\boldsymbol{F}(\boldsymbol{x}, t)$, can be defined inside each linear tetrahedron K as

$$\boldsymbol{F}(\boldsymbol{x}, t) = \eta_1(\boldsymbol{x})\boldsymbol{F}_1(t) + \eta_2(\boldsymbol{x})\boldsymbol{F}_2(t) + \eta_3(\boldsymbol{x})\boldsymbol{F}_3(t) + \eta_4(\boldsymbol{x})\boldsymbol{F}_4(t) = \sum_{j=1}^{4} \eta_j(\boldsymbol{x})\boldsymbol{F}_j(t), \quad \boldsymbol{x} \in K, \tag{5.2}$$

by denoting $\boldsymbol{F}_i(t) = \boldsymbol{F}(\boldsymbol{x}_i, t)$, for $i = 1, 2, 3, 4$, nodes of the tetrahedron.

The way the tetrahedral coordinates η_i, $i = 1, 2, 3, 4$, are defined is by means of the previous interpolating relation together with the summation constraint. This is, when one aims to define the tetrahedron geometry and calculate any position inside the tetrahedron $\boldsymbol{x} = [x_1 \quad x_2 \quad x_3]^\top$, we compute

$$\begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_{1,1} & x_{2,1} & x_{3,1} & x_{4,1} \\ x_{1,2} & x_{2,2} & x_{3,2} & x_{4,2} \\ x_{1,3} & x_{2,3} & x_{3,3} & x_{4,3} \end{bmatrix} \begin{bmatrix} \eta_1(\boldsymbol{x}) \\ \eta_2(\boldsymbol{x}) \\ \eta_3(\boldsymbol{x}) \\ \eta_4(\boldsymbol{x}) \end{bmatrix} \quad \therefore \quad \begin{bmatrix} \eta_1(\boldsymbol{x}) \\ \eta_2(\boldsymbol{x}) \\ \eta_3(\boldsymbol{x}) \\ \eta_4(\boldsymbol{x}) \end{bmatrix} = \frac{1}{6v} \begin{bmatrix} 6v & a_1 & b_1 & c_1 \\ 6v & a_2 & b_2 & c_2 \\ 6v & a_3 & b_3 & c_3 \\ 6v & a_4 & b_1 & c_4 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{5.3}$$

in order to obtain the tetrahedral coordinates system where the coefficients of the inverted matrix are given by

$$\begin{aligned}
a_1 &= x_{2,2}x_{43,3} - x_{3,2}x_{42,3} + x_{4,2}x_{32,3}, & b_1 &= -x_{2,1}x_{43,3} + x_{3,1}x_{42,3} - x_{4,1}x_{32,3}, & c_1 &= x_{2,1}x_{43,2} - x_{3,1}x_{42,2} + x_{4,1}x_{32,2}, \\
a_2 &= -x_{1,2}x_{43,3} + x_{3,2}x_{41,3} - x_{4,2}x_{31,3}, & b_2 &= x_{1,1}x_{43,3} - x_{3,1}x_{41,3} + x_{4,1}x_{31,3}, & c_1 &= -x_{1,1}x_{43,2} + x_{3,1}x_{41,2} - x_{4,1}x_{31,2}, \\
a_3 &= x_{1,2}x_{42,3} - x_{2,2}x_{41,3} + x_{4,2}x_{21,3}, & b_3 &= -x_{1,1}x_{42,3} + x_{2,1}x_{41,3} - x_{4,1}x_{21,3}, & c_3 &= x_{1,1}x_{42,2} - x_{2,1}x_{41,2} + x_{4,1}x_{21,2}, \\
a_4 &= -x_{1,2}x_{32,3} + x_{2,2}x_{31,3} - x_{3,2}x_{21,3}, & b_4 &= x_{1,1}x_{32,3} - x_{2,1}x_{31,3} + x_{3,1}x_{21,3}, & c_4 &= -x_{1,1}x_{32,2} + x_{2,1}x_{31,2} - x_{3,1}x_{21,2}.
\end{aligned}$$

Here, the abbreviation $\boldsymbol{x}_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j$ has been used, and the volume v can be calculated with the expression

$$6v = x_{21,1}\left(x_{31,2}x_{41,3} - x_{41,2}x_{31,3}\right) + x_{21,2}\left(x_{41,1}x_{31,3} - x_{31,1}x_{41,3}\right) + x_{21,3}\left(x_{31,1}x_{41,2} - x_{41,1}x_{31,2}\right).$$

At this point, it is possible to calculate the spatial derivatives of any interpolated function $\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}, t)$ in terms of the tetrahedral coordinates as

$$\frac{\partial \boldsymbol{F}(\boldsymbol{x}, t)}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial F(\boldsymbol{x},t)}{\partial x_1} \\ \frac{\partial F(\boldsymbol{x},t)}{\partial x_2} \\ \frac{\partial F(\boldsymbol{x},t)}{\partial x_3} \end{bmatrix} = \sum_{j=1}^{4} \begin{bmatrix} \frac{\partial F(\boldsymbol{x},t)}{\partial \eta_j}\frac{\partial \eta_j}{\partial x_1} \\ \frac{\partial F(\boldsymbol{x},t)}{\partial \eta_j}\frac{\partial \eta_j}{\partial x_2} \\ \frac{\partial F(\boldsymbol{x},t)}{\partial \eta_j}\frac{\partial \eta_j}{\partial x_3} \end{bmatrix} = \sum_{j=1}^{4} \begin{bmatrix} \frac{1}{6v}\frac{\partial F(\boldsymbol{x},t)}{\partial \eta_j}a_j \\ \frac{1}{6v}\frac{\partial F(\boldsymbol{x},t)}{\partial \eta_j}b_j \\ \frac{1}{6v}\frac{\partial F(\boldsymbol{x},t)}{\partial \eta_j}c_j \end{bmatrix}, \qquad \boldsymbol{x} \in K. \tag{5.4}$$

The way to calculate the continuous stress-rate tensor field is through the derivation of a continuous version of the velocities. Hence, we calculate the continuous velocity field by means of the FEM, in which linear piece-wise polynomials are used to interpolate the velocity at any spatial position inside the mesh. First, let us explain how to calculate the discrete version of the velocity from the cloud of points. We suppose that the displacement $\boldsymbol{s}_i$ of point p_i in the time interval $\left(t^n, t^{n+1}\right)$ can be defined —without loss of accuracy— as infinitesimal, in the sense of $\boldsymbol{s}_i(t^n) \approx \boldsymbol{x}_i\left(t^{n+1}\right) - \boldsymbol{x}_i(t^n)$. With this supposition in hand, we use the *Taylor* expansion:

$$\boldsymbol{x}_i(t) = \sum_{k=1}^{\infty} \frac{\delta t^k}{k!} \left.\frac{d^k \boldsymbol{x}_i}{dt^k}\right|_{t=t_o} = \boldsymbol{x}_i(t_0) + \left.\frac{d\boldsymbol{x}_i}{dt}\right|_{t=t_o} \delta t + \left.\frac{d^2\boldsymbol{x}_i}{dt^2}\right|_{t=t_o} \frac{\delta t^2}{2} + \cdots, \tag{5.5}$$

in order to calculate the discrete velocity $\boldsymbol{v}_i$ of point p_i as

$$\boldsymbol{v}_i\left(t^n\right) = \left.\frac{d\boldsymbol{x}_i}{dt}\right|_{t=t^n} \approx \frac{\boldsymbol{x}_i\left(t^{n+1}\right) - \boldsymbol{x}_i\left(t^n\right)}{\delta t} = \frac{\boldsymbol{x}_i\left(t^{n+1}\right) - \boldsymbol{x}_i\left(t^n\right)}{\left(t^{n+1} - t^n\right)}, \tag{5.6}$$

where the second (and higher) order terms are neglected. The previous result then generates the continuous version of velocity by replacing it in (5.2).

5.2.3 Elemental strain-rate calculation

Having defined the continuous space of velocities, we can calculate the derivatives along each one of the spatial directions and derive the strain-rate tensor field.

Following the continuum mechanics concepts in [128] and assuming small deformations, the *strain-rate tensor* is calculated as

$$\boldsymbol{E}\left(\boldsymbol{x}, t^n\right) := \frac{1}{2}\left(\nabla\boldsymbol{v}\left(\boldsymbol{x}, t^n\right) + \left(\nabla\boldsymbol{v}\left(\boldsymbol{x}, t^n\right)\right)^\top\right),$$

with $\nabla\boldsymbol{v}$ the gradient of velocity. Each component of the $3\times3-$tensor is developed in Cartesian coordiantes as

$$\begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} = \begin{bmatrix} \frac{\partial v_1(\boldsymbol{x},t^n)}{\partial x_1} & \frac{1}{2}\left(\frac{\partial v_1(\boldsymbol{x},t^n)}{\partial x_2} + \frac{\partial v_2(\boldsymbol{x},t^n)}{\partial x_1}\right) & \frac{1}{2}\left(\frac{\partial v_1(\boldsymbol{x},t^n)}{\partial x_3} + \frac{\partial v_3(\boldsymbol{x},t^n)}{\partial x_1}\right) \\ \frac{1}{2}\left(\frac{\partial v_2(\boldsymbol{x},t^n)}{\partial x_1} + \frac{\partial v_1(\boldsymbol{x},t^n)}{\partial x_2}\right) & \frac{\partial v_2(\boldsymbol{x},t^n)}{\partial x_2} & \frac{1}{2}\left(\frac{\partial v_2(\boldsymbol{x},t^n)}{\partial x_3} + \frac{\partial v_3(\boldsymbol{x},t^n)}{\partial x_2}\right) \\ \frac{1}{2}\left(\frac{\partial v_3(\boldsymbol{x},t^n)}{\partial x_1} + \frac{\partial v_1(\boldsymbol{x},t^n)}{\partial x_3}\right) & \frac{1}{2}\left(\frac{\partial v_3(\boldsymbol{x},t^n)}{\partial x_2} + \frac{\partial v_2(\boldsymbol{x},t^n)}{\partial x_3}\right) & \frac{\partial v_3(\boldsymbol{x},t^n)}{\partial x_3} \end{bmatrix}.$$

Moreover, the six independent components of the strain-rate tensor can be arranged using *Voigt's* notation into a 6-component strain-rate vector as follows:

$$\boldsymbol{E}\left(\boldsymbol{x}, t^n\right) = \begin{bmatrix} E_{11}\left(\boldsymbol{x}, t^n\right) & E_{22}\left(\boldsymbol{x}, t^n\right) & E_{33}\left(\boldsymbol{x}, t^n\right) & \gamma_{12}\left(\boldsymbol{x}, t^n\right) & \gamma_{23}\left(\boldsymbol{x}, t^n\right) & \gamma_{31}\left(\boldsymbol{x}, t^n\right) \end{bmatrix}^\top, \tag{5.7}$$

where $\gamma_{12}\left(\boldsymbol{x}, t\right) = 2E_{12}\left(\boldsymbol{x}, t\right), \gamma_{23}\left(\boldsymbol{x}, t\right) = 2E_{23}\left(\boldsymbol{x}, t\right)$ and $\gamma_{13}\left(\boldsymbol{x}, t\right) = 2E_{13}\left(\boldsymbol{x}, t\right)$ are the *Shear-Rate Strains*. With this notation in hand, the strain-rate tensor can be calculated as

$$\boldsymbol{E}\left(\boldsymbol{x}, t^n\right) = \begin{bmatrix} E_{11}\left(\boldsymbol{x}, t^n\right) \\ E_{22}\left(\boldsymbol{x}, t^n\right) \\ E_{33}\left(\boldsymbol{x}, t^n\right) \\ \gamma_{12}\left(\boldsymbol{x}, t^n\right) \\ \gamma_{23}\left(\boldsymbol{x}, t^n\right) \\ \gamma_{31}\left(\boldsymbol{x}, t^n\right) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} & 0 & 0 \\ 0 & \frac{\partial}{\partial x_2} & 0 \\ 0 & 0 & \frac{\partial}{\partial x_3} \\ \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} & 0 \\ 0 & \frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} \end{bmatrix} \begin{bmatrix} v_1\left(\boldsymbol{x}, t^n\right) \\ v_2\left(\boldsymbol{x}, t^n\right) \\ v_3\left(\boldsymbol{x}, t^n\right) \end{bmatrix}, \tag{5.8}$$

by defining the matrix operator of derivatives over the velocity field. In the case of the right hand side velocities, we can arrange a node-wise vector of discrete velocities in the tetrahedron K, as

$$\boldsymbol{V}\left(K, t^n\right) = \begin{bmatrix} v_{1,1}\left(t^n\right) & v_{1,2}\left(t^n\right) & v_{1,3}\left(t^n\right) & v_{2,1}\left(t^n\right) & v_{2,2}\left(t^n\right) & \dots & v_{4,2}\left(t^n\right) & v_{4,3}\left(t^n\right) \end{bmatrix}^\top.$$

Using the definition of the finite element interpolation of any function (5.2) together with its partial derivatives (5.4), and replacing those in (5.8), we obtain

$$\boldsymbol{E}\left(\boldsymbol{x}, t^n\right) = \frac{1}{6v} \sum_{j=1}^{4} \begin{bmatrix} a_j \frac{\partial}{\partial \eta_j} & 0 & 0 \\ 0 & b_j \frac{\partial}{\partial \eta_j} & 0 \\ 0 & 0 & c_j \frac{\partial}{\partial \eta_j} \\ b_j \frac{\partial}{\partial \eta_j} & a_j \frac{\partial}{\partial \eta_j} & 0 \\ 0 & c_j \frac{\partial}{\partial \eta_j} & b_j \frac{\partial}{\partial \eta_j} \\ c_j \frac{\partial}{\partial \eta_j} & 0 & a_j \frac{\partial}{\partial \eta_j} \end{bmatrix} \begin{bmatrix} \eta_j\left(\boldsymbol{x}\right) V_{j,1}\left(t^n\right) \\ \eta_j\left(\boldsymbol{x}\right) V_{j,2}\left(t^n\right) \\ \eta_j\left(\boldsymbol{x}\right) V_{j,3}\left(t^n\right) \end{bmatrix}. \tag{5.9}$$

Now, the operation $\frac{\partial \eta_i}{\partial \eta_j} F_j = F_i$ since $\frac{\partial \eta_i}{\partial \eta_j} = \delta_{ij}$, with δ_{ij} the Kronecker delta. Hence, $\boldsymbol{E}\left(K, t^n\right)$ can be calculated as the product of the matrix $\mathrm{S}\left(K\right)$ and the vector $\boldsymbol{V}\left(K, t^n\right)$. This is,

$$\boldsymbol{E}\left(K, t^n\right) = \mathrm{S}\left(K, t^n\right) \boldsymbol{V}\left(K, t^n\right), \tag{5.10}$$

with $\boldsymbol{x} \in K$ and the discrete matrix $\mathrm{S}\left(K, t^n\right)$ defined as

$$\mathrm{S}\left(K, t^n\right) = \frac{1}{6v} \begin{bmatrix} a_1 & 0 & 0 & a_2 & 0 & 0 & a_3 & 0 & 0 & a_4 & 0 & 0 \\ 0 & b_1 & 0 & 0 & b_2 & 0 & 0 & b_3 & 0 & 0 & b_4 & 0 \\ 0 & 0 & c_1 & 0 & 0 & c_2 & 0 & 0 & c_3 & 0 & 0 & c_4 \\ b_1 & a_1 & 0 & b_2 & a_2 & 0 & b_3 & a_3 & 0 & b_4 & a_4 & 0 \\ 0 & c_1 & b_1 & 0 & c_2 & b_2 & 0 & c_3 & b_3 & 0 & c_4 & b_4 \\ c_1 & 0 & a_1 & c_2 & 0 & a_2 & c_3 & 0 & a_3 & c_4 & 0 & a_4 \end{bmatrix}. \tag{5.11}$$

Thus, this last matrix can be computed solely in terms of the coordinates of the nodes.

Up to this point, we have demonstrated how to calculate the elemental strain-rate. Now, our purpose is to identify the data cloud transformation throughout the visualization of the strain-rate patterns. This is, we need to identify the extrema strain-rates and their orientations. In a formal sense, this is the well-known *Eigenvalues and Eigenvectors* problem, which is stated as: if T is a linear transformation from a vector space V over a field F into itself, and v is a vector in V that is not the zero vector, then v is an eigenvector of T if $T(v)$ is a scalar multiple of v. Knowing that by definition the second order strain-rate tensor is a linear operator from a vector field into another first-order tensor field, the previous definition applied to the strain-rate tensor leads to:

$$\left[\boldsymbol{E}\left(K, t^n\right) - \boldsymbol{I}\lambda\left(K, t^n\right)\right] \boldsymbol{n}\left(K, t^n\right) = 0, \tag{5.12}$$

where $\boldsymbol{I}$ is the 3×3 identity tensor, $\boldsymbol{n}\left(K, t^n\right) \in \mathbb{R}^3$ is a normalized (non zero), i.e. unit, vector called eigenvector, and $\lambda\left(K, t^n\right) \in \mathbb{R}$ is the eigenvalue associated with the eigenvector. In other words, an eigenvector is a vector that changes by only a scalar factor when the strain-rate tensor is applied to it, resulting in a vector parallel to itself. By solving (5.12) one obtains three different eigenvalues $\lambda_1\left(K, t^n\right), \lambda_2\left(K, t^n\right), \lambda_3\left(K, t^n\right)$, and three eigenvectors $\boldsymbol{n}_1\left(K, t^n\right), \boldsymbol{n}_2\left(K, t^n\right), \boldsymbol{n}_3\left(K, t^n\right)$, associated with each eigenvalue.

The eigenvalues and eigenvectors describe the principal magnitudes and orientations of the strain-rate tensor: since the diagonal components of the strain-rate tensor

$E_{11}(K, t^n)$, $E_{22}(K, t^n)$, and $E_{33}(K, t^n)$ have different values in different reference systems, with the set of eigenvalues one finds the extreme —maximum and minimum— possible values that any of these components may take. Indeed, the maximum and minimum stress-rates —and their orientations— are related with the maximum and minimum eigenvalues. In this work, we follow the notation in which positive values for the eigenvalues represent the extension-rate and negative values represent contraction-rate. Hence, $\lambda_1(K, t^n)$ is the maximum and positive eigenvalue meaning extension-rate, $\lambda_3(K, t^n)$ is the minimum and negative eigenvalue meaning contraction-rate, and $\lambda_2(K, t^n)$ is either extension or contraction rate, but in smaller magnitude.

Hence, with the extrema strain-rates at the elemental level we can reveal the deformation trend of the data cloud, and above all, locating which regions suffer the most abrupt change in the time-span. We also propose to draw the family of curves —trajectories— that are instantaneously tangent to $\lambda_1 \boldsymbol{n}_1(K, t^n)$, $\lambda_2 \boldsymbol{n}_2(K, t^n)$, and $\lambda_3 \boldsymbol{n}_3(K, t^n)$ in the complete mesh Ω. This results in a graph plot for each one of the $\lambda \boldsymbol{n}(K, t^n)$ fields that can be easily represented in a two-dimensional framework. These tangent lines can be colored according to the eigenvalue magnitude such that, they completely illustrate the main patterns of change inside the data cloud. Note, nevertheless, that $\lambda \boldsymbol{n}(K, t^n)$ is the composition of a strain vector using the tensor components. Those differ in formal definition, but we use this concept merely for visualization purposes.

Nevertheless, we also perform some temporal statistics of the strain-rate states, where the principal strain-rates for each time-step $\lambda_i \boldsymbol{n}_i(K, t^n)$ are accounted as the temporal events: each strain-rate state is accounted as a single observation. Specifically, we calculate the time-average of the eigenspace components $\overline{\lambda_i}(K)$ and $\overline{\boldsymbol{n}_i}(K)$, the maximum extension-rate in the time span $L^{\infty}(\lambda_1(K))$, and the maximum contraction-rate in the time span $L^{-\infty}(\lambda_3(K))$. In this way, we calculate representative deformation results for the complete temporal interval of analysis, as well as the location of extreme values. Besides, trajectory lines can be plotted using these statistical results.

5.3 Results

In the present section, we demonstrate the application of this methodology to quantify the temporal change of an urban multivariate system (see Figure 5.4). First, we cite the case study that includes the multivariate description of the ten districts of Barcelona, and whose reduced three-dimensional data-set is used as the starting point. Then, we derive the strain-rate state of the data-set, pursuing the extension and contraction patterns visualization. Finally, we close this section with insights about the city transformation implied in the strain-rate state of the data cloud.

5.3.1 Time-dependent data cloud from an urban multivariate description

The time-dependent data cloud comes from the PCA output of a multivariate description of the city of Barcelona. Since 1987, the city has been divided into 10 administrative districts, which are the largest territorial units of the city and can be compared with neighborhoods in a common metropolitan area: Ciutat Vella, Eixample, Gràcia, Les Corts, Sarrià, Sant Andreu,

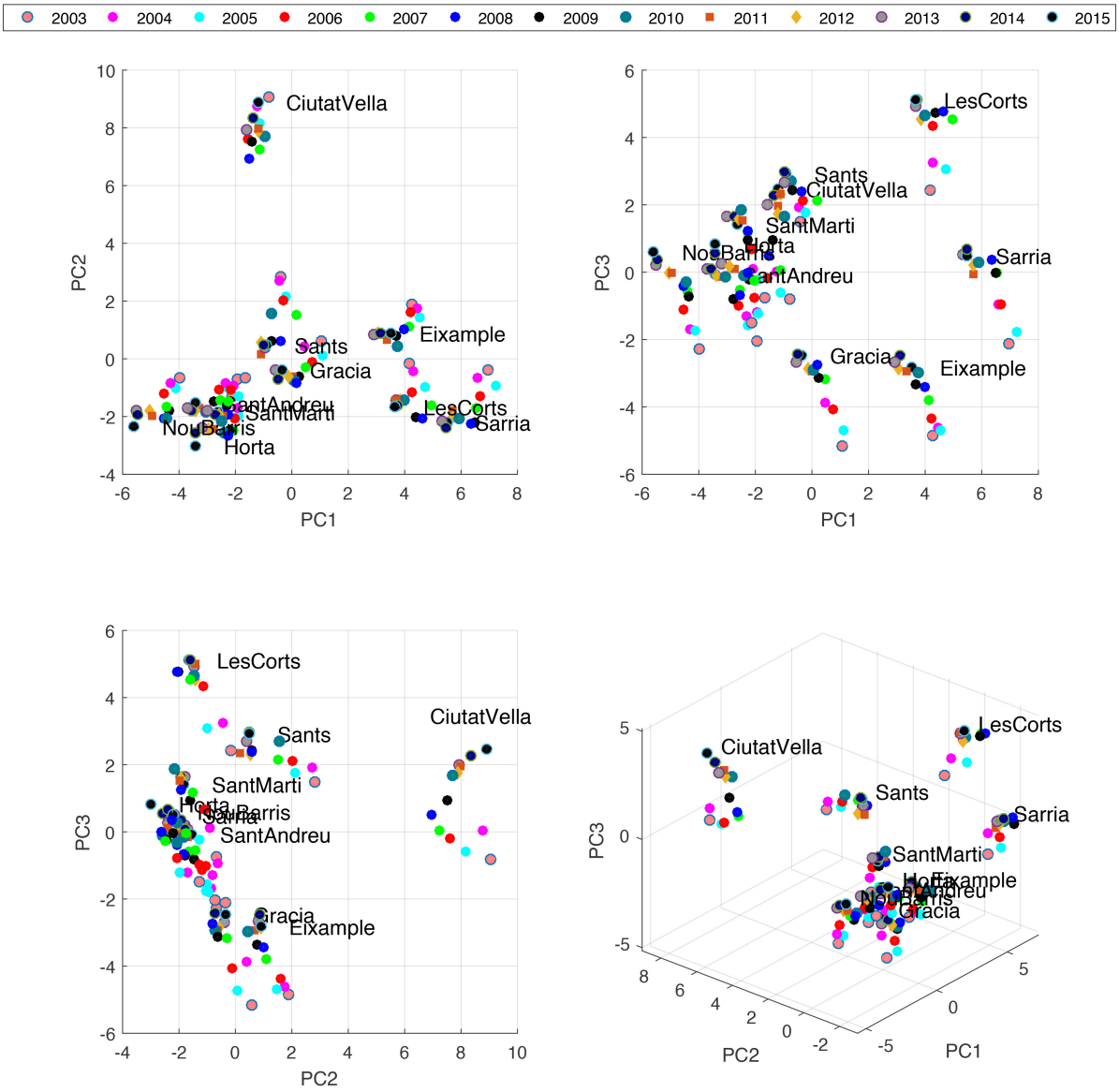

Figure 5.2: Three-dimensional and time-dependent data cloud from the case study.

Table 5.1: Tetrahedral elements derived from the Delaunay Triangulation of the set of points.

Element (id)	First Vertex	Second Vertex	Third Vertex	Fourth Vertex
1	Eixample	LesCorts	Gràcia	Sarrià
2	SantAndreu	Horta	SantMartí	NouBarris
3	Sants	SantAndreu	SantMartí	NouBarris
4	LesCorts	Eixample	Sants	CiutatVella
5	Gràcia	SantMartí	Horta	Sarrià
6	Gràcia	SantMartí	Sarrià	LesCorts
7	Eixample	Gràcia	Sants	CiutatVella
8	CiutatVella	SantAndreu	Sants	NouBarris
9	Sarrià	SantMartí	Horta	LesCorts
10	Gràcia	SantAndreu	Sants	CiutatVella
11	Gràcia	SantAndreu	CiutatVella	NouBarris
12	Sarrià	Eixample	LesCorts	CiutatVella
13	Horta	SantAndreu	Gràcia	NouBarris
14	SantAndreu	SantMartí	Horta	Gràcia
15	LesCorts	Eixample	Gràcia	Sants
16	LesCorts	Gràcia	SantMartí	Sants
17	Gràcia	SantAndreu	SantMartí	Sants

Sant Martí, Horta, Sants-Montjuïc, and Nou Barris. Barcelona has a population of approximately 1.6 million inhabitants living in 10216 ha. The inclusion of all the 10 districts in the multivariate description has been aimed to represent the city at its overall scale and to allow comparisons between them.

The raw multivariate description —from which the PCA is calculated— comprises the data of 40 environmental, economic, and social indicators for the ten districts in the time span of $t^0 = 2003 \leq t^n \leq 2015 = t^N$, $n = 0, 1, .., 12$. Hence, the case study data cloud comes from a PCA reduction of the higher-dimensional multivariate dataset $\mathcal{Y}_n(t^n) \in \mathbb{R}^{40\times12}$, into a lower-dimensional one $\mathcal{X}_n(t^n) \in \mathbb{R}^{3\times12}$ that possesses only three independent dimensions: PC1, PC2, and PC3. The dimensionally-reduced dataset from the application of the PCA is available online at `https://summlabbd.upc.edu/SalazarLlano/PhDThesis/`. Hence, the three-dimensional and time-dependent data cloud is composed by the three-dimensional coordinates $\mathcal{X}_n(t^n)$ of the ten p_i points defined in the sequence of $N = 12$ time-steps from 2003 to 2015, with the time-step size of $\delta t = 1$ year. These points are displayed in Figure 5.2, where all the observations —districts each year— in the time-span are included.

As the first step of our methodology, we apply the Delaunay Triangulation (DT) to the data cloud. Specifically, we calculate the DT to the set of coordinates at each time-step $\mathcal{X}_n(t^n)$. This results in a mesh $\mathcal{T}_h(t^n)$ composed by $nel = |K|$ non-overlapping tetrahedron. Table 5.1 expands the resulting triangulation for year 2003, with the vertices information for the $nel = 17$ tetrahedron. This triangulation is also depicted in Figure 5.3 where the following notation for the districts in the vertices is used: C is Ciutat Vella, E is Eixample, G is Gràcia, H is Horta, L is Les Corts, N is Nou Barris, A is Sant Andreu, M is Sant Martí, S is Sants-Montjuïc, and R is Sarrià. Since the position $\boldsymbol{x}_i(t^n)$ of a given point p_i at a later time-step can surpass the initial tetrahedron's circumscribed sphere, we recalculate the mesh triangulation at each time step t^n, $n = 1, .., 11$.

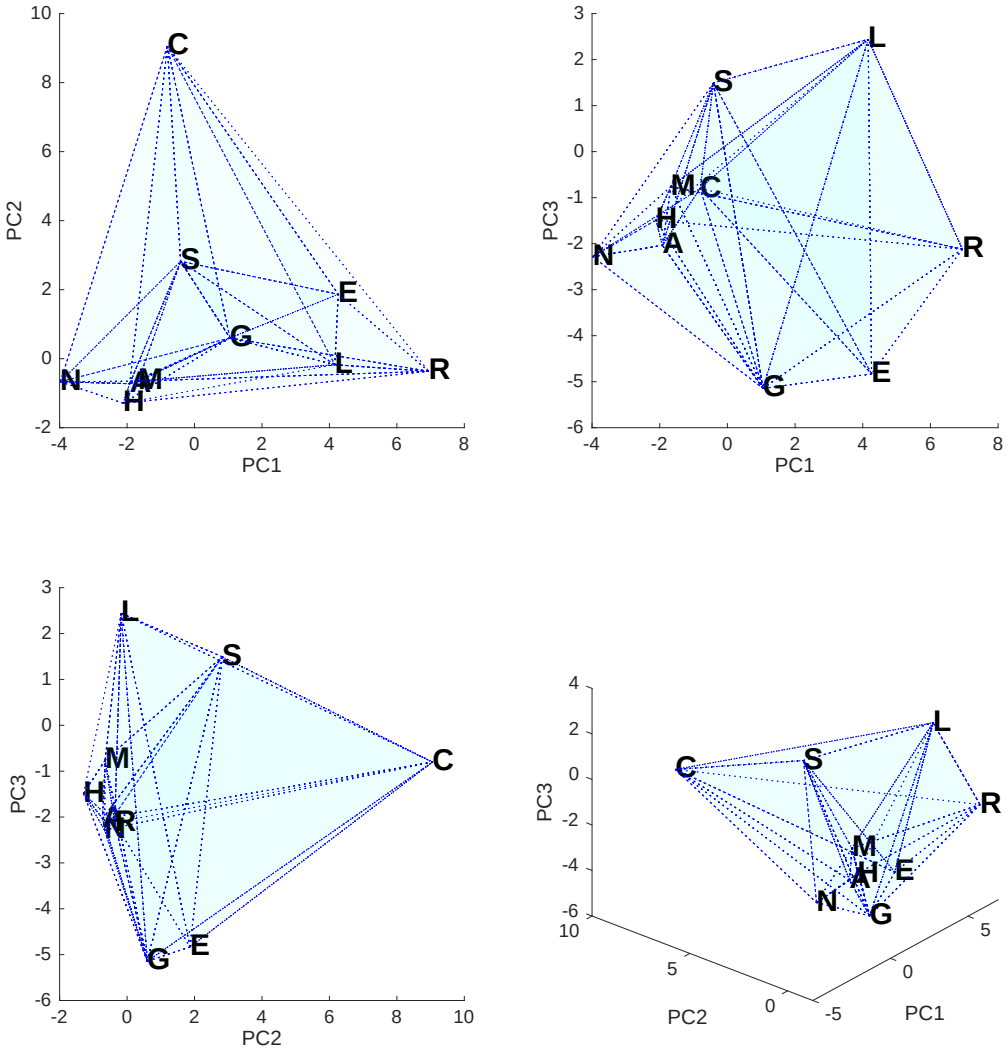

Figure 5.3: Delaunay triangulation of the data cloud from the case study.

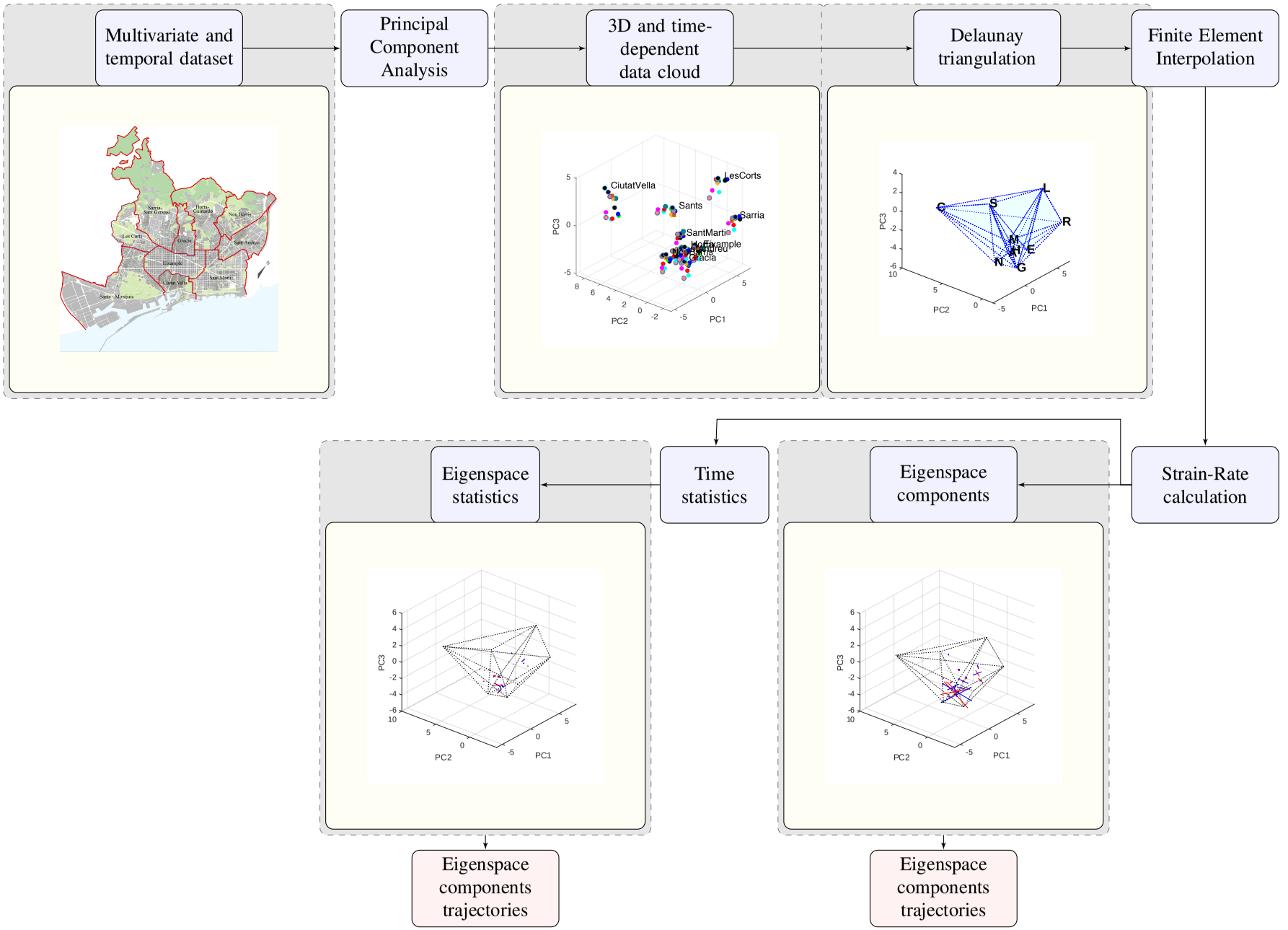

Figure 5.4: Flow chart of the temporal change visualization methodology.

Table 5.2: Principal strain-rate components. Eigenvalues and Eigenvectors of the strain-rate tensor at the year 2003. The extension is denoted by the maximum eigenvalue λ_1, and contraction is denoted by the minimum eigenvalue λ_3.

Element (id)	λ_1 (year^{-1})	λ_2 (year^{-1})	λ_3 (year^{-1})	$\boldsymbol{n}_1^\top$	$\boldsymbol{n}_2^\top$	$\boldsymbol{n}_3^\top$
1	0.53152326	0.03372761	-0.5035305	[-0.4030 -0.7620 0.5069]	[-0.8521 0.1104 -0.5116]	[0.3339 -0.6381 -0.6938]
2	0.82703046	-0.07233917	-3.26208499	[0.0599 0.4379 0.8970]	[0.9979 -0.0488 -0.0428]	[0.0250 0.8977 -0.4399]
3	1.86537595	0.0239402	-1.26311701	[-0.2846 -0.7630 0.5804]	[0.8705 0.0479 0.4898]	[-0.4015 0.6447 0.6505]
4	0.13368209	0.03257601	-0.09645259	[0.3448 0.0267 0.9383]	[0.6847 -0.6909 -0.2320]	[0.6421 0.7225 -0.2565]
5	0.49758344	0.15637449	-0.04820077	[-0.2529 0.9667 0.0396]	[-0.3025 -0.1179 0.9458]	[0.9190 0.2272 0.3223]
6	0.95472589	0.0373597	-0.54006708	[-0.0595 -0.8341 0.5484]	[0.9779 0.0616 0.1998]	[-0.2005 0.5482 0.8120]
7	0.24389104	-0.02462416	-0.21220276	[0.9802 -0.0490 -0.1918]	[0.0873 0.9766 0.1964]	[0.1777 -0.2093 0.9616]
8	0.04516871	-0.00558049	-0.29509916	[-0.5143 -0.5503 -0.6578]	[0.7221 -0.6916 0.0139]	[0.4626 0.4679 -0.7531]
9	0.83841605	0.06226881	-0.72504166	[-0.4963 0.7789 0.3834]	[0.7282 0.1330 0.6723]	[-0.4727 -0.6128 0.6332]
10	0.08250367	-0.007911	-0.1849469	[-0.9511 0.1145 -0.2870]	[0.0253 -0.8969 -0.4415]	[-0.3079 -0.4272 0.8501]
11	0.1403248	- 0.00962214	-0.34760551	[0.1721 -0.7223 -0.6698]	[0.9327 -0.0992 0.3467]	[0.3169 0.6844 -0.6566]
12	0.28802797	-0.02754418	-0.54411227	[-0.1025 -0.4296 -0.8972]	[0.6308 -0.7255 0.2753]	[0.7692 0.5377 -0.3453]
13	5.25376147	-0.1612729	-2.63828711	[-0.2599 -0.7733 0.5783]	[0.9210 -0.0187 0.3890]	[-0.2900 0.6338 0.7171]
14	2.58304682	0.22953554	-2.02948589	[-0.4920 0.6994 0.5184]	[0.7802 0.0900 0.6190]	[-0.3863 -0.7091 0.5900]
15	0.28104681	0.05995401	-0.17975011	[-0.6611 -0.5949 0.4572]	[-0.7488 0.4850 -0.4518]	[-0.0470 0.6410 0.7661]
16	0.12268815	0.06554156	-0.16290843	[-0.0458 0.8952 -0.4432]	[0.9468 0.1805 0.2666]	[-0.3187 0.4074 0.8559]
17	0.17492291	0.11156825	-0.27224695	[0.5895 -0.1016 0.8014]	[-0.2177 0.9354 0.2788]	[-0.7779 -0.3388 0.5292]

5.3.2 Principal strain-rates

We compute the strain-rate tensor of each tetrahedron with the interpolated version of the velocities for the case study, such that linear piece-wise polynomial functions defined inside each tetrahedron are used in the FEM interpolation. Certainly, we suppose that the velocities come from an infinitesimal analysis in which the higher order terms of the displacement are neglected. The gradients inside each tetrahedron are also considered to be constant since the polynomial functions are of first order. Applying (5.10), we compute the strain-rate tensor of every tetrahedron, $\boldsymbol{E}(K, t^n)$ for time-steps $n = 0, \ldots, 11$. Note that displacements cannot be calculated for the last year $t^{12} = 2015$, and that the strain-rate tensor units are year^{-1} (for the case study).

We are interested in the magnitude and orientations of the principal strain-rates —extension and contraction— at the elemental level. Hence, the next step is to solve (5.12) and obtain the eigenspace components (eigenvalues and eigenvectors) of the strain-rate tensor. For the sake of conciseness, we list in Table 5.2 the results of the principal strain-rates for the year 2003 solely.

The application of this methodology to the case study is displayed graphically in Fig. 5.4, beginning with the map of the ten districts of Barcelona as the abstraction of the multivariate and time-dependent dataset. The three-dimensional coordinates arising from the PCA output are displayed next. We also present next the triangulated mesh at the initial year 2003. It is clear from the visual inspection that the quantitative analysis of the temporal transformation is greatly justified, so that we calculate the strain-rate tensor over the FEM interpolation of discrete velocities and compute its principal components.

5.3.2.1 Trajectory patterns of the principal strain-rates

In favor of the analysis, we display the principal strain-rate components in a graphical way. One first approach is to illustrate the patterns of extension-rate and contraction-rate using a

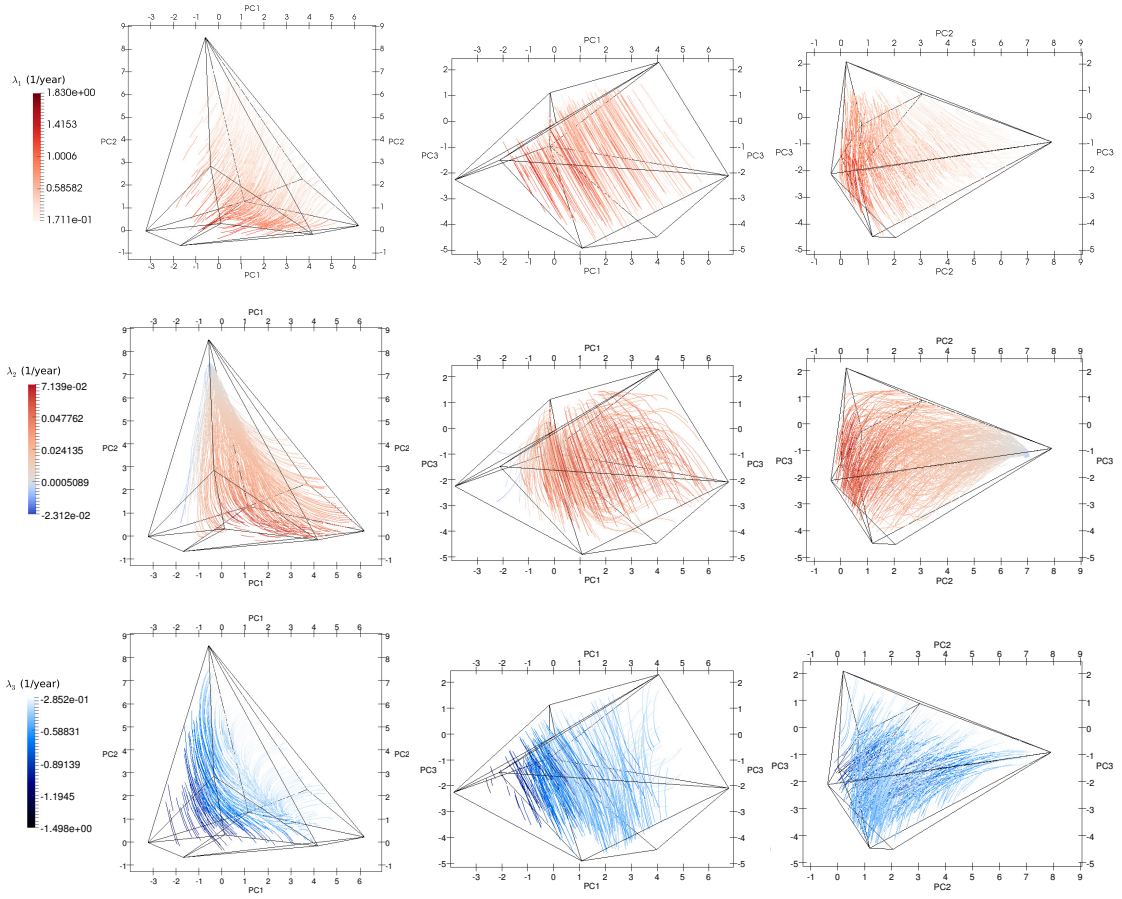

Figure 5.5: The trajectory patterns of the principal strain-rates at the year 2003. First principal strain-rate (top), second principal strain-rate (middle), and third principal strain-rate (bottom).

vector representation, to what is referred as the *Strain-rate diagrams* [244]. In that approach, the centroid of the tetrahedron serves as the location from which the principal components of the strain-rate tensor give a representative result inside the element. We draw the strain-rate diagram of the year 2003 in the sixth step of Fig. 5.4, where extension-rate is represented by symmetric blue vectors $\lambda_1 \boldsymbol{n}_1$ pointing out the centroid, and contraction-rate is represented by the red vectors $\lambda_3 \boldsymbol{n}_3$ pointing in. But, it is hard to visualize the distribution of the principal strain-rates and their three-dimensional orientations using this type of illustration.

Our approach to ease the visualization and understanding of the strain-rate state is to draw the *trajectories* of the principal strain-rate components, as used for displaying stresses in beams and columns in [256]. In the following we demonstrate our findings of the strain-rate state at the year 2003 using the trajectories visualization. In Figure 5.5 we display the principal components trajectories, where the lines are colored by the magnitude of the principal strain-rate and those are parallel to its orientation. From these representations, we can understand the magnitude and orientation of each principal component of the strain-rate tensor. And more importantly, the trajectory patterns overlapped with the coordinates of the districts (in Figure 5.3) provide information about local regions of extension and contraction rates inside the urban description, where extension-rate patterns means differentiation and contraction-rate patterns means clustering —or homogenization. This can be observed from the color intensity: locations of high strain-rate are colored with great intensity, while the locations of low deformations are not visible since the white color represents null strain-rates.

In the case of the first strain-rate component which is shown at the top of Fig. 5.5, we observe that the larger magnitude of extension-rate is localized in between Nou Barris, Sant Andreu, Sant Martí, Horta and Sants-Montjuïc, and that it decreases near Eixample, Les Corts, and Ciutat Vella. Therefore, the main transformation is located at the group of nearby districts: Nou Barris, Sant Andreu, Sant Martí, and Horta. The extension-rate patterns are oriented from this group apart to Ciutat Vella, suggesting that there is a divergence of Ciutat Vella from the grouped districts. Indeed, the main extension pattern is oriented along the PC3 dimension and covers the grouped districts. It is of lesser importance the pattern which comprises the districts of Nou Barris, Sant Martí, and Sants-Montjuïc and ends at Gràcia and Eixample.

Contraction-rate, on the other hand, is expressed by the third principal strain-rate component, which by definition is orthogonal to the first and second principal strain-rates. The third principal strain-rate component is shown at the bottom of Fig. 5.5, where we can appreciate this orthogonality by noticing that the trajectories of the third principal strain-rate are perpendicular to the extension-rate pattern. We observe that the contraction-rate trajectories are mostly homogeneous, with a minor importance between Sant Martí, Sant Andreu, Nou Barris and Horta districts, and completely declining at Sants-Montjuïc and Gràcia. This direct relation between extension and contraction is found in solids' deformations, where it is ruled by the conservation of mass —or Poisson ratio— [128].

Apart from the extension and contraction patterns of the mesh, locations of smaller strain-rates are represented by the second principal component. Considering the middle plots of Fig. 5.5, we recognize that the orientation of this strain-rate component is concentrated in between Les Corts, Sarrià, Horta and Nou Barris, and that it is directed towards Eixample, fading at Ciutat Vella. This principal strain-rate component is certainly orthogonal to the first and third components, but it implies a strain-rate pattern that is two orders of magnitude smaller.

In the previous lines we have demonstrated the application of the trajectories diagrams

Table 5.3: Time-averaged eigenspace components.

Element(id)	$\overline{\lambda_1}$ (year^{-1})	$\overline{\lambda_2}$ (year^{-1})	$\overline{\lambda_3}$ (year^{-1})	$\boldsymbol{n}_1^\top \implies \overline{\lambda_1}$	$\boldsymbol{n}_2^\top \implies \overline{\lambda_2}$	$\boldsymbol{n}_3^\top \implies \overline{\lambda_3}$
1	0.1157	-0.0086	-0.0687	[-0.4941 -0.4365 0.7519]	[0.1770 0.1339 -0.9751]	[0.1682 0.8390 0.5175]
2	0.5087	-0.0085	-0.8859	[-0.5777 0.5570 0.5966]	[-0.4259 -0.1592 -0.8907]	[-0.0636 -0.9309 0.3596]
3	0.3919	-0.0440	-0.1662	[-0.3746 -0.7054 0.6017]	[-0.9963 0.0335 0.0789]	[0.3127 -0.7926 -0.5234]
4	0.0868	0.0041	-0.0193	[0.4684 0.8217 0.3247]	[-0.0619 0.3756 -0.9247]	[-0.7437 -0.4144 -0.5246]
5	0.0435	-0.0326	-0.0659	[-0.2319 0.8799 0.4148]	[-0.0891 -0.1728 0.9809]	[0.2229 -0.2816 -0.9333]
6	0.1834	0.0096	-0.0536	[-0.2271 -0.8440 0.4859]	[-0.6281 0.2260 -0.7446]	[-0.4216 0.8544 0.3038]
7	0.0647	-0.0059	-0.0485	[0.5157 -0.6416 0.5678]	[0.0741 -0.9932 0.0893]	[0.2985 -0.4332 -0.8504]
8	0.1436	-0.0198	-0.0174	[-0.3994 -0.4110 0.8194]	[-0.1094 0.9614 -0.2525]	[0.3984 -0.8684 -0.2952]
9	0.6311	0.0178	-0.1144	[-0.2472 -0.6917 0.6785]	[0.2277 0.2955 0.9278]	[0.1771 0.8523 0.4921]
10	0.0355	0.0098	-0.0745	[0.4275 0.8923 -0.1449]	[0.8566 -0.5130 -0.0551]	[-0.5198 0.2220 -0.8249]
11	0.0452	-0.0239	-0.1039	[0.4742 -0.0948 0.8753]	[-0.7018 -0.6930 0.1650]	[-0.3132 -0.9345 -0.1693]
12	0.2054	0.0336	-0.0703	[-0.5086 -0.8385 -0.1954]	[0.1867 0.6566 0.7307]	[0.0460 -0.4726 -0.8801]
13	1.1642	-0.0218	-0.3192	[-0.4080 -0.6299 0.6609]	[-0.9632 0.2605 0.0654]	[-0.5136 0.5798 -0.6325]
14	0.1923	0.0415	-0.2640	[-0.4758 0.8168 0.3264]	[0.9570 0.2693 -0.1076]	[0.4308 -0.2715 -0.8606]
15	0.0552	0.0085	-0.0363	[-0.2364 0.6160 0.7514]	[0.7518 0.2783 -0.5978]	[0.6817 0.6544 -0.3270]
16	0.0438	0.0175	-0.0059	[0.2794 0.9257 -0.2549]	[0.8449 -0.2441 0.4760]	[-0.5430 -0.2892 -0.7883]
17	0.2113	-0.0196	-0.3129	[-0.0062 0.9656 0.2599]	[0.0760 0.2441 0.9668]	[0.0666 -0.9959 -0.0609]

of the principal strain-rate components as a powerful visualization technique of the three-dimensional strain-rate state of a data cloud. The strain-rate patterns can be used to analyze the system's development, in example, with the identification of regions with a special behavior: although there are some grouped districts in the case study, all of those are separating at a high rate in the first and third principal components (spatial dimensions), as described by the first eigenspace component. Hence, those are differentiating themselves in the PC1 and PC3 description. On the contrary, low strain-rates can be an indication of stagnation, and thus, an expression of inactivity where an abrupt change is not probably to occur. That is specially the case of the Ciutat Vella district, which is separated from the grouped districts but it is neither diverging nor converging to them.

One final remark is that our visualization approach is mesh independent. This means that the principal strain-rate trajectory plots coincide regarding of the triangulated mesh: different triangulations will produce different positions, magnitudes, and orientations of the principal strain-rate components, nevertheless, the trajectory lines that join them coincide for any type of mesh.

5.3.3 Temporal statistics of the principal strain-rates

Plots of the principal strain-rate components trajectories can be completed for the remaining years of the time span, t^n, $n = 1, 2, ..., 11$. Readings of the strain-rate streamline patterns for those years can be completed straightforwardly as discussed in the paragraphs above. Nevertheless, we perform some temporal statistics of the strain-rate states, where the principal strain-rates calculated for each time-step are accounted as the temporal events: each strain-rate state is accounted as a single observation.

The first statistics that we perform is the time-average of the principal strain-rate components, separated as the first, second, and third principal strain-rates. Table 5.3 presents the time-averaged results of the principal strain-rates. Also, in the last schematic of Fig. 5.4, we plot the strain-rate diagram of the time-averaged principal strain-rates at the time-averaged centroid of tetrahedron elements. This figure gives insights about the orientation and magni-

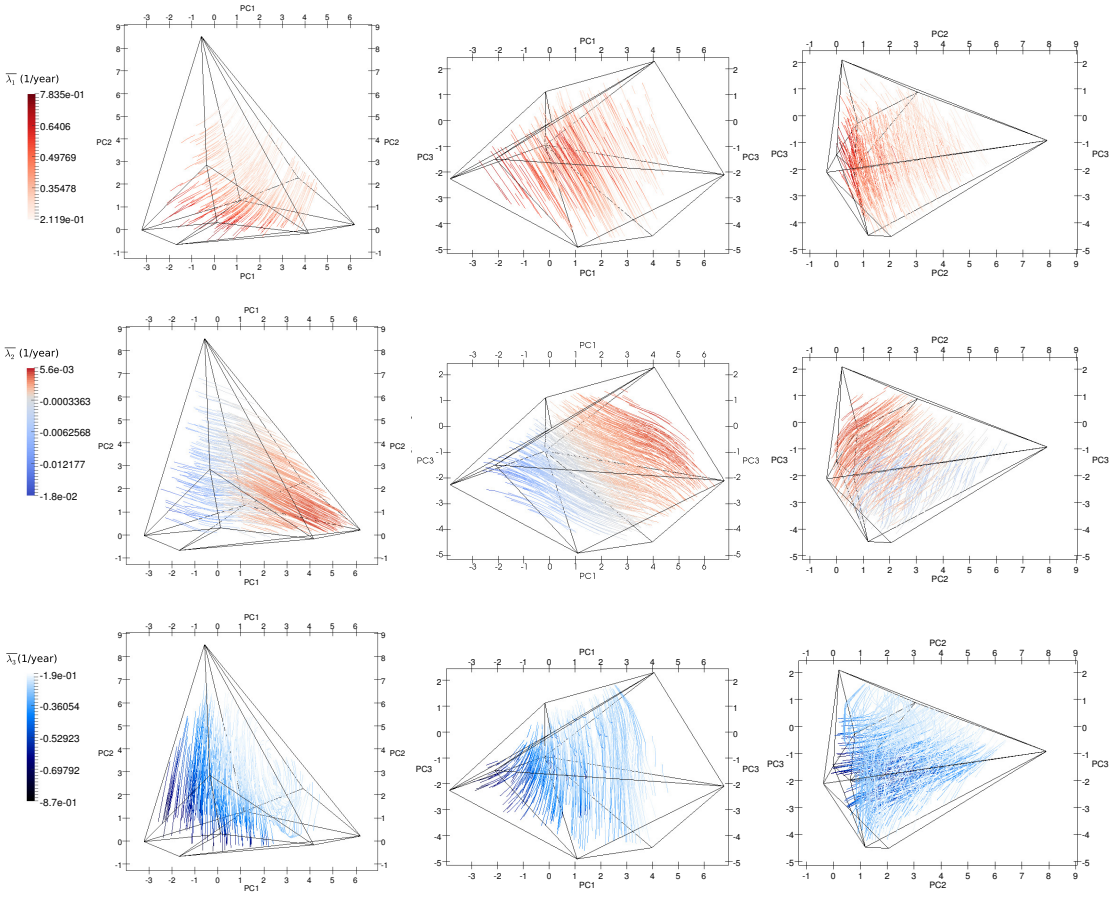

Figure 5.6: The trajectory patterns of the time-averaged principal strain-rates. Averaged first principal strain-rate (top), averaged second principal strain-rate (middle), and averaged third principal strain-rate (bottom).

Table 5.4: Maximum and minimum eigenspace components in the time span.

Element(id)	$L^{\infty}(\lambda_1)$	Year	$\boldsymbol{n}_1^{\top} \implies L^{\infty}(\lambda_1)$	$L^{-\infty}(\lambda_3)$	Year	$\boldsymbol{n}_3^{\top} \implies L^{-\infty}(\lambda_3)$
1	0.5315	2003	[-0.4030 -0.7620 0.5069]	-0.5035	2003	[0.3339 -0.6381 -0.6938]
2	2.0893	2010	[-0.5094 0.8338 0.2130]	-3.2621	2003	[0.0250 0.8977 -0.4399]
3	3.0249	2009	[-0.0766 -0.8956 0.4383]	-1.2631	2003	[-0.4015 0.6447 0.6505]
4	0.4860	2010	[0.5309 0.8024 0.2724]	-0.1901	2005	[0.8017 0.5583 -0.2134]
5	0.6321	2014	[0.0691 0.9893 -0.1286]	-0.3019	2010	[-0.3219 0.5283 0.7857]
6	1.1842	2006	[0.1651 -0.9360 0.3110]	-0.9833	2005	[0.2453 -0.7335 -0.6338]
7	0.2772	2010	[0.0108 -0.9360 0.3518]	-0.2122	2003	[0.1777 -0.2093 0.9616]
8	0.5104	2010	[-0.4614 -0.6584 0.5946]	-0.5374	2009	[-0.2484 -0.6025 0.7585]
9	4.3858	2010	[-0.2509 -0.6991 0.6696]	-2.4530	2010	[0.4203 -0.7018 -0.5752]
10	0.3753	2010	[-0.4800 0.8570 -0.1874]	-0.3477	2009	[0.4401 -0.6943 0.5695]
11	0.5210	2009	[-0.0542 0.3205 0.9457]	-0.5164	2009	[0.3738 0.8847 -0.2784]
12	1.5272	2008	[-0.6804 -0.7328 0.0027]	-0.7379	2007	[0.4304 0.8624 -0.2666]
13	5.2538	2003	[-0.2599 -0.7733 0.5783]	-4.6094	2008	[0.2022 -0.9445 0.2588]
14	2.5830	2003	[-0.4920 0.6994 0.5184]	-2.0295	2003	[-0.3863 -0.7091 0.5900]
15	0.2810	2003	[-0.6611 -0.5949 0.4572]	-0.4221	2010	[0.0927 -0.9519 -0.2919]
16	0.2659	2009	[0.4571 0.7618 -0.4591]	-0.1636	2012	[-0.5501 0.7352 0.3961]
17	1.4269	2009	[-0.0935 0.9952 0.0288]	-1.1715	2010	[0.1262 0.9888 0.0798]

tude of the first and third principal strain-rates, again by plotting the symmetric arrows pointing out for extension and pointing in for contraction. Still, we complete our analysis by drawing the trajectory curves of those time-averaged strain-rate diagrams in Fig. 5.6.

With the aid of Figures 5.3 and 5.6 we analyze the trajectory patterns of the time-averaged strain-rate state of the data cloud. In the case of the first and third strain-rate components, those are comparable to the ones described in the previous paragraphs for the year 2003. We observe a high extension-rate behavior between Nou Barris, Sant Andreu and Horta in the PC1 dimension. The contraction-rate, on the other side, is oriented in the PC2 dimension and it is mostly located in between Gràcia and Eixample and decreases near Sants-Montjuïc. In the case of the contraction-rate in the PC3 dimension, it is mostly present between Nou Barris and Horta, and in smaller magnitude between Horta and Sant Andreu. The contraction-rate between the remaining districts is negligible in all dimensions. Both the extension and contraction rates demonstrate trajectory patterns which are directed from the grouped districts towards the separated district of Ciutat Vella. However, the magnitude of the time-averaged strain-rates is much smaller than the ones obtained for year 2003. In the case of the time-averaged second strain-rate, its magnitude is greater for Eixample, Les Corts, and Sarrià nodes than for the grouped districts. The orientation of the trajectories involving this second strain-rate component is parallel to the one linking Eixample and Les Corts to the grouped districts.

The second temporal statistics that we perform is to calculate $L^{\infty}-$norm of the temporal strain-rate distribution. That is, to calculate the year where the maximum extension and contraction rates occur within each tetrahedron. The results of the application of the $L^{\infty}-$norm to the case study are presented in Table 5.4. We observe that the maximum strain-rates occur at year 2003: either expansion or contraction. Also, that important contraction-rate magnitudes take place between the years 2003 and 2010. This is not the case of the extension-rate magnitudes, which are more prevalent after year 2008.

5.4 Conclusions

In the present chapter, we have quantified and visualized the temporal change of a time-dependent and three-dimensional dataset. Contrary to other approaches [115, 121, 231, 232], we have calculated the three-dimensional strain-rate state of the dataset based on the interpolation of discrete point-wise displacements —or data variations. We have applied a technique in continuum mechanics using a FEM interpolation of non-overlapping linear tetrahedral elements that covers the three-dimensional dataset.

The methodology has demonstrated to exhibit regions of major deformation-rate. Departing from the calculation of the numerical strain-rate values, we have introduced some data-visualization techniques that help to locate the magnitudes and orientations of the strain-rate state in the three-dimensional framework. This is the case of the principal strain-rate trajectories, whose have demonstrated to be more detailed than other possible visualization techniques, e.g. strain-rate diagrams in [131, 247] or Confidence Ellipsoids in [110, 111, 236–238]. The main difference with strain-rate diagrams is the ability of the present approach to display a continuum version of the strain-rate inside the dataset, and to separate the analysis into each principal component of the strain-rate state, while being mesh independent. Compared to the method of Confidence Ellipsoids, the kinematic basis of the proposed methodology makes an accurate description of the deformation patterns in the complete three-dimensional domain, rather than in a compact and partial region that contains the data distribution. It also allows a detailed temporal description of the change between one instant and another. This is particularly not possible using the method of the ellipsoids since that method accounts for the complete temporal distribution in order to evaluate change.

The calculation of the strain-rate state shows that the methodology is suitable for quantifying the temporal change of a reduced three-dimensional dataset describing the social, economic and environmental state of the city of Barcelona. In particular, high strain-rates are associated with the localized deformation of regions that represent the districts of the city in the time span of 13 years. The method not only portrays the similarities and differences—distances—between the districts of the city, like *Cluster Analysis* [205, 257, 258] or *Distance measures* [259, 260] which explore similarities and groups in data. The distinctive attribute of the present work is the feasibility to quantify the districts' differentiation with time: the strain-rate tensor provides quantitative information about local regions of extension and contraction, where extension-rate patterns means differentiation and contraction-rate patterns means clustering —or homogenization. In this regard, the strain-rate methodology can be potentially used as an extension of clustering analysis since it reveals contraction and extension patterns of clusters. Conclusions about the divergence or clustering of districts in time can therefore be stated. For example, it reveals the time, location, and orientation of pressures affecting the inhabitants of certain districts of the city, essentially those which are rapidly diverging from the rest (e.g. the case of Ciutat Vella for the case study). This methodology locates regions where detailed action is necessary, as well as the foretelling of possible ruptures in the system.

Finally, we would like to say that this method is not limited to the study of the data change, but can also be applied to other descriptions: a natural sequel to the present chapter is the study of the time-dependent data cloud as a deforming elastic solid under equilibrium. Solving the inverse problem, namely the identification of the constitutive moduli of the deforming material emerges as the first step for predicting the system's future state given its historical data.

Chapter 6

Conclusions: Towards a better understanding of the urban system

In this thesis, we have studied the city as a complex system composed of multiple subsystems and cross-scale interactions portrayed through statistical and mathematical methods. Certainly, we have integrated several methodologies to analyze the city abstracted as a multivariate information system under the general framework of urban sustainability. Particularly, we have sought for displaying diversity patterns at three different scales: the overall city's diversity, between districts' diversity, and within districts' diversity. This by understanding the concept of diversity from the perspectives of ecology and complex systems theory. We have applied information entropy, multivariate analysis, and computational mechanics-based methods to picture and measure diversity.

6.1 Achievements

We have begun by developing the conceptual framework which supports the arguments made in this work. In this sense, we have surveyed the field of urban ecology, urban sustainability, and complex systems. At the introductory part of the thesis, we have made the effort to conceptualize new ideas in the science of cities. For example, we have proposed an original model that represents the urban system as a dynamic mechanical system that responds to disturbances. This model has been used to investigate the relationship between diversity and resilience, both of which could be quantified. Also, we have proposed a novel approach to the study of urban systems from physics and computer sciences. We have brought robust ideas from physics, not only from mathematics — which are frequently used in the computational field — to be able to compute urban data. The very fact of being able to establish a link between the city as seen from the science of cities, with the most recent advances in computer science opens up an enormous amount of hypothesis to investigate.

In the second chapter of the thesis, Chapter 2, we have applied an information entropy measure to assess the urban metabolism of the system. At that first stage, we introduced the multivariate data system that characterizes the city under the urban metabolism framework, which encompasses multiple social, economic, and environmental variables arranged in four metabolic types or categories: Input Supportive, Output Pressure, Destructive, and Regener-

ative. The link between the measured indicators and the ideal state of the system has been defined using a normalization process that correlates measured and optimal values for each indicator. Precisely, in this process, relies upon its main drawback: that the optimal state of the system has to be chosen *a-priori* without really accounting for the theoretical developments of sustainability. In that methodology, the information entropy has been computed over the occurrence probability matrix of the normalized events: the information entropy is zero exactly when one of the probabilities is one and the others are zero, and reaches its maximum value when all the events are equally likely to occur.

We have applied this information entropy-methodology to analyze the case study of four districts of Barcelona between 2003 and 2012. We have used the information entropy measure to describe the metabolic state of the city's information system: if the data system fitted the expected optimal behavior then the information entropy led to a great value. It has been able to describe the temporal behavior of the urban system: the yearly calculation of the information entropy to the dataset has been able to portray global and local disturbances —such as the economic Spanish crisis of 2008—, in the urban system. It also has contributed to the understanding of the energy, water, matter, and information flows at the district's scale. In particular, we have observed the relationship between the environmental problems of urban ecosystems and the energy and mass flows. Also, the highly dissipative character of the city. However, given that the information entropy model arises with a single —aggregated— measure, a direct interpretation of these shocks and trends concerning the initial set of indicators is not feasible. The information entropy —which has been widely used in the literature to measure diversity within a sample— has also been applied in Chapter 3 to measure the urban structure of the four districts in the case study.

Given the limitations of the information entropy-based methodology as an aggregated measure of diversity, we have next applied in Chapter 3 a multivariate statistical analysis to improve the study of the urban system. The main motivation to implement multivariate statistical analysis techniques has been their ability to provide a moderate trade-off between the large volume of initial disaggregated data and its final aggregation. Expressly, PCA has been the principal multivariate method to reduce the information system's dimensionality and enhance the interpretation of data. Using the proposed exploratory methodology, we have been able to portray and measure the diversity among the components of the urban system. We demonstrated that the dimensionally-reduced description of the city given by PCA could expose both, the diversity of types and the diversity within types (districts). Certainly, we have portrayed diversity using biplot graphs and measured dissimilarities among urban components by mean of Euclidean distances: distanced points in the reduced description indicate diverse attributes of districts.

The results of applying this methodology to the case study of four districts of Barcelona have revealed differentiated neighborhood characters and distinctive functions at the district's scale. Two distinctive groups of districts —or Barcelonas— have emerged: a first group that has been characterized by large public spaces and green and recreational areas, and a second one that has been dense and highly polluted. The differences between the districts have been also quantified employing Euclidean distances among districts' centroids: Ciutat Vella and Les Corts have been the most dissimilar —distant— districts, and Eixample and Gràcia have been the most similar —closest— ones. The application of the exploratory analysis into the measured dataset of the case study has demonstrated to be a suitable methodology to construct

a synthetic and accurate description of the city. It has revealed the interactions among variables and components of the complex system allowing for understanding the city at a macro level and thus supporting urban planning.

With those results in hand, we have extended the exploratory multivariate analysis in Chapter 4. There, the city has been conceptualized as a multivariate system organized in the three pillars of sustainable development: Economic, Environmental, and Social. Since multivariate analysis can be applied to the complete dataset, it is not restricted to a specific categorization of the variables in the information system (i.e. urban metabolic types). Our motivation has been to expand the analysis by including other statistical techniques in the exploratory multivariate analysis. The main idea: to assess the complexity of the urban system, as well as its temporal change.

Consequently, we have applied PCA, MFA, Hierarchical Cluster Analysis, and Confidence Ellipsoids to the complete description of the city of Barcelona in the period between 2003 and 2015. Three groups —or clusters— of districts have emerged from the application of both, PCA and hierarchical clustering. The analysis of all the 10 districts of the city has confirmed the diversity patterns found in the four districts' study in Chapter 3: where Ciutat Vella exhibited a distinctive performance, displaying a particular functional typology —it is a cultural, touristic, administrative, and walkable district, that provides services for the whole city. Furthermore, two dissociated groups of districts have emerged: one which has been characterized by high income and large housing surfaces, and other comprising working-class districts, characterized by low-to-middle income and small housing surfaces.

Regarding differentiation among districts, some relevant similarities and dissimilarities have been found. Horta and Sant Andreu are the most similar —closest— districts. Nevertheless, distances confirmed that Nou Barris, Sant Martí, Horta, and Sant Andreu are highly clustered, constituting a homogeneous continuum in the system. On the other side, Ciutat Vella and Sarrià emerged as the most dissimilar —furthest— districts, followed by Ciutat Vella and Les Corts (which emerged as the most dissimilar in the reduced case study)

We have gone forward in that chapter and exploited the dimensionally-reduced description given by PCA to build a first approach for assessing the system's temporal evolution. Under that approach, the city has been abstracted as a three-dimensional and time-dependent data cloud arising from PCA output. In that representation, points indicate the performance of the districts in a given year as described by the first three principal components from PCA. We have calculated the Confidence Ellipsoids of the scattered observations and diversity could be associated with their volume, whereas the temporal change has been linked with the volume's variation. Although the Confidence Ellipsoids method exposed the temporal transformation of the city, it failed at describing the in-between district's variation.

Thus, in Chapter 5, we have deepened the analysis of the temporal change of the urban system. We have proposed to quantify and visualize the patterns of change by using a computational mechanics approach based on the kinematics of continuum deformable media: the computation of the strain-rate has been aimed to evaluate the temporal transformation of the three-dimensional and time-dependent dataset representing the city. In addition to this calculation, we have proposed a visualization technique that is based on the principal components of the strain-rate. Specifically, we have plotted the trajectories of each principal strain-rate in the data cloud. The important contribution has been to calculate the strain-rate tensor as the Finite Element continuous interpolation of the discrete displacement field. The mandatory in-

gredient has been to produce a mesh of tetrahedral finite elements by applying the Delaunay triangulation to the data cloud.

The strain-rate based method has proved to be a feasible technique in quantifying inter-districts' diversification over time: the strain-rate tensor provides quantitative information about local regions of extension and contraction; where extension-rate portrays differentiation —diversification—; and contraction-rate implies clustering —or homogenization. Also, the trajectories of the principal strain-rate components served as a powerful visualization technique of the strain-rate state of the data cloud, displaying regions of major temporal change. Indeed, some relevant findings regarding the application of the strain-rate methodology to the case study of Barcelona city have demonstrated a large diversification pattern at the working-class group of districts, and the divergence of Ciutat Vella from the grouped districts. Those have also indicated minor homogenization trends between some specific districts, and that the largest variations occurred in 2003: either diversification or homogenization.

6.2 General discussion

Acknowledging the contributions listed above, we first discuss some general features of the proposed methodology and end up debating the concepts that have been developed throughout the document.

One central hypothesis has been the abstraction of the city as a multivariate data system. As remarked by [75, 77, 79], the selection of meaningful indicators for the multivariate description is crucial in assessing urban sustainability and constitutes a research subject itself. Even so, the construction of the multivariate system of indicators for the case study of Barcelona between 2003 and 2015, constitutes a reliable dataset for future research.

Concerning the methodological approach, we have brought concepts and methods from ecology and complexity science to analyze the city. The novelty of this thesis relays on the integration of several statistical and mathematical methods to assess the urban system as a compound of districts. Although we recognize that most of the applied multivariate analysis techniques (i.e. PCA and HCA) are not new, they have been barely used to analyze urban systems, and rarely at the district's scale. Hence, the appropriation of those tools (coming from other scientific disciplines) in the urban analysis constitutes a novel descriptive approach.

In the case of the proposed methodology in Chapter 5, we have developed a general approach to analyze and visualize the deformation of any time-dependent data cloud. Although the visualization methodology has been applied to the analysis of an urban system, it is not restricted to that type of application and can be used in several other disciplines to display the variation of data systems.

Regarding the relevance of this study for Barcelona, results obtained with the proposed methodology highlighted specific patterns that can be easily identified by decision-makers and should be applied in the urban planning of the city. Some policy recommendations have been suggested throughout the document to enhance urban sustainable development. For example, to improve the urban regenerative capacity and reduce the environmental inequality in the city, green-based strategies (i.e. urban gardens, green roofs, alternative energies) must be implemented. Besides, the compactness of the urban structure of Barcelona constitutes an opportunity to strengthen its sustainability. Since this compact urban form favors the mix —

diversity— of land uses and the proximity of services, walkability and sustainable modes of transport should be promoted. This policy would be effective in reducing air and noise pollution, which takes place —mainly— in the compact and dense central districts of the city. Having done the thesis with a single case study limited the extracted conclusions and suggestions to the ones explained before, and did not allow for generalizations. That is precisely its importance, to be able to apply the present methodology in other cities to perform future cross-link analyses.

Having to do with the visualization of the diversity patterns of Barcelona, the proposed methodology has succeeded to portray the diversity of the city: many types of districts and/or groups of districts have emerged. However, this fact implies also that there are many types of citizens, displaying social and economic inequalities and raising some governance concerns. For example; a given dweller in one of the districts in the low-income cluster (i.e. Nou Barris), has an income 71% lower, lives in a surface 47% smaller, consumes 47% less energy, and 28% less water, with respect to a given dweller in one of the districts in the high-income cluster (i.e. Sarrià). This fact revealed that, despite social and renewal programs from the last decade, the right to the city [261] is not equal for all the citizens, and definitely, that it is correlated to the place of residence. Since this type of diversity appeared by contrasting — or separating— districts by their economic performance, then attention should be paid to the feedback of urban diversity strategies. In this scenario, participatory planning and inclusive decision-making processes [18] should be implemented to improve diversity while strengthening equity and sustainability.

Given that diversity and resilience are both scale-dependent concepts [56] that we ratified in the application of our methodology to the case study, some scale issues must be considered in the improvement of the methods. The first is that the city may seem diverse on one scale, but when looking at another scale it may appear as homogeneous. Therefore, the choice of scale influences the overall interpretation of diversity. Another is that resilience is understood as a property of the system that is not subject to a specific scale [74]. Hence, both concepts should be evaluated at multiple scales to deepen on their trade-offs.

Finally, let us mention some insights about the diversity-stability debate [58, 59] that we used in the understanding of the relationship between diversity and resilience —or diversity and sustainability. Even though this debate has led to long discussions, and constitutes a research topic in itself, the mechanical dynamical model from Chapter 1 pointed out a complex response between diversity and stability in the system, suggesting that in some cases an increase in diversity strengths resilience (being a wider concept than stability, remaining functional after disturbances). In this sense, given the diversity of the urban system by having many types of districts performing different urban functions, the services provided by those types can be sustained over a wider range of conditions, and the system will have a greater resilience capacity. However, since diversity have also portrayed some inequalities, special attention should be paid to trade-offs between global diversity and sustainability. Precisely this is one of the challenges when planning for urban resilience and diversity: cross-scale trade-offs among the different strategies that aim at fostering diversity and/or sustainability. Focusing only on one scale could enhance diversity and resilience in a single district without considering the effects in the other districts, and a general perspective may neglect subsystems that can contribute to the diversity and resilience in the whole city.

Acknowledging this cross-scale trade-off implies that in designing for urban diversity and

urban resilience the following questions (built over [50] guidelines) must be addressed during planning.

- Diversity and/or resilience for whom, who benefits with that strategy?
- Which part of the city is going to be more diverse and/or resilient?
- Is the strategy oriented to face short-term disruptions or long-term stress?
- Is the policy limited to a spatial scale, is portraying cross-scalar interactions?
- Which is the final goal of the program?

Future studies on the science of cities and urban resilience (and/or urban sustainability) will help to answer these open questions and will allow us to more fully understand the links between diversity, urban resilience, and their trade-offs with sustainability.

6.3 Ongoing and future lines of research

The following research lines are suggested to expand the debates left by the open questions of this thesis. Most are related to the analysis of the urban system exploiting the methodological framework that has been developed. Some others belong to the application of the present methodology to other urban systems, encouraging the analysis of factors influencing diversity, urban metabolism, and urban resilience.

6.3.1 Exploitation of the mechanical dynamic model

As explained in Chapter 1, we can use the translational mechanical model to obtain displacement results for a wide range of possible configurations. We aim to further complexify the mechanical elements and their relations, such that other definitions of diversity can be studied: we have been thinking in improving the interactions between individuals by considering dissipative elements (e.g. dampers), or by increasing the types of individuals (e.g. rotating elements). Another particular research topic has to do with the stability estimates of the translational system based on analytical techniques. Those estimates may be useful to bound the domain where diversity has a positive relationship with the oscillation amplitude. In any case, the final goal of the mechanical model is to contribute to the current debate about the relationship between diversity and stability, or resilience of the system. Hence, to contribute to the Diversity-Stability-Resilience debate.

6.3.2 Prediction of the urban system's future states

One main goal in urban studies is the foretelling of the future states of a city by auditing the existent information. Producing a methodology in this matter emerges as a novel approach to detect critical states, and to diagnose the responses of public policies improving their relevance. Our idea is to use the elastic mechanical problem to predict the future states of the urban system. Since the discrete displacements of the data-set are defined spaced in time, we suppose

that the deformation is performed slowly such that it can be modeled with the *elastic equation for small-deformations*:

$$\rho\partial_t^2\boldsymbol{u}(\boldsymbol{x},t) + (\lambda+\mu)\nabla(\nabla\cdot\boldsymbol{u}(\boldsymbol{x},t)) + \mu\nabla^2\boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{f}(\boldsymbol{x}) \qquad \boldsymbol{x}\in\partial\Omega, \tag{6.1}$$

where ∂_t is the eulerian derivative in time, ρ is the density of the medium, $\boldsymbol{u}(\boldsymbol{x},t)$ is the displacement, $\boldsymbol{f}$ is the vector of external forces —that we neglect—, and λ and μ are the *Lamè* constants. The previous formulation is commonly referred as the *direct elasticity problem*, where the displacements are determined by knowing the geometry Ω, the Lamè elastic constants (constitutive moduli), and the data of initial and boundary conditions.

In the case of predicting the future deformation states of the urban system using (6.1) one needs first to solve an inverse elastic problem for reconstructing the constitutive moduli that best matches the historic deformations. Indeed, the problem is now related to the identification of the constitutive parameters of the elastic material that is being deformed, so that, those can be used in the direct problem to evaluate possible future states. Hence, we wish to solve the inverse situation to the direct elasticity problem: given the displacement field $\boldsymbol{u}(\boldsymbol{x},t^n)$ in Ω, that we call the measured displacement $\widetilde{\boldsymbol{u}}(\boldsymbol{x}_i,t^n)$, determine the constitutive parameters λ and μ. This problem is commonly known as the *inverse elasticity problem*, and is equivalent to the following minimization:

$$[\lambda^*,\mu^*] = \arg\min_{(\lambda,\mu)\in\mathbb{R}^+}\frac{1}{2}\int_\Omega(\boldsymbol{u}_h(\boldsymbol{x},t)[\lambda,\mu]-\tilde{\boldsymbol{u}}(\boldsymbol{x},t))^2\ d\Omega, \tag{6.2}$$

where one seeks the constitutive parameters of the elastic domain λ^*, and μ^* that minimizes the distance between the measured displacement and a solved displacement $\boldsymbol{u}_h(\boldsymbol{x},t)[\lambda,\mu]$, since the displacement field depends on the parameters λ and μ. Nevertheless, it has been demonstrated that the solution to the previous problem leads to undetermined solutions. Therefore, the actual challenge is to compose a regularization term that imposes some restrictions on λ and μ, and afterward to use these parameters in the calculation of displacements at future times using (6.1).

6.3.3 Resilience metric of the urban system

The concept of resilience comes originally from material science, where it is used as a measure of the energy that a material is capable to absorb up to the point of yielding. In mathematical terms, the expression for the *modulus of resilience* is,

$$U = \int_\Omega\int_0^{|\boldsymbol{\epsilon}|}\boldsymbol{\sigma}[\tilde{\boldsymbol{u}}(\boldsymbol{x})] : \boldsymbol{d\epsilon}[\tilde{\boldsymbol{u}}(\boldsymbol{x})]\ d\Omega, \tag{6.3}$$

where the strain increment $\boldsymbol{d\epsilon}[\tilde{\boldsymbol{u}}(\boldsymbol{x})]$ can be calculated using the ideas in Chapter 5. The stress $\boldsymbol{\sigma}[\tilde{\boldsymbol{u}}(\boldsymbol{x})]$, on the other hand, can be simplified for an *isotropic material*, such that it can be described using only the Lamè constants as

$$\boldsymbol{\sigma}[\tilde{\boldsymbol{u}}(\boldsymbol{x})] = \lambda\mathrm{tr}(\boldsymbol{\epsilon}[\tilde{\boldsymbol{u}}(\boldsymbol{x})])\boldsymbol{I} + 2\mu\boldsymbol{\epsilon}[\tilde{\boldsymbol{u}}(\boldsymbol{x})], \tag{6.4}$$

where $\boldsymbol{I}$ is the second rank identity tensor. This relation measures the total mechanical energy (per unit volume) that is consumed by the material in a straining process [128]. It is a

quantitative indicator of the straining state: if the material is unaffected by the application of some stress, and upon unloading that stress recovers its original shape, then it is considered as an elastic process where the mechanical energy is restored. But, if the stress exceeds the yield point of the material, then, the material will be deformed irreversibly losing some of the mechanical energy and never recovering its original shape.

Hence, we aim to relate engineering resilience with urban resilience. Up to our knowledge, the analogy of the resilience of the urban description with the resilience of a ductile material has not been proposed before. And it makes all the sense: to calculate the strain-rate state of the urban system meaning the diversification patterns in the city and to relate it directly with the energy that the system can absorb from a perturbation (stress). Nevertheless, the important step is to be able to solve the inverse elasticity problem and recover the stresses.

6.3.4 Towards a general methodology for assessing diversity and sustainability in cities

Concerning the future work in the exploratory data analysis of the urban system, two other methods could be used straight away to expand the analysis of the multivariate data system: Multidimensional Scaling (MDS) (also known as Principal Coordinates Analysis (PCoA)), and Rao's quadratic entropy. The first method —somehow equivalent to PCA— has been widely used in the analysis of genetic and microbial diversity [95, 97]. The later, Rao's quadratic entropy, has been used to analyze functional diversity from dissimilarities [93, 262], and patterns of convergence or divergence [263].

In the second chapter of this thesis, we suggested that Barcelona exhibits the four fundamental conditions to urban diversity proposed by [2]; mixed land uses, small blocks, different building ages, and population density. In this sense, the question would be if diversity would be exposed when analyzing several different urban systems?. Findings have revealed that the proposed methodology constitutes a reliable tool for urban analytics, contributing to the science of cities [39, 145], thus, it can be applied to other urban systems. The only advice that we could make here is to carefully design the information system, paying special attention to the construction of the set of indicators, which must respond to the relevant conditions of each case study.

Even though findings suggested that particular aspects of each district, such as age, land uses, location, urban form, and some socio-economic features are influencing overall diversity, to reach accurate conclusions about urban factors fostering diversity and influencing urban resilience, complementary studies are needed. We believe that a useful future research line would be the scaling analysis of those factors in several different urban systems. Cross-city profiles could be developed to expose diversity patterns and causes influencing urban resilience, and the scaling of those particular patterns in several urban systems. This analysis would allow; in the first place, to correlate urban aspects with measures of diversity. Secondly, to relate aggregate indices of diversity with common measures of urban sustainability developed in the literature. And finally, to show up the scaling of those measures concerning the city's size; either by surface or by population. Because the science of cities is a novel area of research, the road ahead is still long. Future studies in the topics developed through this thesis would help to clarify correlations and trade-offs among diversity and sustainability in cities, which are crucial in facing the regional and global urbanization challenges.

Bibliography

[1] United Nations Population Division, "World Urbanization Prospects: The 2014 Revision," United Nations, New York, Tech. Rep., 2014.

[2] J. Jacobs, *The Death and Life of Great American Cities*. New York: Random House, 1961.

[3] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Ku, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 17, pp. 7301–7306, 2007.

[4] M. Batty, "The size, scale, and shape of cities," *Science*, vol. 319, no. 5864, pp. 769–771, 2008.

[5] C. Kennedy, J. Cuddihy, and J. Engel-yan, "The Changing Metabolism of Cities," *Journal of Industrial Ecology*, vol. 11, no. 2, pp. 43 –59, 2007.

[6] N. B. Grimm, S. H. Faeth, N. E. Golubiewski, C. L. Redman, J. Wu, X. Bai, and J. M. Briggs, "Global change and the ecology of cities," *Science*, vol. 319, no. 5864, pp. 756–760, 2008.

[7] J. Wu, "Urban ecology and sustainability: The state of the science and future directions," *Landscape and Urban Planning*, vol. 125, pp. 209–221, 2014.

[8] N. Munier, M. M. Ziara, R Cole, J Curiel, A Esteban, M Ertsen, V Lall, S Lall, T Litman, R Morrison, *et al.*, *Handbook on Urban Sustainability*. Netherlands: Springer, 2007.

[9] R. Camagni, "Sustainable urban development: Definition and reasons for a research programme," *International Journal of Environment and Pollution*, vol. 10, no. 1, pp. 6–27, 1998.

[10] K. Mori and T. Yamashita, "Methodological framework of sustainability assessment in City Sustainability Index (csi): A concept of constraint and maximisation indicators," *Habitat International*, vol. 45, pp. 10–14, 2015.

[11] S. T. Pickett, M. L. Cadenasso, J. M. Grove, C. G. Boone, P. M. Groffman, E. Irwin, S. S. Kaushal, V. Marshall, B. P. McGrath, C. H. Nilon, *et al.*, "Urban ecological systems: Scientific foundations and a decade of progress," *Journal of Environmental Management*, vol. 92, no. 3, pp. 331–362, 2011.

[12] T. McPhearson, S. T. Pickett, N. B. Grimm, J. Niemelä, M. Alberti, T. Elmqvist, C. Weber, D. Haase, J. Breuste, and S. Qureshi, "Advancing urban ecology toward a science of cities," *BioScience*, vol. 66, no. 3, pp. 198–212, 2016.

[13] J. Niemelä, "Ecology and urban planning," *Biodiversity & Conservation*, vol. 8, no. 1, pp. 119–131, 1999.

[14] P. Ferrão and J. E. Fernández, *Sustainable urban metabolism*. Cambridge: MIT press, 2013.

[15] J. Ahern, "From fail-safe to safe-to-fail: Sustainability and resilience in the new urban world," *Landscape and Urban Planning*, vol. 100, no. 4, pp. 341–343, 2011.

[16] S. Meerow, J. P. Newell, and M. Stults, "Defining urban resilience: A review," *Landscape and Urban Planning*, vol. 147, pp. 38–49, 2016.

[17] S. T. Pickett, M. L. Cadenasso, and J. M. Grove, "Resilient cities: Meaning, models, and metaphor for integrating the ecological, socio-economic, and planning realms," *Landscape and Urban Planning*, vol. 69, no. 4, pp. 369–384, 2004.

[18] Y. Jabareen, "Planning the resilient city: Concepts and strategies for coping with climate change and environmental risk," *Cities*, vol. 31, pp. 220–229, 2013.

[19] J. Wu, "Urban sustainability: An inevitable goal of landscape research," *Landscape Ecology*, vol. 25, no. 1, pp. 1–4, 2010.

[20] J. I. Nassauer and P. Opdam, "Design in science: Extending the landscape ecology paradigm," *Landscape Ecology*, vol. 23, no. 6, pp. 633–644, 2008.

[21] A. Wolman, "The metabolism of cities," *Scientific American*, vol. 213, no. 2, pp. 179 –190, 1965.

[22] P. W. G. Newman, "Sustainability and cities: Extending the metabolism model," *Landscape and Urban Planning*, vol. 44, no. 4, pp. 219–226, 1999.

[23] H. R. Sahely, S. Dudding, and C. Kennedy, "Estimating the urban metabolism of Canadian cities: Greater Toronto Area case study," *Canadian Journal of Civil Engineering*, vol. 30, no. 2, pp. 468–483, 2003.

[24] Y. Zhang, Z. Yang, and W. Li, "Analyses of urban ecosystem based on information entropy," *Ecological Modelling*, vol. 197, no. 1-2, pp. 1–12, 2006.

[25] N. Codoban and C. a. Kennedy, "Metabolism of Neighborhoods," *Journal of Urban Planning and Development*, vol. 134, no. 1, pp. 21–31, 2008.

[26] S. Barles, "Urban metabolism of Paris and its region," *Journal of Industrial Ecology*, vol. 13, no. 6, pp. 898–913, 2009.

[27] ——, "Society, energy and materials: The contribution of urban metabolism studies to sustainable urban development issues," *Journal of Environmental Planning and Management*, vol. 53, no. 4, pp. 439–455, 2010.

[28] W. Huang, S. Cui, M. Yarime, S. Hashimoto, and S. Managi, "Improving urban metabolism study for sustainable urban transformation," *Environmental Technology & Innovation*, vol. 4, pp. 62–72, 2015.

[29] C.-L. Huang, J. Vause, H.-W. Ma, and C.-P. Yu, "Using material/substance flow analysis to support sustainable development assessment: A literature review and outlook," *Resources, Conservation and Recycling*, vol. 68, pp. 104–116, 2012.

[30] L. S. Conke and T. L. Ferreira, "Urban metabolism: Measuring the city's contribution to sustainable development," *Environmental Pollution*, vol. 202, pp. 146–152, 2015.

[31] C. S. Holling, "Resilience and stability of ecological systems," *Annual Review of Ecology and Systematics*, vol. 4, no. 1, pp. 1–23, 1973.

[32] C. Folke, "Resilience: The emergence of a perspective for social–ecological systems analyses," *Global Environmental Change*, vol. 16, no. 3, pp. 253–267, 2006.

[33] C. Folke, S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, L. Gunderson, and C. S. Holling, "Regime shifts, resilience, and biodiversity in ecosystem management," *Annual Review of Ecology, Evolution, and Systematics*, vol. 35, pp. 557–581, 2004.

[34] B. Walker and D. Salt, *Resilience Thinking: Sustaining Ecosystems and People in a Changing World*. Washington D.C.: Island Press, 2006.

[35] J. Ahern, "Urban landscape sustainability and resilience: The promise and challenges of integrating ecology with urban planning and design," *Landscape Ecology*, vol. 28, no. 6, pp. 1203–1212, 2013.

[36] X. Zhang and H. Li, "Urban resilience and urban sustainability: What we know and what do not know?" *Cities*, vol. 72, pp. 141–148, 2018.

[37] L. J. Vale and T. J. Campanella, *The resilient city: How modern cities recover from disaster*. Oxford University Press, 2005.

[38] S. Levin, T. Xepapadeas, A.-S. Crépin, J. Norberg, A. De Zeeuw, C. Folke, T. Hughes, K. Arrow, S. Barrett, G. Daily, *et al.*, "Social-ecological systems as complex adaptive systems: Modeling and policy implications," *Environment and Development Economics*, vol. 18, no. 2, pp. 111–132, 2013.

[39] M. Batty, *The new science of cities*. Cambridge: MIT press, 2013.

[40] L. M. A. Bettencourt and G. West, "A Unified Theory of Urban Living," *Nature*, vol. 465, no. 7318, pp. 912–913, 2010.

[41] F. S. Chapin III, S. R. Carpenter, G. P. Kofinas, C. Folke, N. Abel, W. C. Clark, P. Olsson, D. M. S. Smith, B. Walker, O. R. Young, *et al.*, "Ecosystem stewardship: Sustainability strategies for a rapidly changing planet," *Trends in Ecology & Evolution*, vol. 25, no. 4, pp. 241–249, 2010.

[42] C. S. Holling, "Understanding the complexity of economic, ecological, and social systems," *Ecosystems*, vol. 4, no. 5, pp. 390–405, 2001.

[43] L. H. Gunderson, *Panarchy: understanding transformations in human and natural systems*. Washington, D.C.: Island Press, 2001.

[44] J. Anderies, C. Folke, B. Walker, and E. Ostrom, "Aligning key concepts for global change policy: Robustness, resilience, and sustainability," *Ecology and Society*, vol. 18, no. 2, 2013.

[45] E. Jen, *Robust design: a repertoire of biological, ecological, and engineering case studies*. Oxford University Press, 2005.

[46] S. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. New Jersey: Princeton Univeristy Press, 2007.

[47] ——, *Diversity and Complexity*. New Jersey: Princeton Univeristy Press, 2011.

[48] S. Pickett, M. Cadenasso, and J. Grove, "Resilient cities: meaning, models, and metaphor for integrating the ecological, socio-economic, and planning realms," *Landscape and Urban Planning*, vol. 69, no. 4, pp. 369–384, 2004.

[49] S. A. Levin and J. Lubchenco, "Resilience, robustness, and marine ecosystem-based management," *BioScience*, vol. 58, no. 1, pp. 27–32, 2008.

[50] S. Meerow and J. P. Newell, "Urban resilience for whom, what, when, where, and why?" *Urban Geography*, vol. 40, no. 3, pp. 309–329, 2019.

[51] M. Suárez, E. Gómez-Baggethun, J. Benayas, and D. Tilbury, "Towards an urban resilience index: A case study in 50 spanish cities," *Sustainability*, vol. 8, no. 8, p. 774, 2016.

[52] M. Spaans and B. Waterhout, "Building up resilience in cities worldwide– Rotterdam as participant in the 100 Resilient Cities Programme," *Cities*, vol. 61, pp. 109–116, 2017.

[53] R. Biggs, M. Schlüter, D. Biggs, E. L. Bohensky, S. BurnSilver, G. Cundill, V. Dakos, T. M. Daw, L. S. Evans, K. Kotschy, *et al.*, "Toward principles for enhancing the resilience of ecosystem services," *Annual Review of Environment and Resources*, vol. 37, pp. 421–448, 2012.

[54] R. Biggs, M. Schlüter, and M. L. Schoon, *Principles for building resilience: sustaining ecosystem services in social-ecological systems*. Cambridge University Press, 2015.

[55] C. Folke, J. Colding, and F. Berkes, "Synthesis: Building resilience and adaptive capacity in social-ecological systems," *Navigating social-ecological systems: Building resilience for complexity and change*, vol. 9, no. 1, pp. 352–387, 2003.

[56] K. Kotschy, R. Biggs, T. Daw, C. Folke, and P. C. West, "Principle 1 – Maintain diversity and redundancy," in *Principles for Building Resilience: Sustaining Ecosystem Services in Social-Ecological Systems*, R. Biggs, M. Schlüter, and M. L. Schoon, Eds. Cambridge University Press, 2015, 50–79.

[57] B. Walker and D. Salt, *Resilience thinking: sustaining ecosystems and people in a changing world*. Washington, D.C.: Island Press, 2012.

[58] A. R. Ives and S. R. Carpenter, "Stability and Diversity of Ecosystems," *Science*, vol. 317, no. 5834, pp. 58–62, 2007.

[59] K. S. McCann, "The diversity-stability debate," *Nature*, vol. 405, no. 6783, pp. 228–233, 2000.

[60] G. Duranton and D. Puga, "Diversity and Specialisation in Cities: Why, Where and When Does it Matter?" *Urban Studies*, vol. 37, no. 3, pp. 533–555, 2000.

[61] S. S. Fainstein, "Cities and Diversity: Should We Want It? Can We Plan For It?" *Urban Affairs Review*, vol. 41, no. 1, pp. 3–19, 2005.

[62] T. Kemeny, "Cultural diversity, institutions, and urban economic performance," *Environment and Planning A*, vol. 44, no. 9, pp. 2134–2152, 2012.

[63] E. Talen, *Design for diversity: Exploring Socially Mixed Neighbourhoods*. Oxford: Routledge, 2012.

[64] ——, "Neighborhood-level social diversity," *American Planning Association. Journal of the American Planning Association*, vol. 72, no. 4, p. 431, 2006.

[65] ——, "Design that enables diversity: The complications of a planning ideal," *Journal of Planning Literature*, vol. 20, no. 3, pp. 233–249, 2006.

[66] M. Powe, J. Mabry, E. Talen, and D. Mahmoudi, "Jane Jacobs and the value of older, smaller buildings," *Journal of the American Planning Association*, vol. 82, no. 2, pp. 167–180, 2016.

[67] E. Mohareb, S. Derrible, and F. Peiravian, "Intersections of Jane Jacobs' conditions for diversity and low-carbon urban systems: A look at four global cities," *Journal of Urban Planning and Development*, vol. 142, no. 2, p. 05 015 004, 2015.

[68] L. M. Bettencourt, H. Samaniego, and H. Youn, "Professional diversity and the productivity of cities," *Scientific Reports*, vol. 4, p. 5393, 2014.

[69] R. H. Whittaker, "Vegetation of the Siskiyou mountains, Oregon and California," *Ecological Monographs*, vol. 30, no. 3, pp. 279–338, 1960.

[70] R. H. Whittaker, "Evolution and measurement of species diversity," *Taxon*, vol. 21, no. 2-3, pp. 213–251, 1972.

[71] M. J. Anderson, T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J. Sanders, H. V. Cornell, L. S. Comita, K. F. Davies, *et al.*, "Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist," *Ecology Letters*, vol. 14, no. 1, pp. 19–28, 2011.

[72] A. Stirling, "A general framework for analysing diversity in science, technology and society," *Journal of the Royal Society Interface*, vol. 4, no. 15, pp. 707–719, 2007.

[73] S. C. Chapra, R. P. Canale, *et al.*, *Numerical methods for engineers*. Boston: McGraw-Hill, 2010.

[74] T. Elmqvist, E. Andersson, N. Frantzeskaki, T. McPhearson, P. Olsson, O. Gaffney, K. Takeuchi, and C. Folke, "Sustainability and resilience for transformation in the urban century," *Nature Sustainability*, vol. 2, no. 4, p. 267, 2019.

[75] R. Singh, H. Murty, S. Gupta, and A. Dikshit, "An overview of sustainability assessment methodologies," *Ecological Indicators*, vol. 15, no. 1, pp. 281–299, 2012.

[76] L.-Y. Shen, J. J. Ochoa, M. N. Shah, and X. Zhang, "The application of urban sustainability indicators–A comparison between various practices," *Habitat International*, vol. 35, no. 1, pp. 17–29, 2011.

[77] C. Turcu, "Re-thinking sustainability indicators: Local perspectives of urban sustainability," *Journal of Environmental Planning and Management*, vol. 56, no. 5, pp. 695–719, 2013.

[78] U. Pupphachai and C. Zuidema, "Sustainability indicators: A tool to generate learning and adaptation in sustainable urban development," *Ecological Indicators*, vol. 72, pp. 784–793, 2017.

[79] G. A. Tanguay, J. Rajaonson, J. Lefebvre, and P. Lanoie, "Measuring the sustainability of cities: An analysis of the use of local indicators," *Ecological Indicators*, vol. 10, no. 2, pp. 407–418, 2010.

[80] L. Huang, J. Wu, and L. Yan, "Defining and measuring urban sustainability: A review of indicators," *Landscape Ecology*, vol. 30, no. 7, pp. 1175–1193, 2015.

[81] J. M. Klopp and D. L. Petretta, "The urban sustainable development goal: Indicators, complexity and the politics of measuring cities," *Cities*, vol. 63, pp. 92–97, 2017.

[82] A. Cherp and J. Jewell, "Energy Security Assessment Framework and Three Case Studies," in *International Handbook of Energy Security*, Cheltenham, UK - Northampton, MA, USA: Edward Elgar, 2013, pp. 146 –173.

[83] N. Jollands, J. Lermit, and M. Patterson, "Aggregate eco-efficiency indices for New Zealand — a principal components analysis," *Journal of Environmental Management*, vol. 73, no. 4, pp. 293–305, 2004.

[84] W. H. A. Piña and C. I. P. Martínez, "Urban material flow analysis: An approach for Bogotá, Colombia," *Ecological Indicators*, vol. 42, pp. 32–42, 2014.

[85] C. Kennedy, S. Pincetl, and P. Bunje, "The study of urban metabolism and its applications to urban planning and design," *Environmental Pollution*, vol. 159, no. 8-9, pp. 1965–1973, 2011.

[86] C. Kennedy, I. D. Stewart, N. Ibrahim, A. Facchini, and R. Mele, "Developing a multi-layered indicator set for urban metabolism studies in megacities," *Ecological Indicators*, vol. 47, pp. 7–15, 2014.

[87] I. Prigogine and I. Stengers, *Order Out of Chaos*. New York: Bantan Books Inc., 1984.

[88] E. C. Pielou *et al.*, *An introduction to mathematical ecology*. New York: Wiley-Inter-Science, 1969.

[89] L. Jost, "Entropy and diversity," *Oikos*, vol. 113, no. 2, pp. 363–375, 2006.

[90] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[91] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, 1963.

[92] E. H. Simpson, "Measurement of diversity," *Nature*, vol. 163, no. 4148, p. 688, 1949.

[93] Z. Botta-Dukát, "Rao's quadratic entropy as a measure of functional diversity based on multiple traits," *Journal of Vegetation Science*, vol. 16, no. 5, pp. 533–540, 2005.

[94] M. L. Weitzman, "On diversity," *The Quarterly Journal of Economics*, vol. 107, no. 2, pp. 363–405, 1992.

[95] C. A. Lozupone, J. I. Stombaugh, J. I. Gordon, J. K. Jansson, and R. Knight, "Diversity, stability and resilience of the human gut microbiota," *Nature*, vol. 489, no. 7415, p. 220, 2012.

[96] C. A. Lozupone and R. Knight, "Global patterns in bacterial diversity," *Proceedings of the National Academy of Sciences*, vol. 104, no. 27, pp. 11 436–11 440, 2007.

[97] C. Huttenhower, D. Gevers, R. Knight, S. Abubucker, J. H. Badger, A. T. Chinwalla, H. H. Creasy, A. M. Earl, M. G. FitzGerald, R. S. Fulton, *et al.*, "Structure, function and diversity of the healthy human microbiome," *Nature*, vol. 486, no. 7402, p. 207, 2012.

[98] T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, *et al.*, "Human gut microbiome viewed across age and geography," *Nature*, vol. 486, no. 7402, p. 222, 2012.

[99] A. Ramette, "Multivariate analyses in microbial ecology," *FEMS microbiology ecology*, vol. 62, no. 2, pp. 142–160, 2007.

[100] M. Nelis, T. Esko, R. Mägi, F. Zimprich, A. Zimprich, D. Toncheva, S. Karachanak, T. Piskáčková, I. Balaščák, L. Peltonen, *et al.*, "Genetic structure of Europeans: A view from the North–East," *PloS one*, vol. 4, no. 5, e5472, 2009.

[101] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics*, vol. 38, no. 8, p. 904, 2006.

[102] T. J. Vicino, B. Hanlon, and J. R. Short, "A Typology of Urban Immigrant Neighborhoods," *Urban Geography*, vol. 32, no. 3, pp. 383–405, 2011.

[103] N. Foote and R. Walter, "Neighborhood and socioeconomic change in emerging megapolitan nodes: Tracking shifting social geographies in three rapidly growing United States metropolitan areas, 1980–2010," *Urban Geography*, vol. 38, no. 8, pp. 1203–1230, 2017.

[104] A. Owens, "Neighborhoods on the rise: A typology of neighborhoods experiencing socioeconomic ascent," *City & Community*, vol. 11, no. 4, pp. 345–369, 2012.

[105] I. Jolliffe, *Principal Component Analysis*, 2nd. New York: Springer Verlag, 2002.

[106] W. K. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*. Berlin Heidelberg: Springer Verlag, 2015.

[107] H. Abdi and L. J. Williams, "Principal Component Analysis," *WIREs Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[108] H. Abdi, L. J. Williams, and D. Valentin, "Multiple factor analysis: principal component analysis for multitable and multiblock data sets," *WIREs Computational Statistics*, vol. 5, no. 2, 2013.

[109] K. Gabriel, "The biplot graphic display of matrices with application to principal component analysis," *Biometrika*, vol. 58, no. 3, pp. 453–467, 1971.

[110] M. T. Werth, S. Halouska, M. D. Shortridge, B. Zhang, and R. Powers, "Analysis of metabolomic PCA data using tree diagrams," *Analytical Biochemistry*, vol. 399, no. 1, pp. 58–63, 2010.

[111] B. Worley, S. Halouska, and R. Powers, "Utilities for quantifying separation in pca/plsda scores plots," *Analytical Biochemistry*, vol. 433, no. 2, pp. 102 –104, 2013.

[112] J. Josse, S. Wager, and F. Husson, "Confidence areas for fixed-effects pca," *Journal of Computational and Graphical Statistics*, vol. 25, no. 1, pp. 28–48, 2016.

[113] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[114] J. Süßmuth, M. Winter, and G. Greiner, "Reconstructing animated meshes from time-varying point clouds," in *Computer Graphics Forum*, Wiley Online Library, vol. 27, 2008, pp. 1469–1476.

[115] M. N. Ustinin, E Kronberg, S. V. Filippov, V. V. Sytchev, E. V. Sobolev, and R Llinás, "Kinematic visualization of human magnetic encephalography," *Mathematical Biology Bioinformatics*, vol. 5, no. 2, pp. 176–187, 2010.

[116] S.-W. Wang, V. Interrante, and E. Longmire, "Multivariate visualization of 3D turbulent flow data," in *Visualization and Data Analysis 2010*, International Society for Optics and Photonics, vol. 7530, 2010.

[117] M. Kohler and A. Krzyżak, "Nonparametric estimation of non-stationary velocity fields from 3D particle tracking velocimetry data," *Computational Statistics & Data Analysis*, vol. 56, no. 6, pp. 1566–1580, 2012.

[118] A. Aguirre-Pablo, A. B. Aljedaani, J Xiong, R. Idoughi, W. Heidrich, and S. T. Thoroddsen, "Single-camera 3D PTV using particle intensities and structured light," *Experiments in Fluids*, vol. 60, no. 2, p. 25, 2019.

[119] L. Gao, D. G. Heath, B. S. Kuszyk, and E. K. Fishman, "Automatic liver segmentation technique for three-dimensional visualization of CT data.," *Radiology*, vol. 201, no. 2, pp. 359–364, 1996.

[120] M. Teplan *et al.*, "Fundamentals of EEG measurement," *Measurement Science Review*, vol. 2, no. 2, pp. 1–11, 2002.

[121] R. Fuchs and H. Hauser, "Visualization of multi-variate scientific data," in *Computer Graphics Forum*, Wiley Online Library, vol. 28, 2009, pp. 1670–1690.

[122] C. Nastar, B. Moghaddam, and A. Pentland, "Generalized image matching: Statistical learning of physically-based deformations," in *European Conference on Computer Vision*, Springer, 1996, pp. 589–598.

[123] P. Y. Simard, D. Steinkraus, J. C. Platt, *et al.*, "Best practices for convolutional neural networks applied to visual document analysis.," in *Icdar*, vol. 3, 2003.

[124] S. Wang, Y. Zhang, G. Liu, P. Phillips, and T.-F. Yuan, "Detection of alzheimer's disease by three-dimensional displacement field estimation in structural magnetic resonance imaging," *Journal of Alzheimer's Disease*, vol. 50, no. 1, pp. 233–248, 2016.

[125] R. Ibanez, E. Abisset-Chavanne, J. V. Aguado, D. Gonzalez, E. Cueto, and F. Chinesta, "A manifold learning approach to data-driven computational elasticity and inelasticity," *Archives of Computational Methods in Engineering*, vol. 25, no. 1, pp. 47–57, 2018.

[126] V. Mayer-Schönberger and K. Cukier, *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt, 2013.

[127] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer-Verlag, 2006.

[128] G. T. Mase, R. E. Smelser, and G. E. Mase, *Continuum mechanics for engineers*. Boca Raton: CRC Press, 2009.

[129] H. Wang and A. A. Amini, "Cardiac motion and deformation recovery from MRI: A review," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 487–503, 2012.

[130] G. Pedrizzetti, S. Sengupta, G. Caracciolo, C. S. Park, M. Amaki, G. Goliasch, J. Narula, and P. P. Sengupta, "Three-dimensional principal strain analysis for characterizing subclinical changes in left ventricular function," *Journal of the American Society of Echocardiography*, vol. 27, no. 10, pp. 1041–1050, 2014.

[131] J. Cai and E. W. Grafarend, "Statistical analysis of geodetic deformation (strain rate) derived from the space geodetic measurements of BIFROST Project in Fennoscandia," *Journal of Geodynamics*, vol. 43, no. 2, pp. 214–238, 2007.

[132] ——, "Statistical analysis of the eigenspace components of the two-dimensional, symmetric rank-two strain rate tensor derived from the space geodetic measurements (ITRF92-ITRF2000 data sets) in central Mediterranean and Western Europe," *Geophysical Journal International*, vol. 168, no. 2, pp. 449–472, 2007.

[133] P. C. Dodwell, *Visual pattern recognition*. New York: Holt, Rinehart and Winston, 1970.

[134] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[135] K.-S. Shin, T. S. Lee, and H.-j. Kim, "An application of support vector machines in bankruptcy prediction model," *Expert Systems with Applications*, vol. 28, no. 1, pp. 127–135, 2005.

[136] K. Fu, *Sequential methods in pattern recognition and machine learning*. Academic Press, 1968, vol. 52.

[137] M. Bonnet and A. Constantinescu, "Inverse problems in elasticity," *Inverse Problems*, vol. 21, no. 2, R1, 2005.

[138] M. M. Doyley, "Model-based elastography: A survey of approaches to the inverse elasticity problem," *Physics in Medicine & Biology*, vol. 57, no. 3, R35, 2012.

[139] D. D. Woods and J. Wreathall, "Stress-strain plots as a basis for assessing system resilience," *Resilience Engineering Perspectives*, vol. 1, pp. 145–161, 2008.

[140] Y. Zhang, "Urban metabolism: A review of research methodologies," *Environmental Pollution*, vol. 178, pp. 463–473, 2013.

[141] L. M. Bettencourt and J. Lobo, "Urban scaling in Europe," *Journal of The Royal Society Interface*, vol. 13, no. 116, p. 20 160 005, 2016.

[142] L. M. Bettencourt, "The origins of scaling in cities," *Science*, vol. 340, no. 6139, pp. 1438–1441, 2013.

[143] M. Schläpfer, L. M. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, and C. Ratti, "The scaling of human interactions with city size," *Journal of the Royal Society Interface*, vol. 11, no. 98, p. 20 130 789, 2014.

[144] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty, "Constructing cities, deconstructing scaling laws," *Journal of The Royal Society Interface*, vol. 12, no. 102, p. 20 140 745, 2015.

[145] M. Batty, "Building a science of cities," *Cities*, vol. 29, S9–S16, 2012.

[146] N. Mohajeri, J. R. French, and M. Batty, "Evolution and entropy in the organization of urban street patterns," *Annals of GIS*, vol. 19, no. 1, pp. 1–16, 2013.

[147] J. Fariña, "Cálculo de la entropía producida en diversas zonas de Madrid," *Cuadernos de Investigación Urbanística*, vol. 0, no. 10, 1995.

[148] A. Gudmundsson and N. Mohajeri, "Entropy and order in urban street networks," *Scientific Reports*, vol. 3, p. 3324, 2013.

[149] N. Mohajeri and A. Gudmundsson, "The evolution and complexity of urban street networks," *Geographical Analysis*, vol. 46, no. 4, pp. 345–367, 2014.

[150] A. Yeh, "Measurement and monitoring of urban sprawl in a rapidly growing region using entropy," *Photogrammetric Engineering & Remote Sensing*, vol. 67, no. 1, pp. 83–90, 2001.

[151] X. Li and A. G.-O. Yeh, "Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS," *Landscape and Urban Planning*, vol. 69, no. 4, pp. 335–354, 2004.

[152] Z. Lin and B. Xia, "Sustainability analysis of the urban ecosystem in Guangzhou City based on information entropy between 2004 and 2010," *Journal of Geographical Sciences*, vol. 23, no. 3, pp. 417–435, 2013.

[153] Departament d'Estadística, *Index Statistical yearbooks of Barcelona city*, 2017. [Online]. Available: `http://www.bcn.cat/estadistica/angles/dades/anuaris/index.htm` (visited on 10/14/2016).

[154] Barcelona City Council, *ED50. 1:5,000 CITY MAP scale in CAD format*, 2015. [Online]. Available: `http://w20.bcn.cat/cartobcn/default.aspx?lang=en`.

[155] Agència de Salut Pública de Barcelona, "Informe d'avaluació de la qualitat de l'aire a la ciutat de Barcelona," Barcelona, Tech. Rep., 2016. [Online]. Available: `http://www.aspb.cat/wp-content/uploads/2016/05/Qualitat{\_}aire{\_}2016.pdf`.

[156] P. Érdi, *Complexity explained*. Berlin Heidelberg: Springer-Verlag, 2007.

[157] M. Mitchell, *Complexity: A guided tour*. New York: Oxford University Press, 2009.

[158] L. Chelleri, J. J. Waters, M. Olazabal, and G. Minucci, "Resilience trade-offs: Addressing multiple scales and temporal aspects of urban resilience," *Environment and Urbanization*, vol. 27, no. 1, pp. 181–198, 2015.

[159] M. P. Feldman and D. B. Audretsch, "Innovation in cities:: Science-based diversity, specialization and localized competition," *European Economic Review*, vol. 43, no. 2, pp. 409–429, 1999.

[160] J. a. Griffith, E. a. Martinko, and K. P. Price, "Landscape structure analysis of Kansas at three scales," *Landscape and Urban Planning*, vol. 52, no. 1, pp. 45–61, 2000.

[161] S. A. Cushman, K. Mcgarigal, and M. C. Neel, "Parsimony in landscape metrics: Strength, universality and consistency," *Ecological Indicators*, vol. 8, no. 5, pp. 691–703, 2008.

[162] L. Salvati, M. Zitti, and M. Carlucci, "Territorial systems, regional disparities and sustainability: Economic structure and soil degradation in Italy," *Sustainability*, vol. 6, no. 5, pp. 3086–3104, 2014.

[163] P. Taylor and D. Walker, "World cities: A first multivariate analysis of their service complexes," *Urban Studies*, vol. 38, no. 1, pp. 23–47, 2001.

[164] H. Mirshojaeian and S. Kaneko, "Dynamic sustainability assessment of countries at the macro level: A principal component analysis," *Ecological Indicators*, vol. 11, no. 3, pp. 811–823, 2011.

[165] L. Salvati, "Towards a polycentric region? the socio-economic trajectory of Rome, an 'Eternally Mediterranean' City," *Tijdschrift Voor Economische en Sociale Geografie*, vol. 105, no. 3, pp. 268–284, 2014.

[166] P. Bolcárová and S. Kološta, "Assessment of sustainable development in the EU 27 using aggregated SD index," *Ecological Indicators*, vol. 48, pp. 699–705, 2015.

[167] M. Cecchini, S. Cividino, R. Turco, and L. Salvati, "Population age structure, complex socio-demographic systems and resilience potential: A spatio-temporal, evenness-based approach," *Sustainability*, vol. 11, no. 7, p. 2050, 2019.

[168] J. Pagès, *Multiple factor analysis by example using R*. Boca Raton: CRC Press, 2014.

[169] B. Escofier and J. Pagès, *Analyses factorielles simples et multiples. Objectifs méthodes et interprétation*. Paris: Dunod, 2008.

[170] J. Pagès, "Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley," *Food Quality and Preference*, vol. 16, no. 7, pp. 642–649, 2005.

[171] M. Bécue-Bertaut and J. Pagès, "Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data," *Computational Statistics & Data Analysis*, vol. 52, no. 6, pp. 3255–3268, 2008.

[172] L. Perrin, R. Symoneaux, I. Maître, C. Asselin, F. Jourjon, and J. Pagès, "Comparison of three sensory methods for use with the Napping® procedure: Case of ten wines from Loire Valley," *Food Quality and Preference*, vol. 19, no. 1, pp. 1–11, 2008.

[173] L. Dooley, Y.-s. Lee, and J.-F. Meullenet, "The application of check-all-that-apply (cata) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping," *Food Quality and Preference*, vol. 21, no. 4, pp. 394–401, 2010.

[174] T Feuillet, H Charreire, C Roda, M. B. Rebah, J. D. Mackenbach, S Compernolle, K Glonti, H Bárdos, H Rutter, I. D. Bourdeaudhuij, M Mckee, J Brug, J Lakerveld, and J Oppert, "Neighbourhood typology based on virtual audit of environmental obesogenic characteristics," *Obesity Reviews*, vol. 17, pp. 19–30, 2016.

[175] L. Salvati and M. Carlucci, "A composite index of sustainable development at the local scale: Italy as a case study," *Ecological Indicators*, vol. 43, pp. 162–171, 2014.

[176] G. Boeing, "Urban spatial order: Street network orientation, configuration, and entropy," *Applied Network Science*, vol. 4, no. 1, p. 67, 2019.

[177] I. Cerdà, "Memoria del Anteproyecto de Ensanche de Barcelona, 1855," in *TCC. Cerdà & Barcelona*, INAP, Ed., vol. 1, Madrid, 1991, pp. 51 –106.

[178] ——, "Teoría de la Construcción de las Ciudades aplicada al Proyecto de Reforma y Ensanche de Barcelona, 1859," in *TCC. Cerdà & Barcelona*, INAP, Ed., vol. 1, Madrid, 1991, pp. 107 –455.

[179] ——, "Ordenanzas Municipales de Construcción para la Ciudad de Barcelona y pueblos comprendidos por el Ensanche, 1859," in *TCC. Cerdà & Barcelona*, INAP, Ed., vol. 1, Madrid, 1991, pp. 513 –548.

[180] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: `https://www.R-project.org/`.

[181] T. Wei and V. Simko, *R package "corrplot": Visualization of a correlation matrix*, (Version 0.84), 2017. [Online]. Available: `https://github.com/taiyun/corrplot`.

[182] F. Husson, S. Le, and J. Pages, "Principal Component Analysis (PCA)," in *Exploratory Multivariate Analysis by Example Using R*, Boca Raton: CRC Press, 2011, pp. 1 –59.

[183] S. Lê, J. Josse, F. Husson, *et al.*, "Factominer: An R package for multivariate analysis," *Journal of Statistical Software*, vol. 25, no. 1, pp. 1–18, 2008.

[184] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.

[185] A. Kassambara and F. Mundt, "Package 'factoextra'," *Extract and Visualize the Results of Multivariate Data Analyses*, vol. 76, 2017.

[186] I. Gallego-Alvarez, M. Vicente-Galindo, M. Galindo-Villardón, and M. Rodríguez-Rosa, "Environmental performance in countries worldwide: Determinant factors and multivariate analysis," *Sustainability*, vol. 6, no. 11, pp. 7807–7832, 2014.

[187] N. A. Bokulich, S. Subramanian, J. J. Faith, D. Gevers, J. I. Gordon, R. Knight, D. A. Mills, and J. G. Caporaso, "Quality-filtering vastly improves diversity estimates from illumina amplicon sequencing," *Nature Methods*, vol. 10, no. 1, p. 57, 2013.

[188] M. Hartmann, B. Frey, J. Mayer, P. Mäder, and F. Widmer, "Distinct soil microbial diversity under long-term organic and conventional farming," *The ISME Journal*, vol. 9, no. 5, p. 1177, 2015.

[189] T. P. Makhalanyane, A. Valverde, N.-K. Birkeland, S. C. Cary, I. M. Tuffin, and D. A. Cowan, "Evidence for successional development in Antarctic hypolithic bacterial communities," *The ISME Journal*, vol. 7, no. 11, p. 2080, 2013.

[190] C. J. ter Braak, "Principal components biplots and alpha and beta diversity," *Ecology*, vol. 64, no. 3, pp. 454–462, 1983.

[191] C. A. Lozupone, J. Stombaugh, A. Gonzalez, G. Ackermann, D. Wendel, Y. Vázquez-Baeza, J. K. Jansson, J. I. Gordon, and R. Knight, "Meta-analyses of studies of the human microbiota," *Genome Research*, vol. 23, no. 10, pp. 1704–1714, 2013.

[192] L. J. Barwell, N. J. Isaac, and W. E. Kunin, "Measuring β-diversity with species abundance data," *Journal of Animal Ecology*, vol. 84, no. 4, pp. 1112–1122, 2015.

[193] S. Champely and D. Chessel, "Measuring biological diversity using euclidean metrics," *Environmental and Ecological Statistics*, vol. 9, no. 2, pp. 167–177, 2002.

[194] J. Rius Ulldemolins, "Culture and authenticity in urban regeneration processes: Place branding in central Barcelona," *Urban Studies*, vol. 51, no. 14, pp. 3026–3045, 2014.

[195] R. Lagonigro, J. C. Martori, and P. Apparicio, "Environmental noise inequity in the city of Barcelona," *Transportation Research Part D: Transport and Environment*, vol. 63, pp. 309–319, 2018.

[196] O. Marquet and C. Miralles-Guasch, "Walking short distances. the socioeconomic drivers for the use of proximity in everyday mobility in Barcelona," *Transportation Research Part A: Policy and Practice*, vol. 70, pp. 210–222, 2014.

[197] S. Szabó, L. Bertalan, Á. Kerekes, and T. J. Novák, "Possibilities of land use change analysis in a mountainous rural area: A methodological approach," *International Journal of Geographical Information Science*, vol. 30, no. 4, pp. 708–726, 2016.

[198] Y. Wu, W. Que, Y.-g. Liu, J. Li, L. Cao, S.-b. Liu, G.-m. Zeng, and J. Zhang, "Efficiency estimation of urban metabolism via emergy, dea of time-series," *Ecological Indicators*, vol. 85, pp. 276–284, 2018.

[199] S. Arbaci and T. Tapada-Berteli, "Social inequality and urban regeneration in Barcelona city centre: Reconsidering success," *European Urban and Regional Studies*, vol. 19, no. 3, pp. 287–311, 2012.

[200] I. A. Papachristou and M. Rosas-Casals, "Unveiling connectivity patterns of categories in complex systems: An application to human needs in urban places," *The Journal of Mathematical Sociology*, vol. 40, no. 4, pp. 219–238, 2016.

[201] D. Arribas-Bel, K. Kourtit, and P. Nijkamp, "Benchmarking of world cities through self-organizing maps," *Cities*, vol. 31, pp. 248–257, 2013.

[202] D. Arribas-Bel, P. Nijkamp, and H. Scholten, "Multidimensional urban sprawl in Europe: A self-organizing map approach," *Computers, Environment and Urban Systems*, vol. 35, no. 4, pp. 263–275, 2011.

[203] V. Pawlowsky-Glahn, J. J. Egozcue, and R. Tolosana-Delgado, *Modeling and analysis of compositional data*. John Wiley & Sons, 2015.

[204] X. Liu, B. Derudder, and P. Taylor, "Mapping the evolution of hierarchical and regional tendencies in the world city network, 2000–2010," *Computers, Environment and Urban Systems*, vol. 43, pp. 51–66, 2014.

[205] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009.

[206] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20 150 202, 2016.

[207] T. W. Liao, "Clustering of time series data—a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[208] E. Strano, M. Viana, L. da Fontoura Costa, A. Cardillo, S. Porta, and V. Latora, "Urban street networks, a comparative analysis of ten European cities," *Environment and Planning B: Planning and Design*, vol. 40, no. 6, pp. 1071–1086, 2013.

[209] Q. Wu, J. Cheng, G. Chen, D. J. Hammel, and X. Wu, "Socio-spatial differentiation and residential segregation in the Chinese city based on the 2000 community-level census data: A case study of the inner city of Nanjing," *Cities*, vol. 39, pp. 109–119, 2014.

[210] P. J. Taylor and B. Derudder, *World city network: a global urban analysis*. Routledge, 2015.

[211] C. A. Lozupone, M. Hamady, S. T. Kelley, and R. Knight, "Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities," *Applied and Environmental Microbiology*, vol. 73, no. 5, pp. 1576–1585, 2007.

[212] M. E. Garber, O. G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, M. Van De Rijn, G. D. Rosen, C. M. Perou, R. I. Whyte, *et al.*, "Diversity of gene expression in adenocarcinoma of the lung," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13 784–13 789, 2001.

[213] H. Y. Chang, J.-T. Chi, S. Dudoit, C. Bondre, M. van de Rijn, D. Botstein, and P. O. Brown, "Diversity, topographic differentiation, and positional memory in human fibroblasts," *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12 877–12 882, 2002.

[214] R. Cervero and M. Duncan, "Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay Area," *American Journal of Public Health*, vol. 93, no. 9, pp. 1478–1483, 2003.

[215] M. Zitti, C. Ferrara, L. Perini, M. Carlucci, and L. Salvati, "Long-term urban growth and land use efficiency in Southern Europe: Implications for sustainable land management," *Sustainability*, vol. 7, no. 3, pp. 3359–3385, 2015.

[216] D. E. Sands, "Correlation and covariance," *Journal of Chemical Education*, vol. 54, no. 2, p. 90, 1977.

[217] D. Wang, T. Abdelzaher, L. Kaplan, R. Ganti, S. Hu, and H. Liu, "Exploitation of physical constraints for reliable social sensing," in *2013 IEEE 34th Real-Time Systems Symposium*, IEEE, 2013, pp. 212–223.

[218] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[219] F. Murtagh and P. Legendre, "Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion?" *Journal of Classification*, vol. 31, no. 3, pp. 274–295, 2014.

[220] R. V. Jones, A. Fuertes, and K. J. Lomas, "The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings," *Renewable and Sustainable Energy Reviews*, vol. 43, pp. 901–917, 2015.

[221] Y. G. Yohanis, J. D. Mondol, A. Wright, and B. Norton, "Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use," *Energy and Buildings*, vol. 40, no. 6, pp. 1053–1059, 2008.

[222] L. Lamsal, R. Martin, D. D. Parrish, and N. A. Krotkov, "Scaling relationship for NO2 pollution and urban population size: A satellite perspective," *Environmental Science & Technology*, vol. 47, no. 14, pp. 7855–7861, 2013.

[223] B. Gurjar, T. M. Butler, M. G. Lawrence, and J. Lelieveld, "Evaluation of emissions and air quality in megacities," *Atmospheric Environment*, vol. 42, no. 7, pp. 1593–1606, 2008.

[224] C. Borrego, H. Martins, O. Tchepel, L Salmim, A. Monteiro, and A. I. Miranda, "How urban structure can affect city sustainability from an air quality perspective," *Environmental Modelling & Software*, vol. 21, no. 4, pp. 461–467, 2006.

[225] US Census Bureau, *QuickFacts New York County*, 2018. [Online]. Available: `https://www.census.gov/quickfacts/newyorkcountymanhattanboroughnewyork` (visited on 05/22/2019).

[226] J. E. Stiglitz, "The theory of 'screening', education, and the distribution of income," *The American Economic Review*, vol. 65, no. 3, pp. 283–300, 1975.

[227] A. Ferrer-i Carbonell, "Income and well-being: An empirical analysis of the comparison income effect," *Journal of Public Economics*, vol. 89, no. 5-6, pp. 997–1019, 2005.

[228] F. M. White and I. Corfield, *Viscous fluid flow*. New York: McGraw-Hill, 2006, vol. 3.

[229] A. F. Bower, *Applied mechanics of solids*. Boca Raton: CRC Press, 2009.

[230] T. McLoughlin, R. S. Laramee, R. Peikert, F. H. Post, and M. Chen, "Over two decades of integration-based, geometric flow visualization," in *Computer Graphics Forum*, Wiley Online Library, vol. 29, 2010, pp. 1807–1829.

[231] D. A. Keim, "Information visualization and visual data mining," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 1–8, 2002.

[232] C. Wang, H. Yu, and K.-L. Ma, "Importance-driven time-varying data visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1547–1554, 2008.

[233] F Husson, S Lê, and J Pagès, "Confidence ellipse for the sensory profiles obtained by principal component analysis," *Food Quality and Preference*, vol. 16, no. 3, pp. 245–250, 2005.

[234] M. Cadoret and F. Husson, "Construction and evaluation of confidence ellipses applied at sensory data," *Food Quality and Preference*, vol. 28, no. 1, pp. 106–115, 2013.

[235] H. Abdi, L. J. Williams, and M. Béra, "Barycentric discriminant analysis," *Encyclopedia of Social Network Analysis and Mining*, pp. 1–20, 2017.

[236] B. Wang, W. Shi, and Z. Miao, "Confidence analysis of standard deviational ellipse and its extension into higher dimensional euclidean space," *PloS one*, vol. 10, no. 3, e0118537, 2015.

[237] J. Josse, S. Wager, and F. Husson, "Confidence areas for fixed-effects PCA," *Journal of Computational and Graphical Statistics*, vol. 25, no. 1, pp. 28–48, 2016.

[238] P. Cederholm, "Deformation analysis using confidence ellipsoids," *Survey Review*, vol. 37, no. 287, pp. 31–45, 2003.

[239] J. E. Marsden and T. J. Hughes, *Mathematical foundations of elasticity*. New York: Dover Publications, 1994.

[240] P. Xu and E. Grafarend, "Statistics and geometry of the eigenspectra of three-dimensional second-rank symmetric random tensors," *Geophysical Journal International*, vol. 127, no. 3, pp. 744–756, 1996.

[241] S. Wdowinski, Y. Sudman, and Y. Bock, "Geodetic detection of active faults in S. California," *Geophysical Research Letters*, vol. 28, no. 12, pp. 2321–2324, 2001.

[242] C. Straub, H.-G. Kahle, and C. Schindler, "GPS and geologic estimates of the tectonic activity in the Marmara Sea region, NW Anatolia," *Journal of Geophysical Research: Solid Earth*, vol. 102, no. B12, pp. 27 587–27 601, 1997.

[243] P. Moin, *Fundamentals of engineering numerical analysis*. Cambridge University Press, 2010.

[244] M Hackl, R Malservisi, and S Wdowinski, "Strain rate patterns from dense GPS networks," *Natural Hazards and Earth System Sciences*, vol. 9, no. 4, pp. 1177–1187, 2009.

[245] B Mastrolembo and A Caporali, "Stress and strain-rate fields: A comparative analysis for the Italian territory.," *Bollettino di Geofisica Teorica ed Applicata*, vol. 58, no. 4, 2017.

[246] N. Houlie, J. Woessner, D. Giardini, and M. Rothacher, "Lithosphere strain rate and stress field orientations near the Alpine arc in Switzerland," *Scientific Reports*, vol. 8, no. 1, 2018.

[247] X. Su, L. Yao, W. Wu, G. Meng, L. Su, R. Xiong, and S. Hong, "Crustal deformation on the northeastern margin of the Tibetan plateau from continuous GPS observations," *Remote Sensing*, vol. 11, no. 1, p. 34, 2019.

[248] I. N. Sneddon, "The distribution of stress in the neighbourhood of a crack in an elastic solid," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 187, no. 1009, pp. 229–260, 1946.

[249] M. Williams, "The bending stress distribution at the base of a stationary crack," *Journal of Applied Mechanics, volume=28, pages=78–82, year=1961*,

[250] B. Aktuğ, E. Parmaksız, M. Kurt, O. Lenk, A. Kılıçoğlu, M. A. Gürdal, and S. Özdemir, "Deformation of Central Anatolia: GPS implications," *Journal of Geodynamics*, vol. 67, pp. 78–96, 2013.

[251] E. Grafarend, "Criterion matrices for deforming networks," in *Optimization and Design of Geodetic networks*, Springer, 1985, pp. 363–428.

[252] E. W. Grafarend, "Three-dimensional deformation analysis: Global vector spherical harmonic and local finite element representation," *Tectonophysics*, vol. 130, no. 1-4, pp. 337–359, 1986.

[253] A Dermanis and E. Grafarend, "The finite element approach to the geodetic computation of two-and three-dimensional deformation parameters: A study of frame invariance and parameter estimability," in *International Conference "Cartography-Geodesy", Maracaibo, Venezuela*, 1992.

[254] M. De Berg, M. Van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*. Berlin Heidelberg: Springer-Verlag, 2008.

[255] C. Marot, J. Pellerin, and J.-F. Remacle, "One machine, one minute, three billion tetrahedra," *International Journal for Numerical Methods in Engineering*, vol. 117, no. 9, pp. 967–990, 2019.

[256] J. M. Gere and B. J. Goodno, "Mechanics of Materials, Brief Edition," *Cengage Learning*, 2012.

[257] S. Clemants and G. Moore, "Patterns of species diversity in eight northeastern United States cities," *Urban Habitats*, vol. 1, 2003.

[258] C. Raudsepp-Hearne, G. D. Peterson, and E. M. Bennett, "Ecosystem service bundles for analyzing tradeoffs in diverse landscapes," *Proceedings of the National Academy of Sciences*, vol. 107, no. 11, pp. 5242–5247, 2010.

[259] E. Laliberté and P. Legendre, "A distance-based framework for measuring functional diversity from multiple traits," *Ecology*, vol. 91, no. 1, pp. 299–305, 2010.

[260] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *City*, vol. 1, no. 2, p. 1, 2007.

[261] D. Harvey, *Rebel cities: From the right to the city to the urban revolution*. New York: Verso Books, 2012.

[262] S. Pavoine, S. Ollier, and D. Pontier, "Measuring diversity from dissimilarities with rao's quadratic entropy: Are any dissimilarities suitable?" *Theoretical population biology*, vol. 67, no. 4, pp. 231–239, 2005.

[263] C. Ricotta and M. Moretti, "CWM and Rao's quadratic diversity: A unified framework for functional ecology," *Oecologia*, vol. 167, no. 1, pp. 181–188, 2011.